

Signs of the times: medieval punctuation, diplomatic encoding and rendition

Elisa Cugliana^{†,††}, Gioele Barabucci^{}*

[†] Ca' Foscari University of Venice

^{††} University of Cologne

^{*} NTNU – Norwegian University of Science and Technology

Abstract

Digitally managing punctuation in the editions of medieval manuscripts is one of those issues that initially look like a minor detail, but later reveal themselves as a tangled web of problems spanning from computer science (how to represent punctuation signs?) to philology (what types of signs do exist?) through epistemology (is the processing of punctuation a mere technical transformation or a valuable part of the scholarship?). The aim of this paper is to address the theoretical aspects of these questions and their practical implications, providing a couple of solutions fitting the paradigms and the technologies of the TEI.

This paper describes how we dealt with the encoding and transformation of the punctuation in the Early New High German edition of Marco Polo's travel account. Technically, we implemented a set of general rules (as XSLT templates) plus various exceptions (as descriptive instructions in XML attributes), and applied in an automated fashion (using XProc pipelines). In addition to this, we discuss the philological foundation of this method and, contextually, we address the topic of the transformation of a single original source into different transcriptions: from a highly diplomatic edition to an interpretative one, going through a spectrum of intermediate levels of normalisation. We also reflect on the separation between transcription and analysis, as well as on the role of the editor when the edition is the output of a semi-automated process

Keywords: normalisation; punctuation; automated workflow; pipelines; computable edition; formalisation of editorial rules; reproducibility.

1 Introduction

Since its inception, the critical edition of the Early New High German Marco Polo¹ (version DI² of the *Devisement dou Monde*) has been planned as a multiple edition. For each witness the editor would manually edit a master TEI file and, from that file, three editions would be generated, each at a different level of faithfulness/readability: a purely diplomatic edition (where aspects like colours,

¹ A philological introduction to the project can be found in Cugliana (2019).

² Originally “DI” stood for “Deutsche Inkunabel”, but nowadays it just denotes the Early New High German version of Marco Polo's work, preserved in five witnesses.

letter sizes, glyph variance and layout are preserved), a fully interpretative edition (presenting a more modernised text) and a semi-diplomatic edition (with a text which still maintains relevant features from the source, but which is already more readable and fluent, also thanks to a minimal system of punctuation).

These issues quickly arose while editing:

- the master TEI file became too big and its structure too complex, thus too hard to navigate and maintain, even when using advanced XML editors such as oXygen XML;
- normalising punctuation revealed itself as a complex task that required profound changes to the structure of the edited text.

The first issue, the overgrowth of the master TEI file, is due to the need of transcribing and encoding three variants of each word of the witness (one for each level of faithfulness). Not only is this process extremely time consuming, but it also requires the addition of a lot of markup (e.g. `<choice>` elements) that stiffens the text hierarchy and makes it cumbersome to encode complex linguistic phenomena or critical annotations spanning over multiple words.

The second issue, the complexity of normalising punctuation, emerged from the combination of two facts:

- it was not possible to find an underlying system governing the use of punctuation in the source;
- using the standard approach to normalisation in TEI (using `<choice>` with, e.g., `<orig>` and `<reg>`) turned out to be impracticable for two main reasons: the verbosity of the encoding and the lack of correspondences between the medieval and the modern systems (`<orig>` would very often be empty), both complicated by the presence of three levels of edition (diplomatic, semi-diplomatic and interpretative).

To address these issues, we devised and implemented a novel approach that made it possible for a single scholar to edit each witness in three different levels of edition in the allotted time span of 18 months, while keeping the master TEI file lean in its structure and readable even by scholars that are not well versed in XML technologies.

The key innovation of this approach is the encoding of many textual phenomena (and of how these phenomena should be handled) as formal, machine readable rules, separate from the edited text. An automated workflow is then used to apply these rules on the master file and to produce each of the three levels of edition, depending on which set of rules has been selected. In contrast with other similar systems, in our approach there is a clear distinction between rules and exceptions. This is particularly important in the ENHG Marco Polo edition because the language it is written in has never been standardised and thus the system of rule has been organically derived during the editorial work (as described in detail in Section 2).

This paper provides an overview of our approach (Section 3) and shows how our approach addresses both issues. In particular we show how normalising punctuation represents a dramatic step beyond the more classical normalisation of words. The current implementation of our approach, based on XProc and XSLT, is also presented in Section 3.

Moving to an editorial workflow with such a level of automation requires a reevaluation of the role of the editor, from wordsmith to formaliser of rules (and exceptions). In Section 4 we discuss how our approach fits the recent edition-as-a-program paradigm and which changes in perspective a wide-spread adoption of this methodology could entail for future critical editors.

2 Background and issues

2.1 Marco Polo and ENHG

The *Devisement dou Monde* is a travel narrative written by Rustichello da Pisa and Marco Polo, most probably in 1298, while they were both in prison, in Genoa. It narrates about Marco Polo's travels to and within the Eastern World: the traveller was a Venetian who, following his father and uncle on a business trip to Asia, ended up spending almost 25 years in the East, at the court of Kublai Khan. His account of this travel is preserved in more than 140 manuscripts which were produced in the time span of two centuries. The original version composed by the two prison companions is unfortunately lost. The text was translated twice into German, in the Middle Ages. While one of these translations (version VG3) was already edited by Tschärner in 1935, version DI is being edited for the first time in this project, which serves as a frame for the present study.

Under the name *ENHG Marco Polo* we therefore refer to an edition of three main witnesses of version DI of the *Devisement dou Monde*, all written in Early New High German: a manuscript (München, BSB Cgm 696), an incunabulum (München, BSB-Ink P-671) and a short fragment (München, BSB Cgm 252). These three witnesses are considered the most important from a stemmatic point of view (Cugliana, 2019) and all date back to the second half of the 15th century.

The output of the ENHG Marco Polo project is not a critical edition which aims at presenting a reconstructed text, but rather a collection of editions, in which each witness is edited separately: for each witness, three different editions will be produced, one for each of three targeted levels of renditions: diplomatic, semi-interpretative, and interpretative. Because each witness is treated as a separate entity, no critical apparatus in the traditional sense will be produced, and in its stead, the edited texts will be accompanied by a set of commentaries³. The commentaries will, in part, be shared by all editions (for example the commentaries on the named entities, which are central to Marco Polo research), and, in part, be edition-specific (for example, the edition of the incunabulum will contain a commentary on the style of the writing typical of the first printed texts, while the edition of the fragment will contain a description of the conservative traits of the language it is written in⁴).

Our use of the term “edition” matches that of the Italian school of textual

³ These commentaries are not a critical apparatus in the traditional sense (see Macé and Roelli, 2015, entry: Apparatus) as no critical text was established for this edition: the reasons of this choice depend on the history of transmission of the text and are illustrated in Cugliana (2019).

⁴ The conservative traits of the language in which these works are written have already been noted by Benedetto (1928), one of the most important scholars in the field.

scholarship⁵, according to which the term edition refers to a specific rendition of the text (diplomatic, interpretative, etc.). This use of the term “edition” is also conform to the definition of scholarly edition provided by Sahle (2016:23): the act of rendering the text in, say, an interpretative way, is namely “erschließend” in its nature, as some specific aspects of the text are presented there in such a way that the reader is granted more informed access to them.

The edition will be published online using a specifically tailored version of EVT (Edition Visualization Technology⁶) and will present, on the one hand, each witness in its continuum from facsimile to multiple levels of normalisation and, on the other hand, the three main witnesses in synopsis. From each module of the edition and from each of the texts composing the editorial project it will be possible to access a twofold commentary, that is, specific notes referring to the named entities and the *realia* appearing in the text and philological notes referring either to all of the three witnesses or to one witness in particular.

The language in which the three witnesses are written is Early New High German. ENHG represents a quite problematic phase of the German language, already as far as the definition of its time span is concerned. As a matter of fact, there have been different proposals for a definition of its periodisation (see for instance Grimm 1822, Scherer 1878 and Wells 1990), but what is relevant here is that ENHG spans over a long time, going through Gutenberg’s media revolution, which had a huge impact on the written German language (Bosco Coletsos, 2007: 187-191). This stage of the German language is therefore extremely diverse, which has quite a few consequences for the scholarly editor approaching the texts written, or printed, in the period.

2.2 Normalisation

“Normalisation takes place when the orthography (or other aspects of the language) of a text is changed in order to make it correspond to a certain standard” (Macé and Roelli, 2015, entry: Normalisation).

The need for normalisation in scholarly editions arises from the fact that pre-modern texts present a high degree of variation when it comes to orthography and punctuation.⁷ This clashes with the desire of many scholarly editors to produce readable texts where orthographic, morphological and dialectal variance have been regularised and levelled off.

This desire has had different motivations in the history of editorial practice. When philology was still considered an ancillary discipline, the role of the editors was to prepare the texts for historians and literary scholars, who were mainly interested in the contents. So for instance diacritics and other linguistic peculiarities were considered on a par with “ein verdammtes Unkraut” (*a darn*

⁵ See for instance Buzzoni (2016), and the bibliography cited there, for an introduction to the Italian school of thought, and the works by other philologists such as Gianfranco Contini, Alfredo Stussi, Domenico de Robertis, Giorgio Pasquali, Raul Mordenti and Paolo Trovato, to mention a few.

⁶ Rosselli Del Turco, Roberto, et al. Edition Visualization Technology: EVT is “A light-weight, open source tool specifically designed to create digital editions from XML-encoded texts” <http://evt.labcd.unipi.it>.

⁷ On the variation characterising pre-modern texts see Contini (2017†:45) and Pierazzo (2016:41). As for punctuation, the following manuals are used as a reference in the present project: Parkes (2016), Garavelli (2008), Rinas (2017) and Kirchhoff (2017).

weed) in a tidy flower garden, as Julius Weizsäcker wrote in 1867⁸.

With Lachmann,⁹ normalisation became necessary for the philologists themselves: needing to summarise a whole textual tradition in an apparatus at the bottom of the page, it was no longer feasible to take into account all variants and punctuation. In particular, collation implies normalisation: the establishment of a traditional stemma is based on Leitfehler, not on equivalent variants.¹⁰ The formal appearance of the reconstructed text (e.g. the dialectal variety chosen for the text) would then be given the shape considered to be the closest to the language of the author.¹¹ Generally, however, it is commonly assumed that editors of medieval texts are free to decide when, how and how much their texts should be normalised.¹²

For many languages the exact procedures used to normalise a text have been established a posteriori by historical linguists. For instance, normalised Middle High German was first established by Karl Lachmann and then accepted by the scholarly community.

However, no such normalisation standards currently exist for Early New High German. As Erbert et al. (1993: 7) write: “Until the 16th c. and for many aspects also later, there is no variety of Early New High German that [...] had a generally accepted and undisputable prestige over all the other and which can, on the basis of such sociolinguistic fact, be considered as a reference variety for the synchronic writing and speaking or for the diachronic use”.¹³

A survey of existing editions of ENHG texts shows that, while internal coherence in normalisation practices is desired and documented, not only normalisation standards are missing, they are also not actively sought upon. Let’s take, for instance, the *Ingelheimer Haderbücher* project: although the editors were extremely accurate in their outline of their principles for the transcription of the texts, they did not provide any specific reference in support of their decisions on how the texts were transcribed and normalised. In fact, they only tangentially address this concern by writing that “for the orthography of the single letters and the creation of the transcriptions we mostly welcomed and observed the suggestions of germanists, linguists and paleographers”.¹⁴

⁸ Quote from Sahle (2013, vol.1: 71)

⁹ For a systematisation of Lachmann’s method see Maas (1950).

¹⁰ But things are different in digital stemmatology: see in this respect Andrews (2020).

¹¹ After a long discussion on the importance of formal variance and on the overlapping between form and content, Contini (2017†:50) gives concrete indications on how formal variance should be dealt with in a critical edition: “a parità di condizioni, si adotta costantemente la forma di un testimone, scelto (ma per di solito apoditticamente) per ragioni o di antichità o di congruenza regionale o di sorvegliata organicità. [...] Non ci si sottrae all’impressione che la forma passi in seconda linea innanzi alla sostanza” (“all things being equal, one constantly adopts the form of one witness, chosen (but usually apodictically) for reasons either of antiquity, or regional consistency, or surveilled organicity. [...] One does not escape the impression that the form is in the second place with respect to the substance”).

¹² “The extent of normalisation, as well as the rules followed by the editors, depends on their judgement and the methods they adopt.” (Buzzoni, 2020: 140)

¹³ “Es gibt bis ins 16. Jahrhundert und unter vielen Aspekten selbst darüber hinaus keine Varietät des Frühneuhochdeutschen, die [...] ein zeitgenössisch allgemein anerkanntes, unbestreitbar über alle anderen Varietäten hinausragendes Prestige gehabt hätte und der man aufgrund eines solchen sprachsoziologischen Faktums die Funktion einer Leitvariante für synchrones Schreiben und Sprechen und für die Diachronie zuschreiben könnte”. This horizontal variation is in fact “eines der markantesten Kennzeichen des Frühneuhochdeutschen”.

¹⁴ “Bei der Schreibung der einzelnen Buchstaben wurden vor allem Hinweise von Germanisten, Sprachforschern und Paläographen aufgenommen und bei der Erstellung der Transkription berücksichtigt”. <https://www.haderbuecher.de/projekt/editionsrichtlinien/>

An interesting case is the *Frühneuhochdeutsches Wörterbuch*, a dictionary of ENHG, started in the 1980s and expected to be completed in 2027.¹⁵ The editors of the FWB did propose a set of rules, a necessary precondition for such a large-scale project. The principle at the basis of the orthography chosen for the lemmata in the dictionary is a phonological one: after having established an ideal phonological system of ENHG, the corresponding graphemes were chosen. The editors themselves, however, state very clearly that such a systematisation represents only an ideal (in their words “ausgezeichnet”) language.¹⁶ The reason why we are not using the FWB as a reference for our normalisation is, firstly, because this ideal language is very distant from the actual language and orthography we find in our witnesses and, secondly, because the dictionary is still not complete (many letters are completely missing, and many are incomplete: only six letters of the alphabet are fully represented).

2.2.1 Normalisation choices are editorial choices

For a scholar editing a medieval text, “to normalise” means making a series of editorial decisions.¹⁷ We can provide some examples:

- how to substitute original textual features with their modern correspondents (e.g. if and when to substitute the “long s” (ſ) with the modern “s”);
- how to change certain signs into others (e.g. when to turn a virgula into a comma, when to turn it into a full stop or when to eliminate it completely);
- when to omit features that are today considered unnecessary (such as decorations);
- when to add new features that were not yet fully developed in medieval times (like capitalisation rules or spaces between different words or sections of text).

To an external observer, the act of normalising a text can be seen as a sort of correction being imparted on the text. Scholars, in this respect, have often distinguished between *normalisation* and *correction*, between mechanical and less mechanical interventions: the former, normalisation, being seen as an almost objective process, while the latter, correction often considered as a subjective act that involves the editor’s *iudicium*.¹⁸

In our view there is no strict dichotomy between normalisation and correction, but rather a continuum. No normalisation is entirely objective, no correction is entirely subjective. For example, one can argue that substituting “v” with “u” and vice versa is quite unproblematic and rather objective because

¹⁵ <https://fwb-online.de/>

¹⁶ In their words, the lemmata appear in “einer ausgezeichneten Wortbildungsform in ausgezeichneter Lautung und ausgezeichneter Graphie. Diese Form muß alle Vorkommensformen in einem bestimmten Sinne repräsentieren.” <https://fwb-online.de/einleitung/positionen-des-woerterbuchartikels-i-das-lemma>

¹⁷ For a guide on how to normalise medieval texts see Stussi (1988: 161) and, for a concrete example of normalisation see Renzi and Andreose (2009: 272-282)

¹⁸ See for example the difference between substantial and formal variants and between operations *ope codicum* and *ope ingenii* (Stussi, 1988; Renzi and Andreose, 2009).

it follows an established rule of the art in a certain school of philology. However, in our opinion, choosing which set of rules to follow is also a subjective choice of the scholar that cannot, and should not, be seen as independent of their work. In addition, we argue that also in the simplest cases the philologist runs the risk of concealing linguistic features of great importance, taking them for scribal negligence.

In our opinion no intervention pertaining to the normalisation process can be considered objective or mechanical. The lack of clear line of distinction suggests that as much information as feasibly possible should be preserved during the preparation of a scholarly edition.

Preserving information has, however, a cost: the editor must put in the effort to record every single correction, even the most minute. The method we present in Section 3 aims at making it possible to preserve a vast amount of information with little effort from the side of the editor.

2.2.2 Normalisation in TEI

In TEI most of these normalisation actions are represented using `<choice>` elements. For example, the expansion of “p” into “per” would be encoded as:

```
<w>
  <choice>
    <abbr>p</abbr>
    <expansion>per</expansion>
  </choice>
</w>
```

It is also possible to use `<choice>` to encode more than one normalisation form for the same word, for example, using the ad-hoc elements described by the Menota handbook:¹⁹

```
<w>
  <choice>
    <mc:facsimile>wan</mc:facsimile>
    <mc:diplomatic>warn</mc:diplomatic>
    <mc:normalization>waren</mc:normalization>
  </choice>
</w>
```

There are many TEI elements that are routinely used in combination with `<choice>` to express different kinds of normalisation intents: for instance `<abbr>`, `<expansion>`, `<sic>`, `<corr>`, `<orig>`, `<reg>`, `<supplied>` and, for punctuation `<pc>`.

The use of `<choice>` has two main drawbacks: first, it causes a 20x expansion in character-count (a five-letter word with two normalisations requires more than 100 characters to be encoded), and, second, it forces what would be a single `<w>` element to grow into a 8-tag, 2-level deep markup structure. When each word in a manuscript needs such a treatment (as it is the case in medieval texts), these two drawbacks quickly lead to the creation of a TEI file that is hard to navigate and to maintain.

2.3 Punctuation

Punctuation is one of the many aspects of a text that undergoes normalisation during the creation of a scholarly edition. In most cases normalising punctu-

¹⁹ https://www.menota.org/HB3_ch4.xml

ation is considered a secondary task and often very quickly brushed off in the introduction of many editions.²⁰

In our opinion, in contrast, the normalisation of punctuation is a critical step in the creation of any non-diplomatic edition. Punctuation is not only a reading aid: it delivers meaning, suggests interpretations, creates structures, changes the importance of words. It is the responsibility of the editor of a non-diplomatic edition to translate ancient punctuation systems into something that a modern reader can fully understand.

From an operational point of view, the normalisation of punctuation marks is profoundly different from the normalisation of words, although the two are often treated in the same way.

The normalisation of words is usually quite simple and has only minimal impact on the markup structure (when done without preserving information, as described in the previous section). In the most common case, some letters in a word are changed requiring the replacement of that word. The more convoluted cases require splitting or merging words, but those still have at most linear effects on the markup structure (for example turning `<w>aberes</w>` into `<w>aber</w><w>es</w>`²¹).

In contrast, the normalisation of punctuation is often more complex, leads to modifications in the hierarchical structure of the markup, and even has a ripple effect on the surrounding text. For example, normalising the middle dot in `<s>[...] geben von stund er gesund wirt [...]</s>` into a full stop will not only change the markup structure, but also require the capitalisation of the following word, now the first word of the following sentence: `<s>[...] geben.</s><s>Von stund er gesund wirt [...]</s>`²².

The additional complexity of normalising punctuation is especially prominent when dealing with medieval texts. As a matter of fact, punctuation in the Middle Ages responded to needs that were completely different from what we are used to in modern times, and different rules were applied based on many factors: the *scriptorium* where the manuscripts were produced, the single scribe, the language, the destination of the text and the local customs.

This is the case in medieval German, where punctuation was usually not driven by syntactic principles, but instead often had a rhetoric function: it was meant as an aid to those reading the text aloud. (Digilio 2008:373) For instance, an important element in the clause could be put between two commas to signal that it was to be read louder; today this would have the opposite result. Take for example, “Mary, his sister, is a doctor”: if we were to follow some medieval punctuation conventions, like the ones often found in the witnesses of the edition

²⁰ An example of the traditional way to address punctuation is the edition of *Erek* by Hammer et al. (2017) which is based on manuscript A (the authors state that “in order to account for the dominance of A in the context of the textual transmission [...], we put this text in the focus of our edition and we represent it in the form (except for the interventions noted below) in which it came to us” [our translation]) p. XXXV). In the introduction to the edition, only six lines are devoted to punctuation: there, it is stated that “[w]e insert punctuation according to modern rules and to our understanding of the text. Generously at the end of the verse, much more careful inside the verse” [our translation, p. XXXVIII]. The original reads “[i]nterpunktion setzen wir nach modernen Regeln und entsprechend unserem Verständnis des Textes. Am Verseende großzügig, im Versinneren deutlich vorsichtiger”. This is no criticism of the edition itself, which is extremely valuable, but a matter-of-fact observation capturing the common way of dealing with punctuation in state-of-the-art editions.

²¹ From München, BSB Cgm 696, f. 145r.

²² From München, BSB Cgm 696, f. 265r.

of the ENHG Marco Polo, “his sister” would be the focus of the sentence, while according to the modern use of commas, the same string would be of secondary importance. An example of this is the following sentence, from the rubric of chapter 3 in München, BSB Cgm 696: “die zwen prüder [...] darnach chomen, zu dem groffen hern, der ganczen tartarey”.²³ Despite being the pragmatic focus of the sentence, “zu dem groffen hern” is here between commas: the commas here are therefore clearly emphasising the string in question. Sometimes, moreover, modern punctuation imposed on medieval texts can obscure important linguistic phenomena: in the case of the ENHG Marco Polo, this regards the presence or absence of structures similar to the so-called relative nexus in Latin and with the syntactic parameter of the null subject.

Unfortunately, things are even more complex than that; the sentence we have just analysed proceeds in fact as follows: “die zwen prüder [...] darnach chomen, zu dem groffen hern, der ganczen tartarey, gnant 8 grofe cham, kayfer von Chatay”: the second part of the sentence, which translates “called the Great Kaan, emperor of Chatay”, would seem to reflect instead a more modern use of commas.

The underlying truth is that the systems regulating punctuation in medieval texts are not yet fully understood. This is true for the German Middle Ages in general, as Digilio (2008: 373) observes, and for ENHG in particular. In this respect, Ebert et al. (1993:29) writes that “how far it is possible to describe rules and how much rulelessness reigns here, remains to be determined”²⁴. The very meaning of the signs is often ambiguous: *punctus* and *virgula*, the two signs that appear in the witnesses at the basis of the edition, are, particularly in the 14th and 15th century, “polyfunctional” (Ebert et al. 1993:29).

This lack of systematic rules and the high degree of ambiguity makes it hard to understand which set of rules was being used in the first place and, in turn, it makes it almost impossible to transpose the punctuation system used by a manuscript into one of the many currently used, in a fully mechanical way. Contini supports our observations when he writes that “the frictions due to the change of system are particularly visible in the case of punctuation [that in old texts mixes semantic and melodic functions], so that it is only rarely possible to insert or omit one and the same sign in the same place” (Contini: 2017†:23). So instead of a translation from one system into another (which is the definition of normalisation given by the scholar), in the case of punctuation one would have to substitute the system *in toto*.

We are not the first to notice that normalising punctuation differs profoundly from normalising words. In 2014, for instance, Elena Pierazzo wrote in a discussion on the TEI mailing list on the subject of normalisation of punctuation²⁵:

The point is that punctuation is not like spelling and orthography. For the latter, you have words and sequences of characters that evolve from one accepted format to another accepted format. [...] However, punctuation does not work like this: you cannot trace the evolution of a medieval or early modern comma into, say, a semi-

²³ “The two brothers [...] then come, to the great lord, of whole Tatarstan”.

²⁴ “Inwieweit hier Regeln beschreibbar sind und inwieweit hier Regellosigkeit herrscht, bleibt noch festzustellen”

²⁵ <http://tei-1.970651.n3.nabble.com/Original-and-editorial-punctuation-td4025791.html#a4025812>

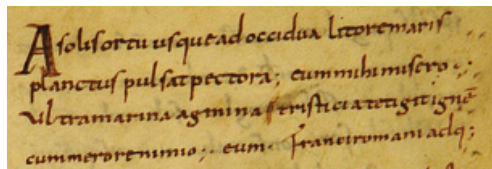
colon, because they are completely different systems that are practically unrelated to each other. [...] So when editing a medieval/early modern text, **if providing a modern punctuation is a requirement of the project, then first you strip the text from all the punctuation and then you add a completely new one** (emphasis added).

2.3.1 Possible approaches

Among the different attempts on the part of scholarly editors to address this issue, we can identify three main tendencies:

1. a complete removal of the original punctuation and replacement with modern one;
2. the decision to maintain punctuation exactly as it was in the original;
3. an endeavour to normalise punctuation, while remaining as close as possible to the source.

An example of the first approach, removal and replacement of the original punctuation, is Stella's *Corpus Rhythmorum Musicum*²⁶, a multimodal edition containing the earliest medieval Latin songs, together with their musical rendition. Figure 1 shows an excerpt from the edition alongside with one of the witnesses.



- 1 A solis ortu usque ad occidua
littora maris planctus pulsat pectora.
Heu mihi misero!
- 2 Ultra marina agmina tristitia
tetigit ingens cum merore nimio.
Heu mihi misero!
- 3 Franci, Romani atque cuncti creduli
luctu punguntur et magna molestia.
Heu mihi misero!

Fig. 1: Left: Manuscript of song “A solis ortu usque ad occidua”.²⁷ Right: Edition of the song from the *Corpus Rhythmorum Musicum*.²⁸ The medieval punctuation present in the original manuscript has been removed in toto and replaced by modern punctuation in the edited text.

An example of the second method, faithful preservation of the original punctuation in the edited text, can be found in Stolz's Parzival Projekt²⁹, depicted in Figure 2. Here the transcription of the witness perfectly resembles the original punctuation: the medieval dot is encoded using a modern full point, fully comparable from the point of view of its visual rendering.

²⁶ <http://www.corimu.unisi.it/public/frontend>

²⁷ Verona, Biblioteca Capitolare, XC (85)

²⁸ <http://www.corimu.unisi.it/public/previewedizione/testo/idschede/1>

²⁹ <http://www.parzival.unibe.ch/cod857/Daten/materialien.html>



Fig. 2: Figure 2. Extract St. Gallen, Stiftsbibliothek, Cod. 857 from the edition of MS D in the Parzival project. The edited text maintains the original punctuation.

An example of the third strategy, normalising punctuation while trying to keep it close to the original, is the *Vercelli Book* by Roberto Rosselli Del Turco. The Vercelli Book has two editions: a diplomatic and an interpretative one. While the diplomatic edition preserves the original punctuation, the interpretative edition presents a simplification of the original punctuation. This simplification replaces most archaic signs with full stops, but still respects most of the features of the original punctuation, in particular its position. This often leads to unnatural looking edited texts, such as that shown in Figure 3. In this example, the dot preceding the string “achine” is maintained in its original position, but the two words composing the string “ac” and “hine” are separated. Having a full stop still adjacent to the first letter of a word, though, is still quite unfamiliar for modern readers, who might have difficulties parsing the text.

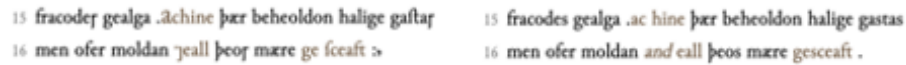


Fig. 3: Figure 3. Extract from the Digital Vercelli Book. Some punctuation is simplified “in place”, leading to unfortunate hard-to-read artefacts that do not match current punctuation conventions (e.g., “.ac hine”).

If, on the one hand, we appreciate the attempt to try and normalise the original punctuation, on the other hand, this approach does not seem to improve the readability of the text, as one expects from an interpretative edition. Moreover, it does not seem to help much in the analysis of the original punctuation either. However, it could represent one step in a potential spectrum of normalisations, showing a first reduction of complexity with respect to the source.

2.3.2 Punctuation in the ENHG Marco Polo

In our project, punctuation will be handled in three different ways: one for each level of faithfulness: diplomatic, semi-diplomatic and interpretative.

At the diplomatic level punctuation will be fully preserved. Each punctuation mark will be encoded in its original position and using the nearest Unicode code-point. Extracting precise information about the punctuation marks from the facsimile and encoding it in the TEI files will also have the positive side-effect of unlocking important analyses in the fields of stylometry and phylogenetics (Darmon et al., 2020).

At the semi-diplomatic level a reduced system of punctuation consisting of two signs, a middle dot (·) and a virgula (|) will be introduced. Similar

systems are often attested in the late Middle Ages (Parkes, 2016:46) and are often adopted in ENHG texts of the 14th and 15th centuries (Hartweg and Wegera, 2005:131). This reduced system based on unfamiliar signs will allow a smooth presentation of the structure of the text, and, at the same time, reduce the intervention of the author on the interpretation of the textual content. A similar approach was used by Mitchell and Robinson (1998), as discussed in the next section. Also in our case, the insertion of this minimal system of punctuation follows the motto “no punctuation where the sense is clear without any” (Mitchell and Robinson, 1998: 313).

At the interpretative level the punctuation will be completely modernised, allowing the general public to easily read the text.

2.4 Multiple editions

Notwithstanding the issues outlined in the previous section, the project behind the Marco Polo edition strives to match and surpass the state of the art for digital editions.

One particular aim of the ENHG Marco Polo is to be able to programmatically produce multiple editions, each with different notes and normalised texts, from a single master TEI file. In practice, this means:

- The witnesses will be encoded using TEI-compliant XML markup.
- The scholar will work by directly editing the TEI/XML master file.
- For each witness, there will be a single master file containing all the related information and critical annotations (including the references for the editor’s notes).
- The same text will be edited in multiple levels of faithfulness/readability: diplomatic (with special characters to show allographic variance, colours, abbreviations, errors etc.), semi-diplomatic (with expanded abbreviations, but still maintaining some relevant graphematic distinctions, a reduced punctuation system and the correction of trivial mistakes) and interpretative (with the correction of all mistakes and the modernisation of graphemes, punctuation and other textual features). Given that the project is based on 3 main witnesses, this means producing 9 editions.
- Multiple editions will be generated automatically from the master, with no manual intervention on the produced files.
- The generated editions files will conform to the TEI subset understood by EVT.

Some of these desiderata clash with each other, like for example, directly editing the XML file and keeping all the information needed to generate the three editions in one master file.

2.4.1 Pre-digital approaches

In pre-digital times, presenting the same text at different levels of rendition would of course mean a lot of work. The philologist, after completing the

diplomatic transcription, would start all over again and proceed with another transcription and repeat the process for each level of normalisation they want to have. Not only is this a very long and demanding workflow, but it also poses some problems: first of all, if an uncertain passage of the text gets clarified later on, the changes need to be made on all the different editions separately. Moreover, inconsistencies can emerge between different edition levels, as they are created independently from each other.

An interesting example, in this respect, are Mitchell and Robinson (1998): in their edition of *Beowulf*, they decided to use modern punctuation. The text, in fact, needed to be accessible to a wide audience. At the same time, however, they expressed their reservations on the matter: modern punctuation would in fact not really suit Old English poetry and it would force the editors to make decisions in order to solve ambiguities, which cannot be solved. Given these considerations, extensively explained in section IIIB of their edition, they wanted to offer a way out, or at least an alternative. So they proposed a different system of punctuation, which actually inspired our intermediate level of normalisation. However, they limited themselves to offering just a taste of their solution in the appendix to the edition (pages 313-318). They edited verses 1 to 114 again, using a different punctuation system and inviting the reader to go back to the beginning of the full edition, where the first three folios of the manuscripts are reproduced in black and white (pages 44-48). Such an approach has many limitations: it is highly impractical, both from the point of view of the reader, who is forced to go back and forth from the beginning to the end of the book and who, in the end, only gets a small portion of the “best” version of the text, and from the point of view of the scholar, who cannot really realise his or her ideas to the fullest, as that would probably cost too much time and space.

2.4.2 Digital approaches

One of the main advantages of digital transcriptions of manuscripts is that they can be reprocessed and transformed quite easily, at least in theory. This lends itself to the idea of having a single container that contains both a diplomatic transcription as well as hints on how to normalise certain pieces of texts.

For example, it is quite common for philologists that prepare their editions using the TEI to make use of elements such as `<choice>`, `<orig>`, and `<reg>` to indicate how to normalise certain words, or `<sic>` and `<corr>` to encode corrections. A software tool will then process these hints and, for example, allow the reader to choose between a version that has been created on the fly by selecting the text marked with `<orig>/<sic>` and another version created by selecting the text in the `<reg>/<corr>` elements.

A digital edition exemplifying this approach is the aforementioned Digital Vercelli Book, by Roberto Rosselli Del Turco: here two levels of edition are offered, a diplomatic and a more interpretative one. The user can compare the two editions visualising them synoptically in the EVT software used for the edition.

3 Our approach

To address the challenges and the desiderata of the ENHG Marco Polo edition (single master file, directly editable, multiple output editions), we developed and implemented a novel approach where normalisation is dealt with using formal rules, explicit exceptions, and automated pipelines to generate ephemeral edition files. This approach is inspired by similar initiatives in the fields of collation (Barabucci and Fischer 2017, van Zundert 2019), textual edition (Barabucci et al 2017), legal analysis (Barabucci et al 2011) and curation of research data (Barabucci 2018).

This methodology is characterised by the following key aspects:

- the master TEI file contains the transcriptions;
- general rules are encoded using a formal language, with a scheme such as “if CONDITION then TRANSFORMATION”;
- rules are stored outside the master TEI file;
- rules are applied to structures, not just to pieces of text;
- the few exceptions (such as manual corrections or cases where a rule should not be applied) are encoded in the master TEI file;
- the edition files, despite being the main concrete output of the editorial project, are ephemeral and never modified directly.

The implementation consists of a series of XSLT transformations, each representing and implementing a single rule, coordinated by three different XProc pipelines, one for each level of edition. The source code is available at <https://marcopolo.cceh.uni-koeln.de/>.

This methodology contrasts with the established editorial practice of mingling transcription, normalisation and critical amendments. Instead of just performing the desired normalisation steps while transcribing and keeping track of them in an introduction to the edition, here the rules are separate artefacts, related but independent from the edited text. In fact, as a nice byproduct of this methodology, the ENHG Marco Polo produced a set of rules for the normalisation of words and punctuation in Early New High German. The scholar community has now the opportunity to evaluate, discuss, critique, modify, and, if desired, adopt this set of rules independently from the edited text.

Another important distinction is the epistemological status of the generated edition files. The three edition files are the main scholarly output of the editorial project, but they are also ephemeral files that can be thrown away and reproduced in their exact form at any time. This is possible due to the fact that all the knowledge needed to create the edition files (both the scholarly and the technical knowledge) is stored either in the master TEI file or in one of the pipelines. No manual “drive-by” modification is performed at any time. With this framework of operation in place, it even becomes possible to reconstruct the chain of reasoning and transformations that lead to a certain result in a specific part of the edition file, giving critical editing the same kind of transparency and mechanical reproducibility common to other sciences.

3.1 Rules

Rules are the way in which the scholar formalises the normalisation process. The kind of rules used in the project are of the form “if CONDITION then TRANSFORMATION”. Examples of rules implemented in ENHG Marco Polo edition are:

- orthographic corrections after changes in punctuation (e.g., if a virgula has been normalised into a full stop, then the first letter of the following word must be capitalised);
- methods to join words split at the end of a line (e.g., if a token is at the end of a line and ends with one of the characters used to denote a joining, then merge this token with the following token).
- replacement rules for archaic letters (e.g., replace all long s with lowercase s);
- expansion rules for abbreviations (e.g., replace all occurrences of “vñ” with “vnd”).

The first two examples demonstrate why one should not think of these rules as simple textual substitutions. As described in previous sections, there is a deep connection between punctuation and the structure of the text. Changing punctuation can, and often does, change the structure of the text. At the same time, it is fundamental to represent such structure in a machine-readable way. For example, without a proper representation of the concept of “sentence” it would not be possible to write rules such as “if a word appears at the beginning of a sentence...”. What follows is that there is a continual interaction between the rules and the structure of the edited text. Rules, and in particular punctuation rules, often modify the structure of the text. A technical system that does not allow for this interaction to happen is not able to deal in a proper way with normalisation in general and punctuation in particular.

Each rule is implemented as a small and self-contained XSLT transformation. At the time of writing, the ENHG Marco Polo project comprises about a hundred rules, grouped in twenty macro categories. On average, the core of each rule is implemented in less than 3 lines of XSLT.

To give the readers an impression of the simplicity of the implementation of the rules, we show here the main parts of the XSLT that implement one of the example rules described above.

3.1.1 Example: Rule to join words split at the end of a line

In ENHG, a punctuation sign that we nowadays call a double oblique hyphen was used to mark that a word has been split at the end of a line. In the diplomatic rendition we want to preserve this word division and the forced line-break, while in other renditions we want to reconstruct the complete word.

The XSLT excerpt in Figure 4 shows how split words are joined when a double oblique hyphen is found. The joining is performed in a lossless way: all information present in the original witness is preserved. This is possible thanks to the fact that this step operates on a *over-textual* structure that contains information about the structure of the text and the positioning of a token (e.g.

@eol='true'). It must be stressed that this last piece of information, and the over-textual structure in general, are not part of the master TEI file and have been added by preceding rules and steps. It should also be noted that this rule does not delete any text, it just marks that two words have been joined and what the result of this operation is. Another rule will take care of modifying the structure and yet another rule will take care of removing the now superfluous partial word in the next line before creating the final edition file. However, while removing and carrying out all these changes, comments about what is being done will be added, as an additional aid to the readers of the final edition file.

```
<xsl:template
  match="mp:TOKEN[@eol='true']
    [ends-with(., $char-double-oblique-hyphen)]">

  <!-- First, the two parts of the split word are identified. -->
  <xsl:variable name="part1-raw" select="string(.)"/>
  <xsl:variable name="part1"
    select="substring-before($part1-raw,
      $char-double-oblique-hyphen)"/>
  <xsl:variable name="part2" select="string(following::p:TOKEN[1])"/>

  <!-- Then the joined word is computed. -->
  <xsl:variable name="joined-parts">
    <xsl:value-of select="$part1"/>
    <xsl:value-of select="$part2"/>
  </xsl:variable>

  <!-- Finally, the content of the original element is replaced by
    the joined word, and the source pieces stored and
    carried along in extra attributes. -->
  <xsl:copy>
    <xsl:apply-templates select="@*"/>

    <xsl:attribute name="joined" select="true()"/>
    <xsl:attribute name="part1" select="$part1-raw"/>
    <xsl:attribute name="part2" select="$part2"/>

    <xsl:value-of select="$joined-parts"/>
  </xsl:copy>
</xsl:template>
```

Fig. 4: XSLT implementation of the rule “Join words separated by a double oblique hyphen”.

The rule in Figure 4 is independent from other rules in the pipeline. The scholar is free to use this rule in the generation of a specific edition or to leave it out. In the case of the Marco Polo project, this rule is employed in the pipelines that produce the semi-diplomatic and the interpretative editions, but not in the diplomatic edition.

Because it is independent from the other rules used in the pipeline, other scholars can reuse this rule in their projects without being forced to adopt the Marco Polo pipelines, and all their rules, in their entirety.

3.2 One-off cases, exceptions and manual corrections

Before a rule can be written, a textual phenomenon must first be recognised as such and then a pattern must be identified. At the beginning of the editorial process, when the textual materials are not yet well known and the normalisation rules are not yet clear, it is simpler and more productive to address one-off textual phenomena in a manual way. This is done by annotating the word to be normalised or the sign to be changed in the master TEI file. Once a pattern is recognised, a rule can be written and these annotations removed from the master TEI file.

Similarly, there are cases where the conditions of a rule would normally apply, but where the scholar does not want the transformation to happen. Rather than complicating the condition in the rule to exclude this specific case, it makes sense to mark this specific occurrence as an exception in the master TEI file.

Finally, the scholars may want to correct an obvious mistake, while keeping track of the original reading. This also is done not by writing a rule, but by placing an annotation in the master TEI file.

All these issues (one-off normalisation, exceptions and manual corrections) are encoded using TEI elements directly in the master TEI file.

One-off normalisations are marked in the master TEI file using one of the non-standard attributes defined by the ENHG ODD. For example attaching `@mp:n1-subst='foo'` to any element will force the substitution of the content of that element with the word “foo”, but only at the normalisation level denoted by N1 (“semi-interpretative”). Such exceptions are marked using `<w>` or `<pc>` elements, together with the already cited project-specific `@mp:nX-subst` attributes, as shown in Figure 5.

```
<pc mp:n1-subst="|" mp:n2-subst=";">,</pc>
```

Fig. 5: Substitution of a comma with a virgula in the semi-diplomatic edition (n1) and with a semicolon in the interpretative edition (n2). This exception is marked in the master TEI file.

Manual corrections are instead encoded using a combination of `<choice>`, `<orig>` and `<reg>` elements, together with the `@mp:use-in-nX` attributes, as illustrated in Figure 6.

```
<persName ref="#marcoPolo">
  <choice>
    <orig>Macho</orig>
    <reg mp:use-in-n1="true">Marcho</reg>
    <reg mp:use-in-n2="true">Marco</reg>
  </choice>
</persName>
```

Fig. 6: Correction of the typo in the name “Macho”, which will appear as “Marcho” in the semi-diplomatic edition and as “Marco” in the interpretative edition, the latter rendition presenting also a normalisation of the “ch”-cluster. This one-off manual correction is marked in the master TEI file.

The ability to quickly deal with occasional exceptions without having to resort to writing a new rule or making the existing ones more complicated greatly simplifies the day-to-day job of the scholar. It also allows for a cleaner separation between rules (theoretically shared by a community) and exceptions (fruit of the editor’s own decisions).

3.3 Pipelines and normalisation levels

Pipelines are used by the scholar to indicate which rules should be applied, and in which order, in order to turn the master TEI file into a complete edition at a specific level of normalisation. In the ENHG Marco Polo project there are three pipelines, one for each target faithfulness and readability level. Achieving a specific level of faithfulness and readability is done by carefully choosing which rules to apply from the catalogue of available rules. Some steps are shared by all pipelines, while others are specific to certain faithfulness levels. Figure 7 illustrates this approach.

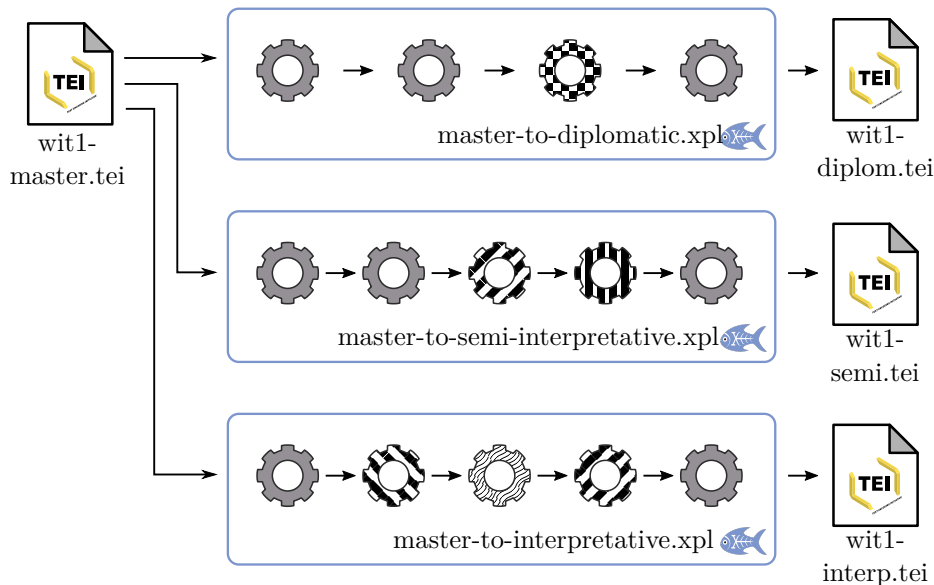


Fig. 7: High-level representation of the pipeline-based workflow. The gray cogs indicate steps shared by all pipelines. The cogs with patterns identify level-specific steps.

Each pipeline is implemented as a XProc pipeline. All the pipelines are simple linear flows (i.e., the output of a rule is the input for the next rule). From a methodological point of view, the XProc pipeline is a record of all the operations that the scholar performs on the transcription. The creation of an edition is equivalent to playing this record again. Figure 8 shows an excerpt of the XProc pipeline used to generate the semi-diplomatic edition.

It is important to note that pipelines comprise three kinds of steps:

1. infrastructural steps, e.g. the `tokenize` step that creates the textual superstructure to which all other steps will refer to;

2. rule-based steps, e.g. the ‘adjust-capitalisation’ step that implements the set of rules to fix the capitalisation of words after changes in punctuation; and
3. exception-related steps, e.g. the apply-punct-n1-subst step that handles all the exceptions to the punctuation rules that are relevant for the N1 level (i.e. semi-diplomatic).

```

<p:declare-step xmlns:p="http://www.w3.org/ns/xproc" version="1.0">
  <p:input port="source">
    <p:document href="marco-polo-master.tei.xml"/>
  </p:input>

  <p:xslt><p:input port="stylesheet">
    <p:document href="tokenize.xsl"/>
  </p:input></p:xslt>
  [...]
  <p:xslt><p:input port="stylesheet">
    <p:document href="remove-line-breaks.xsl"/>
  </p:input></p:xslt>
  [...]
  <p:xslt><p:input port="stylesheet">
    <p:document href="normalize-letters-n1.xsl"/>
  </p:input></p:xslt>
  <p:xslt><p:input port="stylesheet">
    <p:document href="normalize-letters-common.xsl"/>
  </p:input></p:xslt>
  <p:xslt><p:input port="stylesheet">
    <p:document href="apply-punct-n1-subst.xsl"/>
  </p:input></p:xslt>
  <p:xslt><p:input port="stylesheet">
    <p:document href="adjust-capitalization.xsl"/>
  </p:input></p:xslt>
  [...]
  <p:xslt><p:input port="stylesheet">
    <p:document href="untokenize.xsl"/>
  </p:input></p:xslt>

  <p:store>
    <p:with-option name="href" select="'marco-polo-n1.tei.xml'"/>
  </p:store>
</p:declare-step>

```

Fig. 8: Excerpt of the XProc pipeline used to generate the semi-diplomatic edition. Steps marked with ④ are steps that implement rules; those marked with ⑤ is a step that takes care of exceptions.

The fact that the editorial workflows for all the editions are formalised in XProc pipelines makes it possible, for instance, to compare these pipelines and see in detail, and with the utmost precision, how they differ, what is, in practice, the difference between the process of establishing a diplomatic edition and that of creating a semi-diplomatic or an interpretative edition, at least within the project. Breaking down the traditional analogue processes into unambiguous discrete steps can contribute to the scholarly debate on edition typology.

4 Conclusions and future work

In this paper we described the workflow and the methodological approach behind a digital scholarly edition of the ENHG translation of Marco Polo's work. The key idea behind this methodological approach is the encoding of all the knowledge of the scholar (transcribed text, normalisation rules for words and punctuation, exceptions, corrections) in a formalised way. Thanks to this approach it is possible to produce multiple editions at different levels of faithfulness and readability from a single master TEI file while still keeping this master file lean and readable. In addition, the fact that the whole process is recorded and no manual modifications are performed in the generated edition files, allows the scholar to rapidly make fixes or test hypotheses. Lastly, the sets of rules, being separated from the transcriptions, can be debated by the scholarly community independently from the edition.

Compared to other approaches, this pipeline-based approach greatly simplifies the programming of rules that modify the structure of the text, leading to simpler and more succinct rules. First-class support for structural modifications is necessary to properly handle the normalisation of punctuation. While textual normalisation can in most cases be carried out through simple textual substitutions, many cases of punctuation normalisation require, instead, changes to the structure of the text, as well as follow-up adjustments, for instance when full stops are inserted and sentences split.

The adoption of our approach, as well as other similar programming-based approaches, forces a reflection on the role of the editor, or more precisely, on the many new roles that the editor must embrace. First, the editor stops being a transcriber and becomes a documenter of textual phenomena, thanks to each normalisation action being explicitly recorded in the source files. With a collection of recorded phenomena in place, the editor can turn these single occurrences into systematic rules and catalogues of rules, becoming thus a recogniser of patterns and synthesizer of knowledge. Once various such catalogues of rules have been established (e.g., for different languages, for different time periods, for different scriptoria) and publicly shared, we envision future editors starting their editions by picking and choosing their preferred set of rules, therefore explicitly signalling in which philological tradition they are placing their work while maintaining the technical support needed to express their scholarly freedom to devise new exceptions and new rules. The publication of used rules and workflows (in the form of steps and pipelines) would then become part of the expected content of a digital scholarly edition.

In the future we would like to test out various possible improvements. First we would like to experiment with creating *declarative rule generators*. Many rules are repetitive in their nature (for example the normalisation of single characters) and it should be possible to express them in a declarative fashion. These abstract rules would then be compiled down into XSLT transformations. Another aspect we would like to reflect on is how the transformation process directed by the pipelines influences the various levels of abstraction of the document being transformed, drawing parallels with stratified document models such as CMV+P (Barabucci 2019). A last thing we would like to test is the replacement of the XProc pipelines with pure XSLT pipelines (Birnbaum, 2017). Replacing XProc with XSLT pipelines would reduce the number of technologies that other scholars have to be familiar with in order to understand the editorial

process in its entirety.

Another future development that we envision is the deconstruction of the visualisation of the edition into a series of small, explicit steps, taking place one after the other, just like their counterparts in the pipelines: one click would show the effects of the normalisation of the allographs, another click would expand the abbreviations, another one would introduce a different system of punctuation, etc, until the last click would finally reveal the scholarly edition in its completeness. Our approach, in which all the knowledge of the editor is formalised, recorded and made actionable, would be a solid base for such a future fractally detailed edition.

We hope to be of inspiration to other edition projects. We firmly believe that all medieval manuscripts should be accessible in this way. We need to get out of the constraints of paper editions and to start thinking differently. As Sahle famously wrote, a digital scholarly edition is really digital if it cannot be printed without a loss of the original contents and functionalities. Only editions that fully embrace the digital medium and its tools, including formalisation of knowledge and programming, can satisfy this condition.

In this environment where we have no space limits, why should we limit ourselves to the presentation of a predefined number of rendition levels? Why only diplomatic, semi-diplomatic and interpretative? We believe, in fact, that the granularity can increase even further.

Bibliography

Andrews, Tara, Transcription and collation. In *Handbook of Stemmatics*, edited by Philipp Roelli, Berlin/Boston: De Gruyter, 2020, pp. 166-181.

Barabucci, Gioele, The CMV+ P Document Model, Linear Version. In *Versioning Cultural Objects: Digital Approaches*. SIDE 13. BoD. 2019

Barabucci, Gioele, Funktionale und deklarative Programmierung-basierte Methode für nachhaltige, reproduzierbare und verifizierbare Datenkuratation. DHd Konferenz 2018: Kritik der digitalen Vernunft; Köln, 26. Februar bis 2. März 2018. 2018.

Barabucci, Gioele and Franz Fischer, The formalisation of textual criticism: bridging the gap between automated collation and edited critical texts. *Advances in Digital Scholarship Editing: Papers Presented at the DiXiT Conferences in The Hague, Cologne and Antwerp*. Sidestone Press, 2017.

Barabucci, Gioele, Elena Spadini and Magdalena Turska, Data vs Presentation. What is the core of a Scholarly Digital Edition? *Advances in Digital Scholarship Editing: Papers Presented at the DiXiT Conferences in The Hague, Cologne and Antwerp*. Sidestone Press, 2017.

Barabucci, Gioele, Monica Palmirani, Fabio Vitali, and Luca Cervone. Long-term preservation of legal resources. In *International Conference on Electronic Government and the Information Systems Perspective*, LNCS 6866. Springer, Berlin, Heidelberg, 2011. DOI: 10.1007/978-3-642-22961-9_7

Benedetto, Luigi F., Marco Polo, Il “Milione”, prima edizione integrale a cura di L.F. Benedetto, Firenze: Leo S. Olschki, 1928.

Birnbaum, David J., Patterns and antipatterns in micropipelining within XSLT, in *Proceedings of Balisage: The Markup Conference 2017*. Balisage Series on Markup Technologies, vol. 19 (2017).

Bosco Coletsos, Sandra, *Storia della lingua tedesca*, Torino: Rosenberg & Sellier: 2007.

Buzzoni, Marina, A Protocol for Scholarly Digital Editions? The Italian Point of View, in *Digital Scholarly Editing Theories and Practices*, edited by Matthew J. Driscoll and Elena Pierazzo, Cambridge (UK): Lightning Source for Open Book Publishers, 2016.

Buzzoni, Marina, (ed), Towards the construction of a stemma, in *Handbook of Stemmatology*, edited by Philipp Roelli, Berlin/Boston: De Gruyter, 2020, pp. 139-207.

Contini, Gianfranco, *Filologia*, edited by Lino Leonardi, Bologna: Il Mulino, 2017[†].

Cugliana, Elisa, Exploring a German version of Marco Polo's "Devisement dou Monde", *Ecdotica*, 16 (2019), pp. 9-32.

Darmon, Alexandra N. M., et al., Pull out all the stops: Textual analysis via punctuation sequences, in *European Journal of Applied Mathematics* (2020), pp. 1-37.

Digilio, Maria Rita, Scrittura e interpunzione in testi del medioevo tedesco e gotico, in *Storia della punteggiatura in Europa*, a cura di B. Mortara Garavelli, Bari/Roma: Laterza, 2008, pp. 359-373.

Ebert, Robert Peter et al., *Frühneuhochdeutsche Grammatik*, Tübingen: Max Niemeyer Verlag, 1993.

Garavelli, Bice Mortara (ed), *Storia della punteggiatura in Europa*, Bari: Laterza, 2008.

Goebel et al., *Frühneuhochdeutsches Wörterbuch*, Berlin: De Gruyter, 1986.

Grimm, Jacob, *Deutsche Grammatik*, Bd. 1, Göttingen: Dieterich, 1822.

Hartweg, Frédéric and Klaus-Peter Wegera, *Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*, Berlin: De Gruyter, 2013.

Hammer, Andreas, Victor Millet and Timo Reuvekamb-Felber, *Hartmann von Aue Erech Textgeschichtliche Ausgabe mit Abdruck sämtlicher Fragmente und der Bruchstücke des mitteldeutschen ‚Erech‘*, Berlin/Boston: De Gruyter, 2017.

Kirchhoff, Frank, *Von der Virgel zum Komma: Die Entwicklung der Interpunktion im Deutschen*, Germanistische Bibliothek Bd. 61, Heidelberg: Universitätsverlag Winter Heidelberg, 2017.

Maas, Paul, *Textkritik*, Leipzig: B. G. Teubner, 1950.

Mitchell, Bruce and Fred C. Robinson, *Beowulf. An edition*, Oxford: Blackwell Publishing, 1998 [first ed.].

Parkes, Malcolm B., *Pause and effect: an introduction to the history of punctuation in the west*, New York: Routledge, 2016.

Pierazzo, Elena, *Modelling Digital Scholarly Editing: From Plato to Heraclitus*, in *Digital Scholarly Editing, Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Cambridge (UK): Lightning Source for Open Book Publishers, 2016.

Renzi, Lorenzo, and Alvise Andreose, *Manuale di linguistica e filologia romanza*, Bologna: Il Mulino, 2009.

Rinas, Karsten, *Theorie der Punkte und Striche*, Germanistische Bibliothek Bd. 62, Heidelberg: Universitätsverlag Winter Heidelberg, 2017.

Sahle, Patrick, *Digitale Editionsformen*, 3 voll. *Schriften des Instituts für Dokumentologie und Editorik*, voll. 7–9, Norderstedt: BoD, 2013.

Sahle, Patrick, *What is a Digital Scholarly Edition?* In *Digital Scholarly Editing, Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Cambridge, UK: Open Book Publishers, 2016.

Scherer, Wilhelm, *Zur Geschichte der deutschen Sprache*, Berlin: Weidmann, 1878.

Stussi, Alfredo, *Nuovo avviamento agli studi di filologia italiana*, Bologna: Il Mulino, 1988.

Tscharner, Eduard, H, *Der mitteldeutsche Marco Polo nach der Admonter Handschrift. Mit einer Tafel in Lichtdruck*, Berlin: Weidmannsche Buchhandlung, 1935.

Van Zundert, Joris J., “Why the Compact Disc Was Not a Revolution And «Cityfish» Will Change Textual Scholarship, or What Is a Computational Edition?” *Ecdotica* 15: 129–156. 2019. Wells, Christopher J., *Deutsch: eine Sprachgeschichte bis 1945*, Tübingen: Niemeyer, 1990.

Online Projects

Akademie der Wissenschaft zu Göttingen, *Frühneuhochdeutsches Wörterbuch*: <https://fwb-online.de>.

Haugen, Odd Einar (ed), *Menota Handbook* (v. 3.0), 2019: https://www.menota.org/HB3_index.xml

Macé, Caroline and Philipp Roelli, *Parvum Lexicon Stemmatologicum*: <https://wiki.helsinki.fi/display/stemmatology>, 2015.

Marzi Werner, *Ingelheimer Haderbücher*: <https://www.haderbuecher.de>.

Michael Stolz, *Parzival Projekt*, <http://www.parzival.unibe.ch/home.html>.

Rosselli Del Turco, Roberto, *The Digital Vercelli Book*: <http://vbd.humnet.unipi.it/beta/#104v>.

Rosselli Del Turco, Roberto, et al. *Edition Visualization Technology*: <http://evt.labcd.unipi.it>.

Stella, Francesco. *Corpus Rhythmorum Musicum*: <http://www.corimu.unisi.it/public/frontend>.

Authors' Biographies

Elisa Cugliana is doing a joint PhD in Germanic philology at the Ca' Foscari University of Venice and the Cologne Center for eHumanities at the University of Cologne. Her dissertation aims at producing a digital scholarly edition of a medieval German translation of Marco Polo's *Devisement dou Monde*, which is preserved in five not-yet-edited witnesses. In her project, she explores the possibilities offered by the digital medium: both the visualisation and use of scholarly editions, as well as editorial methodologies and workflows, for instance by transforming the traditional analogue continuum spanning between facsimile and interpretative transcription(s) into a supervised (and repeatable) automatic process.

Gioele Barabucci is an associate professor in Computer Science at the Department of Design of the Norwegian University of Science and Technology (NTNU). He received his PhD in Computer Science from the University of Bologna and later went on to pursue his research interests in digital humanities as a Marie Curie Experienced Researcher and as a research fellow the Cologne Center for eHumanities. His main research topics are the design of knowledge and the evolution of information. Concretely, this means studying comparison algorithms, devising versioning systems, formalising document models, as well as researching ontologies, legal documents and multilingual systems.