



Journal of the Text Encoding Initiative

Issue 14 | April 2021 - March 2023
Selected Papers from the 2019 TEI Conference

Signs of the Times: Medieval Punctuation, Diplomatic Encoding and Rendition

Elisa Cugliana and Gioele Barabucci



Electronic version

URL: <https://journals.openedition.org/jtei/3715>
DOI: 10.4000/jtei.3715
ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Elisa Cugliana and Gioele Barabucci, "Signs of the Times: Medieval Punctuation, Diplomatic Encoding and Rendition", *Journal of the Text Encoding Initiative* [Online], Issue 14 | April 2021- March 2023, Online since 24 May 2022, connection on 03 March 2023. URL: <http://journals.openedition.org/jtei/3715> ; DOI: <https://doi.org/10.4000/jtei.3715>

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

Signs of the Times: Medieval Punctuation, Diplomatic Encoding and Rendition

Elisa Cugliana and Gioele Barabucci

ABSTRACT

Digitally managing punctuation in the editions of medieval manuscripts is one of those issues that initially looks like a minor detail, but later reveals itself as a tangled web of problems spanning from computer science (how to represent punctuation signs?) to philology (what types of signs 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 them 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 normalization. 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.

INDEX

Keywords: normalization, punctuation, automated workflow, pipelines, computable edition, formalization of editorial rules, reproducibility

1. Introduction

- 1 Since its inception, the critical edition of the Early New High German (ENHG) Marco Polo¹ (version DI² of the *Devisement dou Monde*) has been planned as an edition with multiple layers. For each witness the editor would manually edit a master TEI file and, from that file, three edition levels would be generated, each at a different level of closeness to the source: a purely diplomatic edition (where aspects like colors, letter sizes, glyph variance, and layout are preserved), a fully interpretative edition (presenting a more modernized text), and a semidiplomatic 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).
- 2 These issues quickly arose while editing:
 1. 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;
 2. normalizing punctuation revealed itself as a complex task that required profound changes to the structure of the edited text.
- 3 The first issue, the overgrowth of the master TEI file, is due to the need to transcribe and encode three variants of each word of the witness (one for each level of edition). 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 that span multiple words.

- 4 The second issue, the complexity of normalizing punctuation, emerged from the combination of two factors:
1. it was not possible to find an underlying system governing the use of punctuation in the source;
 2. using the standard approach to normalization 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, semidiplomatic, and interpretative).
- 5 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 eighteen months, while keeping the master TEI file lean in its structure and readable even by scholars who are not well versed in XML technologies.
- 6 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 process the master file according to the rules 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, there is a clear distinction between rules and exceptions in our approach. This is particularly important in the ENHG Marco Polo edition because the language it is written in has never been standardized and thus the system of rules has been organically derived during the editorial work (as described in detail in section 2).
- 7 This paper provides an overview of our approach (section 3) and shows how our approach addresses both issues. In particular we show how normalizing punctuation represents a dramatic step beyond the more classical normalization of words. The current implementation of our approach, based on XProc and XSLT, is also presented in section 3.
- 8 Moving to an editorial workflow with such a level of automation requires a reevaluation of the role of the editor, from wordsmith to formalizer 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 widespread adoption of this methodology could entail for future critical editors.

2. Background and issues

2.1 Marco Polo and ENHG

- 9 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 Marco Polo's travels to and within the Eastern World: the traveler was a Venetian who, following his father and uncle on a business trip to Asia, ended up spending almost twenty-five years in the East, at the court of Kublai Khan. The account of his journey is preserved in more than 140 manuscripts which were produced over two centuries. The original version composed by the two prison companions is unfortunately lost. The text was translated into German twice during the Middle Ages. While one of these translations (version VG3) was already edited by [Tscharner \(1935\)](#), version DI is being edited for the first time in our project, which serves as the backdrop for the present study.
- 10 Under the name ENHG Marco Polo we refer to an edition project including the 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 fifteenth century.
- 11 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 rendition: diplomatic, semidiplomatic, 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 commentaries.³ 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).

- 12 Our use of the term *edition* matches that of the Italian school of textual scholarship,⁵ according to which the term edition refers to a specific rendition of the text (diplomatic, interpretative, etc.). This use of the word *edition* also conforms to the definition of scholarly edition provided by Sahle (2016, 23), although the scholar uses this term to include all parts of an edition project: the act of rendering the text in, say, an interpretative way, is *erschließend* (revealing, opening up to the text) in its nature, as some specific aspects of the text are presented in such a way that the reader is granted more informed access to them.
- 13 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 normalization 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: 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.
- 14 The language in which the three witnesses are written is Early New High German. ENHG represents quite a problematic phase of the German language, beginning with the definition of its time span. As a matter of fact, there have been different proposals for a definition of its periodization (see, for instance, Grimm 1822, Scherer 1878, and Wells 1990), but what is relevant here is that the ENHG phase covers a long time span, going through Gutenberg's media revolution, which had a huge impact on the written German language (Bosco Coletsos 2003, 187–91). This stage of the German language is extremely diverse, which has quite a few consequences for the scholarly editor approaching the texts written or printed during the period.

2.2 Normalization

- 15 “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, s.v. <https://wiki.helsinki.fi/display/stemmatology/Normalisation>).
- 16 The need for normalization in scholarly editions arises from the fact that premodern 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 regularized and eroded. 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. Diacritics, for instance, and other linguistic peculiarities were considered on a par with *ein verdammtes Unkraut* (a damned weed) in a tidy flower garden, as Julius Weizsäcker wrote in 1867.⁸

- 17 With Karl Lachmann,⁹ normalization became a necessity for philologists themselves: needing to summarize 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. Collation, in particular, implies normalization: the establishment of a traditional stemma is based on Leitfehler (“guiding errors”), not on equivalent variants.¹⁰ The formal appearance of the reconstructed text (for example, the dialectal variety chosen for the text) would then be given the shape considered to be 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 normalized.¹²
- 18 For many languages the exact procedures used to normalize a text have been established *a posteriori* by historical linguists. For instance, normalized Middle High German was first established by Karl Lachmann (see [Maas 1950](#)) and then accepted by the scholarly community.
- 19 However, no such normalization standards currently exist for ENHG. As [Ebert et al. \(1993, 7\)](#) write: “Until the sixteenth century 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 others and which can, on the basis of such sociolinguistic fact, be considered as a reference variety for synchronic writing and speaking or for diachronic use.”¹³
- 20 A survey of existing editions of ENHG texts shows that, while internal coherence in normalization practices is desired and documented, not only are normalization standards missing, they are also not actively sought after. Focusing on the Ingelheimer Haderbücher project ([Marzi and Sprenger n.d.](#)), for instance: although the editors were extremely accurate in outlining their principles for textual transcription, they did not provide any specific bibliographical reference in support of their decisions on how the texts were transcribed and normalized. 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

- 21 An interesting case is the *Frühneuhochdeutsches Wörterbuch* (FWB), 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 systematization represents only an ideal (in their words *ausgezeichnet*) language.¹⁶ We are not using the FWB as a reference for our normalization, first, because this ideal language is very distant from the actual language and orthography we find in our witnesses, and, second, 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 Normalization Choices Are Editorial Choices

- 22 For a scholar editing a medieval text, “to normalize” means making a series of editorial decisions.¹⁷ Some examples of the kinds of decisions that need to be made are:

how to substitute modern equivalents for original textual features (e.g., if and when to substitute the modern *s* for the *long s* [ſ]);

how to change certain signs to others (e.g., when to turn a *virgula* into a comma, when to turn it into a full stop, and 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 capitalization rules or spaces between different words or sections of text).

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

- 24 In our view there is no strict dichotomy between normalization and correction, but rather a continuum. No normalization is entirely objective, no correction is entirely subjective. For example, one can argue that substituting *u* for *v* and vice versa is quite unproblematic and rather objective because 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 even in the simplest cases the philologist runs the risk of concealing linguistic features of great importance, taking them for scribal negligence.
- 25 In our opinion no intervention pertaining to the normalization process can be considered objective or mechanical. The lack of a clear line of distinction suggests that as much information as feasibly possible should be preserved during the preparation of a scholarly edition.
- 26 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 to make it possible to preserve a vast amount of information with little effort on the part of the editor.

2.2.2 Normalization in TEI

- 27 In TEI most of these normalization actions are represented using `<choice>` elements. For example, the expansion of *p̄* into *per* would be encoded as

Example 1. Expansion of an abbreviation with `<choice>`.

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

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

Example 2. Normalisation in the Menota project.

```
<w>
  <choice>
    <me: facs>wa#n</me: facs>
```

```

    <me:dipl>warn</me:dipl>
    <me:norm>waren</me:norm>
  </choice>
</w>

```

- 29 There are many TEI elements that are routinely used in combination with <choice> to express different kinds of normalization intents: for instance, <abbr>, <expan>, <sic>, <corr>, <orig>, <reg>, <supplied>, and, for punctuation, <pc>.
- 30 The use of <choice> has two main drawbacks: first, it causes a twentyfold expansion in character count (a five-letter word with two normalizations requires more than one hundred characters to be encoded), and, second, it forces what would be a single <w> element to grow into an eight-tag, two-level-deep markup structure. When each word in a manuscript needs such a treatment (as 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

- 31 Punctuation is one of the many aspects of a text that undergoes normalization during the creation of a scholarly edition. In most cases normalizing punctuation is considered a secondary task and very quickly brushed off in the introductions of many editions.²⁰
- 32 In our opinion, the normalization of punctuation is a critical step in the creation of any nondiplomatic edition. Punctuation is not only a reading aid: it delivers meaning, suggests interpretations, creates structures, and 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.
- 33 From an operational point of view, the normalization of punctuation marks is profoundly different from the normalization of words, although the two are often treated in the same way.
- 34 The normalization 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, a linear effect on the markup structure (for example, turning `<w>aberes</w>` into `<w>aber</w><w>es</w>`²¹).

- 35 In contrast, the normalization 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, normalizing 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 capitalization of the following word, now the first word of the following sentence: “`<s>... geben.</s><s>Von stund er gesund wirt ...</s>`.”²²
- 36 The additional complexity of normalizing 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 scribe, the language, the destination of the text, and the local customs.
- 37 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 represented in the ENHG Marco Polo, “his sister” would be the focus of the sentence, while according to the modern use of commas, the same phrase 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 clearly emphasize the phrase in question. Moreover, modern punctuation imposed on medieval texts can obscure important linguistic phenomena: in the case of the ENHG Marco Polo, this relates to the presence or absence of structures similar to the so-called relative nexus in Latin and with the syntactic parameter of the null subject.

- 38 Unfortunately, things are even more complex than that; the sentence we have just analyzed proceeds as follows: “die zwen prüder ... darnach chomen, zu dem groffen hern, der ganczen tartarey, gnant δ 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.
- 39 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) write 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 fourteenth and fifteenth centuries, “polyfunctional” (Ebert et al. 1993, 29).
- 40 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, almost impossible to transpose in a fully mechanical way the punctuation system used by a manuscript into one of the many currently used. 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 2014†, 23). Instead of a translation from one system into another (which is the definition of normalization adopted by Contini), in the case of punctuation one would have to substitute the system *in toto*.
- 41 We are not the first to notice that normalizing punctuation differs profoundly from normalizing words. In 2014, for instance, Elena Pierazzo wrote in a discussion on the TEI mailing list on the subject of normalization of punctuation:²⁵
- “
- The point is that punctuation is not like spelling and orthography. [F]or 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 semicolon, 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 completely new one. (emphasis added)

”

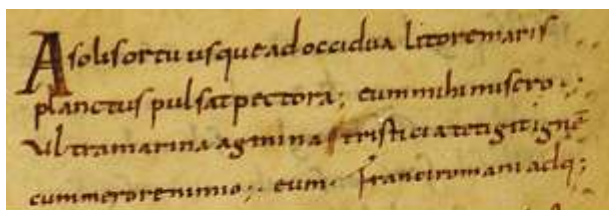
2.3.1 Possible Approaches

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

1. the complete removal of the original punctuation system, replacing it with a modern one;
2. the decision to maintain punctuation exactly as it was in the source document;
3. an endeavor to normalize punctuation, while remaining as close as possible to the source.

43 An example of the first approach, removal and replacement of the original punctuation, is found in *Corpus Rhythmorum Musicum* (Stella 2020),²⁶ a multimodal edition containing the earliest medieval Latin songs, together with their musical rendition. Figure 1 shows an excerpt from the edition alongside one of the witnesses:

Figure 1. Left: Manuscript of the 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.



- 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!

- 44 An example of the second method, the faithful preservation of the original punctuation in the edited text, can be found in the *Parzival-Projekt* (Stolz 2020),²⁹ depicted in figure 2. Here the transcription of the witness perfectly matches the original punctuation: the medieval dot is encoded using a modern full point, fully comparable from the point of view of its visual rendering.

Figure 2. Extract St. Gallen, Stiftsbibliothek, Cod. 857 (f. 10a, verses 18.30–19.03) from the edition of MS D in the *Parzival-Projekt*. The edited text maintains the original punctuation.



- 45 An example of the third strategy, normalizing punctuation while trying to keep it close to the original, is the *Digital Vercelli Book* by Roberto Rosselli Del Turco (n.d.). 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 quite unfamiliar for modern readers, who might have difficulties parsing the text.

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



46 If, on the one hand, we appreciate the attempt to normalize the original punctuation, on the other hand, this approach does not seem to improve the readability of the text, as one would expect 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 normalizations, a first step in reducing the complexity present in the original source.

2.3.2 Punctuation in the ENHG Marco Polo

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

48 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).

49 At the semidiplomatic level a reduced system of punctuation consisting of two signs, a middle dot (·) and a *virgula* (|), will be introduced. Similar systems are often used in the late Middle Ages (Parkes 1992, 46) and are often adopted in ENHG texts of the fourteenth and fifteenth centuries (Hartweg and Wegera 2005, 131). This reduced system based on unfamiliar signs will allow a smooth presentation of the text's structure, and, at the same time, reduce the intervention of the editor on the interpretation of the textual content. A similar approach was used by Mitchell and Robinson (1998), as discussed in the next section. In addition, 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).

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

2.4 Multiple Editions

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

- 52 One particular aim of the ENHG Marco Polo is to be able to programmatically produce multiple editions, each with different notes and normalized 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 related information and critical annotations (including the references for the editor's notes).

The same text will be edited in multiple levels of closeness to the source: diplomatic (with special characters to show allographic variance, colors, 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 modernization of graphemes, punctuation, and other textual features). Given that the project is based on three main witnesses, this means producing a total of nine different editions.

Multiple editions will be generated automatically from the master TEI file, with no manual intervention on the resulting files.

The generated editions files will conform to the TEI subset understood by EVT.

- 53 Some of these desiderata clash with each other. For instance, the desire to directly edit the XML file makes it hard and error-prone to keep in a single master file all the information needed to generate the three editions. It would have been hardly possible to respect the desired level of explicitness in the normalisation process due to the complexities emerging from the high level of granularity of the transcriptions combined with the goal of producing multiple levels of normalisation.

2.4.1 Predigital Approaches

- 54 In predigital 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 normalization they wished to produce. Not only is this a very long and demanding workflow, but it also poses a number of problems: first of all, if an uncertain passage of the text gets

clarified later on in the editing process, 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.

- 55 An interesting example, in this respect, is that of 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 issue: modern punctuation would not suit Old English poetry and it would force the editors to make decisions in order to resolve ambiguities which cannot be resolved. Given these considerations, extensively explained in section IIIB of their edition, they wanted to offer a way out, or at least an alternative. They proposed a different system of punctuation, which inspired our intermediate level of normalization. However, they limited themselves to offering a mere taste of their solution in the appendix to the edition (pp. 313–18). In the appendix they edited verses 1–114 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 a black and white facsimile (pp. 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 realize their ideas to the fullest, as that would probably cost too much time and space.

2.4.2 Digital Approaches

- 56 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 includes both a diplomatic transcription and hints on how to normalize certain pieces of texts.
- 57 For example, it is quite common for philologists who prepare their editions using TEI to make use of elements such as `<choice>`, `<orig>`, and `<reg>` to indicate how to normalize 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.

- 58 A digital edition exemplifying this approach is the aforementioned *Digital Vercelli Book*, by [Roberto Rosselli Del Turco \(n.d.\)](#): here two levels of edition are offered, a diplomatic and a more interpretative one. The user can compare the two editions visualizing them synoptically in the EVT software used for the edition.

3. Our Approach

- 59 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 normalization 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, Spadini, and Turska, 2017](#)), legal analysis ([Barabucci et al. 2011](#)), and curation of research data ([Barabucci 2018](#)).
- 60 This methodology is characterized 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.

- 61 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/>.
- 62 This methodology contrasts with the established editorial practice of mingling transcription, normalization, and critical amendments. Instead of just performing the desired normalization steps while transcribing and keeping track of them in an introduction to the edition, here the rules are separate artifacts, related to, but independent from, the edited text. In fact, as a

fortunate byproduct of this methodology, the ENHG Marco Polo produced a set of rules for the normalization of words and punctuation in Early New High German. The scholar community now has the opportunity to evaluate, discuss, critique, modify, and, if desired, adopt this set of rules independently from the edited text.

- 63 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 because 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 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

- 64 Rules are the way in which the scholar formalizes the normalization process. The kind of rule used in the project is 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 normalized into a full stop, then the first letter of the following word must be capitalized);

methods to join words split at the ends of lines (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 a long s (ſ) with a lowercase s);

expansion rules for abbreviations (e.g., replace all occurrences of “vñ” with “vnd”).

- 65 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 essential to represent such structures 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 properly with normalization in general and punctuation in particular.

- 66 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 three lines of XSLT.
- 67 To give the readers an impression of the simplicity of the rule implementation, 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

- 68 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.
- 69 The XSLT excerpt in [Example 3](#) shows how split words are joined when a middle 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 because this step operates on a supertextual 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 supertextual structure in general, are not part of the master TEI file and have been added in the preceding 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 will remove 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 aid to the readers of the final edition file.

Example 3. XSLT implementation of the rule “Join words split with a double oblique hyphen.”

```
<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>

```

- 70 The rule in Example 3 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.
- 71 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

- 72 Before a rule can be written, a textual phenomenon must first be recognized 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 normalization rules are not yet clear, it is simpler and more

productive to address one-off textual phenomena manually. This is done by annotating the word to be normalized or the sign to be changed in the master TEI file. Once a pattern is recognized, a rule can be written and these annotations removed from the master TEI file.

- 73 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 a specific case, it makes sense to mark the specific occurrence as an exception in the master TEI file.
- 74 Finally, the scholar may want to correct an obvious mistake, while keeping track of the original reading. This is done not by writing a rule, but by placing an annotation in the master TEI file.
- 75 All these issues (one-off normalization, exceptions, and manual corrections) are encoded using TEI elements directly in the master TEI file.
- 76 One-off normalizations 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 word “foo” for the content of that element, but only at the normalization 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 [Example 4](#).

**Example 4. Substitution of a comma with a virgula in the semi-diplomatic edition (n1) and with a semicolon in the interpretative edition (n2).**

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

- 77 Manual corrections are encoded using a combination of <choice>, <orig>, and <reg> elements, together with the @mp:use-in-nX attributes, as illustrated in [Example 5](#).

**Example 5. Correction of the typo in the name “Marcho,” which will appear as “Marcho” in the semi-diplomatic edition and as “Marco” in the interpretative edition.**

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

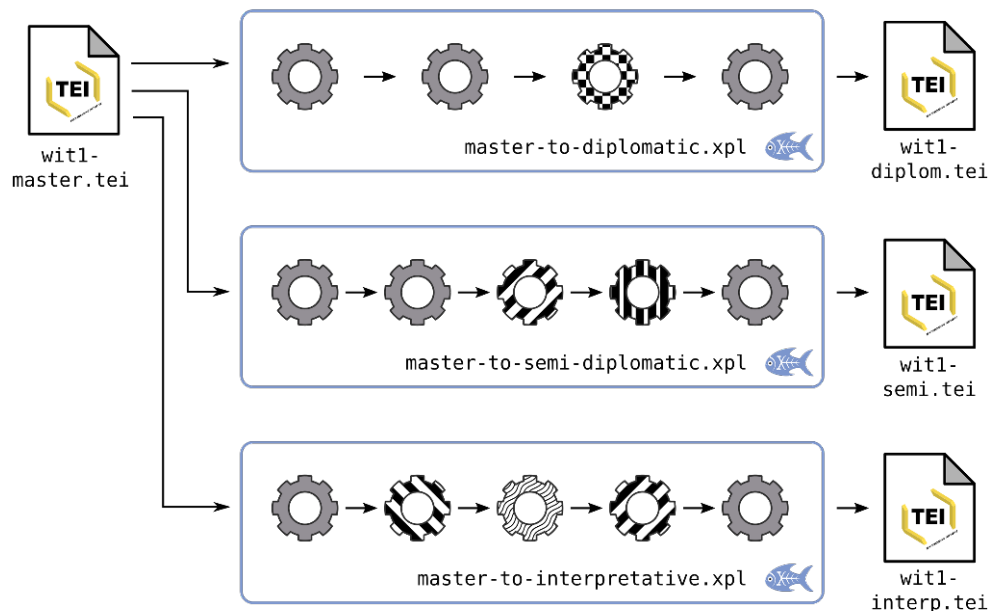
</persName>

- 78 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 (the results of the editor's own decisions).

### 3.3 Pipelines and Normalization Levels

- 79 Pipelines are used by the scholar to indicate which rules should be applied, and in which order, to turn the master TEI file into a complete edition at a specific level of normalization.
- 80 In the ENHG Marco Polo project there are three pipelines, one for each target level of faithfulness and readability. 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 4 illustrates this approach.

Figure 4. 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.



- 81 Each pipeline is implemented as an 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 level is equivalent to replaying this record. [Example 6](#) shows an excerpt of the XProc pipeline used to generate the semidiplomatic edition.
- 82 It is important to note that pipelines comprise three kinds of steps:
1. infrastructural steps: for example, the *tokenize* step that creates the textual super-structure to which all other steps will refer to;
  2. rule-based steps: for example, the *adjust-capitalization* step that implements the set of rules to fix the capitalization of words after changes in punctuation; and
  3. exception-related steps: for example, 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).

**Example 6.** Excerpt of the XProc pipeline used to generate the semi-diplomatic edition. Steps marked A are steps that implement rules; the step marked B takes care of exceptions.

```
<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>

```

- 83 The fact that the editorial workflows for all the editions are formalized in XProc pipelines makes it possible, for instance, to compare these pipelines and see in detail (and with utmost precision) how they differ and what is, in this project, the difference between the processes needed to establish a diplomatic, a semi-diplomatic or an interpretative edition. Breaking down the traditional analogue processes into unambiguous discrete steps can contribute to the scholarly debate on edition typology.

## 4. Conclusions and Future Work

- 84 We have described the workflow and the methodological approach behind a digital scholarly edition of the ENHG translation of Marco Polo's work. The key idea to this methodological approach is the encoding of all the scholar's knowledge (transcribed text, normalization rules for words and punctuation, exceptions, corrections) in a formalized way. Using this approach, it is possible to produce multiple editions at different levels of faithfulness and readability from a single master TEI file while keeping the master file lean and readable. In addition, the fact that the whole process

is logged, and no manual modifications are performed in the generated edition files, allows the scholar to rapidly make fixes or test hypotheses. Finally, the sets of rules, being separated from the transcriptions, can be debated by the scholarly community independently from the edition.

- 85 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 normalization of punctuation. While textual normalization can, in most cases, be carried out through simple textual substitutions, many cases of punctuation normalization instead require changes to the structure of the text, as well as follow-up adjustments, for instance when full stops are inserted and sentences split.
- 86 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 mere transcriber: his/her main task is now to record all textual phenomena (each normalization action is 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 a pattern-spotter and knowledge-synthetizer. Once various such catalogues of rules have been established (e.g., for different languages, different time periods, and different scriptoria) and publicly shared, we envision future editors starting their editions by picking and choosing their preferred set of rules, thereby explicitly signaling 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 the rules and workflows used (in the form of steps and pipelines) would then become part of the expected content of a digital scholarly edition.
- 87 In the future we would like to test various possible improvements. First we would like to experiment with creating *declarative rule generators*. Many rules are repetitive in their nature (for example, the normalization of single characters) and it should be possible to express them in a declarative fashion. These abstract rules would then be translated 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 final 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.

- 88 Another future development that we envision is the deconstruction of the visualization 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 normalization 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 formalized, recorded, and made actionable, would be a solid base for such a future fractally detailed edition.
- 89 We hope to be an inspiration for other edition projects. We firmly believe that all medieval manuscripts should be accessible in this way. We need to escape 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 formalization of knowledge and programming, can satisfy this condition.
- 90 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, semidiplomatic, and interpretative? We believe, in fact, that the granularity can increase even further.

---

## BIBLIOGRAPHY

- Andrews, Tara. 2020. "Transcription and Collation." Chap. 3.3 in *Handbook of Stemmatology*, edited by Philipp Roelli, 160–75. Berlin: De Gruyter. doi:10.1515/9783110684384.
- Barabucci, Gioele. 2018. "Funktionale und deklarative Programmierung-basierte Methode für nachhaltige, reproduzierbare und verifizierbare Datenkuration." Paper presented at DHd Konferenz 2018: Kritik der digitalen Vernunft, Cologne, February 26–March 2, 2018. <https://gioele.io/p/dhd2018>; <http://dhd2018.uni-koeln.de/>.

- . 2019. “The CMV+P Document Model, Linear Version.” In *Versioning Cultural Objects: Digital Approaches*, edited by Roman Bleier and Sean M. Winslow, 153–70. Norderstedt: BoD.
- 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, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk van Hulle, 47–53. Leiden: Sidestone Press.
- Barabucci, Gioele, Elena Spadini, and Magdalena Turska. 2017. “Data vs. Presentation: What Is the Core of a Scholarly Digital Edition?” 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, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk van Hulle, 37–46. Leiden: Sidestone Press.
- Barabucci, Gioele, Monica Palmirani, Fabio Vitali, and Luca Cervone. 2011. “Long-term Preservation of Legal Resources.” In *International Conference on Electronic Government and the Information Systems Perspective*, Lecture Notes in Computer Science 6866: 78–93. Berlin: Springer. doi:10.1007/978-3-642-22961-9\_7.
- Benedetto, Luigi F., ed. 1928. *Il Milione*. By Marco Polo. Prima edizione integrale. Firenze: Leo S. Olschki.
- Birnbaum, David J. 2017. “Patterns and Antipatterns in XSLT Micropipelining.” In *Proceedings of Balisage: The Markup Conference 2017*. Balisage Series on Markup Technologies 19. doi:10.4242/BalisageVol19.Birnbaum01.
- Bosco Coletsos, Sandra. 2003. *Storia della lingua tedesca*. Torino: Rosenberg & Sellier.
- Buzzoni, Marina. 2016. “A Protocol for Scholarly Digital Editions? The Italian Point of View.” In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 59–82. Cambridge: Open Book Publishers.
- Buzzoni, Marina, ed. 2020. “Towards the Construction of a Stemma.” In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 139–207. Berlin: De Gruyter.
- Contini, Gianfranco. 2014†. *Filologia*, edited by Lino Leonardi. Bologna: Il Mulino.
- Cugliana, Elisa. 2019. “Exploring a German Version of Marco Polo’s *Devisement dou Monde*.” *Ecdotica* 16: 9–32.
- Darmon, Alexandra N. M., Marya Bazzi, Sam D. Howison, and Mason A. Porter. 2020. “Pull Out All the Stops: Textual Analysis via Punctuation Sequences.” *European Journal of Applied Mathematics*: 1–37. doi:10.1017/S0956792520000157.
- Digilio, Maria Rita. 2008. “Scrittura e interpunzione in testi del medioevo tedesco e gotico.” In *Storia della punteggiatura in Europa*, edited by Bice Mortara Garavelli, 359–73. Bari: Laterza.
- Ebert, Robert Peter, Oskar Reichmann, Hans-Joachim Solms, and Klaus-Peter Wegera. 1993. *Frühneuhochdeutsche Grammatik*. Tübingen: Max Niemeyer Verlag.

- Grimm, Jacob. 1822. *Deutsche Grammatik*, Bd. 1. Göttingen: Dieterich.
- Hartweg, Frédéric, and Klaus-Peter Wegera. 2005. *Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*. 2nd rev. ed. Berlin: De Gruyter.
- Hammer, Andreas, Victor Millet, and Timo Reuvekamp-Felber, eds. 2017. *Hartmann von Aue, Erech: Textgeschichtliche Ausgabe mit Abdruck sämtlicher Fragmente und der Bruchstücke des mitteldeutschen Erech*. Berlin: De Gruyter.
- Kirchhoff, Frank. 2017. *Von der Virgel zum Komma: Die Entwicklung der Interpunktion im Deutschen*. Germanistische Bibliothek, Bd. 61. Heidelberg: Universitätsverlag Winter.
- Maas, Paul. 1950. *Textkritik*. 2., verb. und verm. Aufl. Leipzig: B. G. Teubner.
- Mitchell, Bruce, and Fred C. Robinson, eds. 1998. *Beowulf: An Edition with Relevant Shorter Texts*. Oxford: Blackwell Publishing.
- Mortara Garavelli, Bice, ed. 2008. *Storia della punteggiatura in Europa*. Bari: Laterza.
- Parkes, Malcolm B. 1992. *Pause and Effect: An Introduction to the History of Punctuation in the West*. New York: Routledge.
- Pierazzo, Elena. 2016. "Modelling Digital Scholarly Editing: From Plato to Heraclitus." In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 41–58. Cambridge: Open Book Publishers.
- Renzi, Lorenzo, and Alvisse Andreose. 2009. *Manuale di linguistica e filologia romanza*. 3rd ed. Bologna: Il Mulino.
- Rinas, Karsten. 2017. *Theorie der Punkte und Striche: die Geschichte der deutschen Interpunktionslehre*. Germanistische Bibliothek, Bd. 62. Heidelberg: Universitätsverlag Winter Heidelberg.
- Sahle, Patrick. 2013. *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. 3 vols. Schriften des Instituts für Dokumentologie und Editorik, vols. 7–9. Norderstedt: BoD.
- . 2016. "What Is a Scholarly Digital Edition?" In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–39. Cambridge: Open Book Publishers.
- Scherer, Wilhelm. 1878. *Zur Geschichte der deutschen Sprache*. 2. ausg. Berlin: Weidmann.
- Stussi, Alfredo. 1988. *Nuovo avviamento agli studi di filologia italiana*. Nuovo ed. Bologna: Il Mulino.
- Tscherner, Eduard Horst von, ed. 1935. *Der mitteldeutsche Marco Polo nach der Admonter Handschrift*. Berlin: Weidmannsche Buchhandlung.
- Van Zundert, Joris J. 2019. "Why the Compact Disc Was Not a Revolution and *Cityfish* Will Change Textual Scholarship, or What Is a Computational Edition?" *Ecdotica* 15: 129–56.
- Wells, Christopher J. 1990. *Deutsch: Eine Sprachgeschichte bis 1945*. Translated by Rainhild Wells. Tübingen: Niemeyer. Originally published as *German: A Linguistic History to 1945* (Oxford: Clarendon Press, 1985).

- Akademie der Wissenschaft zu Göttingen. n.d. *Frühneuhochdeutsches Wörterbuch*. Accessed April 19, 2021. <https://fwb-online.de/>.
- Haugen, Odd Einar, ed. 2019. *The Menota Handbook: Guidelines for the Electronic Encoding of Medieval Nordic Primary Sources*. Version 3.0. Bergen: Medieval Nordic Text Archive. Last modified December 12, 2019. [https://www.menota.org/HB3\\_index.xml](https://www.menota.org/HB3_index.xml).
- Macé, Caroline, and Philipp Roelli, eds. 2015. *Parvum Lexicon Stemmatologicum*. Version 1.0. Last modified May 3, 2018. <https://wiki.helsinki.fi/display/stemmatology>.
- Marzi, Werner, and Kai-Michael Sprenger. n.d. *Ingelheimer Haderbücher*. Accessed April 19, 2021. <https://www.haderbuecher.de/>.
- Stolz, Michael, et al. 2020. *Parzival Projekt*. Last modified May 22, 2020. <http://www.parzival.unibe.ch/home.html>.
- Rosselli Del Turco, Roberto, ed. n.d. *The Digital Vercelli Book*. Beta version. Accessed October 22, 2021. <http://vbd.humnet.unipi.it/beta2/>.
- Rosselli Del Turco, Roberto, et al. 2013. *Edition Visualization Technology*. Accessed April 19, 2021. <http://evt.labcd.unipi.it/>.
- Stella, Francesco, ed. 2020. *Corpus Rhythmorum Musicum*. Last modified July 28, 2020. <http://www.corimu.unisi.it/public/frontend>.

## NOTES

- 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.
- 3 These commentaries are not a critical apparatus in the traditional sense (see [Macé and Roelli 2015, s.v. Apparatus](#)) as no critical text was established for this edition: the reasons for this choice depend on the history of the 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.
- 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 “A light-weight, open source tool specifically designed to create digital editions from XML-encoded texts” (Rosselli Del Turco et al. 2013).

7 On the variation characterizing premodern texts see Contini (2014†, 45) and Pierazzo (2016, 41). As for punctuation, the following manuals are used as a reference in the present project: Parkes (1992), Garavelli (2008), Rinas (2017), and Kirchhoff (2017).

8 As quoted in Sahle (2013, 1:71)

9 For a systematization 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 (2014†, 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”) (our translation).

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” [our translation]. This horizontal variation is in fact *eines der markantesten Kennzeichen des Frühneuhochdeutschen* (one of the most distinctive features of Early New High German).

14 “Bei der Schreibung der einzelnen Buchstaben wurden vor allem Hinweise von Germanisten, Sprachforschern und Paläographen aufgenommen und bei der Erstellung der Transkription berücksichtigt” [our translation] (Marzi and Sprenger (n.d.), “Richtlinien: Transkription und Register,” accessed September 17, 2021, <https://www.haderbuecher.de/projekt/editionsrichtlinien/>).

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 normalize medieval texts, see Stussi (1988, 161), and for a concrete example of normalization, see Renzi and Andreose (2009, 272–82).

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).

19 Haraldur Bernharðsson and Odd Einar Haugen, “Levels of text representation,” ch. 4 in Haugen 2019, [https://www.menota.org/HB3\\_ch4.xml](https://www.menota.org/HB3_ch4.xml).

20 An example of the traditional way to address punctuation is the edition of *Erek* by Hammer, Millet, and Reuvekamb-Felber (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.



- 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 Elena Pierazzo, “Re: Original and editorial punctuation,” TEI-L, May 23, 2014, <https://listserv.brown.edu/cgi-bin/wa?A2=TEI-L;4b7897d6.1405>; <http://tei-l.970651.n3.nabble.com/Original-and-editorial-punctuation-td4025791.html#a4025812>.
- 26 <http://www.corimu.unisi.it/public/frontend>.
- 27 <http://www.corimu.unisi.it/public/manoscritti/manoscritto/idms/7> 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>.

## AUTHORS

### ELISA CUGLIANA

Elisa Cugliana is doing a joint PhD in Germanic philology at the Ca' Foscari University of Venice (Department of Linguistics and Comparative Cultural Studies) and the Cologne Center for eHumanities at the University of Cologne. With her dissertation she aims to produce 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 visualization and use of scholarly editions, as well as editorial methodologies and workflows, for instance by transforming the traditional analog range of editorial options from facsimile to interpretative transcription(s) into a supervised (and repeatable) automatic process.

### GIOELE BARABUCCI

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 the digital humanities as a Marie Curie Experienced Researcher and as a research fellow at 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, and formalizing document models, as well as researching ontologies, legal documents, and multilingual systems.