

TEI encoding in the Hartmann von Aue digital edition projects

Victor Millet / Lorena Pérez Ben

Universidade de Santiago de Compostela

victor.millet@usc.es

<https://orcid.org/0000-0002-7683-8892>

lorena.perez.beno@usc.es

<https://orcid.org/0000-0003-0236-0774>

Received: 02/02/2023; accepted 28/03/2023

DOI: <https://doi.org/10.7203/MCLM.10.26027>

ABSTRACT

Hartmann von Aue was a fundamental and widely recognized German author of the end of the 12th century. Currently we are working on digital editions of Hartmann's four narrative works, which are being hosted in the *heiEditions* digital edition infrastructure of the Universitätsbibliothek Heidelberg. This paper presents the main aspects of the TEI tagging model, developed in conjunction with the German institution and applied homogeneously to the four editions.

KEYWORDS

Hartmann von Aue; Digital Humanities; digital editions; XML/TEI



Magnificat Cultura i Literatura Medievals 10, 2023, 1-29.

<http://ojs.uv.es/index.php/MCLM>


ISSN 2386-8295

RESUMEN

Hartmann von Aue fue un autor fundamental alemán de finales del siglo XII, ampliamente reconocido. Actualmente trabajamos en ediciones digitales de las cuatro obras narrativas de Hartmann, que se hallan alojadas en la infraestructura de edición digital *heiEditions* de la Universitätsbibliothek Heidelberg. Este artículo presenta los aspectos principales del modelo de etiquetado en TEI desarrollado en colaboración con la citada institución alemana y aplicado homogéneamente en las cuatro ediciones.

PALABRAS CLAVE

Hartmann von Aue; Humanidades Digitales; ediciones digitales; XML/TEI

Victor Millet i Lorena Pérez Ben. 2023. 'TEI encoding in the Hartmann von Aue digital edition projects', *Magnificat Cultura i Literatura Medievals*, 10: 1-29, DOI: <https://doi.org/10.7203/MCLM.10.26027> 

This article is part of the I+D+i PID2019-105258RB-I00 project funded by the Spanish MCIN/AEI/10.13039/501100011033/, and of the grant PRE2020-092208 funded by MCIN/AEI/10.13039/501100011033/ and the ESF "The ESF invests in your future".

TABLE OF CONTENTS

1	Introduction	– 3
2	The <i>heiEditions</i> projects	– 4
3	The <code><facsimile></code> element	– 5
4	Data structure in <code><body></code>	– 7
5	Transcribed signs and highlighted letters	– 8
6	Line and 'run over line'	– 9
7	Tokenization	– 10
8	Blank spaces inside and outside of the word	– 11
9	Transpositions	– 13
10	Constructions with <code><choice></code>	– 14
11	Scribal corrections	– 18
12	Punctuation	– 19
13	Editorial treatment of blank spaces	– 20
14	Metamarks	– 20
15	Functions of <code><damage></code>	– 21
16	Uses of <code><gap></code>	– 24
17	<code><gap></code> and blank spaces	– 25
18	Annotation	– 27
19	Conclusion	– 27
20	Works Cited	– 28



1 Introduction

Hartmann von Aue was a German poet who wrote throughout the last quarter of the twelfth century, approximately. Nowadays he is regarded by critics (as he was by his contemporaries) as one of the great classics of that 'golden' German literary period extending from around 1180 to 1230. What is known today from Hartmann's *œuvre* are 18 love songs, one discursive text (the *Klage*) and four narratives. Of the latter, two are relatively short and of a legendary-religious character (*Der Arme Heinrich*, 1520 verses, and *Gregorius*, 4006 verses); the other two, undoubtedly the most important ones, are the first Arthurian novels in German literature, the *Erek* (11100 verses) and the *Iwein* (8166 verses), based on Chrétien de Troyes' romances *Erec et Enide* and *Le Chevalier au lion* (Gentry 2005, Kropik 2021, Lieb 2020, Wolf 2007).

Hartmann's works were first in being critically edited in the early days of German philology. The Grimm brothers edited *Der Arme Heinrich* in 1815 by collating two manuscripts (Grimm-Grimm 1815), and Karl Lachmann produced his first critical edition of a non-Latin text with the *Iwein* in 1827. From that time until the second half of the past century, critical editions in the tradition of Lachmann have dominated German philology, that is: editions which attempt to reconstruct the version of the text closest to the original via critical methods (Benecke-Lachmann-Wolff 1968). The last decades have witnessed a paradigm shift in favour of editions based on the 'lead manuscript method', which had only rarely been applied in the field of classical authors for academic use in German philology (Mertens 2004, Edwards 2007, Krohn 2011, Duca 2014). Simultaneously, new studies on the manuscript transmission of some of the major texts (among them the *Iwein*) have shown that, on the one hand, the Lachmannian principles do not apply to such a transmission situation, and that it is even impossible to construct a reliable *stemma* for them (Stackmann 1964). On the other hand, it has been found that a number of manuscripts of most medieval texts show a state of textual variation which may be regarded as a 'version' (Bumke 1996), although they certainly pretend to be still the 'same text'. In short, what is being illuminated is the *mouvance* of medieval textual tradition and its various gradations.

In this context, digital editions and their capacity for making the entirety of transmitted texts accessible to scholars have acquired great relevance in the Germanist debate.¹ A relevance which contrasts with the small number of existing digital editions, at least of literary texts of a certain importance or centrality (Kragl 2022). The approach of the editions of the *Hartmann von Aue - digital* series (*Der arme Heinrich - digital*, *Iwein - digital*, *Gregorius - digital*, *Erek - digital*) is not to repeat a critical mono-textual edition such as those produced by Lachmann and his followers, insofar as the result would ultimately be very similar. Neither do these projects aim to be limited to an edition based on a single manuscript, since, firstly, this goal has been achieved in most cases; and, secondly, because this approach would not provide access to the diversity of the texts. Our preference has been the multi-text edition with the idea of presenting, for each work, a series of manuscripts from different periods with editorial interventions only regarding standardization, punctuation, and correction, in order to facilitate access to the reading of different historical versions of the text. Regardless of the name of the project, the editions do not seek to reconstruct

1. Clearly, the debate around 'new philology' (Cerquiglini 1989, Nichols 1990) is also in the background, although it has not had much success in the field of German philology (Stackmann 1999).

the author's text; instead, they intend to reflect how these works have been copied and read over the centuries. It is for this reason that the concept of 'history of the text' best defines the approach of our work.

Moreover, the Heidelberg University Library, which hosts the Heidelberg University Publishing Service (heiUP), has in recent years developed an infrastructure for hosting digital editions: *heiEditions* (Effinger *et al.* 2019, Šimek 2022). The fact that the publishing service of that university is part of the library structure has allowed the centralization of projects and has drawn attention to the demand for a solid and useful infrastructure for digital editions in all fields of study. Apropos the digitalization of manuscripts, Heidelberg University Library has a very powerful visualization software called D-Work and they also have developed the principle of a 'virtual library', which entails uniting the images of disperse manuscripts on one digital platform. Being a library ensures that every image and text has a DOI and guarantees persistent preservation of the data. All editions are produced under open access licenses (Effinger *et al.* 2019). As a result, this public institution has become a publisher of digital editions, which in the mid-term will lead to significant stimuli for the dissemination of standards and improvements in the reusability of data for computer-based analyses of all kinds.

We began working on these digital editions in 2016 with Hartmann's *Der Arme Heinrich*, a text of just over 1500 verses and available in only three manuscripts and four fragments (Fernández-Millet 2018). That project enabled *heiEditions* to bring its first edition to completion in 2018. Nevertheless, the coding was comparatively rudimentary. Since then, all the code has been thoroughly revamped and professionalized, the namespace has been created and the visualization modalities of the texts are being modernized. Thus, our second project, the edition of Hartmann's *Iwein*, started in 2018, uses this renewed code, which will be presented here (Millet *et al.* ed. 2019). The code of the *Der Arme Heinrich* edition was transformed into the new one in the summer of 2021 and is pending new developments in visualizations to appear as a second edition. At the same time, the design of these new editions of Hartmann von Aue has attracted critical attention and requests for collaboration. Consequently, in the course of 2022 and in collaboration with other teams, the digital editing projects of Hartmann's other two works, the *Erek* and the *Gregorius*, have been initiated and will soon be visible online as projects under construction. For all these transcriptions and editions, a single encoding manual has been established. Hence, by the end of 2026 all of Hartman's narrative texts will be available as a digital edition, with a complete transcription of all manuscripts and fragments, and homogeneous TEI tagging of high granularity, together with corresponding virtual libraries and additional materials. There will be more than 160,000 transcribed verses and close to 900,000 words.

2 The *heiEditions* projects

As for the encoding of the texts, this is necessarily done following the TEI guidelines, although *heiEditions* has created its own namespace ('*heiEditions Concepts*'),² which primarily serves to define the attribute values. Whilst the basic ontology has been (and partly is still being) developed by the library technicians (mainly by Jakub Šimek, see Šimek 2022), the projects of Hartmann von Aue's editions have been drivers and, in many cases, have been ahead of the curve by raising

2. <<https://heieditions.github.io/concepts/concepts.html>> [accessed: 30-01-2023]. There is also an ODD, not yet published.

problems and demanding answers, and by discussing and proposing coding solutions that were then passed on to the basic code and validation scheme. In some cases, the namespace itself has been used to create certain elements where TEI did not offer a satisfactory solution or where the solution envisaged was too complex. The main example is the `<hei:initial>` element used to tag an initial, which medieval manuscripts have in abundance. The problem was that the only alternative we saw in TEI was to tag the letter with `<hi>`, but `<hi>` is reserved for highlights of all kinds, whereas an initial is not a normal highlighted letter, but rather a letter that has another entity. Some initials occupy large spaces, are historiated, extend across the margin of the text column, etc. For this reason, we finally opted for the creation and description of the `<hei:initial>` tag, which supports attributes of size, colour and type of initial.

Given that it is not our intention to produce a critical text, but rather editions of different singular manuscripts, the basic principle is that the transcription of each manuscript and all editorial operations executed on it are contained in the same XML document and with the minimum repetition of code. In the visualization, a mechanism has been developed which allows to view not only either the transcription or the fully edited text (with expansion of abbreviations, normalizations, punctuation, error correction, etc.), but also allows for progressive intermediate options. This means that a user can view the fully edited text without the punctuation or read the transcription with the abbreviations expanded.

Each manuscript or fragment is encoded in a separate file. As we do not intend to establish a critical text, we do not include an apparatus in these files.³ Rather, we collect variant readings in a stand-off encoding and will point out those we consider relevant, leaving the rest in a long list. Generally speaking, variants will be shown in context and grouped. The specific markup for the variants is still under development. Text comparison will also be assisted by the possibility of parallel visualization of several manuscripts or fragments. Moreover, we have a table showing verse-alignments, transpositions, and additions or omissions of verses.

3 The `<facsimile>` element

In this paper we will leave aside the `<teiHeader>` element, as it has little relevance for the encoding of the text. The synchronization between text and image is not yet implemented at the display level, but it is ready on the encoding side. Between the `<teiHeader>` and `<body>` we insert the `<facsimile>` element that contains a sequence of `<surface>` elements; each page of a manuscript is a `<surface>`.⁴ As far as fragments are concerned, there are cases where we have, e.g., a piece of the upper part of the page and a piece of the lower part. In these cases, the different parts of the same page have been brought together in the same image, so that we are always guided by this `<surface>`. This element carries an `@ana` attribute, whose value in the case of our testimonies is always "hc:Page", and a `@n` number, which always corresponds to the foliation number (or pagination number in paginated manuscripts). The `<surface>` element contains, firstly, an empty `<graphic>` element which carries an `@url` attribute with the address of the corresponding page image; and secondly a `<zone>` element for the general text box. This element in turn contains as many other children `<zone>` elements as there are text zones on the page. A one-column manuscript will generally have only one `<zone>`, a two-column one two, etc. However, possible marginal zones must be added when they

3. For the apparatus toolbox see Burghart 2016-19.

4. In manuscripts with more than one text we only refer to those pages containing our text.

exist, such as marginal gloss zones, claims or sheet numbering zones, among others. The <zone> element carries an @ana attribute indicating the type of zone (usually "hc:TextZone" combined with either "hc:MainColumn" or "hc:MarginalZone").⁵ A page like the one in image 1, of a double-column manuscript with a text at the bottom margin, has the description in the <facsimile> element shown in example 1.

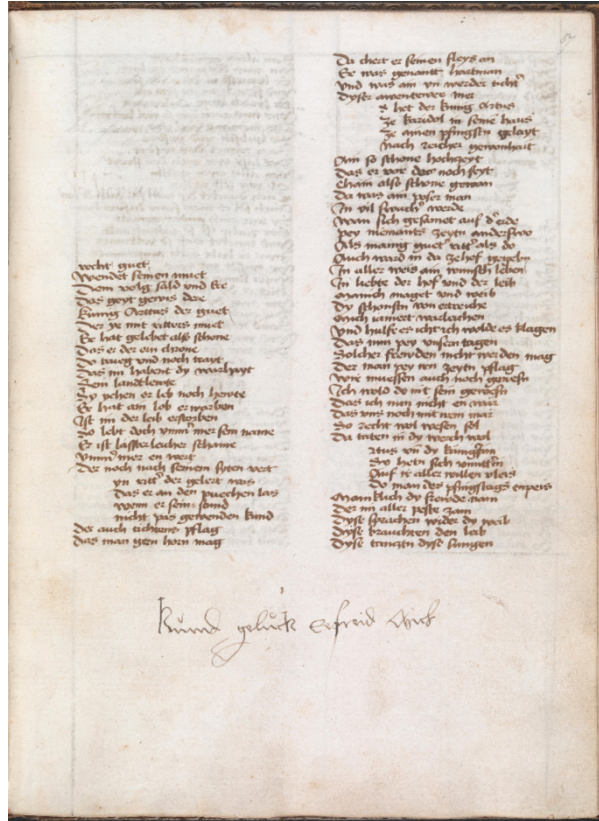


Image 1. London, British Library, Ms. Additional 19554, fol. 57r
<https://doi.org/10.11588/diglit.44385#0001> CCo

```

1. <surface ana="hc:Page" n="57r">
  <graphic url="https://digi.ub.uni-heidelberg.de/diglit/bl_add19554_foll57-100/0001"/>
  <zone>
    <zone ana="hc:MainColumn hc:TextZone" n="a"/>
    <zone ana="hc:MainColumn hc:TextZone" n="b"/>
    <zone ana="hc:MarginalZone hc:TextZone" n="c"/>
  </zone>
</surface>

```

5. The creators of heiEditions, who host editorial projects of all kinds of medieval and modern texts, decided to consistently avoid the attribute @type for classifications, which is commonly used in the TEI-guidelines, but to use @ana instead. @ana points to URIs (Uniform Resource Identifiers) of heiEditions Concepts, which will be listed in the site indicated in note 2.

4 Data structure in <body>

Regarding data structure within the <body>, it should first be noted that, like nearly all German non-religious literature before the 15th century, Hartmann von Aue's works are all composed in rhymed couplets. Rarely does a tercet or quatrain appear; usually these are caused by additions or transpositions of lines. This gives us a basic structure for semantic tagging consisting of lines of verse (tag <l>) grouped in pairs (tag <lg>); see example 2.

```
2. <lg>
   <l>...</l>
   <l>...</l>
</lg>
```

Given that our digital editions are based on several manuscripts and that most of them omit, add, and transpose verses (to a greater or lesser extent, depending on the work), the problem of numbering and identifying the verses arose. If we were to number the verses according to the order in which they appear in each manuscript, a complex system of correspondences between identical verses with different numbers would have to be established. In the end, we have adopted a double numbering system: a standard numbering that corresponds to the one usually used in research and based on some older critical edition, complemented by an alternative numbering that strictly follows the order of the lines in the manuscript; for this alternative numbering, the attribute @hei:altN was created within the namespace. Thus, each verse is given a @n attribute which takes the standard numbering, and a @hei:altN attribute with the sequential numbering particular to each manuscript. Additionally, we include an @xml:id to allow pointing in case of notes and references. We have systematized the values of the @xml:id so that they consist of "element name bearing the xml:id [underscore] standard number of the verse in which it is found (and optionally) [underscore] sequential number". The result is reflected in a structure such as in example 3.

```
3. <lg ana="hc:Couplet">
   <l n="1233" xml:id="l_1233" hei:altN="1207">...</l>
   <l n="1234" xml:id="l_1234" hei:altN="1208">...</l>
</lg>
```

In this way, the verses of two differing versions can easily be aligned or identified, whilst maintaining the option of viewing the verse-numbers per manuscript, should the researcher prefer it.

All of this is logically combined with the structural elements that mark the start of the page <pb>, the start of the column <cb> (even when the page has only one column of text) and the start of the line <lb>. If these beginnings occur between couplets or between verses, we place the tags outside the <lg> or <l> element, following the principle of always looking for the most 'external' position possible, although naturally the tags for the beginning of new pages, columns, and lines can appear in any position, when necessary. If a copyist reverses the position of the lines within the couplet or places a group of lines elsewhere, the transcription maintains the sequence of the original. Thus, for instance, using the above example, the transcription would follow the model shown in example 4, where both the <lb> and the @hei:altN show the sequence of the lines in the manuscript whereas the numbering in @n reflects that the 'normal' order of verses has been altered here.

```
4. <lg ana="hc:Couplet">
  <l n="1234" xml:id="l_1234" hei:altN="1207"><lb n="6"/>...</l>
  <l n="1233" xml:id="l_1233" hei:altN="1208"><lb n="7"/>...</l>
</lg>
```

At the editing stage we will have to decide, depending on the grammatical and semantic correctness, whether to maintain this sequence of verses or to order them following the standard numbering. In the latter case, the verses would be placed in the file in the order desired by the editor; the display in transcription mode orders the verses by the sequence of `<lb>`, not by the sequence of `<l>`, so that the original remains viewable in the transcription-reading mode and is distinguished from the display in edition-reading mode, as in example 5.

```
5. <lg ana="hc:Couplet">
  <l n="1233" xml:id="l_1233" hei:altN="1208"><lb n="7"/>...</l>
  <l n="1234" xml:id="l_1234" hei:altN="1207"><lb n="6"/>...</l>
</lg>
```

5 Transcribed signs and highlighted letters

As these editions offer the possibility of viewing all manuscripts online and even synchronizing them with their transcription, the use of allographs has been avoided. Except for the long *s* (ſ), the divergent forms of *d*, *r*, *z*, etc. have been unified to their modern Unicode sign. The precise transcription of all the allographs, in an edition with so many manuscripts from so many different periods and regions, would have involved a huge amount of work which, on the other hand, did not seem necessary considering how easy it is to verify the original grapheme. We have kept the long *s* because it has occasionally been a source of confusion in the manuscripts themselves. In any case, the allographs can be automatically unified in the visualisation, without the need for special encoding.

For diacritical or superscripted signs as well as for abbreviation signs, we have opted to use entities that replace the Unicode codes they represent; they are easier to introduce and to search for in the document, and they are automatically transformed into the corresponding Unicode sign in the visualization. Our entities are based on those developed by the Medieval Unicode Font Initiative⁶ and its application in the Medieval Nordic Text Archive.⁷ The list of entities is referenced in a special site.⁸ Following these principles, first we transcribe the base letter and then the entity (for instance: ũ → 'u+&osup;'; or: ê → 'e+&combcirc;'. We use the same mechanism to label abbreviations, as in: ñ → 'n+&bar;'; or: ð → 'd+&er;'. For other nonstandard characters we use the element `<g>`.

Highlighted letters in the manuscript are tagged with the element `<hi>` and an attribute `@rendition` that indicates the kind of highlighting. Most manuscripts highlight the first letter of every verse, either by indenting it to the left or by marking it with red color. The former is the method mostly used in parchment manuscripts, in this case the value of the attribute `@rendition` is

6. MUI [\[https://mufi.info\]](https://mufi.info) [accessed: 30-01-2023].

7. MENOTA [\[https://www.menota.org\]](https://www.menota.org) [accessed: 30-01-2023]. Haugen 2008.

8. [\[https://digi.ub.uni-heidelberg.de/schema/tei/heiEDITIONS/declarations/heieditions-entities.txt\]](https://digi.ub.uni-heidelberg.de/schema/tei/heiEDITIONS/declarations/heieditions-entities.txt) [accessed 25-03-2023]

"hc:Versal"; the latter is commonly used in paper manuscripts, the value of @rendition being here "hc:RedStroke".

6 Line and 'run over line'

Many manuscripts of literary texts in couplets, even old ones, prepare the page-layout in such a way as to copy one verse per line, often highlighting the first letter of each line or of each couplet. This produces a coincidence of structural and semantic labelling, as the beginning of the verse (a semantic unit) coincides with the beginning of a line (a structural or visual unit).⁹ German verses are anisosyllabic, which allows for the addition or suppression of new particles or words. Consequently, sometimes the physical line is too short to copy the whole verse. When this happens, the scribe usually finishes the verse at the end of either the line above or below, where there already is another verse. Generally, in these cases one can find a mark separating the main text of that line from the added words, sometimes also a second mark at the end of the unfinished verse connecting it with the words added in the other line. This is called a 'run over line'. In such cases, part of the text of one verse is in a line occupied by another verse; given that this other line is like all lines numbered, we come into conflict with the <lb> numbering, because we would have to jump back and forward in its numbering. To resolve this, we understand that in such cases the line with the excess text has two distinct segments: one in which the main text of that line is found and another in which the secondary text is located, that is, the text that comes from another line.

Since our organization of the data is mainly semantic, i.e., by verse, we tag run over lines within the verse (i.e., within the <|>). But instead of simply introducing a new <lb> in the middle of a verse (which as has been said would produce conflicts with repeated <lb> numbers), what we do is to put a <milestone> with three attributes: the first, @ana, contains the value "hc:LineSegmentBeginning" identifying the line change, and one of the values "hc:RunOverAbove" or "hc:RunOverBelow", which indicate the relative position of the run over. The second attribute will be @n="2", to indicate that it will be the second text segment of that line to which the run over goes (the first segment need not be marked). And thirdly, we add a new attribute created specifically for these cases: @hei:belongsToLine, whose value will be the number of the <lb> in which that excess text is located, marking the absolute position. Hence, a case in which the last word of a line is placed at the end of the following line (image 2) would be coded as in example 6.¹⁰

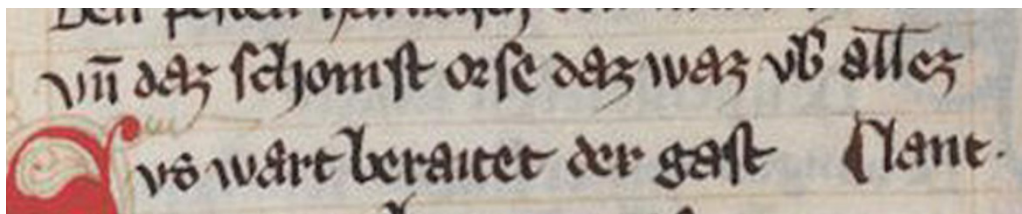


Image 2. Wien, Österreichische Nationalbibliothek, cod. 2779, fol. 56rc, l. 2-3
<https://doi.org/10.11588/diglit.44165#0021> CC-BY-NC

9. In English, 'line' is often used as a synonym of 'verse'. In this paper we strictly distinguish between a 'line', which to us is a visual phenomenon of the page, and a 'verse', which is a semantic or poetic unit.

10. In the following examples, as the cases become more complicated, we sometimes simplify the encoding a bit in order to highlight what is relevant. As a general rule, we often remove @xml:id attributes in words, word content, blank spaces between words and anything else that may be irrelevant to the specific example.

```

6. <l n="3700">
  <lb n="2"/>
  <w/><w/><w/><w/>
  <w>allez</w>
  <milestone ana="hc:LineSegmentBeginning hc:RunOverBelow" n="2" hei:belongsToLine="3"/>
  <metamark ana="hc:RunOverMark">|</metamark>
  <w>lant</w>
</l>

```

This model even allows to mark run over text which spills over two distinct lines; see example 7.

```

7. <l n="419">
  <lb n="26"/>
  <w/><w/><w/><w/>
  <milestone ana="hc:LineSegmentBeginning hc:RunOverAbove" n="2" hei:belongsToLine="25"/>
  <w/>
  <milestone ana="hc:LineSegmentBeginning hc:RunOverAbove" n="2" hei:belongsToLine="24"/>
  <w/>
</l>

```

And in all variants the model is also combinable with word breaks; in this case, however, the `<milestone>` has an additional attribute with a fixed value: `@break="no"`, to prevent the line change from automatically placing a blank space inside the word in the visualization; see example 8.

```

8. <l n="419">
  <lb n="26"/>
  <w/><w/><w/>
  <w>wer
  <metamark ana="hc:Hyphen">=</metamark>
  <milestone ana="hc:LineSegmentBeginning hc:RunOverAbove" n="2" hei:belongsToLine="25" break="no"/>den
  </w>
</l>

```

7 Tokenization

As shown in the above examples, in these editions (as in *heiEditions* projects in general) we opted for a tokenization of the text, i.e., a tagging of all words. The validation scheme checks that there is no text outside the `<w>` tag. In addition, each `<w>` tag is given an `@xml:id` so that it can be referenced. The idea behind this is to be able in the future to carry out stand-off linguistic annotation of these texts in order to improve the possibilities for automated processing and analysis. Nevertheless, this option has not been implemented for the moment, for it is still awaiting the development of a suitable tool. Tokenisation is a semantic tagging that is carried out according to linguistic criteria, not according to the existence or not of blank spaces between words in the original; on blank spaces see chapter 8.

The tokenization of medieval German texts involves certain problems due to the possibility of creating compound words (a phenomenon which, however, was not yet fully developed by 1200), the form of negation, the frequent prepositional adverbs, the separable verbal prefixes, or the possibility of proclitic and enclitic contractions. Compound nouns developed between the 12th and the 16th centuries, so that a late manuscript may regard as a compound what in 13th century writing was not yet one; in this respect we have been guided by dictionaries. German negation consisted, until the beginning of the 14th century, in the particle *en*, *ne* or *n*, which often merged with the verb or with the pronoun in enclitic or proclitic position; however, the negative particle requires its own token. Prepositional adverbs are separated in early medieval writing, which is how we tokenize them, and are joined and considered as one word in the 16th century. When a verb comes with a separable prepositional prefix we tokenize the two elements separately, whereas those cases in which the prefix is not separable are tagged as a single token, even though some scribes might leave a blank space between them. Consequently, not all linguistic words have a blank space between them in manuscripts; in other cases, we are confronted with blank spaces within the linguistic word, for some scribes separate verbal prefixes that linguistics considers inseparable or between two nouns which clearly (as the inflection reveals) form a compound.

8 Blank spaces inside and outside of the word

Tokenization, then, forces a clear distinction between the linguistic unit and the existence or absence of a blank space, whether inside or outside of the word. But XML editors, transformation languages like XSLT, and even HTML tend to remove blank spaces which they consider unnecessary, and this fact produced too many errors in the final visualization results. Therefore, we finally decided to tag spaces using the `<c>` element, the content of which is a blank space. In image 3 you can see a typical line with words and blank spaces.

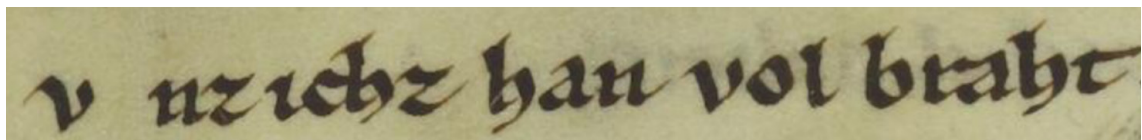


Image 3. Gießen, Universitätsbibliothek, Hs. 97, fol. 19r, l. 2
<https://doi.org/10.11588/diglit.44478#0041> CC-BY-NC

The first letter of the first word is indented to the left; therefore, we will not encode the blank space after it as such, but we encode it with `<hi>` (see chapter 5). The third visual unit *ichz* is a compound of two pronouns, *ich* plus an enclitically added *ez*. Linguistically, they are two different words, so we use two separate `<w>`-elements here, but we do not introduce a blank space between them. On the other side, the last two visual units form one linguistic word, *volbracht*, but there is a clear blank space between the prefix and the root. We would encode the line as in example 9.

```

9. <w><hi>V</hi>nz</w>
   <c> </c>
   <w>ich</w>
   <w>z</w>
   <c> </c>
   <w>han</w>
   <c> </c>
   <w>vol<c> </c>braht</w>

```

That is, wherever the manuscript has a blank space, this should be reflected with a `<c> </c>` element, both inside and outside of the word. We will discuss the editorial treatment of these spaces below.

The introduction of spaces as elements with content also has consequences for additions and deletions. Let us imagine a frequent case of a copyist who writes a line, realises afterwards that he has forgotten a word and adds it between the lines and between the words before and after. We tag this, according to TEI, with the `<add>` tag. However, if we were to encode this by adding only the word, i.e., like in example 10,

```

10. <w>word1</w>
    <c> </c>
    <add>
      <w>word2</w>
    </add>
    <w>word3</w>

```

then the visualization would be lacking a blank space between the words 2 and 3. Consequently, a blank space is needed in the `<add>` tag, as shown in example 11.

```

11. <w>word1</w>
    <c> </c>
    <add>
      <w>word2</w>
      <c> </c>
    </add>
    <w>word3</w>

```

The same happens, but in reverse, in cases where a word is deleted, as the blank space must also be included in the `` tag. But this is not necessary in cases where the copyist deletes a word to replace it with another, because in this case we use the `<subst>` element, which must include a `` and an `<add>`; but the blank space between the deleted word and the following one remains outside the `<subst>`, as it is not affected by this substitution.

Finally, the blank space is also useful in manuscripts that do not separate the verses by line. Because where the verse change occurs within the same line, i.e., without a line break between them, the display in transcription mode results in two words without separation: the last word of the first verse and the first word of the second. For this reason, we introduce a blank space between verses that do not have an `<lb>`, as shown in example 12 at the end of verse 1275.

```

12. <lg>
    <l n="1275">
      <lb n="10"/>
      <w/><w/><w/>
      <pc>.</pc>
      <c> </c>
    </l>
    <l n="1276">
      <w/><w/>
      <lb n="11"/>
      <w/>
      <pc>.</pc>
    </l>
  </lg>

```

9 Transpositions

Blank spaces must also be considered in case of word transpositions. A scribe realises that he has written two words in reverse order and marks this with a sign at the beginning of each word or above them. Although *heiEditions* has not yet introduced phase tagging, in this case we mark the entire process, i.e., we transcribe the words in the order in which they appear and then, by means of the `<listTranspose>` element, we indicate the order of the words that the copyist wanted to fix. But if we were to list only the *words* in the `<listTranspose>` element, that is, the `<w>` elements, then in the automated repositioning the blank space between them would disappear; therefore, we need to include the blank spaces in the `<listTranspose>` element and consequently add an `@xml:id` attribute to the blank spaces affected to be able to point at them from a different position. In image 4 you can see that the scribe writes *Wie ich mit worten gefarn han*, but marks with superscripted numbers that the order of the last two words should be *han gefarn*. The encoding is shown in example 13.

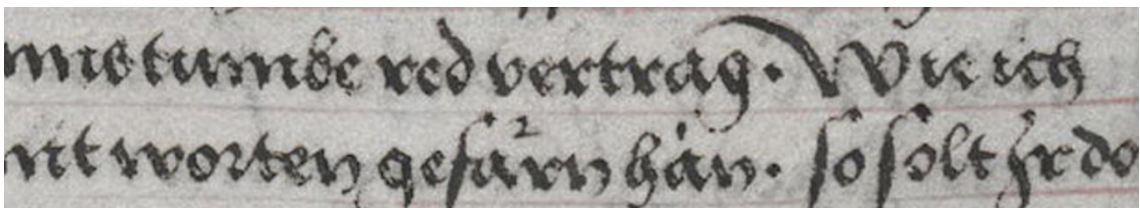


Image 4. Wien, Österreichische Nationalbibliothek, cod. ser. nova 2663, fol. XXIrc, l. 58-59
<https://doi.org/10.11588/diglit.44294#0032> CC-BY-NC

```

13. <w xml:id="w_7685_1">Wie</w>
    <c> </c>
    <w xml:id="w_7685_2">ich</w>
    <c> </c>
    <w xml:id="w_7685_3">mit</w>
    <c> </c>
    <w xml:id="w_7685_4">worten</w>
    <c> </c>
    <w xml:id="w_7685_5">gefarn</w>
    <metamark ana="hc:TranspositionMark" hei:placeRef="hc:AboveLine">1</metamark>
    <c xml:id="c_7685_1"> </c>
    <w xml:id="w_7685_6">han</w>
    <metamark ana="hc:TranspositionMark" hei:placeRef="hc:AboveLine">2</metamark>
    <listTranspose>
      <transpose corresp="#w_7685_5 #c_7685_1 #w_7685_6">
        <ptr target="#w_7685_6"></ptr>
        <ptr target="#c_7685_1"></ptr>
        <ptr target="#w_7685_5"></ptr>
      </transpose>
    </listTranspose>

```

By including the blank space in the transposition list, we avoid the risk of it disappearing in the visualization of the transposed sequence.

10 Constructions with <choice>

Our TEI tagging follows the principle of minimum redundancy or repetition of coded text. Therefore, we have opted for individual markup of each phenomenon within words. This includes for example diacritical marks, which we treat separately from the letter on which they are placed. This procedure allows us to normalize diacritics or resolve abbreviations by acting only on the sign and without the need of repeating the base letter (see chapter 5, second paragraph).

For all kind of editorial treatment of these phenomena we use <choice> constructions, i.e., alternative constructs which allow us to present the user of the editions either one or the other option. Given the high granularity of our tagging, we use these choice constructs very frequently. It has been argued (Materni 2021: 6-7) that these constructions generate very heavy code, and it is true that a simple conjunction such as *vñ* ('and') appears encoded as in example 14.

```

14. <w xml:id="w_123_4"><choice><orig>v</orig><reg>u</reg></choice>n<choice><am>&bar;
    </am><ex>d</ex></choice></w>

```

As it avoids redundancy, though, the extension of the code is unproblematic, at least in the case of our projects.

In a summarized form, we may say that we use the following types of <choice> construction:

a) For normalizations, such as in example 15.

```

15. <choice>
    <orig>v</orig>
    <reg>u</reg>
</choice>

```

Even though its most frequent use is the normalization of spellings such as v/u or j/i, we also use it for the treatment of diacritics, which allows us, as we have stated, not to touch the letter in this process. Thus, if we want to normalize the spelling ⟨û⟩ in the manuscript to present it as ⟨uo⟩, we will put it as in example 16.

```

16. û → uo
    u<choice>
      <orig>&osup;</orig>
      <reg>o</reg>
    </choice>

```

In the case of the spelling ⟨v̇⟩, since the normalization of the letter and of the diacritic are done separately, the structure will have to be duplicated, as shown in example 17.

```

17. v̇ → uo
    <choice>
      <orig>v</orig>
      <reg>u</reg>
    </choice>
    <choice>
      <orig>&osup;</orig>
      <reg>o</reg>
    </choice>

```

This may extend the code, but with the benefit of it being highly specific and of advantage when conducting systematic searches (see example 14).

b) For abbreviations we have followed the same principle, opting for constructions of the kind shown in example 18.

```

18. ē → en
    e<choice>
      <am>&bar;</am>
      <ex>n</ex>
    </choice>

```

Generally, the TEI guidelines recommend the use of dual labelling, such as is seen in example 19.

```

19. <choice>
    <abbr>
      <am></am>
    </abbr>
    <expan>
      <ex></ex>
    </expan>
  </choice>

```

But this model is rather intended for modern abbreviations, where the unit presented is itself an abbreviation and is labelled as such, and the abbreviation mark is labelled within it. Regarding medieval abbreviations, their use differs from modern abbreviations, insofar as one part of the word is written, and another is abbreviated by a sign representing one or more letters. We have therefore chosen to disregard the outer tagging of the TEI guidelines and to limit ourselves to the use of `<am>` and `<ex>`. For instance, see example 20.

```

20. werde → werden
    <w>werde
      <choice>
        <am>&bar;</am>
        <ex>n</ex>
      </choice>
    </w>

```

The code acts in a very specific way on what it wants to indicate in this case and avoids redundancies. When necessary, the letter to which the abbreviation sign is attached, can be normalized separately, as in example 21.

```

21. ũre → verre
    <w>
      <choice>
        <orig>u</orig>
        <reg>v</reg>
      </choice>
      <choice>
        <am>&er;</am>
        <ex>er</ex>
      </choice>re
    </w>

```

The only problem with this method is that it depends on the direction of the text and only works with signs that are placed above the *preceding* character; Unicode calls them 'combining characters'. This means that the model is only useful for abbreviations which expand 'forwards', i.e., where the letters introduced by the expansion are situated *behind* the letter that carries the abbreviation sign. But there are medieval abbreviations in which the expanded letters must be placed *before* the letter that carries the abbreviation sign. Hence, for example, in the above case of *werde* we find that the

abbreviation is often written differently: *werd̄n*.¹¹ This cannot be reproduced with the above model because the abbreviation sign has to be written *after* the letter; by expanding the abbreviation mark, the new letters would appear *after* the one supporting the sign, resulting in *werdne* instead of *werden*. For such cases we have developed a segmental tagging system within `<choice>` that allows the position of the letter supporting the abbreviation sign to be interchanged. See example 22.

```
22. werd̄n → werden
    <w>werd
    <choice>
      <seg ana="hc:AbbreviatedTokenSegment">n<am>&bar;</am></seg>
      <seg ana="hc:ExpandedTokenSegment"><ex>e</ex>n</seg>
    </choice>
  </w>
```

Note that the first segment contains the base letter followed by the abbreviation sign, whereas the second segment puts the expansion first, and then the base letter.

c) In addition, we use constructions with `<choice>` by combining `<orig>` with `<supplied>` within them. This is done in cases where we complete words missing some letter, due to either material loss or the nonexecution of an initial, for instance; see example 23.

```
23. werd- → werden
    <w>werd
    <choice>
      <orig>
        <gap/>
      </orig>
      <supplied>en</supplied>
    </choice>
  </w>
```

This option, logically, is only used as an editorial intervention, not as a codification of the manuscript text.

d) Next, we have the `<choice>` constructions with `<sic>` and `<corr>` for editorial corrections due to an error in the original. This applies both to the correction of letters (example 24),

```
24. mine → minne
    <w>mi<choice><sic>m</sic><corr>nn</corr></choice>e</w>
```

and to the introduction or elimination of letters (by leaving empty `<sic>` or `<corr>`, depending on the case), as shown in example 25.

¹¹ In medieval German writing, the macron rarely, and only in 15th century manuscripts, indicates a doubling of the nasal consonant. As a rule, it is either placed above a vowel to abbreviate the following nasal consonant or, as in this case and as is more frequent in the 13th century, it is placed above the nasal itself to abbreviate the vowel.

```
25. waer → wazer
  <w>wa<choice><sic/><corr>z</corr></choice>er</w>
```

And even to the suppression or introduction of whole words; in these cases, the token is within the <sic> or the <corr>, as the case may be (see example 26).

```
26. stach swert → stach daz swert
  <w>stach</w>
  <choice>
  <sic/>
  <corr>
  <w>daz</w>
  </corr>
  </choice>
  <w>swert</w>
```

e) Finally, there is another alternative construction, which has already been mentioned here, and which is not based on <choice>, but on <subst>. This element must and can contain only the and <add> elements. It has the advantage that it can also be applied to whole words corrected by the copyist, as in example 27.

```
27. rot → bla
  <subst>
  <del rendition="hc:Strikethrough">
  <w>rot</w>
  </del>
  <add hei:PlaceRef="Above">
  <w>bla</w>
  </add>
  </subst>
```

11 Scribal corrections

Scribal corrections sometimes modify a word written in the first instance by mistake or misreading and turn it into a different word. In general, we are only interested in the final result, but there are occasions when the mistake is significant or the correction has been made at a later stage, which may motivate us to tokenize the erroneous word and the corrected word separately. Nevertheless, considering that we want to avoid repeating signs and act only on the corrected letter, the technical team devised a solution in which the correction is made in the first token, and the second token remains empty (i.e., the word is not repeated), but points to the corrected version of the first token. All this, of course, within a <choice> element, so that one or the other option can be selected. Take, for example, a case where a copyist writes *hunt* ('dog') and corrects the ⟨h⟩ into a ⟨b⟩ to obtain the word *bund* ('union'), as in example 28:

```

28. hunt → bunt
  <choice>
    <w xml:id="w_1234_1">
      <subst>
        <del rendition="hc:Strikethrough" corresp="#w_1234_1">h</del>
        <add hei:placeRef="hc:AboveLine" corresp="#w_1234_2">b</add>
      </subst>unt
    </w>
    <w xml:id="w_1234_2" copyOf="#w_1234_1"></w>
  </choice>

```

In the case of editorial interventions, however, if the word in the manuscript has meaning and is attributable to a particular lemma, but the editorial intervention results in a different lemma, this double tokenization should always be effected. For example, in the case where the manuscript reads *uch* ('to you'), but the editors consider that it should read *ouch* ('also'), as in example 29.

```

29. uch → ouch
  <choice>
    <w xml:id="w_145_1">
      <choice>
        <sic corresp="#w_145_1"/>
        <corr corresp="#w_145_2">o</corr>
      </choice>ch
    </w>
    <w xml:id="w_145_2" copyOf="#w_145_1"></w>
  </choice>

```

12 Punctuation

We have already seen here the editorial interventions of both normalisation and correction of errors or substitution of words. The introduction of punctuation, as an editorial revision process, is done in much the same way with `<choice>` constructions with `<orig>` and `<reg>`. Notwithstanding, in these cases, the `<reg>` element will always contain an `@ana` attribute with the value `"hc:PunctuationRegularization"` that allows these normalizations to be identified as belonging to editorial punctuation and to be turned on or off in the visualization. Within `<reg>` the punctuation mark is itself contained within the descendant element `<pc>`; if two punctuation marks coincide, each will be contained in its own `<pc>` element. When the manuscript has no punctuation in a particular position and the editor decides that it should contain, for example, a comma, the `<orig>` element will remain empty, as in example 30.

```

30. <choice>
  <orig/>
  <reg ana="hc:PunctuationRegularization">
    <pc>,</pc>
  </reg>
</choice>

```

If the manuscript does have a sign in that position, for example a middle dot, this could be included in the element `<orig>`; see example 31.

```
31. <choice>
    <orig>
      <pc>.</pc>
    </orig>
    <reg ana="hc:PunctuationRegularization">
      <pc>,</pc>
    </reg>
  </choice>
```

13 Editorial treatment of blank spaces

Blank spaces inside a token (i.e., when a word has been written with a blank space inside it) are treated as editorial normalizations with `<choice>` constructs. We do not consider them as errors, so we operate with `<orig>` and `<reg>`. Inside a token we will only find blank spaces that the editor wants to delete, so the blank space `<c> </c>` will be inside `<orig>`, and `<reg>` will be empty, as shown in example 32.

```
32. <w>ge
    <choice>
      <orig><c> </c></orig>
      <reg/>
    </choice>slafen
  </w>
```

Outside the token, on the other hand, we intervene only when we consider that a blank space is missing. In other words, we do not join two different words when the copyist leaves a blank space between them. But there are cases in which two words are written together without this union being considered a compound word, or a case of enclisis or proclisis. In such cases we intervene in the same way as in the previous example, except that in this case the blank space is inside `<reg>` and the choice construction is between two tokens, as in example 33:

```
33. <w>ze</w>
    <choice>
      <orig/>
      <reg><c> </c></reg>
    </choice>
  <w>swert</w>
```

14 Metamarks

Marks which are neither letters nor numbers nor punctuation characters are tagged with `<metamark>`. They may be hyphens, insertion marks, correction marks, transposition marks, end-of-line marks, number delimitation marks, or marks separating two words. This

element carries the `@ana` attribute indicating the type of metamark. The most common are: `"hc:Hyphen"`, `"hc:RunOverMark"`, `"hc:WordDivider"`, `"hc:CorrectionMark"`, `"hc:InsertionMark"`, `"hc:TranspositionMark"`, `"hc:ReferenceMark"`, `"hc:WordMark"`, `"hc:NumberMark"`, `"hc:DecorativeMark"` and `"hc:LineDelimiter"`. In addition, this element carries a `@hei:PlaceRef` attribute to indicate the position of this mark: whether it is at the margin, above the line, below, etc. The reproduction of these marks can rarely be represented by Unicode signs; we try to systematize wherever possible, but it is quite free. The `<metamarks>` are visible in transcription mode but disappear automatically when viewed in edited text mode.

15 Functions of `<damage>`

One of the most complicated aspects of encoding has been the treatment of `<gap>` and `<damage>`. In the case of physical damage to the testimonies that affects the preservation of the text, but does not prevent it from being read, we use the `<damage>` element. It must always be contained within `<l>`, the letter being the minimum unit to which it can refer.

The element `<damage>` must carry the `@rendition` attribute, whose values describe the type of damage. The following are among the most frequent ones: `"hc:BrokenOff"`, `"hc:CutOff"`, `"hc:Erased"`, `"hc:PastedOver"`, `"hc:PeeledOff"`, `"hc:Perforated"`, `"hc:Stained"` y `"hc:TornOff"`. See example 34.

```
34. <l>
    <lb/>
    <w>vnde</w>
    <damage rendition="hc:PastedOver">
      <w>hat</w>
      <w>vnf</w>
      <w>der</w>
    </damage>
    <w>
      <damage rendition="hc:PastedOver">f</damage>
      <gap/>
    </w>
  </l>
```

When the damage is such that it prevents proposing an incontrovertible reading, a descendant element `<unclear>` is added inside `<damage>` to tag the uncertain part. The former does not need to carry attributes, since they are given by the parent `<damage>` element, as shown in image 5 and example 35.



Image 5. Dr. Jörn Günther Rare Books AG, Schweiz, no shelfmark, fol. 1vb, l. 4
<https://doi.org/10.11588/diglit.47871#0002> CCo

```

35. <l>
    <lb/>
    <w>da</w>
    <w>
      <damage rendition="hc:Stained hc:PastedOver">
        <unclear>b</unclear>
      </damage>if
    </w>
    <w>tu</w>
    <damage rendition="hc:Stained hc:PastedOver">
      <w>aine</w>
      <unclear>
        <w>fculdich</w>
        <w>an</w>
      </unclear>
    </damage>
  </l>

```

Damage to manuscripts is often not isolated, but rather affects large areas. A detailed labelling of all of them would be redundant. This is why it is preferable to include as much text as possible in a single element without exceeding the level of the verse (<l>), as shown in example 36.

```

36. <l>
    <lb/>
    <damage rendition="hc:Mould">
      <unclear>
        <w>daz</w>
        <w>chefet</w>
        <w>ob</w>
        <w>ir</w>
        <w>welt</w>
      </unclear>
    </damage>
  </l>

```

Similarly, we try to avoid using <damage> within the word (<w>) when the whole word is affected. Although it is possible to do so in specific cases where the damage only affects a single letter or a small set of letters, when the damage encompasses the whole word then <w> should be a descendant element of both <damage> and <unclear>, and not vice versa. We apply this principle in order to keep the encoding clean and make it easier for the human eye to read; see example 37.

```

37. <|>
    <|b/>
    <gap rendition="hc:CutOff"/>
    <w><gap rendition="hc:CutOff"/>z</w>
    <w>in</w>
    <damage rendition="hc:Stained">
      <unclear>
        <w>vil</w>
      </unclear>
    </damage>
    <w>vnchvm</w>
    <w>gena
      <damage rendition="hc:Perforated">f</damage>
    </w>
  </|>

```

When the loss of text is impossible to reconstruct with an acceptable level of certainty, we use the element `<gap>`, which already appeared in example 34. Unlike `<damage>` and `<unclear>`, the element `<gap>` has no content. However, where textual stability of the testimony allows us to infer the missing content and the damage is considered to impede the fluidity of the reading, it is possible to reconstruct the loss using a `<choice>` structure in which the source element (equivalent to `<orig>`) is the `<gap>` and the conjecture is included in the element `<supplied>` (see example 23).

We do not use this procedure to restore large textual gaps, but only in those cases where reconstruction is not very invasive and is beneficial for the reader. Nevertheless, it should be borne in mind that we are presenting a multi-text digital edition, which means that the loss of content does not always require a full reconstruction, because it can be inferred by parallel comparison with other testimonies.

The reason for textual loss is indicated by the `@rendition` attribute, whose values can be, in addition to those mentioned above for `<damage>`, `"hc:HiddenInFold"` and `"hc:Illegible"`. We also specify the type of loss with the attribute `@unit` and the values `"character"`, `"cm"`, `"leaf"`, `"line"`, `"mm"` or `"page"`; and the number of units of the loss with `@quantity`. See example 38.

```

38. <w>na
    <choice>
      <orig>
        <gap rendition="hc:TornOff" quantity="2" unit="character"/>
      </orig>
      <supplied>ch</supplied>
    </choice>
  </w>

```

Should it not be possible to calculate what has been lost, instead of `@quantity` we would use the `@extent` attribute, the value of which must always be 'unknown', as in example 39.

```

39. <gap rendition="hc:TornOff" unit="character" extent="unknown"/>

```

The `<gap>` element may be nested within `<damage>` and `<unclear>` when part of the sequence the latter contains is illegible, as in image 6 and example 40.



Image 6. St. Paul im Lavanttal, Stiftsbibliothek, Cod. 7/8, fol. 3va, l. 1

<https://doi.org/10.11588/diglit.45031#0006> CC BY-SA

```

40. <damage>
    <unclear>
      <w>
        <choice>
          <orig>
            <gap rendition="hc:CutOff" unit="character" quantity="2"/>
          </orig>
          <supplied>ku</supplied>
        </choice>nec
      </w>
    <w/><w/>
  </unclear>
</damage>

```

16 Uses of <gap>

Nonetheless, not all damage that a witness may have suffered is limited to affecting the readability of the text; in many instances there is simply material loss of text, without affecting the text around or beside that loss. In such cases, the <gap> element is used independently of <damage>. It is often found as a descendant of <l> pointing to the loss of words. If these form a succession and do not exceed the limit of a verse (<l>), in order to avoid redundancy, we use one single <gap>. When it is easy to deduce what has been lost from the logic of the sentence and from the confirmation of other manuscripts, we can even place the <gap> within <orig> in a construction with <choice> and provide the lost text within the <supplied> element, as in image 5 and example 41.



Image 7. Kassel, Universitätsbibliothek, Landesbibliothek und Murhardsche Bibliothek der Stadt Kassel, 2° Ms. philol. 28[3, fol. 1ra, l. 16

<https://doi.org/10.11588/diglit.44205#0001> CCo


```

41. <w>Der</w>
    <c> </c>
    <w>wirt</w>
    <c> </c>
    <w>m
      <choice>
        <orig>
          <gap rendition="hc:Perforated" unit="character" quantity="3"/>
        </orig>
        <supplied>ich</supplied>
      </choice>
    </w>
<choice><orig/><reg><c> </c></reg></choice>
<choice>
  <orig>
    <gap rendition="hc:Perforated" unit="character" quantity="5"/>
  </orig>
  <supplied>
    <w>an</w>
    <c> </c>
    <w>der</w>
  </supplied>
</choice>

```

Conversely, a single affected part may comprise the loss both of complete words and of some letters of another word. In such cases, there is loss inside and outside the <w>. Thus, we need a <gap> element both inside the token, tagging the lost letters of this word, and another <gap> element outside (be it before or after) the token, representing those lost words, as shown in example 42.

```

42. <l>
    <w>fin
      <gap rendition="hc:Faded" unit="character" quantity="3"/>
    </w>
    <gap rendition="hc:Illegible" unit="character" extent="unknown"/>
  </l>

```

Even though the goal is always to avoid complex and redundant code, when dealing with severely corrupted testimonies the proliferation of <gap> elements is unavoidable.

17 <gap> and blank spaces

It should also be noted that <gap> elements outside the token must be separated from the adjacent <w> by an optional blank space added by the editor (see chapter 13), in order to distinguish between the loss of a part of the word and the loss of other words in the verse. See image 8 and example 43.



Image 8. Klagenfurt, Universitätsbibliothek, Perg.-Hs. 63 /
Schlägl, Stiftsbibliothek, no shelfmark, fol. 1ra, l.1
<https://doi.org/10.11588/diglit.48570#0001> CC-BY-NC

```
43. <I>
  <lb/>
  <gap rendition="hc:CutOff"/>
  <choice>
    <orig/>
    <reg>
      <c> </c>
    </reg>
  </choice>
  <w><gap rendition="hc:CutOff"/>e</w>
  <c> </c>
  <w>ir</w><w>wille</w><w>vñ</w><w>ir</w><w>heil</w>
</I>
```

In short, even if it is the same loss, at the time of encoding it can be divided into several `<gap>` elements which can be placed in different `<I>` elements and which we separate from preserved words by means of optional blank spaces.

If necessary, we also add this separation before and/or after lines whose content is a single `<gap>`. When these phenomena occur, we always code them within a line (`<I>`) and within the couplet to which they correspond. This `<I>` tag may or may not carry the relevant attributes (see chapter 4): if it is possible to infer the identity of the missing line even if it is illegible, we number it accordingly. Otherwise, we leave the `<I>` element without attributes, as shown in image 9 and example 44.



Image 9. Klagenfurt, Universitätsbibliothek und Stiftsbibliothek Schlägl, Perg.-Hs. 63
(Klagenfurt); no shelfmark (Schlägl), fol. 1ra, l. 9
<https://doi.org/10.11588/diglit.48570#0001> CC-BY-NC

```
44. <lg>
  <I>
    <lb n="9"/>
    <gap rendition="hc:Illegible"/>
  </I>
  <I><w><c> </c><w><c> </c><gap/></I>
</lg>
```

Finally, if the loss covers more than one verse, we simply place the `<gap>` element outside of the `<|>` and `<lg>` and identify the loss with the corresponding attributes and values, as in example 45:

```
45. <gap rendition="hc:Lost" quantity="1" unit="leaf"/>
```

18 Annotation

In the field of annotation there are two working procedures: the occasional annotation and the critical apparatus. The occasional annotation is simple, since it consists of a `<note>` element that we place outside the verse to which it refers. This element carries an attribute `@ana` which can have very different values, as it includes both editorial annotation and annotation in the original (glosses, *maniculæ*, source indications, etc.). Since our texts lack complex original annotation (at most a few marginal glosses), we use `<note>` mainly for notes that comment on phenomena or particularities of the text or the medium, or to comment on editorial decisions or make comments on the text. For this reason, the `@ana` values in `<note>` that we use are usually `"hc:TranscriptionNote"` or `"hc:EditorialNote"`, and in some cases `"hc:Gloss"`.

The critical apparatus is a different case, though. The future idea of *heiEditions* is to develop a technology which enables automatizing the generation of the critical apparatus. But it is uncertain when this will be developed. Thus, the idea of the *Hartmann von Aue - digital* projects is to create a manual critical apparatus in a separate file, i.e., as stand-off markup. The encoding of this file has not yet been finalized, but the idea is to take advantage of our tokenization and of the fact that every `<w>` carries an `@xml:id`. In this way, it is possible to point to segments of, for instance, three or four tokens of a manuscript or a set of testimonies, and then point to the manuscripts that have a relevant variant for that segment. The apparatus is intended to be displayed in a side panel.

19 Conclusion

Although we cannot yet offer an encoding system in advanced aspects such as this last one, we do believe that the encoding model we have established together with *heiEditions* will serve as a reference for other projects, both because the projects encompassed in *Hartmann von Aue - digital* will offer the whole set of manuscripts and fragments of the complete works of a classical and canonical author in academic training, and because *heiEditions* is establishing itself as a 'publisher' of digital editions of German (and soon also French) medieval texts. In the medium term, we will probably see the accumulation of two to three million tokens of medieval manuscript transcriptions, which is a considerable text database for future research.

20 Works Cited

- Benecke, Georg Friedrich; Lachmann, Karl (ed.). 1827. *Iwein: Der Ritter mit dem Lewen getihtet von dem Hern Hartman, Dienstman ze Ouwe* (Berlin: Reimer) <<https://doi.org/10.11588/diglit.43793>>
- Benecke, Georg F.; Lachmann, Karl; Wolff, Ludwig (ed.). 1968. *Iwein: eine Erzählung von Hartmann von Aue* (Berlin: de Gruyter), 7th edn
- Bumke, Joachim. 1996. *Die vier Fassungen der 'Nibelungenklage': Untersuchungen zur Überlieferungsgeschichte und Textkritik der höfischen Epik im 13. Jahrhundert* (Berlin: de Gruyter) <<https://doi.org/10.1515/9783110812923>>
- Burghart, Marjorie. 2016. 'The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features', *Journal of the Text Encoding Initiative*, 10 <<https://doi.org/10.4000/jtei.1520>>
- Cerquiglini, Bernard. 1989. *Éloge de la variante: Histoire critique de la philologie* (Paris: du Seuil)
- Duca, Patrick del (ed.). 2014. *Hartmann von Aue Iwein: Édition établie d'après le manuscrit 97 de Giessen (version B)*, Textes Vernaculaires du Moyen Âge, 13 (Turnhout: Brepols)
- Edwards, Cyril (ed.). 2007. *Hartmann von Aue Iwein or The Knight with the Lion: Edited from Manuscript B, Gießen, Universitätsbibliothek Codex Nr. 97 and Translated*, Arthurian Archives, 16 - German Romance III (Woodbridge: D.S. Brewer)
- Effinger, Maria; Maylein, Leonhard; Šimek, Jakub. 2019. 'Von der elektronischen Bibliothek zur innovativen Forschungsinfrastruktur: Digitale Angebote für die Geisteswissenschaften an der Universitätsbibliothek Heidelberg', *Bibliothek Forschung und Praxis*, 43: 311-323 <<https://doi.org/10.1515/bfp-2019-2067>>
- Fernández Riva, Gustavo; Millet, Victor (ed.). 2018. *Hartmann von Aue Der arme Heinrich: Textgeschichtliche elektronische Ausgabe* (Heidelberg: Universitätsbibliothek) <<https://doi.org/10.11588/edition.ahd>>
- Fradejas Rueda, José Manuel. 2009-10. 'La codificación XML/TEI de textos medievales', *Memorabilia*, 12: 219-247 <<https://go.uv.es/omQeXkP>>
- Gentry, Francis G. 2005. *A Companion to the Works of Hartmann von Aue* (Rochester NY: Camden House)
- Grimm, Jacob; Grimm, Wilhelm (ed.). 1815. *Der arme Heinrich von Hartmann von Aue: aus der Straßburgischen und Vatikanschen Handschrift* (Berlin: Realschulbuchhandlung) <<https://doi.org/10.11588/diglit.29724#0005>>
- Haugen, Odd Einar. 2008. *The Menota Handbook: Guidelines for the Electronic Encoding of Medieval Nordic Primary Sources* (Bergen: Clarino Bergen Center, UiB) <https://www.menota.org/HB2_index.xml>
- Kragl, Florian. 2022. 'Jenseits der Textkritik: Überlegungen zu einer Theorie der digitalen Edition', in *Lachmanns Erbe: Editionsmethoden in klassischer Philologie und germanistische Mediävistik*, ed. by Anna Kathrin Bleuler and Oliver Primavesi, *Zeitschrift für deutsche Philologie*, Beiheft 19 (Berlin: E. Schmidt), pp. 575-584

- Krohn, Rüdiger (ed.). 2011. *Hartmann von Aue Iwein* (Stuttgart: Reclam)
- Kropik, Cordula. 2021. *Hartmann von Aue: eine literaturwissenschaftliche Einführung* (Tübingen: Narr Francke Attempto) <<https://doi.org/10.36198/9783838555621>>
- Lieb, Ludger. 2020. *Hartmann von Aue: Erec - Iwein - Gregorius - Armer Heinrich* (Berlin: E. Schmidt)
- Materni, Marta. 2021. 'Les Guidelines TEI à l'épreuve de la complexité textuelle et graphique médiévale', *Magnificat Cultura i Literatura Medievals*, 8: 1-32 <<https://doi.org/10.7203/MCLM.7.16724>>
- Mertens, Volker (ed.) 2004. *Hartmann von Aue Gregorius, Der arme Heinrich, Iwein*, Bibliothek des Mittelalters, 6 (Frankfurt am Main: Deutscher Klassiker Verlag)
- Millet, Victor; González Miranda, Emilio; Pérez Ben, Lorena (ed.). 2019. *Hartmann von Aue Iwein: Textgeschichtliche elektronische Ausgabe* (Heidelberg: Universitätsbibliothek) <<https://doi.org/10.11588/edition.iwd>>
- Nichols, Stephen. 1990. 'Introduction: Philology in a Manuscript Culture', *Speculum*, 65: 1-10 <<https://doi.org/10.2307/2864468>>
- Šimek, Jakub. 2022. 'heiEDITIONS: eine Heidelberger Infrastruktur für Editionen (nicht nur) mittelalterlicher Texte', in *Digitale Mediävistik: Perspektiven der Digital Humanities für die Altgermanistik*, ed. by Elisabeth Lienert et alii (Oldenburg: Universität), pp. 27-46 <<https://doi.org/10.25619/BmE20223193>>
- Stackmann, Karl. 1964. 'Mittelalterliche Texte als Aufgabe', in *Festschrift für Jost Trier zum 70. Geburtstag*, ed. by William Foerste and Karl Heinz Borck (Köln-Graz: Böhlau), pp. 240-267 [repr. in: Stackmann, Karl. 1997. *Mittelalterliche Texte als Aufgabe: Kleine Schriften*, 1 (Göttingen: Vandenhoeck & Ruprecht), pp. 1-25 <<https://tinyurl.com/5yvb72cm>>]
- Stackmann, Karl. 1999. 'Neue Philologie?', in *Modernes Mittelalter: neue Bilder einer populären Epoche*, ed by Joachim Heinze (Frankfurt: Insel), pp. 398-427
- Wolf, Jürgen. 2007. *Einführung in das Werk Hartmanns von Aue* (Darmstadt: Wissenschaftliche Buchgesellschaft)