

Chapter 3. Levels of text representation

3.1 Introduction

A transcription is basically a representation of a primary source in another format, such as paper or the electronic medium. Some transcriptions aim to reproduce the source text as closely as possible, others allow for a certain amount of generalisation. In transcriptions of speech, a distinction is usually drawn between narrow and broad transcriptions, depending on the amount of phonetic detail. The same perspective applies to transcriptions of manuscript texts. Close (or narrow) transcriptions are usually referred to as diplomatic, while regularised transcriptions are often referred to as normalised. This is the basic distinction drawn in e.g. [Wittgenstein's Nachlass: The Bergen Electronic Edition](#) (1998-2000). Here, all texts are available in two versions, a diplomatic transcription and a normalised one. For examples, please refer to [this page](#).

We suggest that medieval Nordic texts may be transcribed on up to three levels. In addition to the **normalised** level, we identify two closer levels. We shall refer to the narrowest level as the **facsimile** level, while the 'medium' level is designated as **diplomatic**. The three levels are exemplified in [ch. 3.2](#) below.

The distinction between three levels of text representation does not mean that a Menota transcription should contain all three levels. Many transcribers will probably choose a single level for their transcription. Our recommendation is to use these levels as a guide, so that a transcription can be described as following one of these levels. This information should be given in the header, and can optionally be given by use of specific elements in the transcription itself, as discussed in [ch. 3.2](#) below. If a transcriber wishes to deviate from any of these levels, and there may be good reasons to do so, we recommend that the deviations are specified in the header.

It is convenient to begin by looking at a Latin text example, *Passio et Miracula Beati Olavi*. An important source for this work is Corpus Christi College, Oxford MS 209, a vellum manuscript from the late 12th century. Below is a low resolution facsimile from the beginning of the *Passio*. For a facsimile of the whole manuscript in high resolution, please refer to [Early Manuscripts at Oxford University](#).

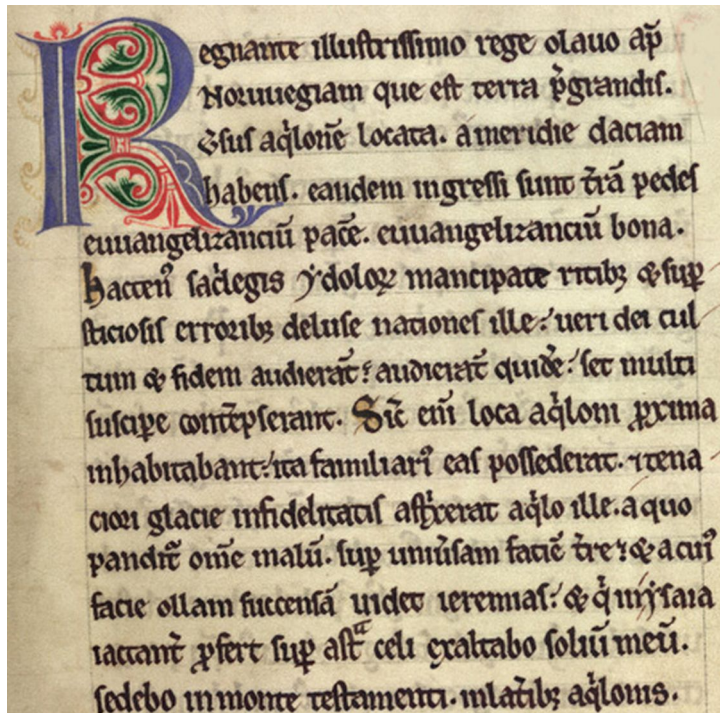


Fig. 3.1 CCC 209, fol. 57r., l. 1-15. © Corpus Christi College, Oxford

The following elements and attributes will be used in the encoding:

| Element / attribute | Contents |
|---------------------|--|
| <head> | contains a heading (or title) |
| <div> | contains a section of the text; can be nested hierarchically |
| @type | specifies the type of section |
| @n | specifies the number of a section |
| <p> | contains a paragraph of text |
| <hi> | contains a highlighted part of the text, e.g. by way of bold, italics, underlining |
| @rend | specifies how the highlighted text is rendered, e.g. by way of italics |

Using these elements and attributes to describe the structure of the text, the manuscript can be transcribed straight away:

```
<head>Passio et miracula beati Olavi</head>

<div type="section" n="1">
  <p><hi rend="blue">R</hi>egnante illustrissimo rege Olauo
  apud Noruuegiam, que est terra pregrandis uersus aquilonem
  locata, a meridie Daciam habens, eandem ingressi sunt terram
  pedes euuangelizantium pacem, euuangelizantium bona.</p>
</div>

<div type="section" n="2">
  <p>Hactenus sacrilegis ydolorum mancipate ritibus et
  supersticiosi erroribus deluse nationes ille ueri Dei cultum
  et fidem audierant; audierant quidem, set multi suscipere
  contempserant.</p>
```

```

</div>

<div type="section" n="3">
  <p>Sicut enim loca aquiloni proxima inhabitabant, ita
    familiarius eas possederat et tenaciori glacie
    infidelitatis astrinxerat aquilo ille, a quo panditur omne
    malum super uniuersam faciem terre, et a cuius facie ollam
    succensam uidet Ieremias, et qui in Ysaia iactanter profert:</p>
</div>

<div type="section" n="4">
  <p><quote>Super astra celi exaltabo solium meum, sedebo in
    monte testamenti in lateribus Aquilonis.</quote></p>
</div>

```

(Adapted from an edition by Lars Boje Mortensen, University of Bergen. Cf. also the edition by [Frederick Metcalfe 1881](#).)

(When the illustrious King Óláfr ruled in Norway, a vast country located towards the north and having Denmark to the south, there entered into that land the feet of them that preach the gospel of peace and bring glad tidings of good things. The peoples of that country, previously subject to the ungodly rites of idolatry and deluded by superstitious error, now heard of the worship and faith of the true God – heard indeed, but many scorned to accept. Living in a region close to the north, it was the same north, from which comes every evil over the whole face of earth, that had possessed them all the more inwardly and gripped them all the more firmly in the ice of unbelief. From its face Jeremiah saw a seething pot; and in Isaiah there is the boaster who says, "I will exalt my throne above the stars of God: I will sit also upon the mount of the congregation, in the sides of the north.") [Translated by [Devra Kunin 2001](#).]

The transcription above is easily readable, even in its ‘raw’ XML format. In fact, if it was stripped for all elements, it would look like a plain ASCII text from any word processor:

Passio et miracula beati Olavi Regnante illustrissimo rege Olauo apud Noruuegiam, que est terra pregrandis uersus aquilonem locata, a meridie Daciam habens, eandem ingressi sunt terram pedes euuangelizantium pacem, euuangelizantium bona. Hactenus sacrilegis ydolorum mancipate ritibus et supersticiosis erroribus deluse nationes ille ueri Dei cultum et fidem audierant; audierant quidem, set multi suscipere contempserant. Sicut enim loca aquiloni proxima inhabitabant, ita familiarius eas possederat et tenaciori glacie infidelitatis astrinxerat aquilo ille, a quo panditur omne malum super uniuersam faciem terre, et a cuius facie ollam succensam uidet Ieremias, et qui in Ysaia iactanter profert: Super astra celi exaltabo solium meum, sedebo in monte testamenti in lateribus Aquilonis.

With the help of an XML style sheet, the text could be displayed with a certain amount of formatting on the basis of the mark-up. For example, the title (<head>) might be shown in bold type, the initial might be rendered with an enlarged capital in blue colour, sections might be set out in separate paragraphs and numbered in bold type, and the Biblical quotation could be given in italics:

Passio et miracula beati Olavi

1. Regnante illustrissimo rege Olauo apud Noruuegiam, que est terra pregrandis uersus aquilonem locata, a meridie Daciam habens, eandem ingressi sunt terram pedes euuangelizantium pacem, euuangelizantium bona.

2. Hactenus sacrilegis ydolorum mancipate ritibus et supersticiosis erroribus deluse nationes ille ueri Dei cultum et fidem audierant; audierant quidem, set multi suscipere contempserant.
3. Sicut enim loca aquiloni proxima inhabitabant, ita familiarius eas possederat et tenaciori glacie infidelitatis astrinxerat aquilo ille, a quo panditur omne malum super uniuersam faciem terre, et a cuius facie ollam succensam uidet Ieremias, et qui in Ysaia iactanter profert:
4. *Super astra celi exaltabo solium meum, sedebo in monte testamenti in lateribus Aquilonis.*

Medieval Nordic texts need not contain any more mark-up than in this example, and they will still be fully valid XML. However, in order to comply with the Menota standard, it should follow the TEI guidelines. A few more elements and attributes will be needed for this:

| Element / attribute | Contents |
|--------------------------------|--|
| <code><TEI></code> | states that the contents of this element is a single TEI conformant document, comprising a header and a text |
| <code><teiHeader></code> | contains structured information on the text according to the recommendations by TEI |
| <code><text></code> | contains the text |
| <code><body></code> | contains the body of the text, excluding and front or back matter |

The basic structure of the file is thus quite simple:

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
  xmlns:me="http://www.menota.org/ns/1.0">
  <teiHeader>
    Here goes structured information on the text and the transcription.
  </teiHeader>
  <text>
    <body>
      Here goes the text as exemplified above.
    </body>
  </text>
</TEI>
```

For an example of a Menota header, please go to [Appendix E](#).

It is important to keep in mind that a transcription may be as straightforward and readable as this, and it would be fully acceptable as a Menota text.

However, not all primary sources are equally straightforward to transcribe. For most vernacular sources, entities will be required to deal with additional characters, and we might also like to transcribe the text at a more diplomatic level than in this example. For example, the last word on the very first line is transcribed as ‘apud’. In the facsimile above, we see that it has been written with the letters ‘ap’ and a superlinear abbreviation mark. Some transcribers might want to record the fact that there is an abbreviation mark at this point, or they might want to show how this abbreviation should be expanded. Two elements will be needed for this:

| Element / attribute | Contents |
|---------------------|---|
| <ex> | contains the text of an expanded abbreviation |
| <am> | contains an abbreviation marker |

The characters that should be added, i.e. the expansion of the abbreviation, will be contained in the <ex> element:

```
ap<ex>ud</ex>
```

Other transcribers would like to encode the actual abbreviation mark being used, in this case a superlinear bar. This might be encoded with the help of an entity such as '&bar;', meaning 'a horizontal bar placed above the preceding character'. The element <am> (for 'abbreviation marker') indicates that the bar is an abbreviation mark, and not, for example, a sign for length (i.e. a macron) or a diacritical sign (like the bar sometimes used above 'u' to distinguish this character from 'n'):

```
ap<am>&bar;i</am>
```

Yet other transcribers would like to encode the fact that the word has been abbreviated with a superlinear bar AND that this abbreviation should be expanded as 'ud' in this particular context. The superlinear bar is highly ambiguous; in this short extract alone, it should be expanded as 'm' in 'terram' (l. 4), 'ut' in 'Sicut' (l. 9), 'ni' in 'enim' (l. 9), 'n' in 'omne' (l. 12).

3.2 Levels of text representation

We believe that there are three focal levels of text representation for medieval Nordic texts and suggest that a transcription should reflect at least one of these levels. Furthermore, a transcription should be easily expandable so as to accommodate one or two additional levels. Since these levels have not been defined by TEI, they have been added in the Menota namespace, 'me':

| Element / attribute | Contents |
|---------------------|--|
| <me:fac> | contains a reading on a facsimile level |
| <me:dipl> | contains a reading on a diplomatic level |
| <me:norm> | contains a reading on a normalised level |

Note: The 'me' prefix can only be used with a RELAX NG schema. It must be left out in texts which will be validated against a DTD.

This time, we shall use a short extract from an Old Icelandic manuscript, AM 233 a fol (first quarter of the 13th century) as an example:

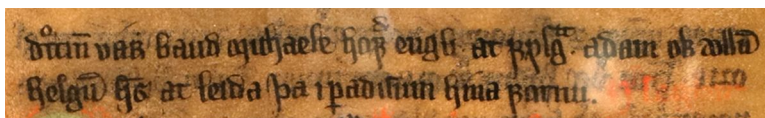


Fig. 3.2 AM 233a fol. 28v, l. 1-2. This is a fragment of *Niðrstigningar saga*, an Old Norse translation of the apocryphal *Evangelium Nicodemi*. A fuller example can be found in [Haugen and Pichler 2005](#), 220-22.

3.2.1 Facsimile level

On this level, the text is transcribed character by character, line by line. Allographic variation is to a great extent reflected in the transcription, and abbreviation marks are copied without any expansion. Thus, the text in fig. 3.2 would be transcribed as

```
&drot; &osup;ttin&bar; vá&rscapdot; bau&drot; michaele ho&fins; &dsup; engli.  
at &fins;ylg&ra; a&drot;am ok &aolig;llu&bar; helgu&bar; &hbar;s at lei&drot;a  
þa i &pbardes;a&drot;i&slong;um hína &fins;ornu.
```

and displayed (subject to an appropriate font) as

ðrottinn várr bauð michaele hofð engli. at fylgð adam ok ællu
helgū h̄s at leida þa i paradifum hína fornu.

Fig. 3.3 Facsimile rendering of the example text in fig. 3.2 using the font Andron.

At the facsimile level, the transcriber ought to encode the manuscript exactly as it reads, even if it contains obvious mistakes. Corrections can be made by inserting a note, or it can be left to the diplomatic or normalised level.

3.2.2 Diplomatic level

On this level, not all types of allographic variation are transcribed, and line divisions are usually not shown in the display of the transcription. In the transcription, expansions are set out by the element `<ex>` and in the display usually by italics. The text would then be transcribed as

```
d<ex>ro</ex>ttin<ex>n</ex> vá&rscapdot; bauð michaele hof<ex>uð</ex> engli.  
at fylg<ex>ia</ex>. adam ok &aolig;llu<ex>m</ex> helgu<ex>m</ex> h<ex>an</ex>s  
at leida þa i p<ex>ar</ex>adi&slong;um hína fornu.
```

and displayed as (now disregarding the line break)

drottinn várr bauð michaele hofuð engli. at fylgja. adam
ok ællum helgum hans at leida þa i paradifum hína fornu.

Fig. 3.4 Diplomatic rendering of the example text in fig. 3.2 (in Andron).

3.2.3 Normalised level

On this level, the orthography is regularised according to the norm found in grammars and dictionaries for the language in question. For Old Icelandic and Old Norwegian texts we recommend the normalisation rules in [AMKO's dictionary](#) (ONP). Abbreviations are expanded silently, and punctuation is regularised as well. Thus, the text in fig. 3.1 would be transcribed as

```
Dróttinn várr bauð Michaele h&oogon;fuðengli at fylgja Adam ok &oogon;llum  
helgum hans at leiða þá í paradísnum hina fornu.
```

and displayed as

Dróttinn várr bauð Michaele hofuðengli at fylgja Adam
ok öllum helgum hans at leiða þá í paradísu hina fornu.

Fig. 3.5 Normalised rendering of the example text in fig. 3.2 (in Andron).

Note that at this level all characters have been encoded using official Unicode code points. So rather than encoding the character ‘ð’ with the entity ‘ð’ it has been encoded simply as ‘ð’, using its code point in Latin-1 Supplement, 00F0. The only exception here is the ‘o ogonek’, which for practical purposes has been encoded with the entity ‘&oogon;’, even if this character, too, has a Unicode code point, 01EB in Latin Extended-B. A suitable keyboard layout is helpful for the actual typing of some of these characters, but in general, all Medieval Nordic texts can be encoded without resorting to entities as long as the text is rendered on a normalised level. Many Old Swedish and Old Danish texts can also be encoded with a minimal amount of character entities.

For a more detailed discussion of the three levels discussed here, please refer to [Haugen 2004](#).

For keyboard layouts and for an overview of Unicode code charts, please see the [MUFI site](#).

3.3 Single-level transcriptions

A transcription of a Medieval Nordic manuscript may be as simple as the Latin example in section 3.1 above. In a typical diplomatic edition, abbreviations are expanded and sometimes proper names are capitalised. The text in fig. 3.2 above could thus be encoded as:

```
<p>Drottinn várr baud Michaele hofud engli. at fylgia. Adam ok &aolig;llum  
helgum hans at leida þa i paradísu hína fornu.</p>
```

Here, abbreviations have been expanded silently and proper names capitalised, but the punctuation and orthography remain unchanged. The small capital ‘R’ with a dot above in the second word has been interpreted as a geminate, ‘rr’, while the ligature of ‘a’ and ‘o’ is transcribed with an entity, since this character was not part of the Unicode Standard as of v. 5.0. A correct display of this character thus requires a specific font with a glyph in the Private Use Area (as explained in [ch. 2](#)).

If the text is going to be annotated on a lexicographical level, we recommend that each word is contained in a `<w>` element. Although not strictly necessary, it is also helpful to identify the level of transcription within each word. This is especially so if the text is going into an archive of texts transcribed on several levels. In the following example, it is clearly stated that each word has been transcribed on a diplomatic level, identified by the `<me:dipl>` element:

```
<w>  
  <me:dipl>d<ex>ro</ex>t<ex>tin<ex>n</ex></me:dipl>  
</w>  
  
<w>  
  <me:dipl>várr</me:dipl>  
</w>  
  
<w>  
  <me:dipl>baud</me:dipl>  
</w>
```

```

<w>
  <me:dipl>michaele</me:dipl>
</w>

<w>
  <me:dipl>hof<ex>ud</ex></me:dipl>
</w>

<w>
  <me:dipl>engli</me:dipl>
</w>

etc.

```

As a rule of thumb, if one removes all mark-up in a single-level transcription, the result is a fully readable text. The text of the example above is thus simply:

```
drottinn v arr baud michaele hofud engli
```

This is equivalent to saving a formatted word processor file in a Text Only format. The text string is unchanged, but all information contained in the mark-up is lost.

3.4 Multi-level transcriptions

The transcriptions in [ch. 3.2](#) each reflect a specific level of text representation. However, we believe that the transcription should be expandable to accommodate more than one level. We recommend using the `<w>` element to group each lexical word in the transcription, as explained in [ch. 2.3](#). Within each `<w>` element, the `<choice>` element should be used to group levels of text representation. Each level is identified by descriptive elements: `<me:fac>` for the facsimile rendering (in which the element `<am>` is used for abbreviations), `<me:dipl>` for the diplomatic rendering (in which the element `<ex>` is used for expansions), and `<me:norm>` for the normalised rendering. This makes for a parallel encoding, in which up to three text strings co-exist within the boundaries of the `<w>` elements. Similarly, punctuation marks appear within the `<me:punct>` element.

| Element / attribute | Contents |
|-------------------------------|---|
| <code><choice></code> | groups a number of alternative encodings for the same point in the text |
| <code><me:punct></code> | contains a punctuation mark |

Note: The ‘me’ prefix can only be used with a RELAX NG schema. It must be left out in texts which will be validated against a DTD.

For the sake of clarity, in the following example we have set out each word in a paragraph of its own:

```

<w>
  <choice>
    <me:fac>&drot ; <am>&osup ; </am>t tin<am>&bar ; </am></me:fac>
    <me:dipl>d<ex>ro</ex>t tin<ex>n</ex></me:dipl>
    <me:norm>Dr ttinn</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>v &rscapdot ; </me:fac>
    <me:dipl>v &rscapdot ; </me:dipl>
  </choice>
</w>

```

```

    <me: norm>várr</me: norm>
  </choice>
</w>

<w>
  <choice>
    <me: facts>bau&drot;</me: facts>
    <me: dipl>baud</me: dipl>
    <me: norm>bauð</me: norm>
  </choice>
</w>

<w>
  <choice>
    <me: facts>michaele</me: facts>
    <me: dipl>michaele</me: dipl>
    <me: norm>Michaele</me: norm>
  </choice>
</w>

<w>
  <choice>
    <me: facts>ho&fins;<am>&dsup;</am> engli</me: facts>
    <me: dipl>hof<ex>ud</ex> engli</me: dipl>
    <me: norm>h&oogon;fuðengli</me: norm>
  </choice>
</w>

<me: punct>
  <choice>
    <me: facts>.</me: facts>
    <me: dipl>.</me: dipl>
    <me: norm></me: norm>
  </choice>
</me: punct>

<w>
  <choice>
    <me: facts>at</me: facts>
    <me: dipl>at</me: dipl>
    <me: norm>at</me: norm>
  </choice>
</w>

<w>
  <choice>
    <me: facts>&fins;ylg<am>&ra;</am></me: facts>
    <me: dipl>fylg<ex>ia</ex></me: dipl>
    <me: norm>fylgja</me: norm>
  </choice>
</w>

<me: punct>
  <choice>
    <me: facts>.</me: facts>
    <me: dipl>.</me: dipl>
    <me: norm></me: norm>
  </choice>
</me: punct>

<w>
  <choice>
    <me: facts>a&drot;am</me: facts>
    <me: dipl>adam</me: dipl>
    <me: norm>Adam</me: norm>
  </choice>
</w>

```

```

<w>
  <choice>
    <me:fac>ok</me:fac>
    <me:dipl>ok</me:dipl>
    <me:norm>ok</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>&aolig;llu<am>&bar;</am></me:fac>
    <me:dipl>&aolig;llu<ex>m</ex></me:dipl>
    <me:norm>&oogon;llum</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>helgu<am>&bar;</am></me:fac>
    <me:dipl>helgu<ex>m</ex></me:dipl>
    <me:norm>helgum</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>h<am>&bar;</am>s</me:fac>
    <me:dipl>h<ex>an</ex>s</me:dipl>
    <me:norm>hans</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>at</me:fac>
    <me:dipl>at</me:dipl>
    <me:norm>at</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>lei&drot;a</me:fac>
    <me:dipl>leida</me:dipl>
    <me:norm>leiða</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>pa</me:fac>
    <me:dipl>þa</me:dipl>
    <me:norm>þá</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac>i</me:fac>
    <me:dipl>i</me:dipl>
    <me:norm>í</me:norm>
  </choice>
</w>

<w>
  <choice>
    <me:fac><am>&pbardes;</am>a&drot;i&slong;um</me:fac>

```

```

      <me:dipl>p<ex>ar</ex>adi&slong;um</me:dipl>
      <me:norm>paradísúsum</me:norm>
    </choice>
  </w>

  <w>
    <choice>
      <me:fac>hína</me:fac>
      <me:dipl>hína</me:dipl>
      <me:norm>hina</me:norm>
    </choice>
  </w>

  <w>
    <choice>
      <me:fac>&fins;ornu</me:fac>
      <me:dipl>fornu</me:dipl>
      <me:norm>fornu</me:norm>
    </choice>
  </w>

  <me:punct>
    <choice>
      <me:fac>.</me:fac>
      <me:dipl>.</me:dipl>
      <me:norm>.</me:norm>
    </choice>
  </me:punct>

```

Note: The sequence ‘hofud engli’ has been analysed as a single word and encoded as suggested in [ch. 2.3.2](#). Punctuation marks have been set out in the **<me:punct>** element; for a fuller discussion, see [ch. 4.8](#).

The display of the transcription is made by style sheets in XML:

- (a) the facsimile level is the content of the **<me:fac>** element, in which the **<am>** element describes abbreviation markers
- (b) the diplomatic level is the content of the **<me:dipl>** element, in which the **<ex>** element describes expanded abbreviations
- (c) the normalised level is the content of the **<me:norm>** element

ðt̃t̃ñ ṽár̃ bauð̃ michaele hof̃ engli. at fylg̃ adam ok ællū
 helgū̃ h̃s at leida þa i paradifum hína fornu.

dr̃ottiñ ṽár̃ bauð̃ michaele hofud̃ engli. at fylgia. adam
 ok ællum̃ helgum̃ hans at leida þa i paradifum hína fornu.

Dróttinn vár bauð Michaele hofuðengli at fylgja Adam
 ok öllum helgum hans at leiða þá í paradísúsum hina fornu.

Fig. 3.6 A display of all levels contained in the multi-level transcriptions above.

As stated above, the elements **<me:fac>**, **<me:dipl>** and **<me:norm>** are not defined in TEI, but are part of the namespace we have defined for Menota texts. Please see the schemas in [Appendix D](#).

The three levels discussed here can be seen as **focal** in the sense that they are typical and often used levels of text representations in Medieval Nordic editions. A number of additional levels can be defined, e.g. a **<me:pal>** level for an even more detailed

paleographical encoding of the text. This level has been included in the Menota schemas, but is not seen as one of the focal levels.