

## Diacritics for medieval studies

Marc Wilhelm Küster and Isabel Wojtovicz

### Introduction

The handling of medieval characters in the context of IT systems is problematic, especially as far as data exchange and publication in electronic media is concerned.

There is as yet no reasonably complete list which covers all characters and special symbols used in modern editions of medieval texts. In particular, there is no complete coverage of the many abbreviations that are typical for medieval manuscripts and no study on the applicability of the character-glyph-model in this context.

This paper concentrates on one special aspect of characters in medieval studies, namely superscript letters which are used as diacritics, e. g. ũ. These we shall call *superscript letter diacritics*. While it also cannot claim completeness, it covers the superscript letter diacritics which occur in major medievalistic internet projects and in a selection of important editions. It concentrates exclusively on modern editions of medieval literature and on grammars of Middle High German.

### Superscript vowels

The by far most frequent case is that of superscript vowels (and *ç*). According to (REICHMANN, 1993) there can be the following letters with a superscript:<sup>1</sup>

- â, â̇, â̈, ẩ, â̊
- ô, ô̇, ô̈, ổ, ô̊
- û, û̇, û̈, û̉, û̊
- î
- ȳ

Most of these represent different pronunciations of the base vowel, usually, but not always, pronunciations that are in between the base vowel and the vowel which is indicated by the superscript, be the result a diphthong or an unlauded letter.<sup>2</sup> Thus we can have *û̈* for the /uo/ diphthong<sup>3</sup> or *û̈* for u umlaut.

<sup>1</sup> (REICHMANN, 1993), §§ L11–17 (short vowels), §§ L18–25 (long vowels) and §§ L27–30 (diphthongs). This list neither is nor claims to be complete.

<sup>2</sup> In some dialects, e. g. middle German (Rheinfränkisch, Moselfränkisch, Sächsisch, Thüringisch, ...) dialects (as opposed to high German), the superscript letter diacritics function as purely graphematic distinctions, e. g. to distinguish *n* from *u*.

<sup>3</sup> This form appears very often, e. g. in (GRUNDMANN, 1958), p. 192, l. 2. This diphthong can still be heard in many German dialects.

The most frequent superscript letter diacritic is the superscript e (°) which usually designates an umlauted letter. It is regularly used in medieval texts. In modern editions it occurs very frequently in the Middle High German parts, whereas text in modern German e. g. in the annotations which appear on the same page is never written this way. Thus we can have words like »rômische rîche« (Roman empire) and »zûversûnende« (confident) on the same page as »Übersetzer«, »übersetzt«, »sinngemäß«, »läßt« etc.<sup>4</sup> Such examples can be found by the hundred, e. g. also in grammars:

So wird der Umlaut von /a/ → /e/ z. T. gar nicht bezeichnet oder als ⟨e, ę, ei, ä, â⟩ angegeben, der des /ā/ → /æ/ als ⟨æ, œ, ê, e⟩.<sup>5</sup>

In spite of this, Middle High German and modern German parts on one page are almost always typeset in the same base font. Font switching is highly unusual.

The superscript letter diacritic o is in the two hundred years of typesetting in this field distinctly regarded as an *o* typographically, not as a ring above. Thus only *zû*, not *zú*, would be an acceptable rendering of the /uo/-diphthong.

To summarize, all of combinations of base letters with the superscript letter diacritics <sup>a</sup>, <sup>c</sup>, <sup>e</sup>, <sup>i</sup>, <sup>o</sup>, <sup>u</sup>, <sup>v</sup> modify the sound value of the base letter in a significant and predictable way. These diacritics were a regular part of both the original Middle High German orthography and (more importantly) of current editorial practice in the complete field of medieval studies and German dialectology. Since quite a number of these projects also are active in electronic data processing and have an urgent need to exchange data, e. g. via the World Wide Web, these are an absolute core set of combining diacritics.

## A few samples

To give a bit more meat to this discussion, here a handful of randomly chosen examples. Many more could easily be found.

- ô: das rômische rîche<sup>6</sup>
- û: zû der Hohen Sen<sup>7</sup>
- ũ: wenn ob sich Pilatus rûmete des Gewaltes<sup>8</sup>
- ǔ: und zu vertriben dem bösen<sup>9</sup>
- ı̇: dı̇ allen vroun ı̇ we gerı̇t<sup>10</sup>
- ũ: si müsten rû ind pine liden<sup>11</sup>

<sup>4</sup> (GRUNDMANN, 1958), p. 192.

<sup>5</sup> (METTKE, 1993), p. 31. Translation: The umlaut from /a/ to /e/ is sometimes not marked at all, sometimes as ⟨e, ę, ei, ä, â⟩. The umlaut from /ā/ to /æ/ as ⟨æ, œ, ê, e⟩.

<sup>6</sup> (GRUNDMANN, 1958), p. 192.

<sup>7</sup> (KOLLER, 1964), p. 57.

<sup>8</sup> (GRUNDMANN, 1958), p. 193.

<sup>9</sup> (KOLLER, 1964), p. 57.

<sup>10</sup> (BACH, 1934), p. 44.

<sup>11</sup> (BACH, 1934), p. 44.

## Consonants

If the case for vowels is clearcut, the case for consonants is somewhat more ambiguous. There are many such examples in (BOXLER, 1996) which is a rich source for this topic (all samples from (BOXLER, 1996)):

- ē: ēift (p. 246)
- ĝ: beĝauen (p. 246)
- ǣ: crǣp (p. 246)
- ǣ: māinys (p. 247)
- ê: ſchêzze (p. 260)
- ħ: begirliĥ (p. 261)
- ṁ: ṁ (p. 279)
- č: čces (p. 431)

This findings were complemented by information which Ute Recker from the University of Trier kindly placed at our disposal. She sent a list of symbols that occur in the fully computerized text archive of the new *Mittelhochdeutsches Wörterbuch*, currently under development. This added a number of further letter-diacritic combinations:<sup>12</sup>

- ǣ: stǣte<sup>13</sup>
- ē: gesemeē<sup>d14</sup>
- ē: dēr<sup>15</sup>
- ô: lôser<sup>16</sup>
- ñ: iuñvrowen<sup>17</sup>
- ĥ: unnaĥer<sup>18</sup>
- ǫ: getǫiv<sup>19</sup>
- ũ: wūrde<sup>20</sup>
- ǰ: herbǰrgen<sup>21</sup>

## Summary

This gives the following list of the following letter-diacritic combinations:

- ǣ ǣ ǣ ǣ ǣ ǣ ǣ
- ĥ ĥ ĥ

---

<sup>12</sup> These samples were culled from the project's as yet not public database.

For further information please contact [recker@uni-trier.de](mailto:recker@uni-trier.de).

<sup>13</sup> (MERTENS, 1971), p. 8.

<sup>14</sup> (DIEMER, 1849), p. 78.

<sup>15</sup> (FRINGS, 1922), p. 683.

<sup>16</sup> (SINGER, 1906), p. 1326.

<sup>17</sup> (FRINGS, 1922), p. 300.

<sup>18</sup> (JUNK, 1905)

<sup>19</sup> (DIEMER, 1849), p. 27.

<sup>20</sup> (KLUGE, 1948), p. 496.

<sup>21</sup> (MERTENS, 1971), p. 18.

- $\overset{a}{d}$
- $\overset{e}{e}$   $\overset{e}{e}$   $\overset{e}{e}$   $\overset{e}{e}$   $\overset{e}{e}$
- $\overset{g}{g}$
- $\overset{h}{h}$
- $\overset{i}{i}$
- $\overset{m}{m}$
- $\overset{n}{n}$
- $\overset{o}{o}$   $\overset{o}{o}$   $\overset{o}{o}$   $\overset{o}{o}$   $\overset{o}{o}$
- $\overset{r}{r}$   $\overset{r}{r}$
- $\overset{u}{u}$   $\overset{u}{u}$   $\overset{u}{u}$   $\overset{u}{u}$   $\overset{u}{u}$
- $\overset{v}{v}$   $\overset{v}{v}$

The same list follows here once more, ordered by the superscript letter diacritic.<sup>22</sup>

- $\overset{a}{d}$   $\overset{e}{e}$   $\overset{o}{o}$
- $\overset{a}{a}$   $\overset{e}{e}$   $\overset{n}{n}$   $\overset{o}{o}$   $\overset{u}{u}$
- $\overset{e}{e}$
- $\overset{a}{a}$   $\overset{h}{h}$   $\overset{i}{i}$   $\overset{o}{o}$   $\overset{u}{u}$   $\overset{v}{v}$
- $\overset{c}{c}$
- $\overset{i}{c}$   $\overset{e}{e}$   $\overset{u}{u}$
- $\overset{m}{a}$
- $\overset{a}{a}$   $\overset{o}{o}$   $\overset{r}{r}$   $\overset{u}{u}$   $\overset{v}{v}$
- $\overset{e}{e}$   $\overset{g}{g}$   $\overset{r}{r}$
- $\overset{m}{m}$
- $\overset{h}{a}$
- $\overset{a}{a}$   $\overset{c}{c}$   $\overset{o}{o}$   $\overset{u}{u}$
- $\overset{x}{a}$

The following superscript letter diacritics are thereby attested in these samples: a, c, d, e, h, i, m, o, r, t, u, v, x ( $\overset{a}{a}$ ,  $\overset{c}{c}$ ,  $\overset{d}{d}$ ,  $\overset{e}{e}$ ,  $\overset{h}{h}$ ,  $\overset{i}{i}$ ,  $\overset{m}{m}$ ,  $\overset{o}{o}$ ,  $\overset{r}{r}$ ,  $\overset{t}{t}$ ,  $\overset{u}{u}$ ,  $\overset{v}{v}$ ,  $\overset{x}{x}$ ).

This list is certainly still incomplete, but can give a first overview about current usage of superscript letter diacritics in the field of German medieval studies.

---

<sup>22</sup> It should be understood that this is an entirely artificial ordering for reference purposes. Such an ordering is not used in medieval studies.

## Required action

Encode superscript letter diacritics. Ideally, all 26 superscript letter diacritics which are possible should be encoded. Should WG2 and the UTC not want to encode all 26 superscript letter diacritics as yet, the thirteen attested letter diacritics should be assigned, but following the example of the Mathematical Alphanumeric Symbols suitable unassigned positions should be reserved for future encoding of newly attested superscript letter diacritics.

The following table works on this assumption:

Glyph	Code-point	Name
ǎ	0x00	COMBINING LATIN SMALL LETTER A ABOVE
	0x01	This position shall not be used
ǎ	0x02	COMBINING LATIN SMALL LETTER C ABOVE
ǎ	0x03	COMBINING LATIN SMALL LETTER D ABOVE
ǎ	0x04	COMBINING LATIN SMALL LETTER E ABOVE
	0x05	This position shall not be used
	0x06	This position shall not be used
ǎ	0x07	COMBINING LATIN SMALL LETTER H ABOVE
ǎ	0x08	COMBINING LATIN SMALL LETTER I ABOVE
	0x09	This position shall not be used
	0x0A	This position shall not be used
	0x0B	This position shall not be used
ǎ	0x0C	COMBINING LATIN SMALL LETTER M ABOVE
	0x0D	This position shall not be used
ǎ	0x0E	COMBINING LATIN SMALL LETTER O ABOVE
	0x0F	This position shall not be used
	0x10	This position shall not be used
ǎ	0x11	COMBINING LATIN SMALL LETTER R ABOVE
	0x12	This position shall not be used
ǎ	0x13	COMBINING LATIN SMALL LETTER T ABOVE
ǎ	0x14	COMBINING LATIN SMALL LETTER U ABOVE
ǎ	0x15	COMBINING LATIN SMALL LETTER V ABOVE
	0x16	This position shall not be used
ǎ	0x17	COMBINING LATIN SMALL LETTER X ABOVE
	0x18	This position shall not be used
	0x19	This position shall not be used

The name COMBINING SUPERSCRIPT LETTER DIACRITIC *x*, where *x* stands for the appropriate letter name, would be equally acceptable.

## Annex A: Current encoding practice

Currently most projects we know of use an encoding for superscript letter diacritics that was developed in Tübingen for the TUSTEP typesetting system, but has spread over into other spheres such as Web publications. This encoding system properly handles these letters as diacritics which can be superscripted to all other letters. This #;eu encodes ũ, #;cn ñ. Other projects use proprietary fonts that usually replace existing but for this purpose superfluous characters with the ones needed for a specific edition. This second approach makes data exchange rather difficult.

## Annex B: Character properties

All superscript letter diacritics have the same character properties except for the character names and the 10646 comment fields. The common properties are:

- General Category: Mark, Non-Spacing (Mn);
- Canonical Combining Classes: Above (230);
- Bidirectionality Category: Non-Spacing Mark (NSM);
- Character Decomposition Mapping: None;
- Decimal digit value: Not applicable;
- Digit value: Not applicable;
- Numeric value: Not applicable;
- Mirrored: Not mirrored (N);
- Unicode 1.0 Name: Not applicable;
- Uppercase Mapping: Not applicable;
- Lowercase Mapping: Not applicable;
- Titlecase Mapping: Not applicable.

The 10646 comment field should read: NON-SPACING LETTER DIACRITIC  $\times$  ABOVE, where  $\times$  corresponds to the name of the superscript letter diacritic. The character name can be found in the table on the preceding page.

## Annex C: Ordering of letters with superscript letter diacritics

There are a variety of ordering schemes in practice. We shall endeavour to consult the relevant experts with the aim to define a universally acceptable default sequence. This sequence will as soon as possible be transmitted to all interested parties.

- (BACH, 1934) Bach, Adolf (ed): *Das Rheinische Marienlob. Eine deutsche Dichtung des 13. Jahrhunderts*. Leipzig: Hiersemann, 1934. Bd.: 281. In: *Bibliothek des Literarischen Vereins in Stuttgart*.
- (BOXLER, 1996) Boxler, Madeleine (ed): »ich bin ein Predigerin und Apostlorin«. *Die deutschen Maria-Magdalena-Legenden des Mittelalters (1330–1550). Untersuchungen und Texte*. Bern: Lang, 1996. Bd.: 22. In: *Deutsche Literatur von den Anfängen bis 1700*.
- (DIEMER, 1849) Diemer, Joseph (ed): *Deutsche Gedichte des XI. und XII. Jahrhunderts. Aufgefunden im regulierten Chorherrenstifte zu Vorau in der Steiermark*. Wien: Braumüller, <sup>1</sup> 1849.
- (FRINGS, 1922) Frings, Theodor; Kuhnt, Joachim (ed): *König Rother*. Bonn: Schroeder, 1922. Bd.: 3. In: *Rheinische Beiträge und Hilfsbücher zur Germanischen Philologie und Volkskunde*.
- (GRUNDMANN, 1958) Grundmann, Herbert; Heimpel, Hermann (ed): *Alexander von Roes. Schriften*. Stuttgart: Hiersemann, 1958. Bd.: 10. In: *Monumenta Germaniae Historica / Scriptorum*.
- (JUNK, 1905) Junk, Victor (ed): *Rudolf von Ems, Willehalm von Orlens. Herausgegeben aus dem Wasserburger Codex der fürstlichen Fürstenbergischen Hofbibliothek in Donaueschingen*. Berlin: Weidmann, 1905. Bd.: 2. In: *Deutsche Texte des Mittelalters*.
- (KLUGE, 1948) Kluge, Reinhold (ed): *Lancelot. Nach der Heidelberger Pergamenthandschrift Pal. Germ. 147*. Berlin: Akademie-Verlag, 1948. Bd.: 1. In: *Deutsche Texte des Mittelalters*.
- (KOLLER, 1964) Koller, Heinrich (ed): *Reformation Kaiser Siegmunds*. Stuttgart: Hiersemann, 1964. Bd.: 6. In: *Monumenta Germaniae Historica / Scriptorum*.
- (MERTENS, 1971) Mertens, Volker (ed): *Das Predigtbuch des Priesters Konrad. Überlieferung, Gestalt, Gehalt und Texte*. München: Beck, 1971. Bd.: 33. In: *Münchener Texte und Untersuchungen zur deutschen Literatur des Mittelalters*.
- (METTKE, 1993) Mettke, Heinz: *Mittelhochdeutsche Grammatik*. Tübingen: Niemeyer, <sup>7</sup> 1993.
- (REICHMANN, 1993) Reichmann, Oskar; Wegera, Klaus-Peter (ed): *Frühneuhochdeutsche Grammatik*. Tübingen: Niemeyer, 1993. In: *Sammlung kurzer Grammatiken germanischer Dialekte*.
- (SINGER, 1906) Singer, Samuel (ed): *Die Werke Heinrichs von Neustadt*. Berlin: Weidmann, 1906. Bd.: 7. In: *Deutsche Texte des Mittelalters*.

