Dr Tarrin Wills
Chair, Medieval Unicode Font Initiative
Associate Professor (Seniorforsker) and Head,
Dictionary of Old Norse Prose
Chair, Medieval Unicode Font Initiative

9 April 2024

To the Unicode scripts panel

RE: Clarification of use and forms of certain combining characters

I have been reviewing material related to certain abbreviation marks covered by the Medieval Unicode Font Initiative (MUFI), and in response to some queries from Harald Tveiten I offer some clarification of the current definitions of supplementary PUA characters in MUFI and some history and actual usage in manuscripts.

Just a little background: I am most familiar with Nordic manuscripts and have been working in my research particularly on abbreviation in Icelandic manuscripts, which was unusually prevalent compared with other vernaculars. My knowledge is therefore based very much on my research, but is supplemented with respect to Latin manuscript writing with reference works such as Cappelli's *Dizionario di abbreviature latine ed italiane*, the standard reference work for medieval abbreviation. My focus here, however, is on my understanding of MUFI's previous work as the current Chair of the group.

I would like to make the following points, firstly regarding MUFI's status and discussions concerning MUFI's recommendations:

- MUFI's official recommendations are to be found on https://mufi.info in the form of the data presented there and the PDFs of the earlier recommendations on the website. No other correspondence or discussions should be considered as having any status with respect to MUFI.
- MUFI works closely with certain font developers, particularly in very recent times with Peter Baker (Junicode's developer). He and current and former members of MUFI have from time to time discussed publicly (e.g. on GitHub) issues to do with character definitions and usage with respect to font development, but it would of course be inappropriate to consider these discussions as official MUFI positions.

Secondly, clarifications regarding particular characters described by MUFI:

- 1DD1 COMBINING UR ABOVE is an abbreviation mark and should not be used to represent non-abbreviating diacritics. The name is specifically given because it is a standard abbreviation for 'ur'. It has several forms which can normally be reduced to two shapes, the most common of which resembles '2' (but which should not be confused with U+1DE3 COMBINING LATIN SMALL LETTER R ROTUNDA, whose similarity of shape is only coincidental in some manuscripts).
- MUFI's recommendation has a PUA character F1C2 COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISKATE FORM for the variant form of 1DD1 COMBINING UR ABOVE which resembles a turned 's'. This form is sufficiently distinct from the '2' form of 1DD1 COMBINING UR ABOVE that it should perhaps be considered by Unicode.
- MUFI's F1CC COMBINING CURLY BAR ABOVE is absolutely not an abbreviation mark but a decorative diacritic functioning similarly to the dot over 'i', to distinguish letters with minims (usually 'u') from surrounding letters. The current glyph forms in MUFI do not accurately represent the normal manuscript forms and should be updated for the relevant fonts. It is unrelated in both shape and use to F1C2 COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISKATE FORM or other forms of the ur-abbreviation.

There is a practical point for keeping abbreviations and other diacritics separate: abbreviations represent letters that are not written, expressed as marks often combining with written letters. Other combining diacritics normally represent modifications of the letters that are actually written. When analysing these texts, it is important to be able to identify marks that represent one or more letters, and those which modify other letters.

It should also be noted that although most abbreviation marks in MUFI are currently defined as combining marks, in many cases they do not appear above the preceding letter, but are spacing. An abbreviation mark is mobile with respect to the character to which it appears closest, because it is largely independent of that character in its meaning. Conversely, diacritics normally modify in some way the character to which they are closest and are therefore less mobile. I believe this is a defining distinction between abbreviation marks and other combining diacritics and perhaps the panel would consider encoding these non-combining abbreviation variants at a future point – I would be happy to supply documentation.

Some detail and examples of the characters mentioned above follows.

U+1DD1 COMBINING UR ABOVE

https://mufi.info/q.php?p=mufi/chars/unichar/7633

This mark, originally used in Latin manuscripts (cf. Hreinn Benediktsson *Early Icelandic Script*, p. 90) is an abbreviation mark and almost always stands for 'ur'. Despite some resemblance to some variants of F1CC COMBINING CURLY BAR ABOVE, this is an abbreviation mark and should not be conflated with F1CC.

Examples in Old Norse mss: GKS 1002 fol 45va/1 (hallmu*n*d*ur*); DKNVSB 41 8° 109/12 (h*e*lld*ur*); AM 713 4° 128/17 (nock*urt*); AM 557 4° 5v/7 (læt*ur*)
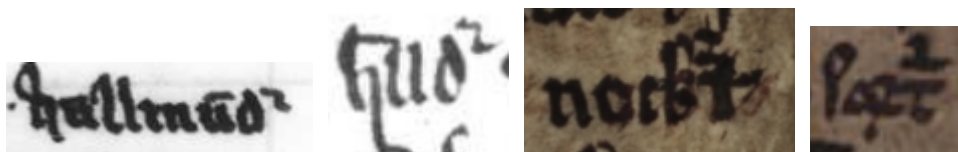
The Unicode glyph example for this character is slightly misleading: it more often has a form similar to '2' (cf. above), and slightly less frequently a sideways 's' and sometimes other forms. Where the preceding character has an ascender it is normally written as spacing or between the previous and next characters, as can be seen in the image. The Junicode glyph represents a good mix of the different forms. Examples from Cappelli show mostly the '2' form but we include here two examples of the 'turned s' form found in Unicode (p. 42 'curabatur', p. 44 'causantur', p. 322 'recurrere', p. 313 'quaeritur'):
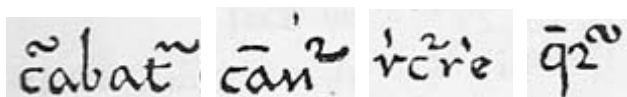
Cf. Cappelli Online and intro xx/xiv, showing the main forms and variants:
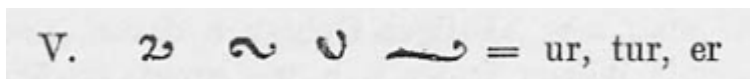
In printed works the character may resemble the current Unicode reference glyph, e.g. https://github.com/psb1558/Junicode-font/issues/11#issuecomment-636282361 where it is described as a 'variant form':

 for 'accipit*ur*'

When this character was accepted into Unicode, it was given a reference glyph that resembles the least common of the three main forms. I believe this was an error, perhaps overlooked due to the large volume of characters that were accepted at the same time. MUFI consequently redefined the character that most resembled this form according to Unicode's standard, but retained the two most common forms in its recommendation for the PUA.

MUFI version 1:

♂	&ur2;	F153	F153	COMBINING ABBREVIATION MARK SUPERSCRIPT UR (2-SIGN)
∞	&ur8;	F1C2		COMBINING ABBREVIATION MARK SUPERSCRIPT UR (8-SIGN)
S	&ur8open;	F1C3		COMBINING ABBREVIATION MARK SUPERSCRIPT UR (OPEN 8-SIGN)

MUFI versions 2 and 3:

♂	&urrot;	F153	F153	COMBINING ABBREVIATION MARK SUPERSCRIPT UR ROUND R FORM
∞	&urlemn;	F1C2		COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISKATE FORM
~	&ur;	F1C3		COMBINING ABBREVIATION MARK SUPERSCRIPT UR TILDE FORM
In v. 1.0 of the MUFI character recommendation the glyph for this character was unfortunate. It should look like a superscript round 'r' lying on the side (as shown here).				

MUFI version 4 (after Unicode acceptance using the third most common form):

♂	&urrot;	F153	COMBINING ABBREVIATION MARK SUPERSCRIPT UR ROUND R FORM
∞	&urlemn;	F1C2	COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISKATE FORM

|				
~		&ur;	1DD1	COMBINING UR ABOVE

MUFI–F1CC COMBINING CURLY BAR ABOVE

https://mufi.info/q.php?p=mufi/chars/unichar/61900

This is frequently used in postmedieval handwriting to distinguish the character 'u' (and sometimes 'n'/'m') from other minims (i, m, n) and thus functionally similar to the dot above the character 'ı'. For semantic reasons, it should not be unified with the similar–looking characters 0303 COMBINING TILDE (Latin script), 0342 COMBINING GREEK PERISPOMENI

(Greek script) or 0DC3 COMBINING SUSPENSION MARK (Glagolitic script), nor with any abbreviation mark such as 1DD1 COMBINING UR ABOVE.

The character has a variety of forms but should be encoded as a single codepoint. Examples (AM 560 c 4°, 2r; AM 426 fol, 25r; AM 165 m fol, 3r; AM 154 fol, 1v; AM 35 fol, 60r)
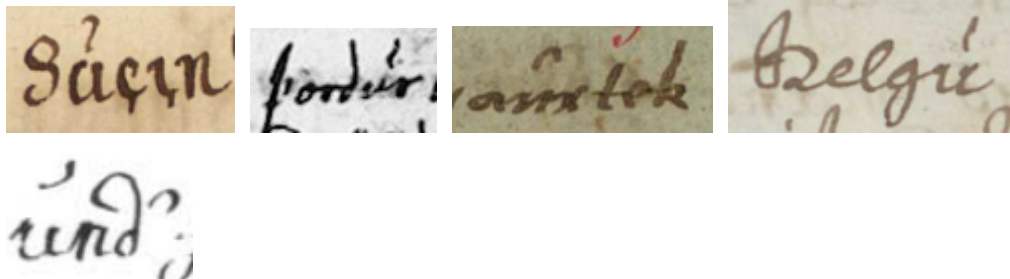
The diacritic can be considered 'soft', as it does not occur on capital forms of 'u', nor in instances where an abbreviation or diacritic is used over the character which would otherwise have this mark, that is in the same manuscripts where it is found on the lower case version (e.g. Thott 972 fol, 415vb; AM 35 fol, 60r; AM 142 fol, 82):
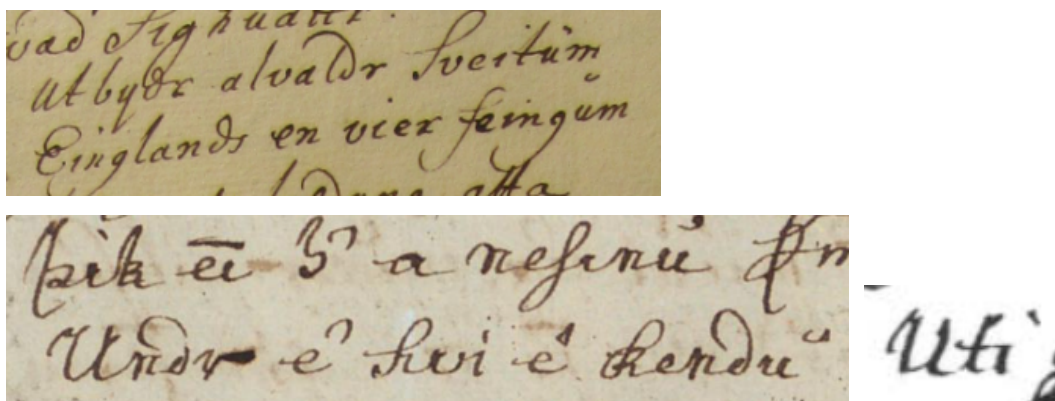
The first two examples show the consistent use of the mark above small 'u' in manuscripts where it is absent in capital 'U'.

In an intermediate version of the MUFI site, this mark was incorrectly labelled 'Combining abbreviation marks'. This was due to an error in importing the PDF MUFI recommendations.

The manuscripts I have surveyed suggest that the example glyph (and name) in MUFI is misleading and should be changed to resemble a combination of the vertical tilde and hook above.

MUFI–F1C2 COMBINING ABBREVIATION MARK SUPERSCRIPT UR LEMNISKATE FORM

https://mufi.info/q.php?p=mufi/chars/unichar/61890

This character is a variant of the 'ur' abbreviation 1DD1 with a sideways 's' form or similar ('lemniskate' is probably misleading) – the Junicode and Palemonas glyphs are the most accurate. Cf. the 2nd and 4th forms in Cappelli's introduction:
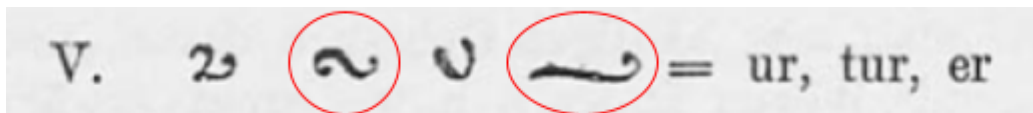
This should not be confused with F1CC COMBINING CURLY BAR ABOVE, which is not an abbreviation mark.

Finally, I would like to thank you all for your continued excellent work on updating Unicode.

Tarrin Wills
Chair, Medieval Unicode Font Initiative