Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Response to UTC/US contribution N3037R, "Feedback on N3027 Proposal to add medievalist characters"**
Source: **Michael Everson (editor), Peter Baker, António Emiliano, Florian Grammel,**
Odd Einar Haugen, Diana Luft, Susana Pedro, Gerd Schumacher, Andreas Stötzner
Status: **Expert Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2006-03-31**

We would like to thank the UTC and the US National Body for their comments on N3027. We trust that the clarifications we offer here will be sufficient to alleviate the concerns described in N3037. We are very pleased to note that there were so few concerns raised.

1. COMBINING LATIN SMALL LETTER Y. We have not found an example of this in use, despite the report that it is used as an abbreviation for *ꝼr fyrr* 'before'. We do not object to removing the character from the proposal, though we note that almost all of the basic Latin letters occur as combining small letters above. It seems certain that this will turn up sooner or later, as all of the other Latin letters have since they were first proposed by the German National Body in N2160, 2000-02-16.

2. COMBINING OGONEK ABOVE is, in fact, an ogonek. The ogonek diacritic is a well-known diacritic which has been used in Nordic orthographies since the beginning. Like the ogonek, the COMBINING OGONEK ABOVE overlaps with its base character. It was created in analogy with the combining ogonek below, and indeed both the COMBINING OGONEK and the COMBINING OGONEK ABOVE co-occur on the same vowel, and it is clear that the mark is not a COMBINING COMMA ABOVE, a COMBINING HOOK ABOVE, or a COMBINING RIGHT-HALF RING ABOVE, none of which overlap with the base character.

3.1.1.3 The nasal long vowels

It is generally believed that vocalic nasality is not orthographically represented in Old Icelandic script, although the First Grammarian advocated the use of a superscript dot to indicate nasality. Hreinn Benediktsson (1972) says, for example: 'nasal vowels are not represented by separated symbols in the orthography, not even in its earliest stage.'

However, the orthography of IHB seems to indicate that at least /õ/ has had a special notation: ȯ. The symbol ȯ occurs 83 times and ǫ̇ 3 times. Perhaps we should include: óroṁ 12v17, óſtyrcþ 16r30, ǫ́roṁ 16v31, ǫ̇ verþa 17v19, and óvine 17v24, and transcribe ȯroṁ, ȯſtyrcþ, ǫ̇roṁ , ǫ̇ verþa, ȯvine, since in these cases the accent mark is bolder than usual (not in 12v17) and fastened to the o. In fact it looks more like half of the diacritical mark in ȯ, than like an accent. ⟨ȯ⟩ and ⟨ǫ̇⟩ stand for /o/, /ǫ/, /ó/ and /ǫ́/ (I do not differentiate here between oral and nasal vowels):

	/o/	/ǫ/	/ó/	/ǫ́/
⟨ȯ⟩	6	4	74	2
⟨ǫ̇⟩		1	4	
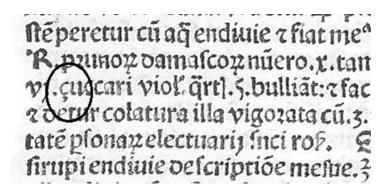
The symbol ǒ occurs 83 times and ǫ 3 times.
', óſtyrcþ 16r30, ǫrom 16v31, ǫ́ verþa 17v19, an
yrcþ, ǫrom , ǫ́ verþa, ǒvine, since in these cases
ιot in 12v17) and fastened to the o. In fact it loo
in ǒ, than like an accent. ⟨ǒ⟩ and ⟨ǫ́⟩ stand for /ɔ

SIGURÐAR SAGA
ÞǪGLA

EDITED BY
M. J. DRISCOLL

STOFNUN ÁRNA MAGNÚSSONAR Á ÍSLANDI

2

3. LATIN LETTER R ROTUNDA. Michael Everson has improved the glyphs: < ꝛ ꝛ >.

4. COMBINING LATIN LETTER SMALL CAPITAL G etc. It is not strictly true to say that there is no reason to make a distinction between capital letters and small capital letters. Consider the following letters: G, h, n, and G. The x-height of the SMALL CAPITAL LETTER G is the same as that of the SMALL LETTER N, and the height of the CAPITAL LETTER G is the same as the height of the SMALL LETTER H. Small capital letters are used in Old Norse orthography to indicate geminate consonants, and the combining characters correspond to those small capitals. If COMBINING LATIN LETTER SMALL CAPITAL G were named COMBINING LATIN CAPITAL LETTER G (or COMBINING LATIN LETTER CAPITAL G), this could cause confusion should genuine combining capital letters be required for other user communities (letters with the height of COMBINING LATIN LETTER SMALL CAPITAL H and not the height of COMBINING LATIN LETTER SMALL CAPITAL N). We prefer the names as proposed.

5. COMBINING ZIGZAG BELOW goes below the letter in the same way that COMBINING ZIGZAG ABOVE goes above the letter. Additional examples are given here from Latin:
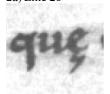
COMBINING ZIGZAG BELOW

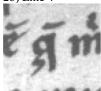1) From Argellata, Petrus de - Chirurgia. Venice: [Printer of Argellata], 1492

2) From a manuscript dated 1256 (IAN/TT, Cabido da Sé de Coimbra, maço 5, nº 14)
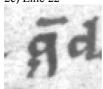
2a) Line 28

2b) Line 4

2c) Line 22

6. COMBINING LATIN SMALL LETTER G and COMBINING LATIN LETTER SMALL CAPITAL G are both meant to go above the letter they follow in the text stream. So they should be shifted to the left in both Figures 99 and 100.

7. LATIN LETTER VY. We disagree with the UTC/US assessment that this isn't a ligature of "v" and "y", because the "basis of its historic origin" is a matter of interpretation. Jón Ólafsson from Grunnavík (1705-1779) is the man who devised this letter. He was the son of a vicar, got a good education, studied and took exam in theology in Copenhagen, but dedicated most of his life to studying ancient Icelandic literature. As a student he worked as a copyist for Árni Magnússon and was later supported by Árni's bequest. He copied many manuscripts, and also wrote many a good deal himself, including biographies of Árni Magnússon and of Páll Vídalín, translations into Latin, essays on Icelandic language and literature. For many years he worked on a book about writing modern Icelandic, but he never decided on a title; one draft title is *Orthographia Islandica. Ritfar eður rjettritun íslenzkrar tungu.* It was never published but the draft is in Reykjavík today, its shelf-mark AM 435 fol.

On pp. 73-87 of his book (*Jón Ólafsson frá Grunnavík*, Safn Fræðafjelagsins 5, Kaupmannhöfn: Möller, 1926), Jón Helgason gives a sketch of this work. In one chapter Jón Ólafsson criticizes the failure of many of his contemporaries to distinguish properly between long and short vowels, and he discusses the ways length can be marked. Jón Ólafsson would prefer the use of acute accents on the vowels in question, but as this isn't customary in the 18th century, he opts to support the use of ligated (doubled) letters as handled by some printers at the time; but as they do not give a mark for "ý", he devises one of his own. Jón Helgason describes this (p. 78):

> Hann aðhyllist hins vegar venju prentuðu bókanna, að tvöfalda stafina og skeyta þeim saman (ꜳ = á, w = ú o. s. frv.), en fer þó lengra en þær að því leyti, að hann eykur við nýjum staf fyrir ý, og er gerð hans eins og v sje felt framan við y.

> 'He subscribes, on the other hand, to the practice of printed books, to double the letters and join them together (ꜳ = á, w = ú, and so on), but still goes further than them in this respect, that he adds a new letter for ý, and its shape is as a v put before y.'

Let us look at these letters:

Jón represents	á	by	ꜳ	LATIN LETTER AA	(phonemic /aː/)
Jón represents	í	by	ꟼ	LATIN LETTER IJ	(phonemic /iː/)
Jón represents	ó	by	ꝏ	LATIN LETTER OO	(phonemic /oː/)
Jón represents	ú	by	w	LATIN LETTER W	(phonemic /uː/) (w < vv of course)
Jón represents	ý	by	ꭣ	LATIN LETTER VY	(phonemic /yː/)

Just as we would not describe "ꟼ" as "Latin letter jj", because it is obviously "ij" (and not "ii" or "ü" or "ÿ"), so we cannot describe "ꭣ" as "Latin letter yy" because it is, as Jón Helgason says, obviously "vy". The Nordicist expectation for this letter is <Ꝡ ꝡ>, and by no means is it a true "yy", which is <�YY ꟼyy>. The name "Latin letter yy" should be reserved for <YY yy>. Therefore, for consistency with

the usual UCS naming conventions, the name and note for the character should be "LATIN LETTER VY (used for ý or /y:/)". Some more examples:

8. LATIN LETTER ET. We disagree with the UTC/US suggestion that the name of this character should be based on its shape. It is true that ET is not the *only* name the character it could have, but it is sufficient. This is one of a group of letters we have called "Letters with syllabic content". All of these letters are named after sounds they may represent: ET can represent *et, m, ue, que, ue, est, eð*; IS can represent *is, ys, es*; CON can represent *con, com, co, us, os*; US can represent *us, os*; DUM can represent *dum, die, dia, dias*; LUM can represent *los*; MUM can represent *mum*; NUM can represent *num*; RUM can represent *rum, rom*; TUM can represent *tum*; UM can represent *um, us, os, un*. And there are doubtless more readings for some of these; we have re-listed only the multiple readings we gave in N3027. The requirement is that the names be unique and that they identify the character.

Indeed, ET is a particularly *good* name for this character because of the mnemonic factor. In English, at least, one says *viz.* which = *viȝ*, where the modern *z* replaces the letter ȝ, and the word abbreviated is *videlicet*. Naming it for a "three" shape, for instance—we note that this has not been proposed by the UTC/US document—would be anachronistic, since the character was used before the Arabic digit 3 arrived in Europe. Furthermore, the letter is in fact a cursive evolution of an earlier sign, which looked like our modern semicolon ";". We also note that the UTC/US did not suggest that we rename US or IS to "shape-based" names (and we would also oppose a suggestion like that!). We prefer the name ET for this character.