

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

Doc Type: Working Group Document

Title: Proposal to encode one Irish phonetic letter in the UCS

Source: Michael Everson

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2003-10-05

Irish phonological texts, like all such texts, make use of a number of special symbols to distinguish different sounds. While many researchers use the International Phonetic Alphabet today, a body of linguistic material exists which makes use of a phonetic notation system which is not unlike the Uralic Phonetic Alphabet. One character from that tradition is not yet encoded in the UCS, and is proposed here.

1D79 ɣ LATIN SMALL LETTER INSULAR G

The following samples are taken from one of Irish linguistics' most important works:

Ó Máille, Tomás. 1927. *Urlabhraidheacht agus Graiméar na Gaedhilge*. Cuid I. Baile Átha Cliath: Comhlucht Oideachais na hÉireann.

VIII. Cúl-chonsuin, nó, Consuin Choguais.

Modh	Fá ghuth	Fá anáil
Fosgailte	ɣ : <i>ghaol, dhall*</i> Gear. <i>sagen</i> ³	χ : <i>chaol; teach</i> Gear. <i>Fach, Acht</i>
Dúinte	g : <i>gaol</i> B. <i>gone</i>	k : <i>caol, cúl</i> Gear. <i>kalt, B. call</i>

Figure 1. Back consonants. Shown are “open voiced” ɣ [ɣ], “open breathed” χ [x], “closed voiced” g [g], and “closed breathed” k [k].

Reference can be made here to N1549 *On the derivation of YOGH and EZH* (1997-04-20), in which it was demonstrated that <ɣ> EZH is derived from <z> via a more zed-like shape <ɣ> and that <ɣ> YOGH was derived from <ɣ> INSULAR G which had become redundant after the Carolingian g was introduced by Norman scribes in England. The INSULAR G as used in Irish phonetics is *not* a YOGH, however, and YOGH glyph shapes are as unacceptable for it as the original insular glyphs shape is for YOGH. It is interesting to note that the Edlund Medievalist font encodes all three <g>, <ɣ>, and <ɣ> as independent glyphs.

XXXIX. Na Comharthaí.

342. Seo athchumair ar na comharthaí atá i n-úsáid sa leabhar:

A. Na Consuin.

- I. Na cúl chonsuin, § 42: (a) fá ghuth, \mathfrak{z} , g , v , t , \mathfrak{x} ; (b) fá anáil, \mathfrak{x} , k , p , t , \mathfrak{v} . Féach l. 9, 10.
- II. Consuin tosaigh, l. 14: (a) fá ghuth, \mathfrak{i} (j), d' , N' , L' , (R'); (b) fá anáil, j , t' , N' , L' , (R').
- III. Consuin charbaid, l. 20: (a) fá ghuth, j , g' , v' ; (b) fá anáil, g , k' , v' .
- IV. Idirfhiacloch, l. 23: (a) fá ghuth, \mathfrak{d} , d , N , L ; (b) fá anáil, θ , t , N , L .
- V. Consuin stuagh-charbaid, l. 27: (a) fá ghuth, \mathfrak{d} , l , n , r ; (b) fá anáil, t , l , n , r . Leath-análach, l , n , r .
- VI. Consuin bhairr-theangan fhiaclocha, l. 31: (a) fá ghuth, \mathfrak{d} , d , n' , l' ; (b) fá anáil, l , t , n' , l' . Leath-análach l' , n' . (Féach § 345 c).
- VII. Consuin bhairr-theangan (stuagh an charbaid), l. 32: (a) fá ghuth, D , n , l , r ; (b) fá anáil, T .
- VIII. Consuin lainne, l. 33, 34: s , \mathfrak{s} , r' , r' . Leath-análach, r' .
- IX. Consuin chaib, l. 35, 36: (a) v' , \mathfrak{v} , v , w , w , b' , m' , b , m ; (b) f' , \mathfrak{v} , p , p' , m' , m , \mathfrak{v} .
- X. Cúl-chonsuin ar na modhú leis na caib, l. 37, 41: g^u , \mathfrak{x}^u . Cab-chonsuin 7 teangaidh, § 161: y , \mathfrak{y} .
- An Anáil: h .

343. B. Na guthaidhthe.

- I. Guthaidhthe tosaigh maola, l. 49: i , ε , \mathfrak{x} , \mathfrak{x} , \bar{i} , $\bar{\varepsilon}$, $\bar{\mathfrak{x}}$.
- II. Guthaidhthe tosaigh géara, l. 56: i , e , \mathfrak{e} , \mathfrak{x} , \bar{i} , \bar{e} , $\bar{\mathfrak{x}}$.
- III. Guthaidhthe tosaigh béal-chruinne, l. 59: \mathfrak{f} (\mathfrak{u} -Fr.), \mathfrak{u} , \mathfrak{p} (láir), \mathfrak{o} (eur), leath-iseal.
- IV. Guthaidhthe maola cúil, l. 60: (a) u , o , \mathfrak{o} , a , \mathfrak{a} , \mathfrak{o} , \mathfrak{o} , \bar{a} ; (b) l. 64, \bar{u} , \mathfrak{y} , $\bar{\mathfrak{o}}$, \mathfrak{o} , \bar{a} , \mathfrak{a} .
- V. Guthaidhthe géara cúil, l. 65, 76: (a) u , \mathfrak{o} , \mathfrak{o} , a , \bar{u} , $\bar{\mathfrak{o}}$, $\bar{\mathfrak{o}}$, \bar{a} ; (b) $\cdot\bar{u}$, $\cdot\bar{\mathfrak{o}}$, $\cdot\bar{a}$.
- VI. Guthaidhthe srónacha, l. 73: \bar{u} , $\bar{\mathfrak{o}}$, $\bar{\mathfrak{o}}$, \bar{a} , (nó \bar{u} : $\bar{\mathfrak{o}}$, \bar{a} : $\bar{\mathfrak{a}}$) \bar{a} , \bar{i} : $\bar{\varepsilon}$, $\bar{\mathfrak{x}}$: $\bar{\mathfrak{x}}$, $\bar{\mathfrak{r}}$.
- VII. Guthaidhthe meadhoin, l. 67: (a) \mathfrak{i} , \bar{i} , \mathfrak{o} , \mathfrak{x} , a , \bar{a} , \mathfrak{e} , \mathfrak{x} , Φ , a , \mathfrak{y} ; (b) \bar{i} , \bar{y} , \mathfrak{o} , \mathfrak{e} , \bar{a} .
- VIII. Na deoghra, § 304, 305: $\bar{u}\mathfrak{o}$, $\bar{u}\mathfrak{x}$, $\bar{i}\mathfrak{o}$, $\bar{i}\mathfrak{x}$, $\mathfrak{o}\mathfrak{l}$, $\mathfrak{o}\bar{i}$, $\mathfrak{e}\mathfrak{i}$, $\mathfrak{o}\mathfrak{y}$ (ord), $\mathfrak{e}\mathfrak{u}$ ($bheadh:v'eu$).

Figure 2. The phonetic symbols used by Ó Máille. Section 342 lists the consonants, and section 343 lists the vowels. All of these but the INSULAR G can be represented in the UCS at present.

One might recommend writing the interdental letters N, L (in 342.IV above) with N U+0274 and L U+029F rather than with capitals; Ó Máille has used small capital versions for the voiceless versions of these in any case. His rather unusual use of bold-face letters for alveolar consonants (in 342.VII above) can be handled with the mathematical symbols in Plane 1, though as these sounds are foreign in Irish in any case it is of little consequence. In 343.III, unusually, Greek PHI \mathfrak{p} is used where we would expect \mathfrak{o} ; this may be a printer's error though it occurs elsewhere in the book, and a CAPITAL PHI Φ is used in 343.VII. Note the TIRONIAN SIGN ET in 342.X.

agas an'shin' Nuair' a v'i 'n' sehaγ fdlə agas ə Loχd ə
 mīç əs i rē l'ē dul an' šql' də hug šē kir' ə d'iN'er' gən v'æn'
 šq' gəs dā f'ar ə (l'ē) hīχd er' bəufd' eg' ə t'r'i ə χlog ə
 tr'hnōnə šən', ənsən' am, də v'i k'ap' zōb. Nuair' a v'i N
 t'am kat'i agas ə d'iN'er' hart dārt' ə šan-ar go rē šē n' am ə
 zol əwəl'ə. Nuair' a fuair' ə kəp't'in' hīs iəd (ēd) eg' ə maird
 eg' içə N'iN'er' də zo šē 'maχ agas də hug šē əurdū gən wātə
 n' sēχ ə χur ən' šql' agas də zo šē šd'aχ ē f'ēn' ər' išd'
 hær' ēš əN ordū šən' ə hōrt' wət. də hug ə mātə əurdū gə
 Nə mārN'ēlt ə zol əs Nə sLata šql' agas Nə šōLta sgl'ə Nūas
 agas də r'iNə šjad šin' əgəs kir'ū 'n' sēχ ən' šōl'. an' šən'
 Nuair' a d'air' i n' šqn-ar šq' nə hæsə ē hēn' agas ə v'qn l'ē gol
 ə wal'ə k'ē wēiχ švəd ēd hēn' aχ er' wəurd sehə, ə Lār Nə
 farəg'ə mōr'ə, tē m'īç gə Nə hārəN'əχt frēšən'.

Figure 3. Sample text of a story called “An Fuadach” (The Kidnapping) told by Peadar Ó Hallmhuráin, who hailed from near Galway.

Agus annsin nuair a bhí 'n soitheach falamh agus
 an locht amuigh is í réidh le dul 'un seoil, do thug sé
 cuireadh dinnéir don bhean seo agus dá fear a (le
 thigheacht ar bord ag an tri a chlog an trathnóna sin,
 insan am do bhí ceaptha dhoibh. Nuair a bhí an t-am
 caithte agus an dinnéar thart, dubhairt an sean-fhear go
 raibh se in' am a dhul abhaile. Nuair a fuair an caiptín
 thíos iad ag an mbord ag ithe a ndinnéir do ghabh sé
 amach agus do thug sé ordú don mháta an soitheach a
 chur un seoil agus do ghabh sé isteach é féin arís, thar
 éis an ordú sin a thabhairt saidh. Do thug an máta
 ordú do na máirnéalaigh a dhul agus na slata seoil agus
 na seolta a sgaoileadh anuas agus do rinne siad sin, agus
 cuireadh an soitheach un seoil. Annsin nuair a d'éirigh
 an sean-fhear seo ina sheasamh é féin agus a bhean le
 dul abhaile, cé bhfaigheadh siad iad féin ach ar bhord
 soithigh i lár na fairrge móire, taobh muigh de na
 hÁrainneacha freisin.

Figure 4. The same text in 1927 Roman-face orthography.

A. Administrative

1. Title

Proposal to encode one Irish phonetic letter in the UCS.

2. Requester's name

Michael Everson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2003-10-05

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical -- General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

Proposed name of script

1b. The proposal is for addition of character(s) to an existing block

Yes.

1b. Name of the existing block

Phonetic extensions.

2. Number of characters in proposal

1

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

4b. Is a rationale provided for the choice?

No.

4c. If YES, reference

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

No.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Spacing letter, category "Lo", bidi category "L" (strong left to right)

1D79

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Nicholas Williams.

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Linguists and phoneticians.

4a. The context of use for the proposed characters (type of use; common or rare)

Phonetic transcription.

4b. Reference

As above.

5a. Are the proposed characters in current use by the user community?

Only in older phonetic texts; most researchers use IPA now.

5b. If YES, where?

By linguists and phoneticians.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Latin phonetic character.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

As noted above, it is similar but not identical to the Latin letter yogh.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?