

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 six punctuation characters in the UCS
Source: Michael Everson
Status: Individual Contribution
Date: 2009-12-05

This proposal requests the encoding of six punctuation characters for linguist, Medievalist and Celticist use. If this proposal is accepted, the following characters will exist:

- | | | | |
|----|------|-------------------------|--|
| ◌̣ | 2E32 | TURNED COMMA | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate nasalization <p>x 060C ARABIC COMMA</p> |
| ◌̤ | 2E35 | TURNED SEMICOLON | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate sudden glottal closure <p>x 061B ARABIC SEMICOLON</p> |
| ◌̥ | 2E36 | DAGGER WITH LEFT GUARD | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate retracted pronunciation |
| ◌̦ | 2E37 | DAGGER WITH RIGHT GUARD | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate advanced pronunciation |
| ◌̧ | 2E38 | TURNED DAGGER | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate retroflex pronunciation |
| ◌̨ | 2E39 | TOP HALF SECTION SIGN | <ul style="list-style-type: none">• used in Palaeotype transliteration to indicate pronunciation on one side of the mouth only |

The TURNED COMMA was used in “Palaeotype”, a pre-IPA phonetic alphabet used by Alexander Ellis in his massive and classic four-volume work on early English pronunciation, published in 1869. The TURNED COMMA indicates nasalization, so (in modern IPA) a. means [ã]. Palaeotype also uses the punctuation marks , (COMMA) and ,, (two COMMAS side by side) and . (FULL STOP) and ., (FULL STOP and COMMA) and ; (SEMICOLON) for various purposes. It is not proposed to re-encode all of these as modifier letters; accordingly, the punctuation mark TURNED COMMA is requested.

The TURNED SEMICOLON was used in “Palaeotype” to indicate a sudden glottal closure. It has also been used as an ordinary punctuation character (perhaps in fallback representation of the medieval *punctus elevatus*).

The DAGGER WITH LEFT GUARD was used in Palaeotype to indicate a “retracted” pronunciation (“nearer to the throat and fruther from the lips”), so b‡ = labiodental [b̠], p‡ = labiodental [p̠].

The DAGGER WITH RIGHT GUARD was used in Palaeotype to indicate an “advanced” pronunciation (“further from the throat and nearer to the lips”), so d‡ = dental [d̠], t‡ = dental [t̠].

The TURNED DAGGER was overlooked when a number of Lhuydian characters were proposed for encoding in N3122, 2006-08-06, “Proposal to add Latin letters and a Greek symbol to the UCS”. In Lhuyd’s 1707 *Archaeologia Britannica*, the dagger is used to indicate words which are attested in the Old Cornish Vocabulary; the inverted dagger is used to indicate words in “modern Cornish”, that is, in Cornish of the late 18th century. In Palaeotype, it indicates retroflexion, so d‡ = [d̠], t‡ = [t̠].

The TOP HALF SECTION SIGN was used in Palaeotype to indicate that the breath was allowed to escape on one side of the tongue or mouth; The section sign was used to indicate that the breath escaped on both sides. Thus l‡h would represent = [ɬ], and l‡h either [ɬ̠] or [ɬ̠̠].

Unicode Character Properties. Character properties are proposed here.

```
2E32;TURNED DAGGER;Po;0;ON;;;;;N;;;;;
2E35;TURNED SEMICOLON;Po;0;ON;;;;;N;;;;;
2E36;DAGGER WITH LEFT GUARD;Po;0;ON;;;;;N;;;;;
2E37;DAGGER WITH right GUARD;Po;0;ON;;;;;N;;;;;
2E38;TURNED DAGGER;Po;0;ON;;;;;N;;;;;
2E39;TOP HALF SECTION SIGN;Po;0;ON;;;;;N;;;;;
```

Bibliography

- Ellis, Alexander J. 1869. *On Early English pronunciation with especial reference to Shakspeare and Chaucer*. 4 vols. London: The Philological Society, The Early English Text Society, and The Chaucer Society.
- Jenner, Henry. 1877. “The History and Literature of the Ancient Cornish Language.” A Paper read before the British Archaeological Association at Penzance, August 19th, 1876. *British Archaeological Journal*, 1877. 8vo.
- Lhuyd, Edward. 1707. *Archaeologia Britannica. Vol. I: Glossography*. Oxford: The Theater

Examples.

- (,) turned apostrophe, nasalize the preceding letter, but not as in F. nasalisation (Δ)
- (;) turned !, attempt to pronounce the preceding letter with inspired breath, (f;, ph;), calling a bird
- (‡) attempt to pronounce the preceding letter with the air in the mouth without inspiring or expiring, click, E. tut = (t‡), E. cl'ck (tj‡t‡)

Figure 1. Example from Ellis 1868 vol 1, showing the TURNED COMMA, called here “turned apostrophe (it is low and in UCS terminology is a turned comma).

Also seen are INVERTED EXCLAMATION POINT used as a sign of ingression and DOUBLE DAGGER used do indicate a kind of click.

- (:) turned semicolon, close the glottis suddenly as in stammering, or suddenly cease any sound, as when startled, leaving a sound half uttered; (H:) is a suddenly checked emission of breath, strongly resembling a click (‡), as in Zulu (ik.H:wa), *Visible Speech*, p. 126.

Figure 2. Example from Ellis 1868 vol 1, showing the TURNED SEMICOLON. The Zulu example ik.H:wa is, I believe, *iklwa* [ik‡wa] ‘broad-bladed spear’.

PLATE I.

LATIN-CORNISH VOCABULARY OF THE LATTER PART OF THE TWELFTH CENTURY.

[N.B.—The words are arranged in columns for the sake of greater convenience. Those that occur in brackets are interlinear and marginal glosses in Welsh added by a later hand. The first and last words are supplied by the preceding and succeeding pages.]

Equus. march	Porcus : hoch
Equa. cassec	Sus : baneu
Asinus. vel asina : asen	Scroffa : guis
Camelus : caurmarch	Aper vel verres : bahet
Onager : asenguills [assenguyll]	Magalis : torch
Elephans : oliphans	Porcellus : porchel
Ursus : ors	Bos : odion
Simia : sim	Vacca : vel Buccula : buch [buwch]
Latrius : doferghi	Vitulus : loch [llo]
Fiber : befer	Iuuenus : deneuoit
Feruncus : yeugen	Ouis : dauat
Mustela : louennan	Aries : horp
Talpa : god	Uernes : mols
Cattus vel murilegus : kat	Agnus : oin
Hyricius vel Erinatius : sort	Pecus vel Iumentum : chal
Clissemus vel mus : vel soorrex : logoden [lhogod]	Animal : mil
Uermis : prif	Canis : ki
Ceruus : caruu ¹	Molusus : guilter
Cerua : euhic	Catulus : coloin
Dama. vel Damula : da	Draco : druic
Hinnulus : loch euhic [lho ewic]	Vipera. vel serpens. vel anguis : nader
Capreolus : kytiorch	Coluber : gorþfel
Caprea : yorch	Rubeta : croinoc
Caper. vel hyrcus : boch	Rana : guilschiu
Capra. vel capella : gauar	Lacerta : wedresif
Hedus : min	Stellio : anaf

Figure 3. Example from Jenner 1877, showing TURNED SEMILCOLON as an ordinary punctuation character.

- † made from †, take the preceding letter nearer the throat and further from the lips, inner position.
- † made from †, take the preceding letter further from the throat and nearer to the lips, outer position.
- † turned †, invert the tongue so that the under part strikes the palate, when pronouncing the preceding letter, see (D, L, N, R, sh, T)
- † protrude the tongue when pronouncing the preceding letter.
- § bi-lateral, allow the breath to escape on both sides of the tongue or mouth, but not over the tip of the tongue or through the middle of the mouth.
- § made from §, uni-lateral, allow the breath to escape on one side of the tongue or mouth only.
- ‡ turned †, trill any free part during the utterance of the preceding consonant.

Figure 4. Example from Ellis 1868 vol 1, showing the DAGGER WITH LEFT GUARD, DAGGER WITH RIGHT GUARD, TURNED DAGGER, U+2020 DAGGER, U+00A7 SECTION SIGH, TOP HALF SECTION SIGN, and U+00BF INVERTED QUESTION MARK.

246	TIT. VI. A CORNISH GRAM	
<p><i>The Passive Voice, or as it may be otherwise called, the Impersonal of this Verb had an Infinitive or Participle, of the Present Tense join'd with it: As Yz eziz a kil huél ymma, I, thou, he, we, ye, or they work here. viz. q. d. Working is perform'd here, whether by me, thee, him or them, &c.</i></p> <p><i>Note that Ma or ymâ [is;] and ymônz [They are] must be always used Narratively and never Interrogatively, nor Negatively; and that they must be always placed before their Nominatrive; but eu or yu, yzi [or † ydzhi] and † asens or † ydzhenz, Either way and in either place: As Yma 'n levar ymma, The book is here; and not An levar ma ymma; Ev eu dên dâ or dên dâ eu, He is a good man; Yu an vez-na gyz hôr? Is that maid your Sister? † Ydzhi 'gyz tâz bêu? Is your Father living?</i></p> <p><i>The Present Tente of the other Verbs, is in the Welsh generally express'd by this Verb Substantive and a Participle of the Present Tense: As Yr yduyv yn 'skrivenny [I am writing] for I write; but in Cornish, as in the English, by the following Verb Gureiz, gil or gry-</i></p>	<p>Gurâ, <i>Do thou</i>; Gurêth and gurez, <i>Let him do</i>; Guraun, <i>Let us do</i>; Gureuh [and gureus] <i>Do ye</i>; Guraenz, <i>Let them do.</i></p> <p>Dhe 'ureiz, dhe 'urithil, dhe 'il and dhe huil, <i>To do or make.</i></p> <p>Another Auxiliary Verb in the Cornish is Menni [or † mynni] <i>To will; which they use also as in the English, before the Future Tente of the Indicative Mood, and as would in the Optative, &c.</i></p> <p>Mennav [and me a ven] <i>I will</i>; Menni or ti a venni, <i>Thou wilt</i>; Men or &c. <i>He will</i>; Mennon, <i>We will</i>; Mennoh, <i>Ye will</i>; Mennan and -anz, <i>They will.</i></p> <p>Mennen, <i>I would or was willing</i>; Mennyz, <i>Thou wouldst or wert willing</i>; Mennaz, <i>He would</i>; Mennen, <i>We would</i>; Menneh, <i>Ye would</i>; Mennenz, <i>They would.</i></p> <p>Menzhon, <i>I would or had been willing</i>; Menzhez, <i>Thou wouldst or hadst been willing</i>; Menzhe, <i>He would or had been willing</i>; Menzhen, <i>We would or had been willing</i>; Menzheh, <i>Ye would or had been willing</i>; Menzhenz, <i>They would, &c.</i></p> <p>Mennav or me a ven, <i>I will</i>; Menni Mennith and Menta, <i>Thou</i></p>	<p>ſe G W ſe T ſe w ue ſe T le: & G L u ſe M M h h I T le: le: le:</p>

Figure 5. Example from Lhuyd 1707, showing DAGGER alongside TURNED DAGGER.

A. Administrative

1. Title

Proposal to encode six punctuation characters in the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2009-12-05

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.

1b. Proposed name of script

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

Yes

1d. Name of the existing block

Supplemental Punctuation

2. Number of characters in proposal

6.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category B-1.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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)?

Yes.

8. 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.

See above.

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?

I am a member of the user community, preparing publications which need these characters.

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, Medievalists, and Celticists.

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

Used historically and in modern editions.

4b. Reference

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

Yes.

5b. If YES, where?

Scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other punctuation characters.

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

No.

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?

Yes.

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

Yes.

9c. If YES, reference

Two look like Arabic characters, but those have LTR directionality.

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

No.

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

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?