

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 for the addition of four Latin characters to the UCS
Source: Michael Everson
Status: Individual Contribution
Date: 2011-04-29

This proposal requests the encoding of four Latin characters comprising one casing pair, and two capital letters which provide casing support for two existing characters. If this proposal is accepted, the following characters will exist:

B̤	A794	LATIN CAPITAL LETTER B WITH FLOURISH
b̤	A795	LATIN SMALL LETTER B WITH FLOURISH
		• Middle Vietnamese
Ʒ	A7AB	LATIN CAPITAL LETTER REVERSED OPEN E
		x 025C Ʒ LATIN SMALL LETTER REVERSED OPEN E
G̤	A7AC	LATIN CAPITAL LETTER SCRIPT G
		x 0261 g̤ LATIN SMALL LETTER SCRIPT G

The B̤ b̤ B WITH FLOURISH is found in the dictionary of Alexandre de Rhodes, which directly led to the modern system of Vietnamese spelling. The de Rhodes' dictionary, the *Dictionarium Annamiticum Lusitanum et Latinum*, used this letter to represent a voiced bilabial fricative [β], a sound which was lost within a century or so, merging with [v], represented by v in modern Vietnamese orthography. To describe Middle Vietnamese, the B WITH FLOURISH IS required.

The Ʒ CAPITAL REVERSED OPEN E serves as an upper-case equivalent of U+025C Ʒ LATIN SMALL LETTER REVERSED OPEN E in the same way as U+0190 Ė LATIN CAPITAL LETTER OPEN E serves as an upper-case equivalent of U+025B ė LATIN SMALL LETTER OPEN E. The absence of this character was noticed in a case-pairing operation involving small-caps. Typically typesetting software makes use of case pairs to generate a small capital in styled text, but where no case-pair exists, this operation fails.

The G̤ CAPITAL SCRIPT G serves as an upper-case equivalent of U+025C g̤ LATIN SMALL LETTER SCRIPT G in the same way as U+2C6D Q̤ LATIN CAPITAL LETTER ALPHA (*alias* SCRIPT A) serves as an upper-case equivalent of U+0251 a̤ LATIN SMALL LETTER ALPHA.

Unicode Character Properties. Character properties are proposed here.

025C;LATIN SMALL LETTER REVERSED OPEN E;Ll;0;L;;;;;N;LATIN SMALL LETTER REVERSED EPSILON;;A7AB;;A7AB
 0261;LATIN SMALL LETTER SCRIPT G;Ll;0;L;;;;;N;;A7AC;;A7AC
 A794;LATIN CAPITAL LETTER B WITH FLOURISH;Lu;0;L;;;;;N;;;;;A794;
 A795;LATIN SMALL LETTER B WITH FLOURISH;Ll;0;L;;;;;N;;A795;;A795
 A7AB;LATIN CAPITAL LETTER REVERSED OPEN E;Lu;0;L;;;;;N;;;;;025C;
 A7AC;LATIN CAPITAL LETTER SCRIPT G;Lu;0;L;;;;;N;;;;;0261;

Bibliography

de Rhodes, Alexandre . 1651. *Dictionarium Annamiticum Lusitanum et Latinum*. Rome: Propaganda Fide.

Carroll, Lewis. [2011, in press]. *Alice’s Adventures in Wonderland in the International Phonetic Alphabet: ‘Ælisɪz ðd’ventʃəz in ‘Wʌndərlænd*. Cathair na Mart: Evertype. ISBN 978-1-904808-72-5.

Examples.

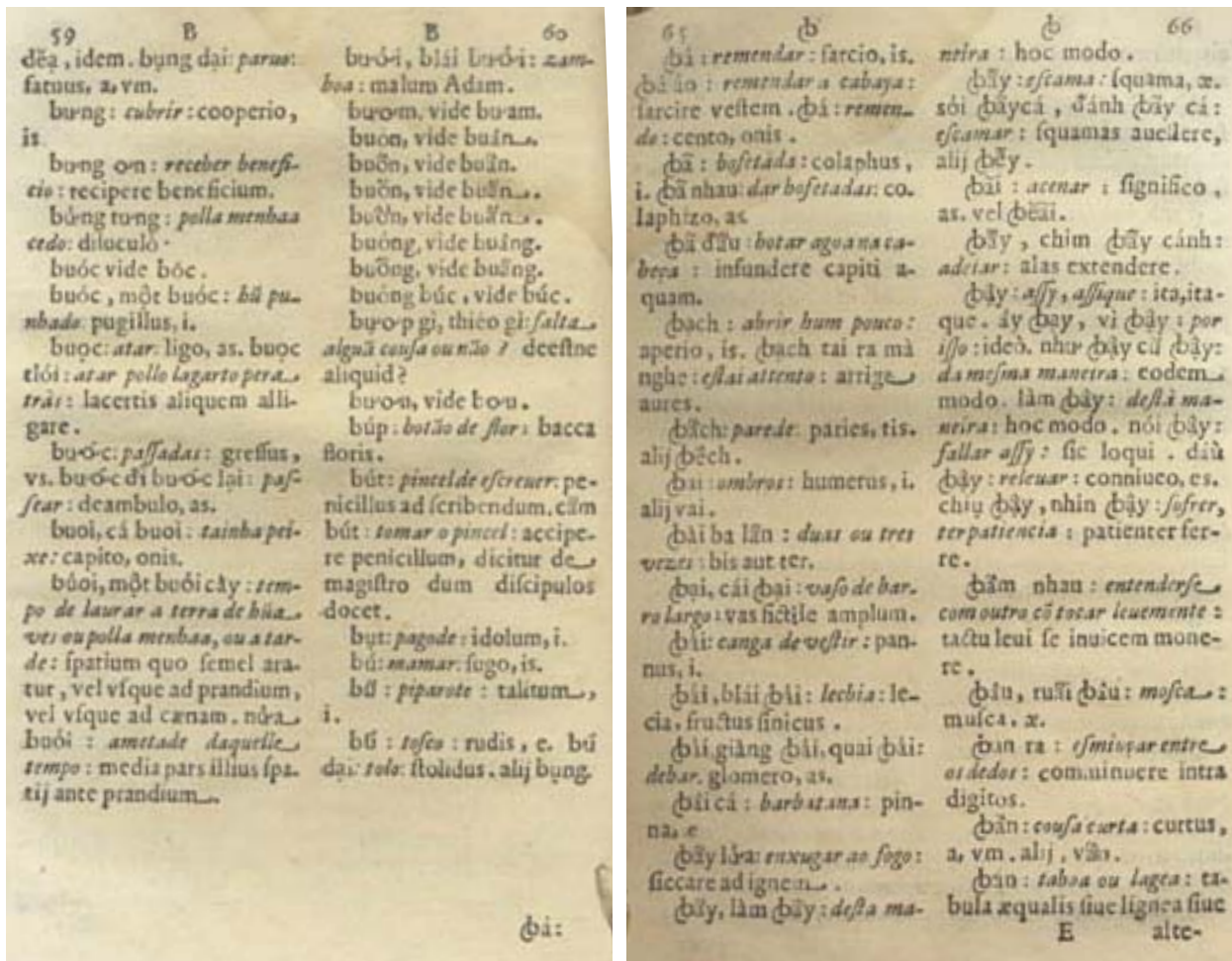


Figure 1. The last page of chapter B and the first page of chapter B of de Rhodes 1651, showing SMALL LETTER B in headwords and CAPITAL LETTER B at the head of the columns and showing SMALL LETTER B WITH FLOURISH in headwords. De Rhodes did not have a CAPITAL LETTER B WITH FLOURISH cut for the dictionary, but it is clear that had he done so it would have appeared at the head of the columns, which use capitals throughout the book.



'F ɔ : w ɜ : d

'L u:ɪs 'Kæɾəl ɪz ə pen neɪm: Tʃɑ:lz 'Lætwidz 'Dɒdsən wɒz ði: 'ɔ:θəz rɪəl neɪm ænd hi: wɒz 'lektʃərəɾ m ,Mæθɪ'mætɪks m Kraɪst Tʃɜ:tʃ 'Dɒksfəd. 'Dɒdsən brɪ'gæən ðə 'stɔ:ri ɒn 4 Dʒu:'laɪ 1862, wen hi: tʊk ə 'dʒɜ:mɪ m ə 'rəʊɪŋ bəʊt ɒn ðə 'rɪvə Temz m 'Dɒksfəd tə'geðə wɪð ðə 'Revərənd 'Rɒbɪnsən 'Dakwɜ:θ, wɪð 'Ælɪs 'Lɪdl (ten jɪəz ɒv eɪdʒ) ðə 'dɔ:tər ɒv ðə Di:n ɒv Kraɪst Tʃɜ:tʃ, ænd wɪð hɜ: tʊ: 'sɪstəz, Ləri:nə ('θɜ:'ti:n jɪəz ɒv eɪdʒ) ænd 'ɪdlθ (ert jɪəz ɒv eɪdʒ). Æz ɪz klə frɒm ðə 'pəʊɪm æt ðə brɪ'gɪnɪŋ ɒv ðə bu:k, ðə θri: gɜ:lz ɑ:skt 'Dɒdsən fə ə 'stɔ:ri ænd rɪ'læktəntli æt fɜ:st hi: brɪ'gæən tə tel ðə fɜ:st 'vɜ:ʃən ɒv ðə 'stɔ:ri tə ðəm. ðeər ɑ: 'meni hæ:f 'hɪdn 'refrənsɪz meɪd tə ðə faɪv ɒv ðəm θru:'aʊt ðə tekst ɒv ðə bu:k ɪt'self wɪtʃ wɒz 'pʌblɪʃt 'faməlɪ m 1865.

ðɪs rɪ'dɪʃən ɒv 'Ælɪsɪz ðə'ventʃəz ɪn 'Wandələænd prɪ'zents ðə tekst m ən ,ɪntə'næʃənl Fəʊ'netɪk 'Ælfəbɪt trəns'krɪpʃən. ðə trəns-'krɪpʃən rɪ'flektɪs ðə 'stændəd 'rɪ:dʒənli 'nʃu:trəl fə:m ɒv 'spəʊkən 'Brɪtɪʃ 'ɪŋɡlɪʃ nəʊn æz "Rɪ'sɪ:vɪd Prə'nansɪ'eɪʃən". ɔ:ɾɑ:ntɪd ðæt məʊst 'ɪŋɡwɪsts ə'gri: ðæt nɒt mætʃ mə: ðæn 4% ɒv ðə ,pɒpjʊ'leɪʃən ɒv 'Brɪtən spɪ:k ɪt tə'deɪ, 'ɑ: 'Pi: wɒz ,nevəðə'les trə'dɪʃnəlɪ beɪst ɒn 'edʒu:kertɪd spɪ:tʃ m 'sʌðən 'ɪŋɡlənd; ɪt ɪz stɪl 'wɑ:dlɪ tɔ:t ænd 'dɪkʃənɪz fə 'neɪtv 'spɪ:kəz ænd 'lɜ:nəz ɒv 'ɪŋɡlɪʃ stɪl meɪk ju:s ɒv ɪt m ðeə trəns'krɪpʃənz.

Tʊ 'prɒdʒu:s ðə tekst hɪər Aɪ fɜ:st 'prəʊsest ðə tekst θru: ðlek'seɪ Vɪnɪ'dɪktəfs "Fəʊnətəɪzə", ə tekst kən'vɜ:ʃən 'prəʊgræm wɪtʃ

Figure 2. Example from Carroll [2011; in press], showing CAPITAL LETTER SCRIPT G.

Note that the word “Foreword” has been transcribed “Fɔ:rwɜd” in Received Pronunciation. If this word were set in small caps or all caps, it should appear as “'Fɔ:RWɜD” or “'Fɔ:RW3D”, but without CAPITAL LETTER REVERSED OPEN E, it would appear as “'Fɔ:RWɜD” or “'Fɔ:RW3D”. (In the typeset example here, Private Use characters have been used for both ɔ and 3.)

ɪm'plɔɪz ði: 'ɑ: 'Pi: træns'krɪpʃən ju:st ɪn ðə 'sevnθ ɪ'dɪʃən ɒv
 Vlə'di:mə Keɪ 'Mɪlləz 'Iŋɡlɪʃ 'Rafən 'Dɪkʃənərɪ ('Mɒskəʊ, 1960).
 'Sʌbsɪkwəntli Aɪ red θru: ðə tekst 'keəfli ænd meɪd ə 'feəli lɑ:dʒ
 'nʌmbər ɒv ə'dʒʌstmənts. ɪn 'meni 'keɪsɪz ðeə wɜ: 'dʒʌdʒmənt kə:lz tə
 bi: meɪd, ænd Aɪ hæʊp ðæt Aɪ hæʊp dʌn ə ,sætɪs'fæktəri dʒɒb ɒv ɪt.
 Məʊst ɒv ði:z 'tʃɔɪsɪz wɜ: 'feəli 'mʌndem, sʌtʃ æz wen tə ju:z [ænd]
 ənd wen tə ju:z [ænd] ə: wen tʊ 'ɪnsət ə 'lɪŋkɪŋ "r" (æz ɪn *her idea*
 [hɜ:ɪ aɪ'diə] 'rɑ:ðə ðæn [hɜ: aɪ'diə]). Aɪ hæʊp nɑ:t ə'temptɪd tʊ: ɪn'sart
 ən ,epən'θetɪk "r" (æz ɪn *her idea of* [hɜ:r aɪ'diə ɒv]) sɪnz ðɪs ɪz
 kən'sɪdərɪd səb'stændɪd. Pə'hæps ɪt ɪz nɒt 'pɒsəbl tʊ ə'tʃɪv kəm'pli:t
 pə'fekʃən ɪn ə træns'krɪpʃən ɒv ə 27,500-wɜ:d 'nɒvəl—bʌt Aɪ brɪ'li:v
 ðæt ðə tekst hɪər ɪz 'feəli kən'sɪstənt ənd 'rɪ:dəbl. Aɪ 'welkəm
 kə'rekʃənz frɒm 'rɪ:dəz hu: wɪʃ tə səb'mɪt ðəm.

Bɪ'kɒz ðɪs ɪz ə 'nɒvəl, ænd ment tə bi: red, Aɪ dɪ'saɪdɪd tə rɪ'tem
 sʌm ,ə:θə'græfɪk 'fɪ:tʃəz wɪtʃ ɑ: nɒt 'nə:məli kept ɪn fəʊ'netɪk
 træns'krɪpʃən: ,pʌŋktjʊ'eɪʃən, ɪ'tæɪsɪs'zeɪʃən ænd kə'pɪtəlaɪ'zeɪʃən.
 Aɪ hæʊp rɪ'temd ə:l ɒv 'Kæɪrəlz ,pʌŋktjʊ'eɪʃən wɪð ði: ɪk'sepʃən ɒv ði:
 ə'pɒstrəfɪ 'mɑ:kɪŋ ðə 'dʒenɪtɪv (bɪ'kɒz *Duchess's voice* dʒʌst lʊkt rɒŋ
 æz '*Dafnis's voice*). Məʊst 'Aɪ 'Pi: 'Eɪ 'kæɪrɪkətəz hæʊp fə'mɪljər ɪ'tæɪk
 ənd 'kæpɪtl fə:mz, bʌt fə ðə kən'vɪ:njəns ɒv ðə 'rɪ:dər Aɪ ɡɪv ðə
 'repətwa: hɪə:

Aa Cc Dd Ææ Bb Dd Ðð Ee Əə Ʒz Ff Gg
 Hh Īi Iɪ Kk Ll Mm Nn Ŋŋ Oo Ɔɔ Pp Rr
 Ss Σς Tt Θθ Uu Uυ Vv Λλ Ww Zz Ʒz

 Aa Cc Dd Ææ Bb Dd Ðð Ee Əə Ʒz Ff Gg
 Hh Īi Iɪ Kk Ll Mm Nn Ŋŋ Oo Ɔɔ Pp Rr
 Ss Σς Tt Θθ Uu Uυ Vv Λλ Ww Zz Ʒz

Figure 4. Example from Carroll [2011; in press], showing CAPITAL LETTER SCRIPT G and CAPITAL LETTER REVERSED OPEN E, alongside their lower-case equivalents, and showing all of the other I.P.A. characters used in this text, all of which have case-pairs. Of these, C, D, and Λ were added relatively recently as case-pairing additions of phonetic characters.

ænd ðə 'kɔnstənt 'hevi 'sɒbiŋ ɒv ðə Mɒk 'Tɜ:tl. 'Ælis wɒz 'veri 'ni:li 'getiŋ ʌp ænd 'seɪŋ, “Oæŋk jʊ, Sɜ:, fə jə: 'ɪntrɪstɪŋ 'stɔ:ri,” bʌt ʃi: kʊd nɒt help 'θɪŋkiŋ ðeə *mast* bi: mæ: tə kʌm, səʊ ʃi: sæt stɪl ænd sed 'nʌθɪŋ.



“Wen wi: wɜ: 'lɪtl,” ðə Mɒk 'Tɜ:tl went ɒn æt lɑ:st, mæ: 'kɑ:mlɪ, ðəʊ stɪl 'sɒbiŋ ə 'lɪtl naʊ ænd ðen, “wi: went tə sku:l ɪn ðə si:. ðə 'mɑ:stə wɒz ən əʊld 'Tɜ:tl—wi: ju:st tə kɔ:l hɪm 'Tɔ:təs—”

“Wai dɪd jʊ kɔ:l hɪm 'Tɔ:təs, ɪf hi: wɒznt wʌn?” 'Ælis ɑ:skt.

90

Figure 4. Example from Carroll [2011; in press], showing in the header the text “'ÆLISIZ ƆÐ'VENTSƆZ IN 'WANDƆLƆND” in small caps.

ÐƏ MĐK 'Tɜ:tlz 'stɔ:ri

“Wi: kə:ld him 'Tɜ:təs br'kɔz hi: tɔ:t əs,” sed ðə Mɔk 'Tɜ:tl 'æŋgrɪl. “Riəlɪ jʊ ə: 'veri dəl!”

“Jʊ ɔ:t tə bi: ə'feɪmɪd ɒv jɔ:'self fər 'ɑ:skɪŋ sətʃ ə 'sɪmpl 'kwɛstʃən,” 'ædɪd ðə 'Gɪrɪfən; ænd ðen ðei bæʊθ sət 'saɪlənt ænd lʊkt æt pʊər 'Ælɪs, hu: felt 'redɪ tə sɪŋk 'ɪntʊ ði: ɜ:θ. Æt lɑ:st ðə 'Gɪrɪfən sed tə ðə Mɔk 'Tɜ:tl “Draɪv ɒn, əʊld 'feləʊ! Dəʊnt bi: ɔ:l ðei ə'baʊt ɪt!” ænd hi: went ɒn ɪn ði:z wɜ:dz:—

“Jes, wi: went tə sku:l ɪn ðə si:, ðəʊ jʊ meɪənt br'li:v ɪt—”

“Aɪ 'nevə sed Aɪ dɪdnt!” ɪntə'ræptɪd 'Ælɪs.

“Jʊ dɪd,” sed ðə Mɔk 'Tɜ:tl

“Həʊld jɔ: tʌŋ!” 'ædɪd ðə 'Gɪrɪfən, br'fɔ:r 'Ælɪs kʊd spɪk ə'gen. Ðə Mɔk 'Tɜ:tl went ɒn.

“Wi: hæd ðə best ɒv ɪdju:'keɪʃənz—ɪn fækt, wi: went tə sku:l 'evrɪ ðei—”

“Aɪv bi:n tʊ ə ðei sku:l, tu:,” sed 'Ælɪs. “Jʊ nɪ:dnt bi: səʊ praʊd æz ɔ:l ðæt.”

“Wɪð 'ekstrəz?” ə:skt ðə Mɔk 'Tɜ:tl, ə 'ɪtl 'æŋkʃəslɪ.

“Jes,” sed 'Ælɪs: “wi: lɜ:nd frentʃ ənd 'mju:zɪk.”

“Ænd 'wɒʃɪŋ?” sed ðə Mɔk 'Tɜ:tl.

“Sɜ:tnlɪ nɒt!” sed 'Ælɪs ɪn'dɪgnəntlɪ.

“O:l! Ðen jɔ:z wɔznt ə 'riəlɪ gʊd sku:l,” sed ðə Mɔk 'Tɜ:tl ɪn ə təʊn ɒv greɪt rɪ'li:f. “Nav, æt 'avəz, ðei hæd, æt ði: end ɒv ðə bɪl, 'frentʃ, 'mju:zɪk, ənd 'wɒʃɪŋ—ekstrə’.”

“Jʊ kʊdnt hæv 'wɔntɪd ɪt mætʃ,” sed 'Ælɪs; “ɪvrɪŋ æt ðə 'bɒtəm ɒv ðə si:.”

“Aɪ kʊdnt ə'fə:d tə lɜ:n ɪt,” sed ðə Mɔk 'Tɜ:tl, wɪð ə saɪ. “Aɪ 'əʊnlɪ tʊk ðə 'regjʊlə kɔ:s.”

“Wɒt wɔz ðæt?” ɪn'kwærəd 'Ælɪs.

“Ri:lɪŋ ənd 'Raɪðɪŋ, ɒv kɔ:s, tə br'gm wɪð,” ðə Mɔk 'Tɜ:tl rɪ'pləɪd; “ænd ðen ðə 'dɪfrənt 'brɑ:ntʃɪz ɒv ʊ'rɪθmætɪk—Æm'bɪʃən, Dɪs'trækʃən, ʌŋlɪfr'keɪʃən, ənd Dɪ'rɪzən.”

“Aɪ 'nevə hɜ:d ɒv 'ʌŋlɪfr'keɪʃən,” 'Ælɪs 'ventʃəd tə seɪ. “Wɒt ɪz ɪt?”

Figure 5. Example from Carroll [2011; in press], showing in the header the text “ÐƏ MĐK 'Tɜ:tlz 'stɔ:ri” in small caps; without a casing pair for REVERSED OPEN E, the third word would be represented as 'Tɜ:tlz. Also shown on this page are three examples of CAPITAL LETTER SCRIPT G.

A. Administrative

1. Title

Proposal for the addition of four Latin characters in the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2011-04-29

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

Latin Extended-D

2. Number of characters in proposal

4.

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 A.

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?

Andrew West, Michael Everson, Wikipedia User Benwing.

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, phoneticians, Carrollians.

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?

Various 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 Latin phonetic 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

Case pairing letters.

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?