Title: Proposal to encode two combining characters in the UCS

State Commission of the Lithuanian Language

( contact: Vladas Tumasonis vladas.tumasonis@maf.vu.lt)

Correspondent Member

Submission date: 2006-03-02

Number of characters in proposal: 2

Proposed level of implementation: 3

Is a repertoire including character names provided? Yes

a. Are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000? Yes

b. Are the character shapes attached in a legible form suitable for review? Yes

Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? True Type, Vilnius University, Lithuania

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Mr. Mindaugas Strockis, (Mindaugas.Strockis@flf.vu.lt), FontLab

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

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

Special encoding issues:

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)?

The characters behave just like the existing combining characters

Additional Information:

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 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.

C. Technical - Justification

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

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?     __________________________
   If YES, with whom? _______________________________________
   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?     Yes
   Reference: Lithuanian Academic community _______________________

4. The context of use for the proposed characters (type of use; common or rare)     Phonetic transcription
   (infrequent, educational use)
   Reference: _______________________________________________________

5. Are the proposed characters in current use by the user community?     Yes
   If YES, where? Reference: In Lithuania, in linguistic literature, dictionaries

6. 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
   If YES, a rationale provided? Yes
   If YES, reference: __________________________

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

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?     No
   If YES, a rationale for its inclusion provided? ______________________
   If YES, reference: __________________________

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?     No
   If YES, a rationale for its inclusion provided? ______________________
   If YES, reference: __________________________

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?     No
    If YES, a rationale for its inclusion provided? ______________________
    If YES, reference: __________________________

11. 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)?     __________________________________________
    If YES, a rationale for such use provided? ______________________
    If YES, reference: __________________________
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? __________________________
    If YES, reference: __________________________

12. Does the proposal contain characters with any special properties such as control function or similar semantics?     No
    If YES, describe in detail (include attachment if necessary) __________________________

13. Does the proposal contain any ideographic compatibility character(s)?     No
    If YES, is the equivalent corresponding unified ideographic character(s) identified? __________________________
    If YES, reference: __________________________
ADDITIONAL INFORMATION

I. PROPOSED CHARACTERS

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+????</td>
<td>ꭜ</td>
<td>COMBINING BREVE-MACRON ABOVE</td>
</tr>
<tr>
<td>U+????</td>
<td>꭛</td>
<td>COMBINING MACRON-BREVE ABOVE</td>
</tr>
</tbody>
</table>

II. RATIONALE

1. Meaning and placing of the accent marks BREVE-MACRON and MACRON-BREVE in the Lithuanian accentuation system.

1.1. Mark types. Phonetic transcription of Lithuanian dialectological texts includes two accent marks, which are not presently included in ISO/IEC 10646:

1.1.1. Breve connected to macron: ꭜ
1.1.2. Macron connected to breve: ꭛

1.2. Meaning of the marks. These two accents mark the intonation of three-mora syllables in Lithuanian dialects. They occur both in the manuscript heritage of Lithuanian dialectological scholarship, and in published books on the subject.

1.2.1. Breve-macron ꭜ:

1.2.1.1. Marks an unstressed syllable of three moras with the rising intonation (for example, in sūnus).
1.2.1.2. Is used to denote the first component of a stressed secondary diphthong (for example, in gūrčius, čērka, barōnka or wienōn ulyczōn).

1.2.2. Macron-breve ꭛ marks unstressed syllables of three moras with falling intonation (for example, in Žmūgus).

1.3. The place of the marks in historical and modern linguistics.

1.3.1. These diacritical marks were used by Antanas Baranauskas (1835–1902), a prominent researcher into Lithuanian dialects. Contemporary linguists use dialectological texts collected and transcribed by him as well as his original poetry in his native dialect as a reference material when conducting diachronic and synchronic analysis of Lithuanian dialects.

1.3.2. These two accent signs are necessary for publishing dialectology works in the exact transcription system used by the original transcriber, both for printing linguistic books and publishing the content on the Internet.
1.4. According the existing rules both mentioned marks can be expressed by combining marks: ć COMBINING BREVE-MACRON ABOVE and ń COMBINING MACRON-BREVE ABOVE. At the moment, ISO/IEC 10646 does not include these two marks. Therefore it is requested that these combining marks be added to ISO/IEC 10646.

2. Use of the accent marks BREVE-MACRON and MACRON-BREVE in linguistic literature.

2.1. Baranauskas’ texts containing these diacritical marks are published in:


2.2. Critical edition:


2.3. The function of these diacritical marks is analysed and described in:

Litauische Mundarten gesammelt von A. Baranowski, Band II, Grammatische Einleitung mit lexikalischem Anhang bearbeitet von Dr. Franz Specht, Leipzig, 1922.

3. Scanned samples of the accent marks BREVE-MACRON and MACRON-BREVE in use.
