ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
(Please read Principles and Procedures Document for guidelines and details before filling this form.)
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html **for latest *Form*.**
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for latest *Principles and Procedures* document.**
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest roadmaps.**

A. Administrative

1. **Title:** **Proposal to encode two combining characters in the UCS**
2. Requester's name: State Commission of the Lithuanian Language
(contact: Vladas Tumasonis *vladas.tumasonis@maf.vu.lt*)
3. Requester type (Member body/Liaison/Individual contribution): Correspondent Member
4. Submission date: 2006-03-02
5. Requester's reference (if applicable):
6. (Choose one of the following:)
This is a complete proposal: Yes
or, More information will be provided later:

B. Technical - General

1. (Choose one of the following:)
a. This proposal is for a new script (set of characters):
Proposed name of script:
b. The proposal is for addition of character(s) to an existing block: Yes
Name of the existing block: Combining Diacritical Marks Supplement
2. Number of characters in proposal: 2
3. Proposed category (see section II, Character Categories): Combining
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): 3
Is a rationale provided for the choice?
If Yes, reference:
5. Is a repertoire including character names provided? Yes
a. If YES, 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
6. 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
7. 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
8. 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
9. 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

[1] Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

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, is 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, is 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, is 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, is 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, is 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

U+????	◌͡◌	COMBINING BREVE-MACRON ABOVE
U+????	◌͜◌	COMBINING MACRON-BREVE ABOVE

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: ◌͡◌

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 ◌͡◌:

1.2.1.1. Marks an unstressed syllable of three moras with the rising intonation (for example, in *sŭ͡mus*).

1.2.1.2. Is used to denote the first component of a stressed secondary diphthong (for example, in *gŏ͡rčius, čĕ͡rka, barŏ͡nka* or *wienŏ͡n ulyczŏ͡n*).

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

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: ◌͡◌ COMBINING BREVE-MACRON ABOVE and ◌͠◌ 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:

Litauische Mundarten gesammelt von A. Baranowski, Band I: Texte. Aus dem Weberschen Nachlass herausgegeben von Dr. Franz Specht, Leipzig, 1920.

2.2. Critical edition:

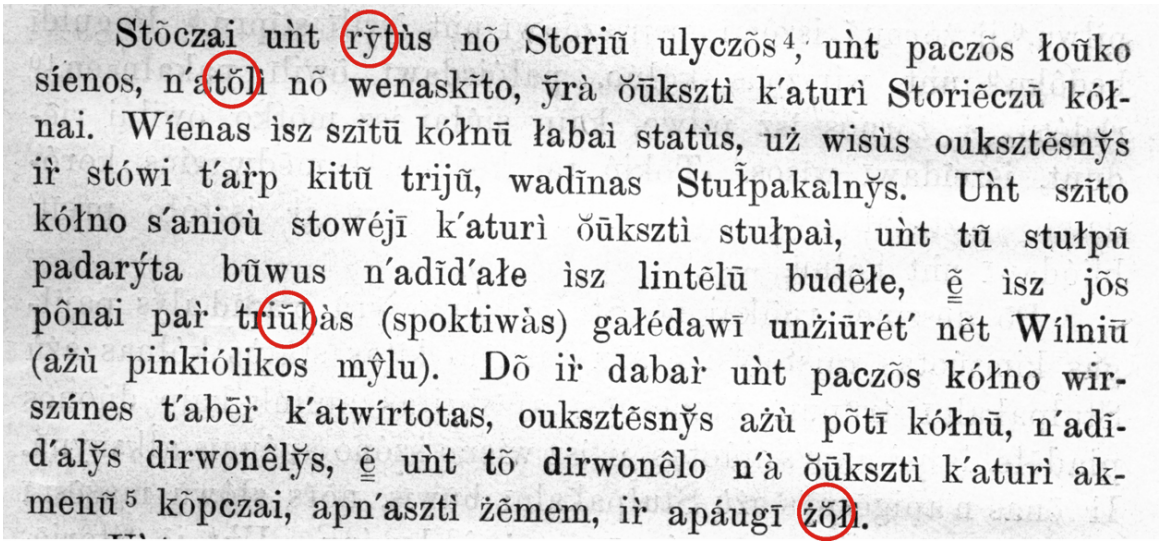
Antanas Baranauskas, *Raštai*, Tomas I. Poezija [Writings, Volume I. Poetry], eds. Regina Mikšytė, Marius Daškus, Vilnius: Baltos lankos, 1995.

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.

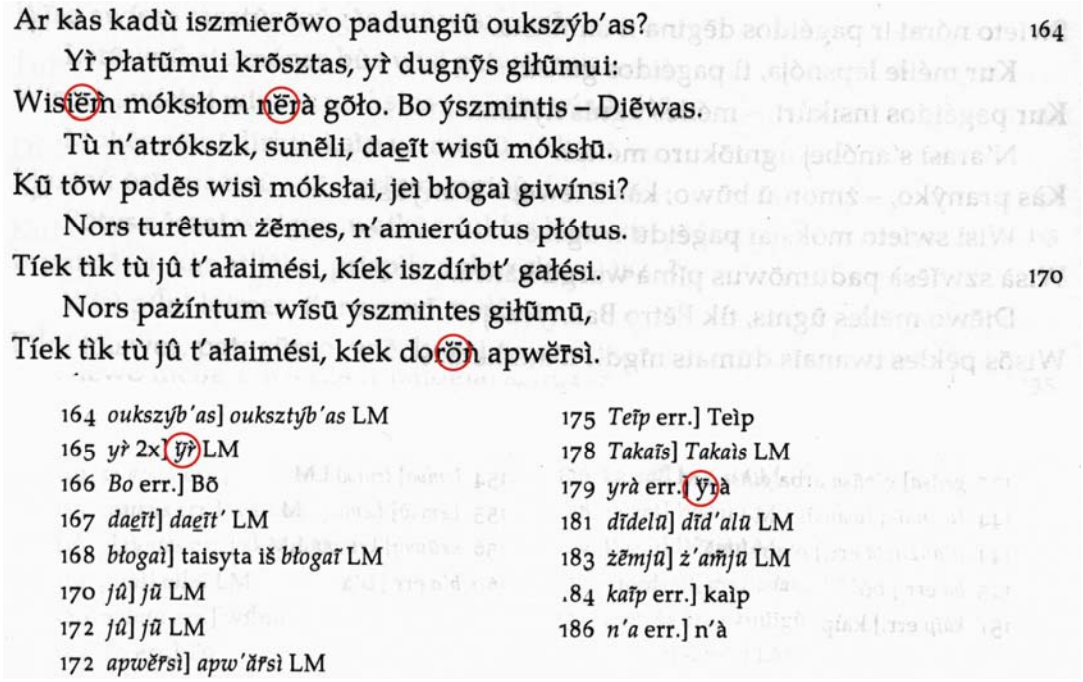
3. Scanned samples of the accent marks BREVE-MACRON and MACRON-BREVE in use.

3.1. *Litauische Mundarten gesammelt von A. Baranowski*, Band I: Texte. Aus dem Weberschen Nachlass herausgegeben von Dr. Franz Specht, Leipzig, 1920.



Stõczai ùnt rỹtùs nõ Storiũ ulyczõs[4], ùnt paczõs łoũko síenos, n'atõlì nõ wenaskĩto, ỹrà ŏũkszti k'aturì Storiẽczũ kół-nai. Wíenas ìsz szĩtũ kółnũ łabaì statùs, už wisùs ouksztẽsnỹs ìr stówi t'arp kitũ trijũ, wadĩnas Stułpakalnỹs. Ùnt szĩto kółno s'anioù stowéjĩ k'aturì ŏũkszti stułpaì, ùnt tũ stułpũ padarýta bũwus n'adĩd'ałe ìsz lintẽlũ budẽłe, ẽ ìsz jõs pônai paì triũbàs (spoktiwàs) gałédawĩ unżiũrét' nẽt Wílniũ (ażù pinkiólikos mŷlu). Dõ ìr dabaì ùnt paczõs kółno wir-szúnes t'abẽì k'atwírtotas, ouksztẽsnỹs ażù põtĩ kółnũ, n'adĩ-d'alỹs dirwonêlỹs, ẽ ùnt tõ dirwonêlo n'à ŏũkszti k'aturì ak-menũ[5] kõpczai, apn'aszti żẽmem, ìr apáugĩ żŏłi.

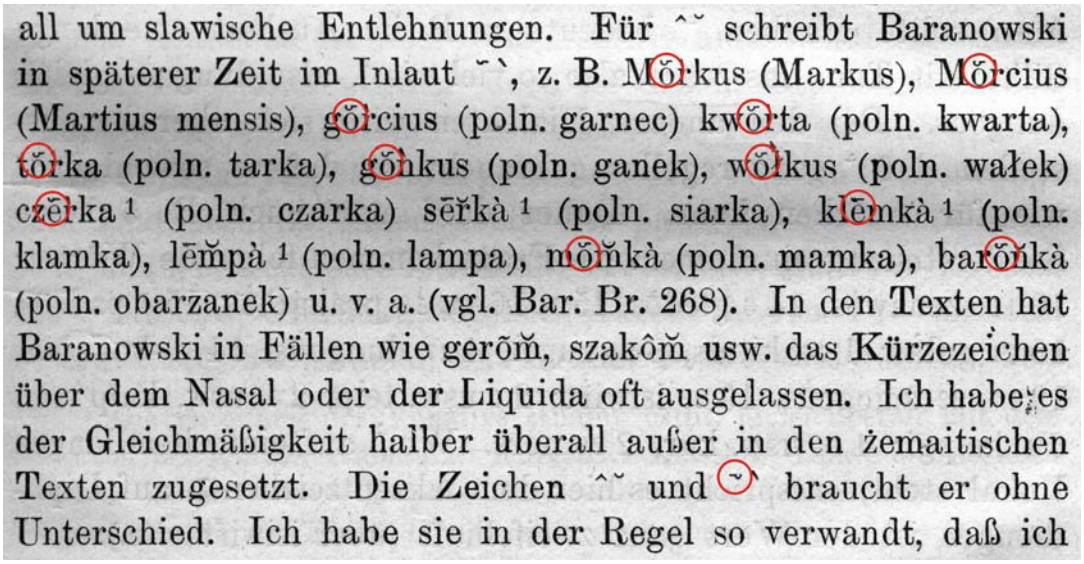
3.2. Antanas Baranauskas, *Raštai*, Tomas I. Poezija [Writings, Volume I. Poetry], eds. Regina Mikšytė, Marius Daškus, Vilnius: Baltos lankos, 1995.



Ar kàs kadù iszmierõwo padungiũ oukszýb'as? 164
Yr̀ płatũmui krõsztas, yr̀ dugnỹs giłũmui:
Wisiẽm̀ mókszłom nẽr̀à gõło. Bo ýszmintis – Diẽwas.
Tù n'atrókszk, sunẽli, daęīt wisũ mókszłũ.
Kũ tõw padẽs wisì mókszłai, jeì błogaì giwínsi?
Nors turêtum żẽmes, n'amierúotus płótus.
Tíek tìk tù jû t'ałaimési, kíek iszdírbt' gałési. 170
Nors pażíntum wīsũ ýszmintes giłũmũ,
Tíek tìk tù jû t'ałaimési, kíek dorõm̀ apwẽrsì.

164 *oukszýb'as*] *oukształb'as* LM
165 *yr̀* 2x] *ỹr̀* LM
166 *Bo* err.] Bõ
167 *daęīt*] *daęīt'* LM
168 *błogaì*] taisyta iš *błogaĩ* LM
170 *jû*] *jũ* LM
172 *jû*] *jũ* LM
172 *apwẽrsì*] *apw'ãrsì* LM
175 *Teĩp* err.] Teìp
178 *Takaĩs*] *Takaìs* LM
179 *yrà* err.] *ỹrà*
181 *dĩdelũ*] *dĩd'alũ* LM
183 *żẽmjũ*] *ż'amjũ* LM
184 *kaĩp* err.] kaìp
186 *n'a* err.] n'à

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



all um slawische Entlehnungen. Für ˆ˘ schreibt Baranowski in späterer Zeit im Inlaut ˘`, z. B. Mŏ̀rkus (Markus), Mŏ̀rcius (Martius mensis), gŏ̀rcius (poln. garnec) kwŏ̀rta (poln. kwarta), tŏ̀rka (poln. tarka), gŏ̀nkus (poln. ganek), wŏ̀lkus (poln. wałek) czẽrka [1] (poln. czarka) sẽrkà [1] (poln. siarka), klẽmkà [1] (poln. klamka), lẽmpà [1] (poln. lampa), mŏ̀mkà (poln. mamka), barŏ̀nkà (poln. obarzanek) u. v. a. (vgl. Bar. Br. 268). In den Texten hat Baranowski in Fällen wie gerõm̆, szakôm̆ usw. das Kürzezeichen über dem Nasal oder der Liquida oft ausgelassen. Ich habe es der Gleichmäßigkeit halber überall außer in den żemaitischen Texten zugesetzt. Die Zeichen ˆ˘ und ˘` braucht er ohne Unterschied. Ich habe sie in der Regel so verwandt, daß ich