

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹
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@ff.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? If YES explain _____	_____ <u>No</u> _____
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? Reference: <i>Lithuanian Academic community</i> _____	_____ <u>Yes</u> _____
4. The context of use for the proposed characters (type of use; common or rare) Reference: _____	_____ <u>Phonetic transcription</u> (infrequent, educational use)
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <i>In Lithuania, in linguistic literature, dictionaries</i> _____	_____ <u>Yes</u> _____
6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: _____	_____ <u>Yes</u> _____ _____ <u>Yes</u> _____
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? If YES, is a rationale for its inclusion provided? If YES, reference: _____	_____ <u>No</u> _____ _____
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: _____	_____ <u>No</u> _____ _____
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: _____	_____ <u>No</u> _____ _____
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? If YES, describe in detail (include attachment if necessary)	_____ <u>No</u> _____ _____
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: _____	_____ <u>No</u> _____ _____ _____ _____

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 *sūnus*).

1.2.1.2. Is used to denote the first component of a stressed secondary diphthong (for example, in *gōrčius*, *čērka*, *barōnka* or *wienōn ulyczōn*).

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

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:

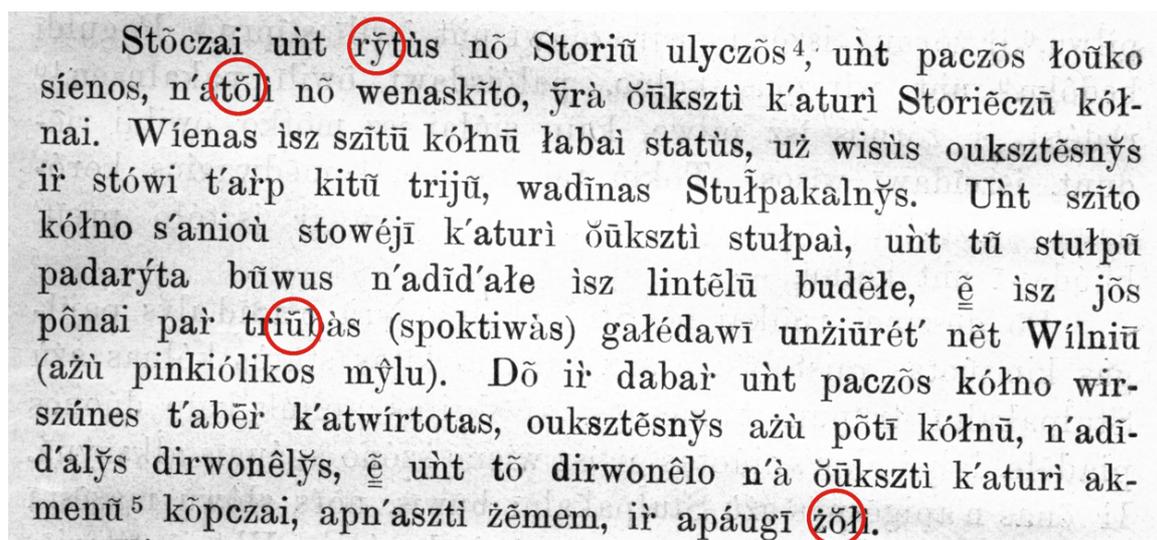
Antanas Baranauskas, *Raštai*, Tomas I. Poezija [Writings, Volume I. Poetry], eds. Regina Mikšytė, Marius Daškus, Vilnius: Baltos lankos, 1995.

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.



Stōczai ùnt rýtūs nō Storiū ulyczōs⁴; ùnt paczōs łouko
sienos, n'atōli nō wenaskīto, yrà ōuksztī k'aturi Storiēcziū kól-
nai. Wīenas isz szītū kólnū labai statūs, už wisūs ouksztēsnyš
ir stōwi t'arp kitū trijū, wadīnas Stūlpakalnys. ùnt szīto
kólno s'aniou stowėjī k'aturi ōuksztī stūlpai, ùnt tū stūlpū
padaryta būwus n'adīd'ale isz lintēlū budēle, ē isz jōs
pōnai par triūlās (spoktiwās) galédawī unziūrēt' nēt Wīlniū
(ażū pinkiólīkos mýlu). Dō ir dabr̄ ùnt paczōs kólno wir-
szūnes t'abēr k'atwīrtotas, ouksztēsnyš ažū pōtī kólnū, n'adī-
d'alys dirwonēlys, ē ùnt tō dirwonēlo n'à ōuksztī k'aturi ak-
menū⁵ kōpczai, apn'asztī zēmēm, ir apaugī žōl.

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 oukszyb'as? 164
 Yr p̃latũmui krõsztas, yr dugnys giũm̃ui:
 Wis̃iẽn mõksłom ñiã gõłõ. Bo ýszmintis – Diẽwas.
 Tũ n'atrõkszk, suñeli, daẽit wisũ mõksłũ.
 Kũ tõw padẽs wis̃i mõksłai, jei błõgai giwĩnsi?
 Nors turẽtum zẽmes, n'amierũotus plõtus.
 Tĩek tik tũ jũ t'ałaim̃esi, kĩek iszdĩrbt' gal̃esi. 170
 Nors pažĩntum wisũ ýszmintes giũmũ,
 Tĩek tik tũ jũ t'ałaim̃esi, kĩek dorõn apw̃ẽrsi.

164 oukszyb'as] oukszyb'as LM	175 Teip err.] Teip
165 yr 2x] yr̃ LM	178 Takais] Takais LM
166 Bo err.] Bõ	179 yrã err.] yr̃ã
167 daẽit] daẽit' LM	181 d̃idelũ] d̃id'alu LM
168 błõgai] taisyta iš błõgai LM	183 zẽmjũ] z'am̃ju LM
170 jũ] jũ LM	.84 kaip err.] kaip
172 jũ] jũ LM	186 n'a err.] n'ã
172 apw̃ẽrsi] apw'ãrsi LM	

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 $\overset{\sim}{\text{~}}$ schreibt Baranowski in späterer Zeit im Inlaut $\overset{\sim}{\text{`}}$, z. B. Mõrkus (Markus), Mõrcius (Martius mensis), gõrcius (poln. garnec) kwõrta (poln. kwarta), tõrka (poln. tarka), gõnkus (poln. ganek), wõlkus (poln. wałek) czõrka¹ (poln. czarka) s̃iřkà¹ (poln. siarka), kl̃iẽmkà¹ (poln. klamka), l̃iẽmpà¹ (poln. lampa), mõĩnkà (poln. mamka), barõĩnkà (poln. obarzanek) u. v. a. (vgl. Bar. Br. 268). In den Texten hat Baranowski in Fällen wie gerõĩ, szakõĩ usw. das Kürzezeichen über dem Nasal oder der Liquida oft ausgelassen. Ich habe es der Gleichmäßigkeit halber überall außer in den zemaitischen Texten zugesetzt. Die Zeichen $\overset{\sim}{\text{~}}$ und $\overset{\sim}{\text{~}}$ braucht er ohne Unterschied. Ich habe sie in der Regel so verwandt, daß ich