Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

L2/03-219

Doc Type: **Working Group Document**
Title: **Proposal to encode one combining character in the UCS**
Source: **Lithuanian Standards Board (LST)**
Status: **Correspondent Member Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2003-06-25**

<table>
<tr><td colspan="2">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 <i>Form</i>.
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for latest <i>Principles and Procedures</i> document.
See http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest roadmaps.</td></tr>
</table>

A. Administrative

1. **Title:**   <u>**Proposal to encode one combining character in the UCS**</u>
2. Requester's name: <u>Lithuanian Standards Board (LST)</u>
3. Requester type (Member body/Liaison/Individual contribution):   <u>Correspondent Member</u>
4. Submission date: <u>2003-06-25</u>
5. Requester's reference (if applicable): _____
6. (Choose one of the following:)
This is a complete proposal: <u>Yes</u>
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: <u>Yes</u>
        Name of the existing block: <u>Combining Diacritical Marks</u>
2. Number of characters in proposal: <u>1</u>
3. Proposed category (see section II, Character Categories):   <u>Combining</u>
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000):   <u>3</u>
    Is a rationale provided for the choice?   _____
        If Yes, reference: _____
5. Is a repertoire including character names provided?   <u>Yes</u>
    a. If YES, are the names in accordance with the 'character naming guidelines
        in Annex L of ISO/IEC 10646-1: 2000?   <u>Yes</u>
    b. Are the character shapes attached in a legible form suitable for review?   <u>Yes</u>
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
    publishing the standard? <u>True Type, Vilnius University, Lithuania</u>
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
    used: <u>Mr. Mindaugas Strockis, (Mindaugas.Strockis@flf.vu.lt), FontLab</u>
    _____
    _____

7. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? <u>Yes</u>
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
        of proposed characters attached?   <u>Yes</u>
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)?
    <u>The character behaves just like the existing combining characters</u>
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
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: _____

To properly describe the accentuation system of the Lithuanian dialects, it is necessary to add the COMBINING ZIGZAG ABOVE to UCS:

⚡	U+????	**COMBINING ZIGZAG ABOVE**

Substantiation of the COMBINING ZIGZAG ABOVE in UCS

1. Meaning and place of the COMBINING ZIGZAG ABOVE in the accentuation system of the Lithuanian language

There are three main accents in the accentuation system of the Lithuanian language: acute, circumflex and grave. Beside these, four variants of the main accents are used in Lithuanian dialects. Only one of these accent variants has no corresponding sign in the UCS.

1.1. The Lithuanian language has two **main** tones for long syllables: a rising tone and a falling tone. The stressed short syllable, even though it has no tone, is marked by a corresponding grave accent; therefore there are three main accents: the acute, the circumflex and grave.

Note on Lithuanian circumflex. The circumflex accent in Lithuanian has a wavy shape, and looks identical to tilde (compare the polytonic Greek circumflex, *perispomeni*). The UCS combining tilde sign therefore can be used as the Lithuanian circumflex. The angled Latin circumflex has a different meaning in Lithuanian (see 1.2.1 below).

1.1.1. The **falling tone** is marked by an acute accent (corresponding to COMBINING ACUTE ACCENT in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable with falling tone	Acute	COMBINING ACUTE ACCENT	0301	◌́

1.1.2. The **rising tone** is marked by a circumflex accent (corresponding to COMBINING TILDE in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable with rising tone	Circumflex	COMBINING TILDE	0303	◌̃

1.1.3. The **short stressed syllable** is marked by a grave accent (corresponding to COMBINING GRAVE ACCENT in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed short syllable	Grave	COMBINING GRAVE ACCENT	0300	◌̀

1.2. Accentuation of the Lithuanian dialects. In addition to the aforementioned tones, four further variants for dialects are distinguished: broken, middle, abrupt and general accents, which are marked as follows.

1.2.1. The **broken tone** is marked by a broken accent (corresponding to COMBINING CIRCUMFLEX ACCENT in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable with broken tone	Broken accent	COMBINING CIRCUMFLEX ACCENT	0302	

1.2.2. The **middle tone** is marked by a middle accent (corresponding to COMBINING VERTICAL TILDE in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable with middle tone	Middle accent	COMBINING VERTICAL TILDE	030E	

1.2.3. The **abrupt tone** is marked by an abrupt accent with a "zigzag" form (there is no corresponding mark in UCS, but it could be named "COMBINING ZIGZAG ABOVE"):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable with abrupt tone	Abrupt accent	COMBINING ZIGZAG ABOVE	????	

1.2.4. In those cases where the **tone** is almost **indefinable** and it is difficult to define it precisely, the general stress is used, represented by a vertical short line (corresponding to COMBINING VERTICAL LINE ABOVE in UCS):

Lithuanian language		UCS		
Description	Name of accent	Name	Code	Sign
Stressed syllable, tone unspecified	General stress	COMBINING VERTICAL LINE ABOVE	030D	

1.3. Summary of accentuation system of the Lithuanian language. For marking all tones of the Lithuanian language and its dialects, seven accents marks are needed. Three of them are main accents, and the other four variants:

MAIN ACCENTS		
Sign	Code	Name
	0301	COMBINING ACUTE ACCENT
	0303	COMBINING TILDE
	0300	COMBINING GRAVE ACCENT

VARIANTS OF ACCENTS		
Sign	Code	Name
	0302	COMBINING CIRCUMFLEX ACCENT
	????	**COMBINING ZIGZAG ABOVE**
	030E	COMBINING VERTICAL TILDE
	030D	COMBINING VERTICAL LINE ABOVE

1.4. UCS lacks only one combining mark of this system: there is no COMBINING ZIGZAG ABOVE,

which takes this form: ⟨⟩. It is absolutely necessary because it marks the abrupt tone, which is characteristic of the northern-eastern Lithuanian dialects (in the surroundings of Biržai, Kupiškis, Anykščiai and other towns). In these dialects the abrupt tone replaces the falling tone in certain positions: on long vowels and especially on monophthongised diphthongs at the ends of words. A syllable bearing this tone is pronounced rather short, with an abrupt stop of the voice, as if chopping off the remainder of the syllable.

2. Usage of the ZIGZAG ABOVE in linguistic literature

The abrupt accent is represented by the mark of ZIGZAG in the principal standard works of Lithuanian dialectology, university student manuals and chrestomathies of the dialects, such as:

(1) GRINAVECKIENĖ, ELENA, MORKŪNAS, KAZYS (eds), *et al.* 1970: *Lietuvių kalbos tarmės. Chrestomatija,* Vilnius: Mintis. 518 pp. (*The Lithuanian Dialects. A Chrestomathy*)

(2) MORKŪNAS, KAZYS (ed), *et al.* 1982: *Lietuvių kalbos atlasas 2. Fonetika,* Vilnius: Mokslas. 148 pp. + 112 maps. (*Lithuanian Language Atlas. Vol. 2. Phonetics)*

(3) ZINKEVIČIUS, ZIGMAS 1966: *Lietuvių dialektologija. Lyginamoji tarmių fonetika ir morfologija*, Vilnius: Mintis. 542 pp. + 75 maps. (*Lithuanian Dialectology. Comparative Phonetics and Morphology of Lithuanian Dialects)*

(4) ZINKEVIČIUS, ZIGMAS 1978: *Lietuvių dialektologija.* Vilnius: Mokslas. 192 p. (*Lithuanian Dialectology)*

3. Samples of usage of the ZIGZAG ABOVE

From (1), p. 277:

From (2), p. 48:

ιų šnektos plote (žr. p. 44 – 45 ir žemėl. nr.

iiame žodžio gale, išskyrus atvejus, kai tas ė

alsių k, g, š, ž, č, j, tariama *å*, plg. *duktå*,

bě̃, v'ersmå̃., ž'ę.vå̃ ~ *žievẽ*, bet *apuš'ė̃*,

7, 144) var'š'kė̃ . Dėl kitų tarmių ir literatū-

įtakos jaunesnioji karta neretai šio dėsnio

pavyzdžiui, taria *kat'ė̃* 207, *g'ę.s'mė̃* ~

ƺar'bė̃, *katė̃* 272a.

e vietoj lk. *ė* yra dvejopas: kirčiuotas ir

From (3), p. 37:

šnektose, turinčiose žodžio galo dvibalsių monoftongizaciją, tiek priartėja prie trum-
pinės, kad netgi sunku atskirti nuo pastarosios. Antai Biržų šnektos dvibalsinės
kilmės monoftongai galūnėse -*ė̃(s)*<-*aĩ(s)*, -*ė̃(s)*<-*eĩ(s)* ir -*iẽ*, -*ỏ̃(s)*<-*aũ(s)* ir -*uõ*
tariami trumpai, lyg būtų -*ė̀(s)*, -*ė̀(s)*, -*ỏ̀(s)*. Maždaug taip čia taria ir kai kuriuos
senuosius monoftongus, ypač galūnėje prieš -*s*, pvz., gen. sing. *g'erklė̃s* 'gerklės',
kat'ė̃s 'katės', *galvỏ̃s* 'galvos' ir nom. sing. *g'erklå* 'gerklė', *kat'ė* 'katė'. Tarmėse, kur
nėra žodžio galo dvibalsių monoftongizacijos, kirstinė priegaidė nėra taip labai pri-
artėjusi prie trumpinės: čia balsiai yra pusilgiai (*g'erklė̂s, kat'ė̂s*...).

§ 27. Cirkumfleksas vietoj dėsningo akūto kartais atskirose šnektose gali at-
sirasti ekspresyvioje kalboje dėl sakinio intonacijos. Tai atsitinka dažniausiai ko

From (4), p. 165:

Papilys

*ƺi.vå.no. kaȓ·tu. bię̆dnas žmɔ.gę̆.l̂i.s ‖ n̂ iẽkas jɔ. žå̆.ma.j n̂eá·ugdavɔ. |
t̂ik v̂iėn̂i ɔ̃.kmanas (~ ãkmenes* 'akmenys') *‖ ɔ. p̂íevɔ.j tai v̂iėni̇ ƙælmâî|
n̂ę.ku ȓ dal̂·gæ.s pasú.kî ‖ vaȓ·ksta kaȓ·tu. žmɔ.gę̆.l̂i.s tɔ.s ƙå̆·lmus ráudamas
ir nušiƙ ę́.iƙ æ. ‖ žû·ȓa | p̂ȓlešais v̂el̂n̂ú.kas atbá·ga ‖ kɔ́·m tu man̂î. šauƙ æî |
kl̂á·uŝa ‖ ɔ. žmɔ́·gus jɔ́·m ir sɔ̃.kɔ. | nùƺi p̂ȓiá·ugɔ. mɔ́·n̂ rɔ́·p̂u. ba tvarkɔ̃̂s
(= labai daug) ‖ rá·unu ir n̂egal̂ù gɔ̃.lɔ. gá·uî ‖ rá·uƙum (= raukime) unt
pú.sa.s* (= per pusę, pusiau).

LKT, Nr. 413 – 429, 431 – 433; LiDi, p. 72 – 76, 100 – 105; Gauthiot R. Le par-