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
ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

Please fill all the sections A, B and C below.
(Please read Principles and Procedures Document for guidelines and details before filling this form.)

A. Administrative

1. Title: Proposal to encode one combining character in the UCS
2. Requester's name: Lithuanian Standards Board (LST)
3. Requester type (Member body/Liaison/Individual contribution): Correspondent Member
4. Submission date: 2003-06-25
5. Requester's reference (if applicable): _____________________________________________________________
6. (Choose one of the following):
   a. This is a complete proposal: _______Yes
   b. 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:
      Name of the existing block: Combining Diacritical Marks
      Is a rationale provided for the choice? Yes
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 character behaves 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
   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:

![COMBINING ZIGZAG ABOVE](image)

**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):

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable with falling tone</td>
<td>Acute</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable with rising tone</td>
<td>Circumflex</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed short syllable</td>
<td>Grave</td>
</tr>
</tbody>
</table>

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):

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable with broken tone</td>
<td>Broken accent</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable with middle tone</td>
<td>Middle accent</td>
</tr>
</tbody>
</table>

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”):

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable with abrupt tone</td>
<td>Abrupt accent</td>
</tr>
</tbody>
</table>

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):

<table>
<thead>
<tr>
<th>Lithuanian language</th>
<th>UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Name of accent</td>
</tr>
<tr>
<td>Stressed syllable, tone unspecified</td>
<td>General stress</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Sign</th>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>🆃️</td>
<td>0301</td>
<td>COMBINING ACUTE ACCENT</td>
</tr>
<tr>
<td>🆃️</td>
<td>0303</td>
<td>COMBINING TILDE</td>
</tr>
<tr>
<td>🆃️</td>
<td>0300</td>
<td>COMBINING GRAVE ACCENT</td>
</tr>
</tbody>
</table>

**VARIANTS OF ACCENTS**

<table>
<thead>
<tr>
<th>Sign</th>
<th>Code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>🆃️</td>
<td>0302</td>
<td>COMBINING CIRCUMFLEX ACCENT</td>
</tr>
<tr>
<td>???</td>
<td></td>
<td>COMBINING ZIGZAG ABOVE</td>
</tr>
<tr>
<td>🆃️</td>
<td>0305</td>
<td>COMBINING GRAVE ACCENT</td>
</tr>
<tr>
<td>🆃️</td>
<td>030D</td>
<td>COMBINING VERTICAL LINE ABOVE</td>
</tr>
</tbody>
</table>
1.4. UCS lacks only one combining mark of this system: there is no COMBINING ZIGZAG ABOVE, which takes this form: \[\text{\textcopyright} \]. 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:


3. Samples of usage of the ZIGZAG ABOVE

From (1), p. 277:

\[\text{Udra varžoje} \]

393. kė\[\text{nevėže}\] [Nevėžė = Nevėžis] nekas\[\text{tų bū\textasciicirc}va, kū\textdash daba\textasciicirc} Šv\textdash ndras, anes\[\text{mē\textasciicirc}tes tē\textasciicirc}e*\] į\[\text{drug pagā\textasciicirc}va: in\textdash tli\textasciicirc}da vār\textasciicirc}on, p\textdash ri\textasciicirc}ja žuvi\textasciicirc}s, pro\[\text{ingerku iš\textasciicirc}št nebegā\textasciicirc}li tē\textasciicirc}e nuvā\textasciicirc}ja vār\textasciicirc}as krait\textasciicirc}t – neiš\textasciicirc}trā\textasciicirc}kį| z\textasciicirc}ū\textasciicirc}ri – kō\textasciicirc}jas žusi\textasciicirc}s, nō\textasciicirc}se* – śuni\textasciicirc}s, vadegō – ke\textasciicirc}p lā\textasciicirc}pes| splav\textasciicirc} bū\textasciicirc}va tō\textasciicirc}ki kē\textasciicirc}p lā\textasciicirc}pes, – gē\textasciicirc} nebea\textasciicirc}menų| nuplē\textasciicirc}šė skū\textasciicirc}ru*, p\textdash ri\textasciicirc}la pelanūt; iē\textasciicirc}dži\textasciicirc}na ir pa\textasciicirc}dave]
From (2), p. 48:

ų šnektos plote (žr. p. 44 – 45 ir žemėl. nr. niame žodžio gale, išskyrus atvejus, kai tas ė alsių k, g, š, ž, č, j, tariama ḏ, plg. dukštā, ḍbē, v’ersmā, ž’e.vā. ~ žievē, bet apuš’e, 7, 144) varš’k(ē). Dėl kitų tarmių ir literatūrėtakos jaunėsnioji karta neretai šio dėsnio pavyzdžiui, taria kat’e. 207, g’e.s’m(e) ~ garb’b(e), kat’e 272a.

e vietoj lk. è yra dvejopas: kirčiuotas ir

From (3), p. 37:

šnektose, turinčiose žodžio galio dvibalsių monofoqtongizaciją, tiek priartėja prie trumpinės, kad netgi sunku atskirti nuo pastarosios. Antai Biržų šnektos dvibalsinės kilmės monofoqtongai galūnėse (s)<-ai(s), (s)<-ei(s) ir -(i), (s)<-au(s) ir -(u) tariami trumpiai, lyg būtų -(e(s), -(s), -(s). Maždaug taip čia taria ir kai kuriuos senosius monofoqtongus, ypač galūnėje prieš -s, pvz., gen. sing. gerkš, 'gerklės', kata, 'katės', galvos 'galvos' ir nom. sing. gerkš'gerklė, katė 'katė'. Tarmėse, kur nėra žodžio galio dvibalsių monofoqtongacijos, kirstinė prieiga nėra taip labai priartėjusi prie trumpinės: čia balso i bar pulgi (gerkš, kats...).

§ 27. Cirkumfleksas vietoj dėsningo akūto kartais atskirose šnektose galio atsirasti ekspresyvioje kalboje dėl sakimio intonacijos. Tai atsitinka dažniausiai ko

From (4), p. 165:

Papilys

gi.vā.na. kař’tu. biėdas žmo.ge.ši.s || niėkas jo. žė.maj ŭe-augdavo. | ũk vieniši š.kmanas (~ akmenys ‘akmenys’) || o. lieroj tai vieni kėlma|| nė.kuflu dař.gae.s pasū.kft || vař-ksta kař’tu. žmo.ge.ši.s to.s kėlmas rā’udamas ir nušikė.ičė. || žū’ra || priešais velniukas atbā’ga || kō’m tu mańi šauką | klauša || o. źm̅’gus jō’m ir sło.ko. || niugi pria’ugo. mō’ń rū’pü. ba tvark̅’s ( = labai daug) || ra’unu ir ſeikal̅’ gō.šo. gā’u’i || ra’ukum (= raukime) unt pū.sas (= per pusę, pusiau).

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