

DATE: 7 June 2013
DOC TYPE: Working Group Document
TITLE: Lithuanian Accents - Follow-Up
SOURCE: Deborah Anderson, Script Encoding Initiative, UC Berkeley
STATUS: Liaison contribution
ACTION: For consideration by WG2
DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

In the “Lithuanian Ad-hoc Report” from Feb. 2012 (<http://std.dkuug.dk/JTC1/SC2/WG2/docs/n4242.pdf>), one suggestion on how to proceed was:

Consider the potential role that that the Unicode CLDR project (<http://cldr.unicode.org/>) might have in capturing requirements for Lithuanian text processing, in driving awareness of those requirements, and in improving implementations in products that make use of CLDR data. Note: In relation to this, Lithuanian experts are encouraged to actively participate in the Unicode CLDR project.

Peter Edberg of Apple Inc. submitted the Lithuanian accented letter combinations to CLDR on 3 May 2012 and they were contained as part of the auxiliary exemplar set for Lithuanian in CLDR release 22 (released September 2012 – see screenshots on page 2).

The UNICODE LOCALE DATA MARKUP LANGUAGE (LDML) spec, which describes the format used in CLDR, characterizes the main and auxiliary exemplar sets as follows:

The *main* [exemplar] set should contain the minimal set required for users of the language [which usually matches the basic alphabet taught in school], while the *auxiliary* exemplar set is designed to encompass additional characters: those non-native or historical characters that would customarily occur in common publications, dictionaries, and so on.

(Source: http://www.unicode.org/reports/tr35/tr35-general.html#Character_Elements)

Two typical usages for the main and auxiliary exemplars are:

- determining the extent to which a font can support a given language (ranking the support level): Does it have glyphs & mapping info for the main exemplars, punctuation exemplars and number systems used by the locale? Can it also support the auxiliary exemplars?
- determining the extent to which a keyboard layout can support input of a given language or determining the set of languages that can be typed using a given key layout.

Some implementations may use them for:

- a very basic language filter
- determining which characters to allow without flagging (e.g., in identifiers)
- determining which non-Unicode code pages support a language.

The information contained in CLDR is used widely by various companies, including Apple, Google, IBM, etc. (See further: <http://cldr.unicode.org/#TOC-Who-uses-CLDR->).

Below is a screenshot of the Lithuanian accented letter combinations in the CLDR Survey Tool (http://st.unicode.org/cldr-apps/v#/lt/Alphabetic_Information/).

The green entries in the “Proposed” column are the currently approved Lithuanian values.

We are aware of stability issues. For now, brief restart at 02:00:10:00:18:00 GMT.

» CLDR 24 Survey Tool » Lithuanian = Lithuanian (Lithuania) » Core Data » Alphabetic Information

Login... | Manage | Instructions | Review Coverage: Automatic (Currently: Modern)

This locale, Lithuanian supplies the *default content* for **Lithuanian (Lithuania)**. Please make sure that all the changes that you make here are appropriate for **Lithuanian (Lithuania)**. If there are multiple acceptable choices, please try to pick the one that would work for the most sublocales.

Code	English	A	Proposed	Others
Characters In Use				
Main Letters	[a b c d e f g h i j k l m n o p q r s t u v w x y z]	✓	a a b c c d e e f f g h i j y j k l m n o p r s s t u u v z z	no value
Others: auxiliary	[á à å ä å ä æ ç é è ē ē ē ē i i i i i ñ ó ò ô õ ö ø ò œ ú û ü ü ü ý]	✓	á á á {a\u0301} {a\u0303} é è ē {e\u0301} {e\u0303} {è\u0301} {è\u0303} {i\u0307\u0301} {i\u0307\u0300} {i\u0307\u0303} {i\u0301} {i\u0303} {j\u0307\u0301} {j\u0307\u0303} {l\u0303} {m\u0303} ñ ó ò ô q {r\u0303} ú û ü {u\u0301} {u\u0303} {ū} {ū\u0303} w x	no value
Others: index	[A B C D E F G H I J K L M N O P Q R S T U V W X Y Z]	✓	A A B C C D E E F F G H I I Y J K L M N O P R S S T U U V V Z Z	
Others: punctuation	[! " # \$ % & ' () * + , - . : ; < = > ? [\] ^ _ ` { } ~ ¡ ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ ® ¯ ° ± ² ³ ´ µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿	✓	! " # \$ % & ' () * + , - . : ; < = > ? [\] ^ _ ` { } ~	! " # \$ % & ' () * + , - . : ; < = > ? [\] ^ _ ` { } ~

Close-up of the auxiliary set:

á à å ä å ä æ ç é è ē
 {e\u0301} {e\u0303} {è\u0301}
 {è\u0303} {i\u0307\u0301} í
 {i\u0307\u0300} i
 {i\u0307\u0303} i {i\u0301} {i
 \u0307\u0301} {i\u0303} {j
 \u0307\u0303} {j\u0303}
 {j\u0307\u0301} {l\u0303}
 {m\u0303} ñ ó ò ô q {r\u0303} ú
 û ü {u\u0301} {u\u0303} {ū
 \u0301} {ū\u0303} w x

To file additions or corrections, contact CLDR via: <http://unicode.org/cldr/trac/newticket>.