L2/02-148

To: UTC

Re: **General Category for Combining Grapheme Joiner**

From: **Mark Davis**

Date: **2001-04-19**

There was one more item that came up late in the development of U3.2. I propose we do it for U3.2.1 (or 4.0, whichever is sooner).

Background

1. If a combining mark occurs after an *enclosing combining mark* (Me), then it applies to the entire grapheme cluster (clarified in UAX #28, Section 3.1 Conformance).
2. The *combining grapheme joiner* (CGJ) is categorized currently as a *non-spacing mark* (Mn), as in UnicodeData.txt:

       034F;COMBINING GRAPHEME JOINER;Mn;0;NSM;;;;;N;;;;;
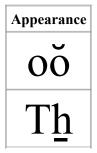
3. Currently, the CGJ doesn't have any use or effect if it is not right before a base character; it is simply invisible.

Proposal

If we make one small change to the general category for the CGJ, changing from Mn to Me, it allows the CGJ to be used for double diacritics. This change does not affect normalization, and it does not affect the other usage of the character. Thus one could encode the following:

Appearance	Encoding				
o͡o	o 006F	CGJ 034F	o 006F	CGJ 034F	˘ 0306
T̲h̲	T 0054	CGJ 034F	h 0068	CGJ 034F	̱ 0331
a̸m̸	a 0061	CGJ 034F	m 006D	CGJ 034F	̸ 0338

For comparison, in each row above if either of the CGJs were absence then the diacritic would only apply to the last character:

Appearance
oŏ
Th̲

aʋ́

Of course, this change alone does not automatically change fonts so that they support the correct rendering, but it does enable a general mechanism for expressing the form, one that font designers can support if desired for particular cases. Such forms appear in dictionary transcriptions and bibliographic use.