

Use of CGJ in Latin script diacritics

Bob Hallissy and Lorna Priest
 SIL International
 2005-04-18

Abstract

Presents a case for using U+034F COMBINING GRAPHEME JOINER within sequences of diacritics in Latin text and thus encode text that could not be represented with Unicode 4.0.

Introduction

The Unicode Standard (TUS) 4.0, section 7.7, discusses the use of double diacritics such as U+0360 COMBINING DOUBLE TILDE (see figures 7-6 and 7-7). As noted just below Figure 7-7:

“As a consequence [of canonical reordering], there is currently no mechanism for representing the occasionally seen orthographic convention of using a dot above a ligature tie—that is, U+0361 COMBINING DOUBLE INVERTED BREVE.”

In the quote above, TUS 4.0 suggests there is need to represent U+0307 COMBINING DOT ABOVE above a ligature tie. Additionally, de Peña¹ has several instances of a diacritic (in this case U+0301 COMBINING ACUTE ACCENT) used above a ligature tie:

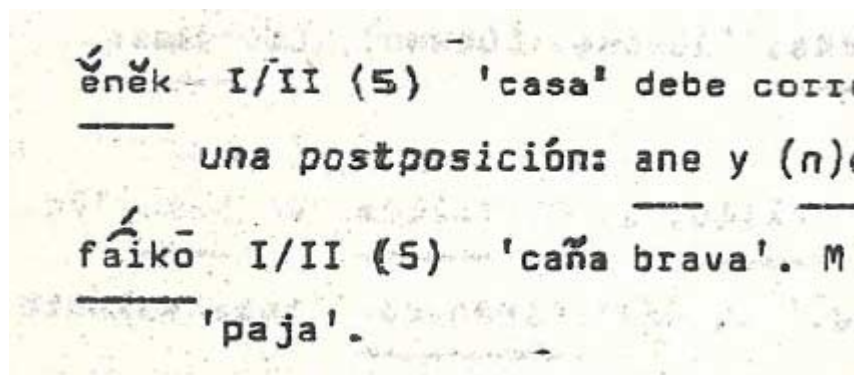
Em(e) eñcontramos en varios saludos:

a) ēme sūinām I/II (5) '¿cómo amaneció?'

b) ēmēs nūima I/II (3) '¿cómo ha amanecido?'

áēng II (4) 'yuca'. M. ti
 clinada arrengó. Cf.

áep I/II (4, 5) 'aji'. I



Unicode 4.1 change relating to CGJ

[Unicode 4.1](#) modifies the semantics of U+034F COMBINING GRAPHEME JOINER to permit its use to prevent canonical reordering of adjacent combining marks. Though the example given, and also the main motivation for this change, relates to Hebrew text, the change itself is written in general terms not restricted to Hebrew.

These new semantics can therefore be used for Latin text to permit encoding of diacritics above a double diacritic, e.g:

$$\begin{array}{ccccccc}
 \text{u} & + & \overset{\circ}{\text{}} & + & \langle \text{cgj} \rangle & + & \overset{\circ}{\text{i}} & + & \text{i} & \longrightarrow & \overset{\circ}{\text{ui}} \\
 0075 & & 0361 & & 034F & & 0301 & & 0069 & &
 \end{array}$$

¹ de Peña, Gertrud Schumacher. 1991. *El vocabulario Mochica de Walter Lehmann (1929), comparado con otras fuentes lexicais*. Lima: Instituto de Investigación de Lingüística Aplicada (CILA), Universidad Nacional Mayor de San Marcos