

DOC TYPE: Working Group Document

TITLE: Comments on N3694 "Rationale for encoding LATIN LETTER MIDDLE DOT"

SOURCE: Deborah Anderson, Script Encoding Initiative, UC Berkeley

STATUS: Liaison member contribution

ACTION: For consideration by WG2

DISTRIBUTION: ISO/IEC JTC1/SC2/WG 2

Document N3694 "Rationale for encoding LATIN LETTER MIDDLE DOT" brings up a number of points that require clarification.

1. The middle dot of Bernard Karlgren, which derives from Johan August Lundell, has been handled in a font from the Sino-Tibetan Etymological Dictionary and Thesaurus Project by U+00B7 since 1999. The font has published documentation and a published code converter for Karlgren's sign. This middle dot only appears in legacy STEDT databases now, as Karlgren's orthography has largely been replaced in STEDT materials by IPA. The comments on the font read-me text state: "Karlgrén's symbols for initial glottal stop, the raised dot, was reduced in size. It now matches Karlgrén's original symbol more closely and is also suitable for indicating half-long vowels." (Sources: <http://stedt.berkeley.edu/stedtfont/downloads/STEDTFontReadMe5.1.1.txt> and http://linguistics.berkeley.edu/~rscook/pdf/HPTB_Appendix_B.pdf) Hence there is data in which an existing character has been used for East Asian transcription and transliteration, and it can also be used for marking lengthened vowels.
2. The middle dot that is often used to designate syllabification in dictionaries is U+2027 HYPHENATION POINT, see <http://www.unicode.org/reports/tr14/#Dictionary>.
3. In an ideal world, characters would fall into neat, clearly defined categories, and users would dutifully abide by these categorizations in their orthographies, making the task of character encoding easier. However, this is not the case: orthographies can be messy, using whatever characters are available, so that the distinction between "letters" and "marks" is blurred.

The use of the middle dot to designate a long mark is not its sole use in linguistic materials, nor is the use of the middle dot for glottal stop restricted to East Asia. In Garo, a Sino-Tibetan language used by 700,000 people in northeastern India and in

Bangladesh, the orthography uses a raised dot or an apostrophe for the glottal stop (*The Sino-Tibetan Languages* by Graham Thurgood and Randy LaPolla, p. 388).

The middle dot also is found as a pharyngealized glottalized stop in Nootka, which appears as a middle dot with a raised comma above (“The Rival Whalers, A Nitinat Story [Nootka Text with Translation and Grammatical Analysis]”, by Edward Sapir, *IJAL*, 1924, p. 82):

The ' of -'att combines with preceding stop (*p, t, k, kw, q, qw, ts, tc, tt*) into glottalized stop (*p', ', k', k'w, ', ', t's, t'c, t't* respectively; origi-

4. A central concern for the use of the middle dot, which appears in the minutes from the Dublin WG2 meeting, is searching, and whether searching is possible with U+00B7 or U+02D1. As described in N3678, search results for U+02D1 were better than U+00B7 at present. As noted in that document, both triangular and circular glyphs are attested for U+02D1. The character U+02D1 remains the character of choice to this author.