RESPONSE TO PRI-37

Date :	August 8, 2004.
Author:	Gautam Sengupta
Address:	Centre for Applied Linguistics & Translation Studies
	University of Hyderabad, Hyderabad 500076
	India
Email:	gsghyd@icqmail.com

If the recommendations in PRI-37 were to be accepted, there would be four distinct uses of the ZWJ in Indic:

    A   <C1 virama zwj> to encode a half form of C1 in isolation,
    B   <C1 virama zwj C2> to encode a conjunct with the half form of C1
    C   <zwj virama C2> to encode a C2 conjoining form of C2 in isolation, and
    D   <C1 zwj virama C2> to encode a C2-conjoining conjunct

Let us observe, in passing, that all of the uses enumerated above are "illegitimate" in so far as they encode rendering variations rather than differences in abstract character sequences. TUS 4.0 is categorically specific about its mandate of encoding characters rather than glyphs or rendering variants. The distinction between a क्ष and its half-form variant in Devanagari that calls for the usage in (B), is undoubtedly one that involves a difference in glyphs rather than abstract characters. While it might be necessary to live with (B) on account of its having been around for long enough to be considered a *fait accompli*, it would be quite perverse to proliferate the undesirable practice of encoding rendering variants and give it a stamp of legitimacy in the face of explicit and categorical prohibitions against it in the Standard. The UTC should carefully avoid creating an impression that the principles and directives of the Standard are of little consequence and can be flouted at will. A principle does not cease to be one by virtue of being violated on one occasion. The use of zwj to encode half forms in Devanagari is at best an undesirable departure from a fundamental principle of TUS. Any recurrence of such practices should be strictly forbidden.

While all of the usages in A-D run contrary to the spirit of TUS, (B) is the most innocuous among them. In (B) the syntax and semantics of virama and zwj are not trashed in ways that they are in A, C and D. Admittedly the function of virama and zwj can be described as follows:

> The ZERO WIDTH JOINER denotes a non-visible letter that presents linking or cursive joining behavior on either side (that is, to the previous or following letter). … The effect of the ZWJ is to control how a dead consonant and a consonant following it will be presented.

> … the VIRAMA has the effect of killing the inherent vowel of the preceding consonant, creating a dead consonant.[1]

But there is no linking or cursive joining behavior in (A), no consonant at all following the dead consonant. The same is true of (C) where, in addition, there is also no consonant to be silenced by the virama. In (D) there is cursive joining but no consonant immediately preceding the virama. A virama has no meaning unless it is immediately preceded by a consonant. A virama preceded by a zwj is both syntactically ill-formed and semantically anomalous. As a matter of fact, a virama and a preceding consonant constitute a single character. Separation of the two is just a consequence of a particular graphological analysis.[2] The virama is not a distinct "character" of any known Indic alphabet.

It is interesting to observe that in A-D, the zwj behaves like an invisible consonant, a consonant with no content other than an inherent default vowel indicating its token consonanthood. Let us try to understand and justify this observation.

(a) Like any other <C1 virama C2> sequence, <C1 virama D> (where D is a dummy consonant) must result in a conjunct. The possibility of a conjunct ligature is ruled out since no font will have such a ligature in its repertoire. The possibility of a C2-conjoining conjunct is ruled out since D has no shape, C2-conjoining or otherwise. The only remaining possibility is a C1-conjoining conjunct, which, in this case is just a half-form of C1, since D must be realized as null.

(b) We have seen that <C1 virama D> results in a half-form of C1. This followed by a C2 can only result in a C1-conjoining conjunct of C1 and C2.

(c) Like any other <C1 virama C2> sequence, <D virama C2> results in a conjunct. For reasons that are obvious from the preceding paragraphs, this cannot be a conjunct ligature or C1-conjoining conjunct. It can only be a C2-conjoining conjunct, which in this case takes the shape of an invisible character followed by the C2-conjoining form of C2. In other words, the sequence is rendered simply as a C2-conjoining form of C2.

(d) We have already seen that <D virama C2> is rendered as a C2-conjoining form of C2. This preceded by a C1 can only result in a C2 conjoining conjunct of C1 and C2.

What we need then is an "invisible consonant"[3] and NOT a control character like zwj/zwnj that has no consonantal properties. The use of zwj in the Brahmi-derived scripts has very little justification. While it might be necessary to live with it in contexts such a (B) for the sake of backward compatibility, there is no justification whatsoever for extending the range and scope of its use any further.

---

[1] PRI-31, page 4.
[2] An alternative analysis is presented in Dasgupta & Sengupta (2003)
[3] For a similar idea see WG2 document number N2822 submitted by Michael Everson, Peter Constable, Rick McGowan, &Ken Whistler.