ZWJ in emoji sequences as hint for single glyph

To: *UTC*

Date: *2015 January 29*

From: *Peter Edberg & Emoji Ad-hoc Committee*

(This proposal was discussed in the emoji ad-hoc committee but that committee does not have consensus regarding it, and would like feedback from the UTC)

As discussed elsewhere, developing emoji usage conventions include the tendency to send and read certain kinds of emoji sequences as a unit; the various elements of the sequence may be intended to represent a group, or some elements may function as "adjectives" modifying previous elements. Draft UTR #51 explicitly discusses this mechanism as a way to indicate various kinds of family groups (see the end of section 2.2).

This proposal discusses the use of U+200D ZERO WIDTH JOINER (ZWJ) to more formally indicate a sequence intended to be read as a single unit, so that rendering systems that may have a single ligature glyph for such a sequence have a clear hint about when that glyph should be used.

For example, suppose on some system a font includes the following "Halloween in North America" glyph (all images in this document courtesy of and © iDiversicons):

There is no Unicode character specifically for this. However, there are characters for GHOST and JACK-O-LANTERN. A sequence of these using ZWJ indicates a preference for a "more joined" presentation such as a ligature. A keyboard or input palette on such a system could provide for entry of "Halloween in North America" as a single unit, showing this glyph and generating the sequence GHOST + ZWJ + JACK-O-LANTERN on entry, and using the ZWJ as a display hint to present the sequence as a single glyph. If the sequence is sent to a system that does not support a single ligature glyph for the sequence, it degrades gracefully to appearing as a sequence of separate glyphs. The possibilities get more interesting for sequences of "portrait emoji" or body part emoji; more on this below.
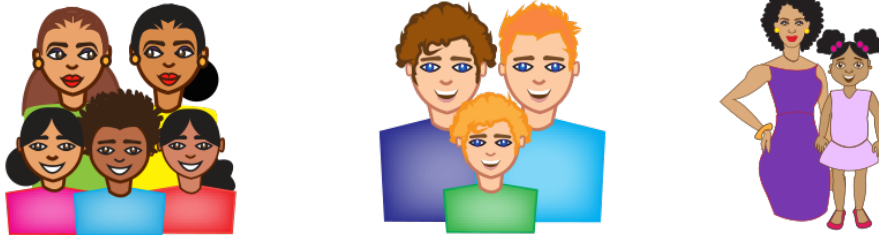
This behavior is, of course, completely conformant with the existing definition of ZWJ (see extracts at end of document) and does not require Unicode to define any new properties or character behavior. The three considerations for this proposal are:

- ■ Should UTR #51 explicitly call attention to this mechanism as a way to hint that a sequence should be displayed using a "more joined" glyph when possible? This could be added to the discussion about family groupings, for example.

- If so, UTR #51 should recommend an ordering for sequences that may include both ZWJ and EMOJI MODIFIER FITZPATRICK TYPE-n characters. (Draft 5 of 2015-01-24 already does so)
- Should UTR #51 provide a list of recommended ZWJ sequences for implementations to support?

The following examples suggest some of the possibilities and issues.

Custom people groupings. Support of single glyphs for families consisting of same-sex or single parents and/or with different numbers of children than what is shown for FAMILY (the following images indicate some of the possibilities):

These could be suggested using sequences with ZWJ:
- WOMAN + ZWJ + WOMAN + ZWJ + GIRL + ZWJ + BOY + ZWJ + GIRL
- MAN + ZWJ + MAN + ZWJ + BOY
- WOMAN + ZWJ + GIRL

People groupings with specific skin tones. Here UTR #51 should recommend that ZWJ not be used between an emoji character and any EMOJI MODIFIER FITZPATRICK TYPE-n intended to affect it; instead the ZWJ should be used between one emoji + modifier combination and the next emoji in the sequence (Draft 5 of 2015-01-24 already says this). For example, the use of the single image below could be requested with the following sequence using ZWJ:
- WOMAN + TYPE-1-2 + ZWJ + BOY + TYPE-3 + ZWJ + GIRL + TYPE-1-2 + ZWJ + MAN + TYPE-4

From *The Unicode Standard, Version 7.0*: "The zero width joiner and non-joiner request a rendering system to have more or less of a connection between characters than they would otherwise have. Such a connection may be a simple cursive link, or it may include control of ligatures. The zero width joiner and non-joiner characters are designed for use in plain text...they are essentially requests for the rendering system to take into account when laying out the text; while a rendering system should consider them, it is perfectly acceptable for the system to disregard these requests...U+200D ZERO WIDTH JOINER is intended to produce a more connected rendering of adjacent characters than would otherwise be the case, if possible. In particular: If the two characters could form a ligature but do not normally, ZWJ requests that the ligature be used…"