

The Trouble with “Top of Head” Emoji

Author: Charlotte Buff

Mail: irgendeinbenutzername@gmail.com

Submitted: 2017-10-12

1. Background

In June I submitted a proposal to the UTC ([L2/17-193](#), ‘*Alternative Encoding Model for Emoji Hair Variations*’) in reaction to the Committee’s plans to encode four characters representing people with different hair styles. While I am glad that the Committee dropped these characters in favour of a more extensible and flexible hair mechanism, the current draft does not reflect what I had proposed and in fact would pose several problems if it were implemented as presently planned. The four characters representing human scalps are inadequate for their intended role as hair modifiers.

2. Fallback Behaviour

The main issue with the proposed hair characters is not unique to this specific set, but also shared by many other ZWJ sequences:

Their fallback display is incomprehensible to unaware users.

Many of the original ZWJ sequences that originated on iOS work so well because their unligated appearance is more or less identical in meaning to their intended joined display. The glyph for COUPLE WITH HEART for instance simply shows two people side by side with a heart in the middle, so the sequence ‘<person>, HEAVY BLACK HEART, <person>’ without a combined glyph is pretty much equivalent for all intents and purposes. The special rendering is just a nice bonus rather than being vitally important to the understanding of the message. This is how ligation should work.

Increasingly, however, the UTC has opted to encode emoji that would work perfectly fine as atomic characters in the form of sequences of characters that more often than not bear little to no resemblance to the desired designs in unconnected form, resulting in possibly the worst offender to this date: Rainbow Flag. Nobody who is not already aware of the intricate technical details of Unicode Emoji would ever conclude that the sequence ‘WAVING WHITE FLAG, RAINBOW’ is meant to be a single waving flag with a rainbow pattern.

There are good reasons for using sequences of characters as emoji and I am not advocating against the idea in general, but it must be done with caution. How a ZWJ sequence looks to users without proper support for them must always be considered carefully. In my proposal to encode generic hair modifiers I deliberately recommended the characters be abstract, simplified symbols that merely suggest a hair style because of two principles that I consider crucial to the mechanism’s success. If an application does not support the ligation, but does support the modifier characters on their own, then:

1. Users must be able to deduce that the characters are somehow related to hair, and:
2. Users must not come to the wrong conclusion that the unconnected display of the sequence ‘<base emoji>, <modifier>’ has a different meaning in its own right.

The top of head candidate emoji easily fulfil the first criterion, but I see problems with the second one. Let’s compare the current hair modifier draft to two other kinds of existing emoji modifiers that *do* adhere to both principles reasonably well: Skin tone and gender modifiers.

Consider the following three hypothetical messages, each containing the character POLICE OFFICER with a modifier applied. Assume for the sake of this comparison that the hypothetical device these messages are viewed on has font support for all individual characters, but not for any of the special sequences they can form.

Skin Tone:

Got stopped by   today 

Gender:

Got stopped by   today 

Hair Style:

Got stopped by   today 

The Fitzpatrick modifier in the first example is just a plain, skin-coloured square. Such a symbol has no pictographic meaning on its own. It is relatively clear that the modifier character implies some attribute of the police officer, and even if a user cannot determine this relationship between the symbols they will likely just be confused and assume that the strange extra character is the result of some kind of glitch or a bizarre typo rather than trying to interpret it as its own deliberately placed semantic element. The first message cannot be mistaken to mean ‘*Got stopped by a police officer’s skin today.*’ or anything silly like that because the Fitzpatrick modifier is too abstract.

The gender modifier doesn’t work quite as well because it could be taken to mean ‘*the manliness police*’, whatever that might be, but the far more likely scenario is that the recipient will correctly conclude that the author of the message got stopped by a male police officer.

Meanwhile, because it is too realistic and graphic, the ginger scalp leaves open a lot of room for metaphorical interpretation. *You got stopped by police hair today? You got stopped by the hair police today? Is that what you meant, your hair looks bad today?*

Imagine if the five Fitzpatrick modifiers were not colour swatches, but realistic depictions of human skin with visible pores and little hairs growing out. Imagine if the characters MAN and WOMAN replaced MALE SIGN and FEMALE SIGN in gendered ZWJ sequences. These situations would be analogous to using human heads with different hair styles as hair modifiers.

The current list of draft candidates categorizes the four head tops as simple body parts, not much different from the other proposed emoji LEG and FOOT. This baffles me, since it implies that – despite their emoji component property – the UTC somehow intends these characters to be used as stand-alone symbols in addition to their modification functions. I cannot conceive of any circumstance in which anyone (without malicious intent) would want to write a tweet or message containing a pictograph of a person’s head cut off below the hair line, just like nobody would ever want to tweet a piece of human skin in isolation. These hair emoji have no meaning on their own. There is no point in making them proper body parts. There are better emoji to represent the concept of hair by itself, like U+1F487 HAIRCUT, or just any emoji with a hair modifier applied.

The potential confusion is amplified if an application supports skin tone but not hair modifiers, which will be very likely considering the inherent open-endedness of hair modifier sequences. The exact details of the mechanism are not yet available to the public, but the annotations in the emoji candidate list imply that the head tops themselves will not take on Fitzpatrick modifiers. This leads to the following being a valid message that users with insufficient font coverage could receive:

Got stopped by   today 

While at least some savvy users without Unicode experience will be able to deduce that a police officer followed by a separate ginger head is meant to be a police officer *with* red hair, I consider it highly unlikely that a police officer followed by a human head *with entirely different skin colour* will be read as anything but two unrelated emoji. Applying Fitzpatrick modifiers to both the base emoji and the hair characters on the other hand only wastes another four bytes without adding anything to the meaning of the sequence while also introducing more opportunity for implementations to get the mechanism wrong.

I also want to bring to the UTC’s attention the fact that some users could mistake TOP OF HEAD WITH NO HAIR for certain other spherical body parts, especially in less detailed fonts, which probably wasn’t the intention.

Hair modifiers are not like skin tone and gender modifiers. Every human-form emoji that doesn’t depict more than one person is an emoji modifier base by default and (almost) all emoji showing a human face or head can form ZWJ sequences with ♀ and ♂, so encountering a sensible, but non-RGI sequence in the wild is very unlikely. With hair, however, there is much more room for variation between vendors. Platforms with more advanced font technology like Microsoft Windows might implement hair modification sequences for *all* suitable emoji in their fonts rather than just for the as-of-yet unknown subset recommended for general interchange, which I suspect will only contain a small selection of characters initially. Or perhaps some vendor will decide not to support any hair sequences at all because even a few would inflate the number of glyphs that need to be designed too much, or because they don’t see the necessity in this kind of glyphic variation.

There is no expectation that any given vendor will support each and every emoji in the canonical Unicode list and the whole point of ZWJ sequences is that vendors can create their own emoji without Unicode’s prior consent, so seeing unligated hair sequences will be almost inevitable. UTS 51 does recommend special rendering for unrecognised ZWJ emoji which would dampen the impact of seeing the scalps used in messages, but it is unrealistic to assume that most or even just some vendors would eventually implement such a mechanism. The most common and in fact the worst-case scenario will be head tops in apparent isolation, so the fallback behaviour must be carefully designed around this.

3. Cultural Problems

The current draft calls the hair modifiers ‘top of head’ and their classification as body parts means that vendors will essentially be implementing them as actual human scalps to varying degrees of realism. This is an issue. The practice of scalping has been recorded from all over the world as part of trophy-taking during times of war. The victorious party would often violently cut and yank off their victims’ upper head skin including the hair and display these scalps as proof of their triumph. This was performed both on living and dead subjects.

Not everyone will automatically associate the head tops with scalping, but for Native American users these emoji could become genuinely insulting. European settlers in North America frequently offered high bounties and rewards for Native scalps or mutilated Natives defeated in battle by taking their scalps, even up to the second half of the 19th century. Nobody wants to be reminded of that when reading their friends’ innocent text messages.

Even ignoring the possibility of unsupported ligatures, offering scalps as colourful emoji – whether included on a keyboard or not – will mean that bigots will be able to directly use these characters to mock and offend Native Americans, either by invoking the associations mentioned previously and threatening them with the horrible things that were done to their ancestors, or by making fun of them as some kind of wild, feral beasts who cut off each other’s body parts as personal trophies, which unfortunately is a common stereotype that partly still persists today in popular culture.

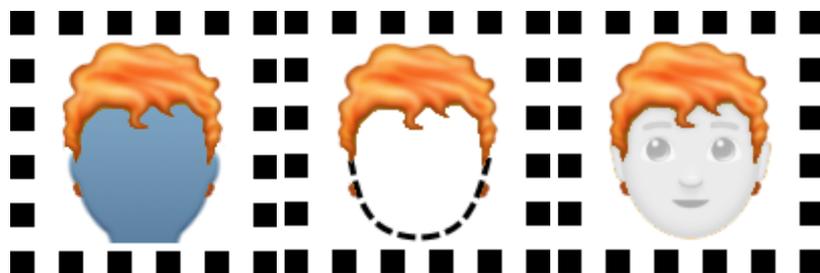
Emoji that depict the top halves of human heads are of no use to the regular user but could enable some people to do real harm.

4. Proposed Solution

I propose the four characters I proposed last time to replace U+1F9B0–U+1F9B3. I do not propose the mechanism I devised again because it was already rejected by the UTC. These modifier characters should instead be used in simple ZWJ sequences as is already planned. A General Category value of *Modifier Symbol (Sk)* is recommended to emphasize their status.

The characters should bear the name EMOJI HAIR MODIFIER <BLANK> or some variation thereof to make it extra clear that these are not body parts to be used as stand-alone pictographs. Their sole and only function is the modification of other emoji. With more abbreviated formal names like RED HAIR, WHITE HAIR etc. vendors may be inclined to draw their glyphs as just the hair pieces. Shorthands like these should be used as CLDR names instead.

As glyphs I propose an abstract human head surrounded by a dashed or dotted box. Another possible design is a square with diagonal hatching for the head icon to rest on. The head may be a monochrome silhouette (similar to U+1F464 BUST IN SILHOUETTE), an outline drawing – perhaps also dashed or dotted –, or a faded greyscale version of U+1F9D1 ADULT. Only the hair shall be clearly visible and thus stand out from the rest of the head; in case of a bald only the very top part of the head should be coloured, possibly in the neutral default skin tone or another bright highlighting colour. Ultimately it is up to each vendor to find the style that best compliments the general look of their font.



These designs could not make it any more obvious even to unknowing users that these characters are not normal emoji and are not supposed to mean anything on their own. Other approaches were considered but none proved a worthy alternative. For example, the hair modifiers cannot be squares filled with a hair texture akin to Fitzpatrick modifiers because then the bald modifier would just be

skin with no connection to hair, and less detailed fonts would make these squares hard to identify as hair as well. Theoretically there is also the option of using generic colour modifiers like those proposed in [L2/17-355](#) (Kendal, ‘*Proposal for New Emoji*’). This would not work for the selection of hair styles picked by the UTC, however, because ‘curly’ and ‘bald’ are obviously not colours. Also, if modifiers like these are ever encoded users might want to utilize them to modify other aspects like clothing or eye colour instead of hair, so there would be great potential for ambiguity.



```
1F9B0;EMOJI HAIR MODIFIER RED COLOUR;Sk;0;ON;;;;;N;;;;;
1F9B1;EMOJI HAIR MODIFIER WHITE COLOUR;Sk;0;ON;;;;;N;;;;;
1F9B2;EMOJI HAIR MODIFIER CURLY;Sk;0;ON;;;;;N;;;;;
1F9B3;EMOJI HAIR MODIFIER BALD;Sk;0;ON;;;;;N;;;;;
```

Abstract hair modifiers cannot be confused for anything else because they do not resemble anything else. They cannot be mistaken for regular body part emoji because of the prominent boxes drawn around them. They unambiguously represent hair because the entire head is visible in their glyph rather than just the top half. They do not cause problems in sequence with Fitzpatrick-modified emoji because they themselves do not show any skin, or just a very minuscule amount of skin depending on the implementation. Because they do not work to convey any meaning in isolation, users are more likely to conclude that they are somehow connected to the emoji immediately preceding them.

Users will probably submit requests for additional hair styles in the future. It would be wise to keep all hair modifiers in a contiguous range of codepoints, so some of the ten empty spots U+1F9A6–U+1F9AF should be informally reserved for such additions.