

Reactivate reduced form of UTS 52 mechanism in UTR 51

Date: 2016-11-09

To: UTC

From: Mark Davis, Peter Edberg

Draft: [link](#)

This is a proposal to reactivate a portion of the tag mechanism originally proposed in [draft UTS 52](#), but in modified form and with reduced scope. The proposal here is to integrate the mechanism into UTR51 rather than have a separate UTS52, so that there is a single document that specifies all of the emoji sequences.

The initial support is only for subregion flags, although the mechanism is general enough to expand later. The reduction in scope is motivated by the fact that a different mechanism (with better fallback) is already in use for gender variants, while the direction and hair color tags in the original proposal were deemed of lower priority than the other proposed tags. Moreover, with the ZWJ sequence mechanism now widely supported, cases where that works (like hair color) should use ZWJ sequences instead.

Note: there is a proposal to have a “reduced” format for just flags that should be discussed at the same time as this document [L2/16-234](#). That reduced format has the disadvantage that it is more complicated for segmentation.

Proposed Changes to: http://unicode.org/reports/tr51/#Emoji_Sequences

Renumber old [ED-15a](#). emoji zwj element to [ED-15b](#).

Add new definition

ED-15a. emoji tag sequence (ETS)

A sequence consisting of an emoji core sequence followed by one or more non-terminating TAG characters, followed by tag_term.

```
emoji_tag_sequence := tag_base tag_value+ tag_term
tag_base           := emoji_character | emoji_modifier_sequence
tag_term           := \x{E007F} // CANCEL TAG
tag_value          := [\x{E0020}-\x{E0040}\x{E005B}-\x{E007E}]+
                  // TAGs - tag_A..tag_Z - tag_term
```

OLD

```
tag_key_value_pair := tag_key tag_value
tag_key             := [\x{E0041}-\x{E005A}]+ // tag_A..tag_Z
```

The above defines a well-formed emoji tag sequence. Each tag_value defines a particular visual variant to be applied to the tag_base character(s). The values [\x{E0041}-\x{E005A}] are reserved for future extensions.

The meaning and validity criteria for emoji_tag_sequences and expected visual variants for tag_values are determined by the documentation in *Annex C. Emoji Tag Sequences*. Whenever an emoji_tag_sequence is not well-formed according to ED15a or not valid according to Annex C, it should be shown as the base character with an overlay, as in [Flag_Presentation](#).

Add emoji_tag_sequence to: [ED-17](#). emoji sequence

Add Review Note below it

Review Note: If emoji tag sequences are to be allowed within ZWJ sequences, then emoji_tag_sequence would be added not to ED-17, but to [ED-15b](#). emoji zwj element. Feedback is welcome on this issue.

Add new Annex C.

Annex C. Emoji Tag Sequences

While the syntax of emoji tag sequences is defined in ED15a, only valid emoji tag sequences are to be displayed. The only valid sequences in this version of Unicode Emoji are defined by sections in this annex. Whenever an emoji tag sequence is not well-formed according to ED15a or not valid according to Annex C, it should be shown as the base character with an overlay, as in [Flag_Presentation](#).

Only the combinations of <tag_base> characters and <tag_value> sequences in this section are valid.

In examples, underlined ASCII characters represent the corresponding tag characters, while \blacklozenge represents the tag_term.

C.1 Flag emoji tag sequences

A valid flag emoji tag sequence must satisfy the following constraints:

- a. The tag_base and tag_value are limited to the following:

tag_base	U+1F3F3 WAVING WHITE FLAG
tag_value	(Tag_0..Tag_9, Tag_a..Tag_z)+




- b. Let SD be the result of mapping each character in the TAG value to [0-9a-z] by subtracting 0xE0000.
 - i. SD must then be a specification as per [\[CLDR\]](#) of either a Unicode [subdivision_id](#) or a 3-digit [unicode_region_subtag](#), and
 - ii. SD must have CLDR idStatus equal to "regular" or "deprecated".

Notes:

1. The deprecated values are only included for compatibility, and should not be used.
2. Note that there is no hyphen in the tag_value, unlike ISO subdivisions like “GB-SCT”.
3. Like the emoji flag sequences, these are used to request an image for whatever is currently the flag of the specified subregion. They are *not* intended to provide a mechanism for versioned representations of any particular flag image.
4. The choice of which emoji extended flag sequences to support is entirely up to platforms and programs. There is no requirement that any be supported.

Review Note: To promote interoperability, we expect to catalog those subdivision sequences that are commonly supported on platforms, as is done with ZWJ sequences.

Examples:

Name	Image	Encoding
Flag for England		🇬🇧 <u>eng</u> ♦
Flag for Scotland		🇬🇧 <u>sct</u> ♦
Flag for Wales		🇬🇧 <u>l</u> s ♦