(working draft) Proposed Update Unicode® Technical Report #51

UNICODE EMOJI

Version	2.0
Editors	Mark Davis (Google Inc.), Peter Edberg (Apple Inc.)
Date	2015-10-29
This Version	http://www.unicode.org/reports/tr51/tr51-4.html
Previous Version	http://www.unicode.org/reports/tr51/tr51-3.html
Latest Version	http://www.unicode.org/reports/tr51/
Latest Proposed Update	http://www.unicode.org/reports/tr51/proposed.html
Revision	4

Summary

This document aims to improve the interoperability of emoji characters across implementations by providing guidelines and data.

- *design guidelines for improving interoperability across platforms and implementations*
- *data for*
  - *which characters normally can be considered to be emoji*
  - *which of those should be displayed by default with a text-style versus an emoji-style*
  - *displaying emoji with a variety of skin tones*

Status

*This document is a **proposed update of a previously approved Unicode Technical Report**. This document may be updated, replaced, or superseded by other documents at any time. Publication does not imply endorsement by the Unicode Consortium. This is not a stable document; it is inappropriate to cite this document as other than a work in progress.*

A Unicode Technical Report (UTR) *contains informative material. Conformance to the Unicode Standard does not imply conformance to any UTR. Other specifications, however, are free to make normative references to a UTR.*

Please submit corrigenda and other comments with the online reporting form [Feedback]. Related information that is useful in understanding this document is found in the References. For the latest version of the Unicode Standard, see [Unicode]. For a list of current Unicode Technical Reports, see [Reports]. For more information about versions of the Unicode Standard, see [Versions].

Contents

---

1 Introduction

Emoji are pictographs (pictorial symbols) that are typically presented in a colorful cartoon form and used inline in text. They represent things such as faces, weather, vehicles and buildings, food and drink, animals and plants, or icons that represent emotions, feelings, or activities.

Emoji on smartphones and in chat and email applications have become extremely popular worldwide. As of 2015, for example, Instagram reported that “in March of this year, nearly half of text [on Instagram] contained emoji.” Individual emoji also vary greatly in popularity (and even by country), as described in the SwiftKey Emoji Report. See emoji press page for details about these reports and others.

Emoji are most often used in social media—in quick, short messages where they connect with the reader and add flavor, color, and emotion. Emoji do not have the grammar or vocabulary to substitute for written language. In social media, emoji make up for the lack of gestures, facial expressions, and intonation that are found in speech. They also add useful ambiguity to messages, allowing the writer to convey many different possible concepts at the same time. Many people are also attracted by the challenge of composing messages in emoji, and puzzling out emoji messages.

The word *emoji* comes from the Japanese:

絵 (e ≅ picture) 文 (mo ≅ writing) 字 (ji ≅ character).

Emoji may be represented internally as graphics or they may be represented by normal glyphs encoded in fonts like other characters. These latter are called *emoji characters* for clarity. Some Unicode characters are normally displayed as emoji; some are normally displayed as ordinary text, and some can be displayed both ways.

There’s been considerable media attention to emoji since they appeared in the Unicode Standard, with increased attention starting in late 2013. For example, there were some 6,000 articles on the emoji appearing in Unicode 7.0, according to Google News. See the emoji press page for many samples of such articles, and also the Keynote from the 38th Internationalization & Unicode Conference.

Emoji became available in 1999 on Japanese mobile phones. There was an early proposal in 2000 to encode DoCoMo emoji in Unicode. At that time, it was unclear whether these characters would come into widespread use—and there wasn't support from the Japanese mobile phone carriers to add them to Unicode—so no action was taken.

The emoji turned out to be quite popular in Japan, but each mobile phone carrier developed different (but partially overlapping) sets, and each mobile phone vendor used their own text encoding extensions, which were incompatible with one another. The vendors developed cross-mapping tables to allow limited interchange of emoji characters with phones from other vendors, including email. Characters from other platforms that could not be displayed were represented with 〓 (U+3013 GETA MARK), but it was all too easy for the characters to get corrupted or dropped.

When non-Japanese email and mobile phone vendors started to support email exchange with the Japanese carriers, they ran into those problems. Moreover, there was no way to represent these characters in Unicode, which was the basis for text in all modern programs. In 2006, Google started work on converting Japanese emoji to Unicode private-use codes, leading to the development of internal mapping tables for supporting the carrier emoji via Unicode characters in 2007.

There are, however, many problems with a private-use approach, and thus a proposal was made to the Unicode Consortium to expand the scope of symbols to encompass emoji. This proposal was approved in May 2007, leading to the formation of a symbols subcommittee, and in August 2007 the technical committee agreed to support the encoding of emoji in Unicode based on a set of principles developed by the subcommittee. The following are a few of the documents tracking the progression of Unicode emoji characters.

Emoji Proposals

Date	Doc No.	Title	Authors
2000-04-26	L2/00-152	NTT DoCoMo Pictographs	Graham Asher (Symbian)
2006-11-01	L2/06-369	Symbols (scope extension)	Mark Davis (Google)
2007-08-03	L2/07-257	Working Draft Proposal for Encoding Emoji Symbols	Kat Momoi, Mark Davis, Markus Scherer (Google)
2007-08-09	L2/07-274R	Symbols draft resolution	Mark Davis (Google)
2007-09-18	L2/07-391	Japanese TV Symbols (ARIB)	Michel Suignard (Microsoft)
2009-01-30	L2/09-026	Emoji Symbols Proposed for New Encoding	Markus Scherer, Mark Davis, Kat Momoi, Darick

2009-03-05	L2/09-025R2	Proposal for Encoding Emoji Symbols	Tong (Google); Yasuo Kida, Peter Edberg (Apple)
2010-04-27	L2/10-132	Emoji Symbols: Background Data	
2011-02-15	L2/11-052R	Wingdings and Webdings Symbols	Michel Suignard

To find the documents in this table, see UTC Documents.

In 2009, the first Unicode characters explicitly intended as emoji were added to Unicode 5.2 for interoperability with the ARIB (Association of Radio Industries and Businesses) set. A set of 722 characters was defined as the union of emoji characters used by Japanese mobile phone carriers: 114 of these characters were already in Unicode 5.2. In 2010, the remaining 608 emoji characters were added to Unicode 6.0, along with some other emoji characters. In 2012, a few more emoji were added to Unicode 6.1, and in 2014 a larger number were added to Unicode 7.0.

Here is a summary of when some of the major sources of pictographs used as emoji were encoded in Unicode. These sources include other characters in addition to emoji.

Major Sources

Source	Abbr	L	Dev. Starts	Released	Unicode Version	Sample Character			
						B&W	Color	Code	Name
Zapf Dingbats	ZDings	z	1989	1991-10	1.0	✏		U+270F	pencil
ARIB	ARIB	a	2007	2008-10-01	5.2	☔		U+2614	umbrella with rain drops
Japanese carriers	JCarrier	j	2007	2010-10-11	6.0	😎		U+1F60E	smiling face with sunglasses
Wingdings & Webdings	WDings	w	2010	2014-06-16	7.0			U+1F336	hot pepper

Unicode characters can correspond to multiple sources. The L column contains single-letter abbreviations for use in charts [emoji-charts] and data files [emoji-data]. Characters that do not correspond to any of these sources can be marked with Other (x).

For a detailed view of when various source sets of emoji were added to Unicode, see *emoji-versions-sources* [emoji-charts]. The data file [JSources] shows the

correspondence to the original Japanese carrier symbols.

The Selected Products table lists when Unicode emoji characters were incorporated into selected products. (The Private Use characters (PUA) were a temporary solution.)

Selected Products

Date	Product	Version	Encoding	Display	Input	Notes, Links
2008-01	GMail mobile		PUA	color	palette	モバイル Gmail が携帯絵文字に対応しました
2008-10	GMail web		PUA	color	palette	Gmail で絵文字が使えるようになりました
2008-11	iPhone	iPhone OS 2.2	PUA	color	palette	Softbank users, others via 3rd party apps. CNET Japan article on Nov. 21, 2008.
2011-07	Mac	OSX 10.7	Unicode 6.0	color	Character Viewer	
2011-11	iPhone, iPad	iOS 5	Unicode 6.0	color	+emoji keyboard	
2012-06	Android	Jelly Bean		B&W	3rd party input	…Quick List of Jelly Bean Emoji…
2012-09	iPhone, iPad	iOS 6	+ variation selectors			
2012-08	Windows	8	Unicode only; no emoji variation	desktop/tablet: b&w; phone: color	integrated in touch keyboards	

			sequences			
2013-08	Windows	8.1	Unicode only; emoji variation sequences	all: color	touch keyboards; phone: text prediction features (e.g. “love” -> ❤)	Color using scalable glyphs (OpenType extension)
2013-11	Android	Kitkat		color	native keyboard	…new, colorful Emoji in Android KitKat

People often ask how many emoji are in the Unicode Standard. This question does not have a simple answer, because there is no clear line separating which pictographic characters should be displayed with a typical emoji style. For a complete picture, see Which Characters are Emoji.

The colored images used in this document and associated charts [emoji-charts] are for illustration only. They do not appear in the Unicode Standard, which has only black and white images. They are either made available by the respective vendors for use in this document, or are believed to be available for non-commercial reuse. Inquiries for permission to use vendor images should be directed to those vendors, not to the Unicode Consortium. For more information, see *Rights to Emoji Images*.

1.1 Emoticons and Emoji

The term *emoticon* refers to a series of text characters (typically punctuation or symbols) that is meant to represent a facial expression or gesture (sometimes when viewed sideways), such as the following.

;-)

Emoticons predate Unicode and emoji, but were later adapted to include Unicode characters. The following examples use not only ASCII characters, but also U+203F (‿), U+FE35 (︵), U+25C9 (◉), and U+0CA0 (ಠ).

^‿^

◉︵◉

ಠ_ಠ

Often implementations allow emoticons to be used to input emoji. For example, the

emoticon ;-) can be mapped to 😉 in a chat window. The term *emoticon* is sometimes used in a broader sense, to also include the emoji for facial expressions and gestures. That broad sense is used in the Unicode block name *Emoticons*, covering the code points from U+1F600 to U+1F64F.

1.2 Encoding Considerations

Unicode is the foundation for text in all modern software: it's how all mobile phones, desktops, and other computers represent the text of every language. People are using Unicode every time they type a key on their phone or desktop computer, and every time they look at a web page or text in an application. It is very important that the standard be stable, and that every character that goes into it be scrutinized carefully. This requires a formal process with a long development cycle. For example, the 🕶 *dark sunglasses* character was first proposed years before it was released in Unicode 7.0.

Characters considered for encoding must normally be in widespread use as elements of text. The emoji and various symbols were added to Unicode because of their use as characters for text-messaging in a number of Japanese manufacturers' corporate standards, and other places, or in long-standing use in widely distributed fonts such as Wingdings and Webdings. In many cases, the characters were added for complete round-tripping to and from a source set, *not* because they were inherently of more importance than other characters. For example, the 🕿 *clamshell phone* character was included because it was in Wingdings and Webdings, not because it is more important than, say, a "skunk" character.

In some cases, a character was added to complete a set: for example, a 🏉 *rugby football* character was added to Unicode 6.0 to complement the 🏈 *american football* character (the ⚽ *soccer ball* had been added back in Unicode 5.2). Similarly, a mechanism was added that could be used to represent all country flags (those corresponding to a two-letter unicode_region_subtag), such as the 🇨🇦 *flag for Canada*, even though the Japanese carrier set only had 10 country flags.

This document describes a new set of selection factors used to weigh the encoding of prospective candidates, in Annex C: Selection Factors.

That annex also points to instructions on submitting character encoding proposals. People wanting to submit emoji for consideration for encoding should see that annex. It may be helpful to review the Unicode Mail List as well.

For a list of frequently asked questions on emoji, see the Unicode Emoji FAQ.

1.3 Goals

This document provides:

- design guidelines for improving interoperability across platforms and implementations
- background information about emoji characters, and long-term alternatives
- data for
    - which characters normally can be considered to be emoji

    - which of those should be displayed by default with a text-style versus an emoji-style
    - displaying emoji with a variety of skin tones
- pointers to [CLDR] data for
    - sorting emoji characters more naturally
    - annotations for searching and grouping emoji characters

It also provides background information about emoji, and discusses longer-term approaches to emoji.

As new Unicode characters are added or the “common practice” for emoji usage changes, the data and recommendations supplied by this document may change in accordance. Thus the recommendations and data will change across versions of this document.

Additions beyond Unicode 7.0 are being addressed by the Unicode Technical Committee: as any new characters are approved, this document will be updated as appropriate.

1.4 Definitions

The following provide more formal definitions of some of the terms used in this document. Readers who are more interested in other features of the document may choose to continue from *Section 2 Design Guidelines* .

> ***ED-1. emoji*** — A colorful pictograph that can be used inline in text. Internally the representation is either (a) an image or (b) an encoded character. The term *emoji character* can be used for (b) where not clear from context.
>
> ***ED-2. emoticon*** — (1) A series of text characters (typically punctuation or symbols) that is meant to represent a facial expression or gesture such as ;-) (2) a broader sense, also including emoji for facial expressions and gestures.

1.4.1 Emoji Levels

> ***ED-3. emoji character*** — A character that is recommended for use as emoji.
>
> - These are the characters listed in [emoji-data].
>
> ***ED-4. level 1 emoji character*** — An emoji character that is among those most commonly supported as emoji by vendors at present.
>
> - These characters have an [emoji-data] Field 2 value of “L1”.
>
> ***ED-5. level 2 emoji character*** — An emoji character that is not a level 1 emoji character.
>
> - These characters have an [emoji-data] Field 2 value of “L2”.

Review Note: We are considering deprecating the levels, since the need for them

appears to have been superseded. The simplest approach may be to just fix L1 as it is and add any additional characters into L2. Feedback is appreciated.

For more details about level 1 and level 2 emoji, see *Section 3 Which Characters are Emoji* .

1.4.2 Emoji Presentation

ED-6. default emoji presentation character — A character that, by default, should appear with an emoji presentation, rather than a text presentation.

- These characters have an [emoji-data] Field 1 value of “emoji”.

ED-7. default text presentation character — A character that, by default, should appear with a text presentation, rather than an emoji presentation.

- These characters have an [emoji-data] Field 1 value of “text”.

For more details about emoji and text presentation, see *2 Design Guidelines* and *Section 4 Presentation Style* .

1.4.3 Emoji Variation Sequences

ED-8. emoji variation selector — Either of the two variation selectors used to request a text or emoji presentation for an emoji character:

- U+FE0E VARIATION SELECTOR-15 (VS15) for a text presentation
- U+FE0F VARIATION SELECTOR-16 (VS16) for an emoji presentation

ED-9. emoji variation sequence — A variation sequence listed in [VSData] that contains an emoji variation selector.

- For a chart of these, see [VSChart] .

ED-10. emoji base variation sequence — An emoji variation sequence that starts with an *emoji modifier base*.

1.4.4 Emoji Modifiers

ED-11. emoji modifier — A character that can be used to modify the appearance of a preceding emoji in an *emoji modifier sequence*.

- These characters have an [emoji-data] Field 3 value of “modifier”.

ED-12. emoji modifier base — A character whose appearance can be modified by a subsequent emoji modifier in an *emoji modifier sequence*.

- These characters have an [emoji-data] Field 3 value of either “primary” or “secondary”, and are listed in Characters Subject to Emoji Modifiers.

ED-13. emoji modifier sequence — A sequence of the following form:

```
(emoji_modifier_base | emoji_base_variation_sequence) emoji_modifier
```

For more details about emoji modifiers, see *Section 2.2 Diversity.*

1.4.5 Emoji Sequences

ED-14. emoji core sequence — A sequence of the following form:

```
(emoji_character | emoji_variation_sequence | emoji_modifier_sequence |
                emoji_flag_sequence) non_spacing_mark*
```

ED-15. emoji sequence — A sequence of the following form:

```
core_sequence ( joiner core_sequence )?
```

ED-16. emoji zwj sequence — An emoji sequence with at least one joiner character.

2 Design Guidelines

Unicode characters can have many different presentations as text. An "a" for example, can look quite different depending on the font. Emoji characters can have two main kinds of presentation:

- an *emoji presentation,* with colorful and perhaps whimsical shapes, even animated
- a *text presentation,* such as black & white

More precisely, a text presentation is a simple foreground shape whose color which is determined by other information, such as setting a color on the text, while an emoji presentation determines the color(s) of the character, and is typically multicolored. In other words, when someone changes the text color in a word processor, a character with an emoji presentation will not change color.

Any Unicode character can be presented with a text presentation, as in the Unicode charts. For the emoji presentation, both the name and the representative glyph in the Unicode chart should be taken into account when designing the appearance of the emoji, along with the images used by other vendors. The shape of the character can vary significantly. For example, here are just some of the possible images for U+1F36D LOLLIPOP, U+1F36E CUSTARD, U+1F36F HONEY POT, and U+1F370 SHORTCAKE:

While the shape of the character can vary significantly, designers should maintain the same “core” shape, based on the shapes used mostly commonly in industry practice. For example, a U+1F36F HONEY POT encodes for a pictorial representation of a pot of honey, not for some semantic like "sweet". It would be unexpected to represent U+1F36F HONEY POT as a sugar cube, for example. Deviating too far from that core shape can cause interoperability problems: see accidentally-sending-friends-a-hairy-heart-emoji. Direction (whether a person or object faces to the right or left, up or down) should also be maintained where possible, because a change in direction can change the meaning: when sending 🐊🔫👮 “crocodile shot by police”, people expect any recipient to see the pistol pointing in the same direction as when they composed it. Similarly, the U+1F6B6 *pedestrian* should face to the left 🚶, not to the right.

General-purpose emoji for people and body parts should also not be given overly specific images: the general recommendation is to be as neutral as possible regarding race, ethnicity, and gender. Thus for the character U+1F64B *happy person raising one hand*, the recommendation is to use a neutral graphic like 🙋 instead of an overly-specific image like 🙋‍♀️. This includes the emoji modifier base characters listed in Characters Subject to Emoji Modifiers. The representative glyph used in the Unicode charts, or images from other vendors may be misleading: for example, the construction worker 👷 may be male or female. For more information, see the Unicode Emoji FAQ.

Names of symbols such as BLACK MEDIUM SQUARE or WHITE MEDIUM SQUARE are not meant to indicate that the corresponding character must be presented in black or white, respectively; rather, the use of “black” and “white” in the names is generally just to contrast **filled** versus **outline** shapes, or a darker color fill versus a lighter color fill. Similarly, in other symbols such as the hands U+261A BLACK LEFT POINTING INDEX and U+261C WHITE LEFT POINTING INDEX, the words “white” and “black” also refer to outlined versus filled, and do not indicate skin color.

However, other color words in the name, such as YELLOW, typically provide a recommendation as to the emoji presentation, which should be followed to avoid interoperability problems.

Emoji characters may not always be displayed on a white background. They are often best given a faint, narrow contrasting border to keep the character visually distinct from a similarly colored background. Thus a Japanese flag would have a border so that it would be visible on a white background, and a Swiss flag have a border so that it is visible on a red background.

Current practice is for emoji to have a square aspect ratio, deriving from their origin in Japanese. For interoperability, it is recommended that this practice be continued with current and future emoji.

Flag emoji characters are discussed in Annex B: Flags.

Combining marks may be applied to emoji, just like they can be applied to other characters. When that is done, the combination should take on an emoji presentation. For example, a 1️⃣ is represented as the sequence "1" plus an emoji variation selector plus U+20E3 COMBINING ENCLOSING KEYCAP. Systems are unlikely, however, to support arbitrary combining marks with arbitrary emoji. Aside from U+20E3, the ~~following can be used~~ most likely to be supported is:

- ~~U+20E4 COMBINING ENCLOSING UPWARD POINTING TRIANGLE to indicate a warning~~
- U+20E0 COMBINING ENCLOSING CIRCLE BACKSLASH, as an overlaid 🚫, to indicate a prohibition or “NO”

For example: ~~⚠ (pedestrian crossing ahead) can be represented as 🚶 + U+20E4, and 🚳 (no bicycles allowed) can be represented as 🚲 + U+20E0.~~

- 🎙⃠ <U+1F399 U+20E0> *no microphones*
- 📸⃠ <U+1F4F8 U+20E0> *no flashes*
- 🔫⃠ <U+1F52B U+20E0> *no guns*

2.1 Gender

The following emoji have explicit gender, based on the name and explicit, intentional contrasts with other characters.

U+1F466 boy
U+1F467 girl
U+1F468 man
U+1F469 woman
U+1F474 older man
U+1F475 older woman
U+1F46B man and woman holding hands
U+1F46C two men holding hands
U+1F46D two women holding hands
U+1F6B9 mens symbol
U+1F6BA womens symbol

U+1F478 princess
U+1F46F woman with bunny ears
U+1F470 bride with veil
U+1F472 man with gua pi mao
U+1F473 man with turban
U+1F574 man in business suit levitating
U+1F385 father christmas

All others should be depicted in a gender-neutral way.

2.2 Diversity

People all over the world want to have emoji that reflect more human diversity, especially for skin tone. The Unicode emoji characters for people and body parts are meant to be generic, yet following the precedents set by the original Japanese carrier images, they are often shown with a light skin tone instead of a more generic (nonhuman) appearance, such as a yellow/orange color or a silhouette.

Five symbol modifier characters that provide for a range of skin tones for human emoji were released in Unicode Version 8.0 (mid-2015). These characters are based on the six tones of the Fitzpatrick scale, a recognized standard for dermatology (there are many examples of this scale online, such as FitzpatrickSkinType.pdf). The exact shades may vary between implementations.

Emoji Modifiers

Code	Name	Samples	
U+1F3FB	EMOJI MODIFIER FITZPATRICK TYPE-1-2		
U+1F3FC	EMOJI MODIFIER FITZPATRICK TYPE-3		
U+1F3FD	EMOJI MODIFIER FITZPATRICK TYPE-4		
U+1F3FE	EMOJI MODIFIER FITZPATRICK TYPE-5		
U+1F3FF	EMOJI MODIFIER FITZPATRICK TYPE-6		

These characters have been designed so that even where diverse color images for human emoji are not available, readers can see what the intended meaning was.

The default representation of these modifier characters when used alone is as a color swatch. Whenever one of these characters *immediately* follows certain characters (such as WOMAN), then a font should show the sequence as a single glyph corresponding to the image for the person(s) or body part with the specified skin tone, such as the following:

However, even if the font doesn't show the combined character, the user can still see that a skin tone was intended:

This may fall back to a black and white stippled or hatched image such as when colorful emoji are not supported.

When a human emoji is not *immediately* followed by a emoji modifier character, it should use a generic, *non-realistic* skin tone, such as:

- RGB #FFCC22 (one of the colors typically used for the smiley faces)
- RGB #3399CC
- RGB #CCCCCC

For example, the following set uses gray as the generic skin tone:

As to hair color, dark hair tends to be more neutral, because people of every skin tone can have black (or very dark brown) hair—however, there is no requirement for any particular hair color. One exception is PERSON WITH BLOND HAIR, which needs to have blond hair regardless of skin tone.

To have an effect on an emoji, an emoji modifier must immediately follow that emoji. There is only one exception: there may be an emoji variation selector between them. The emoji modifier automatically implies the emoji presentation style, so the variation selector is not needed. However, if the emoji modifier is present it must come immediately after the modified emoji character, such as in:

<U+270C VICTORY HAND, FE0F, TYPE-3>

Any other intervening character causes the emoji modifier to appear as a free-standing character. Thus

2.2.1 Multi-Person Groupings

Emoji for multi-person groupings present some special challenges:

- **Gender combinations.** Some multi-person groupings explicitly indicate gender: MAN AND WOMAN HOLDING HANDS, TWO MEN HOLDING HANDS, TWO WOMEN HOLDING HANDS. Others do not: KISS, COUPLE WITH HEART, FAMILY (the latter is also non-specific as to the number of adult and child members). While the *default* representation for the characters in the latter group should be gender-neutral, implementations may desire to provide (and users may desire to have available) multiple representations of each of these with a variety of more-specific gender combinations.
- **Skin tones.** In real multi-person groupings, the members may have a variety of skin tones. However, this cannot be indicated using an emoji modifier with any

single character for a multi-person grouping.

The basic solution for each of these cases is to represent the multi-person grouping as a sequence of characters—a separate character for each person intended to be part of the grouping, along with characters for any other symbols that are part of the grouping. Each person in the grouping could optionally be followed by an emoji modifier. For example, conveying the notion of COUPLE WITH HEART for a couple involving two women can use a sequence with WOMAN followed by an emoji-style HEAVY BLACK HEART followed by another WOMAN character; each of the WOMAN characters could have an emoji modifier if desired.

This makes use of conventions already found in current emoji usage, in which certain sequences of characters are intended to be displayed as a single unit.

Review note: the text following was moved to Section 2.3 Emoji ZWJ Sequences.

2.2.2 Implementations

Implementations can present the emoji modifiers as separate characters in an input palette, or present the combined characters using mechanisms such as long press.

The emoji modifiers are not intended for combination with arbitrary emoji characters. Instead, they are restricted to the following characters, in two separate sets. Of these characters, it is strongly recommended that the Primary set for combination be supported. No characters outside of these two sets are to be combined with emoji modifiers. These sets may change over time, with successive versions of this document.

Characters Subject to Emoji Modifiers

Type	Images
Primary Set (26 code points)	
Secondary Set (39 code points)	

Review Note: The "smiley" faces from the secondary set have been removed, and two characters have been added (person with ball and weight lifter). We are also considering merging the primary and secondary sets, and only having one set, since the need for the distinction does not appear to be relevant any more. Feedback on

these two issues is welcome.

The following chart shows the expected display with emoji modifiers, depending on the preceding character and the level of support for the emoji modifier. A supported sequence with no visible skin should show no visible emoji modifier. The “Unsupported” rows show how the character would typically appear on a system that doesn't have a font with that character in it: with a missing glyph indicator.

Expected Emoji Modifiers Display

Support Level	Emoji Type	Sequence	Display Color	Display B&W
Fully supported	primary / secondary			
	primary / secondary no skin			
	other			
Fallback	primary / secondary			
	other			
Unsupported	primary / secondary			
	other			

Review Note: the change for the skier is for discussion at the UTC. One option, if nobody should show skin for the skier is to remove it (and like characters) from the primary set.

The interaction between variation selectors and emoji modifiers is specified as follows:

Emoji Modifiers and Variation Selectors

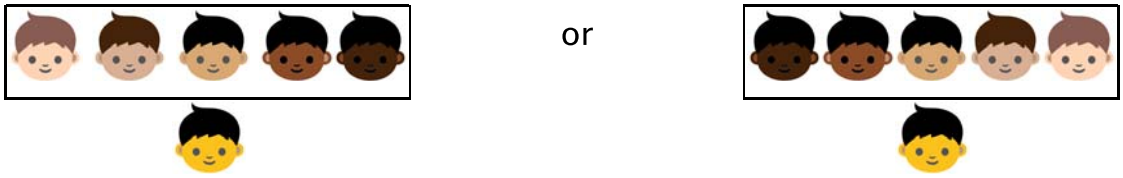
Variation	Emoji	Result	Comment

Selector	Modifier		
None	Yes	Emoji Presentation	In the absence of other information, the emoji modifier implies emoji appearance.
Emoji (U+FE0F)			The emoji modifier base and emoji variation selector must form a valid variation sequence, and the order must as specified in emoji modifier sequence—otherwise support of the variation selector would be non-conformant.
Text (U+FE0E)		Text Presentation	

2.2.3 Emoji Modifiers in Text

A supported emoji modifier sequence should be treated as a single grapheme cluster for editing purposes (cursor moment, deletion, etc.); word break, line break, etc. For input, the composition of that cluster does not need to be apparent to the user: it appears on the screen as a single image. On a phone, for example, a long-press on a human figure can bring up a minipalette of different skin tones, without the user having to separately find the human figure and then the modifier. The following shows some possible appearances:

Minipalettes



Of course, there are many other types of diversity in human appearance besides different skin tones: Different hair styles and color, use of eyeglasses, various kinds of facial hair, different body shapes, different headwear, and so on. It is beyond the scope of Unicode to provide an encoding-based mechanism for representing every aspect of human appearance diversity that emoji users might want to indicate. The best approach for communicating very specific human images—or any type of image in which preservation of specific appearance is very important—is the use of embedded graphics, as described in Longer Term Solutions.

2.3 Emoji ZWJ Sequences

The U+200D ZERO WIDTH JOINER (ZWJ) can be used between the elements of a sequence of characters to indicate that a single glyph should be presented if available. An implementation may use this mechanism to handle such an emoji zwj sequence as a single glyph, with a palette or keyboard that generates the appropriate sequences for the glyphs shown. So to the user, these would behave like single emoji characters, even though internally they are sequences.

When an emoji zwj sequence is sent to a system that does not have a corresponding

single glyph, the ZWJ characters would be ignored and a fallback sequence of separate emoji would be displayed. Thus an emoji zwj sequence should only be supported where the fallback sequence would also make sense to a recipient.

For example, the following are possible displays:

ZWJ Sequence Display

Sequence	Display	Combined glyph?
		Yes
		No

See also Annex E: Existing Use of ZWJ Sequences.

Review note: fix this to point directly at the emoji charts page.

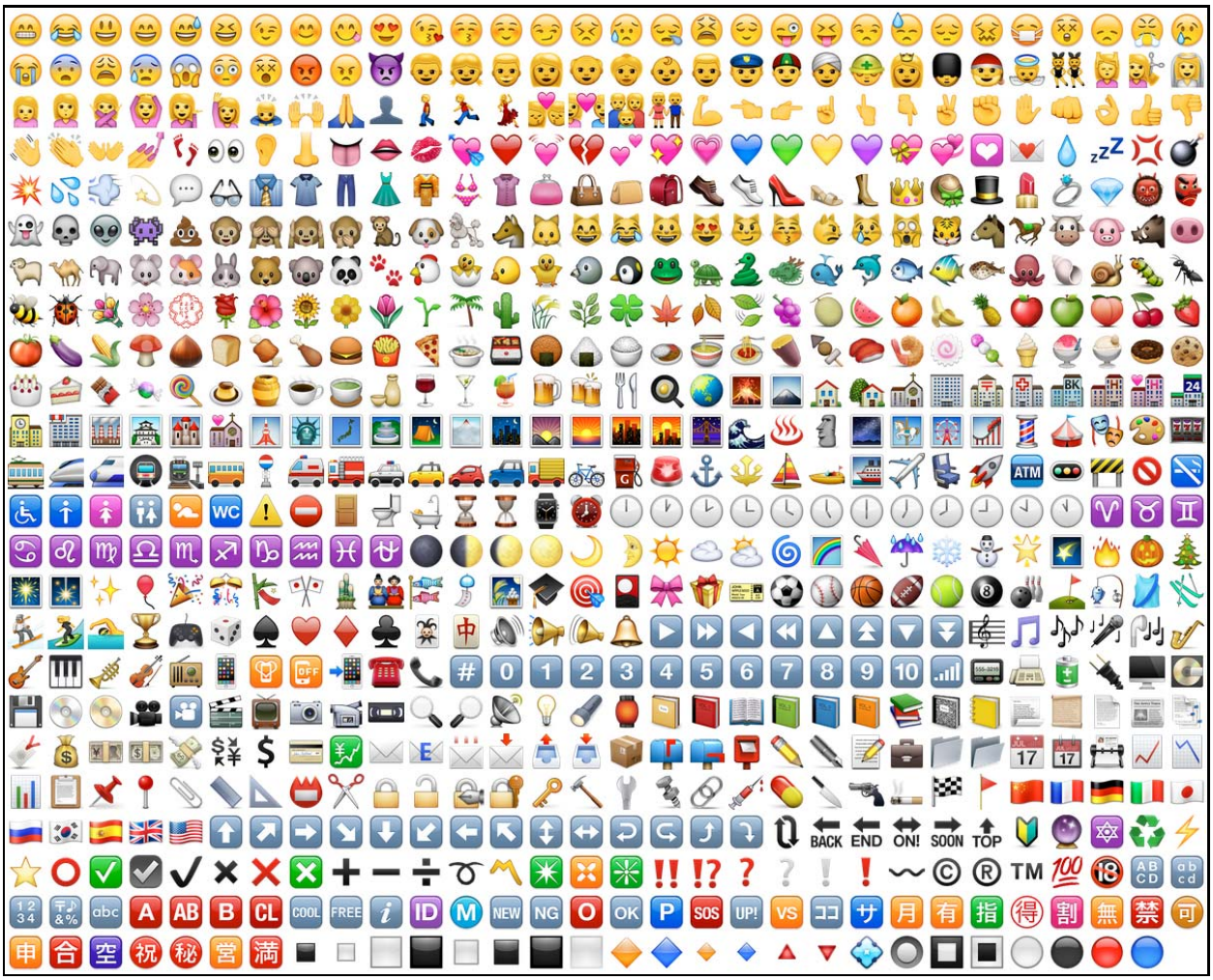
In a sequence of characters connected using ZWJ, it is recommended that the entire sequence have an emoji presentation if any character in the sequence has explicit or default emoji presentation.

3 Which Characters are Emoji

3.1 Level 1 Emoji

There are 722 Unicode emoji characters corresponding to the Japanese carrier sets. Three of these are space characters in Unicode, which cannot have an emoji presentation. This leaves 719 characters, shown below.

Japanese Carrier Emoji



In addition, most vendors support another 126 characters (from Unicode 6.0 and 6.1):

Common Additions

The **Japanese Carrier Emoji** plus the **Common Additions** comprise the 845 characters of ***level 1*** emoji.

3.2 Level 2 Emoji

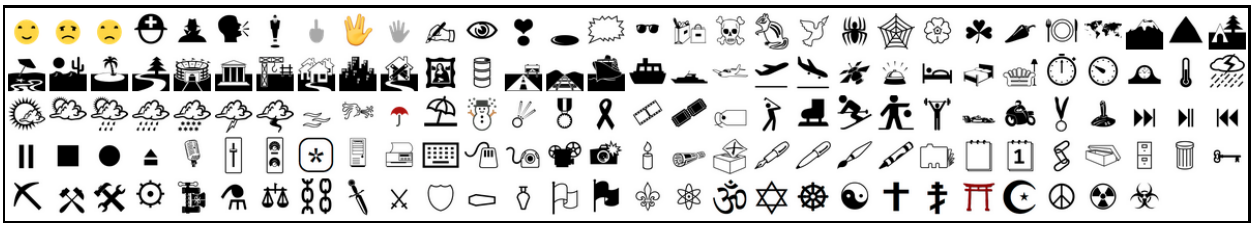
There are another 247 flags (aside from the 10 from the Japanese carrier sets) that can be supported with Unicode 6.0 characters.

Other Flags

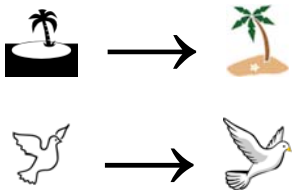
Some of these flags use the same glyphs. For more about flags, see Annex B: Flags.

The following additional 148 characters are recommended as emoji. Most, but not all, of these are new in Unicode 7.0.

Standard Additions



Thus vendors that support emoji should provide a colorful appearance for each of these, such as the following:



The 41 characters that were new in Unicode 8.0 are listed below. For details, including sample colorful images, see Annex D: Standard Additions ~~for Unicode 8.0~~ .

Review Note: point directly at the emoji/charts page instead of Annex D.

Standard Additions 8.0

Candidate emoji for future versions of Unicode are also found under Annex D: Standard Additions.

Review Note: adjust the counts and wording once the Unicode 9.0 repertoire is final.

Probably collapse the Standard Additions 8.0 into the Standard Additions.

Review Note: point directly at the emoji/charts page instead of Annex D.

The **Other Flags**, **Standard Additions**, and **Standard Additions 8.0** comprise the 436 characters of ***level 2*** emoji. Thus there is a total of 1,281 level 1 and 2 emoji characters in Unicode 8.0.

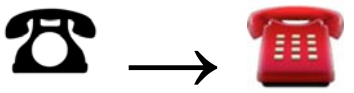
3.3 Methodology

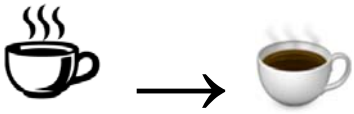
This document provides data files [emoji-data], described in the section Data Files, for determining the set of characters which are expected to have an emoji presentation, either as a default or as a alternate presentation. While Unicode conformance allows any character to be given an emoji representation, characters that are not listed in [emoji-data] should not normally be given an emoji presentation. For example, pictographic symbols such as keyboard symbols or math symbols (like ANGLE) that should never be treated as emoji. These are current recommendations: existing symbols can be added to this list over time.

This data was derived by starting with the characters that came from the original Japanese sets, plus those that major vendors have provided emoji fonts for. Characters that are similar to those in shape or design were then added. Often these characters are in the same Unicode blocks as the original set, but sometimes not.

This document takes a functional view regarding the identification of emoji: pictographs are categorized as emoji when it is reasonable to give them an emoji presentation, and where they are sufficiently distinct from other emoji characters. ~~Symbols with a graphical form that people may treat as pictographs, such as U+2615 HELM SYMBOL (introduced in Unicode 4.0) may be included.~~

~~This document takes a functional view as to the identification of emoji, which is that pictographs—or symbols that have a graphical form that people may treat as pictographs—are categorized as emoji~~, such as U+260E BLACK TELEPHONE (introduced in Unicode 1.1) or U+2615 HOT BEVERAGE (introduced in Unicode 4.0):





The data does not include *non*-pictographs, except for those in Unicode that are used to represent characters from emoji sources, such as:

 or

Game pieces, such as the dominos (🁢 🁣 🁤 ... 🁫 🁬), are currently not included as

emoji, with the exceptions of U+1F0CF (🃏) PLAYING CARD BLACK JOKER and U+1F004 (🀄) MAHJONG TILE RED DRAGON. These are included because they correspond each to an emoji character from one of the carrier sets.

4 Presentation Style

Certain emoji have defined variation sequences, where an emoji character can be followed by one of two invisible emoji variation selectors:

- U+FE0E for a text presentation
- U+FE0F for an emoji presentation

This capability was added in Unicode 6.1. Some systems may also provide this distinction with higher-level markup, rather than variation sequences. For more information on these selectors, see [VSChart] .

Implementations should support both styles of presentation for the characters with variation sequences, if possible. Most of these characters are emoji that were unified with preexisting characters. Because people are now using emoji presentation for a broader set of characters, it is anticipated that more such variation sequences will be needed.

However, even where the variation selectors exist, it has not been clear for implementers whether the *default* presentation for pictographs should be emoji or text. That means that a piece of text may show up in a different style than intended when shared across platforms. While this is all a perfectly legitimate for Unicode characters —*presentation style is never guaranteed*—a shared sense among developers of when to use emoji presentation by default is important, so that there are fewer unexpected and "jarring" presentations. Implementations need to know what the generally expected default presentation is, to promote interoperability across platforms and applications.

There has been no clear line for implementers between three categories of Unicode characters:

1. **emoji-default:** those expected to have an emoji presentation by default, but can also have a text presentation
2. **text-default:** those expected to have a text presentation by default, but could also have an emoji presentation
3. **text-only:** those that should only have a text presentation

The data files [emoji-data], described in the section Data Files, provide data to distinguish between the first two categories: see the **Default** column of full-emoji-list. The data assignment is based upon current usage in browsers for Unicode 6.3 characters. For other characters, especially the new 7.0 characters, the assignment is based on that of the related emoji characters. For example, the “vulcan” hand 🖖 is marked as *emoji-default* because of the emoji styling currently given to other hands like ✋. The *text-only* characters are all those not listed in the data files.

In general, emoji characters are marked as *text-default* if they were in common use and predated the use of emoji. The characters are otherwise marked as *emoji-default*. For

example, the negative squared A and B are text-default, while the negative squared AB is emoji-default. The reason is that A and B are part of a set of negative squared letters A-Z, while the AB was a new character. The default status may change over time, however, if common usage changes.

The presentation of a given emoji character depends on the environment, whether or not there is an emoji or text variation selector, and the default presentation style (emoji vs text). In informal environments like texting and chats, it is more appropriate for most emoji characters to appear with a colorful emoji presentation, and only get a text presentation with a text variation selector. Conversely, in formal environments such as word processing, it is generally better for emoji characters to appear with a text presentation, and only get the colorful emoji presentation with the emoji variation selector.

Based on those factors, here is typical presentation behavior. However, these guidelines may change with changing user expectations.

Emoji vs Text Display

Example Environment	with Emoji VS	with Text VS	with no VS	
			text-default	emoji-default
word processing	☕	☕	☕	☕
plain web pages	☕	☕	☕	☕
texting, chats	☕	☕	☕	☕

5 Ordering and Grouping

Neither the Unicode code point order, nor the standard Unicode Collation ordering (DUCET), are currently well suited for emoji, since they separate conceptually-related characters. From the user's perspective, the ordering in the following selection of characters sorted by DUCET appears quite random, as illustrated by the following example:

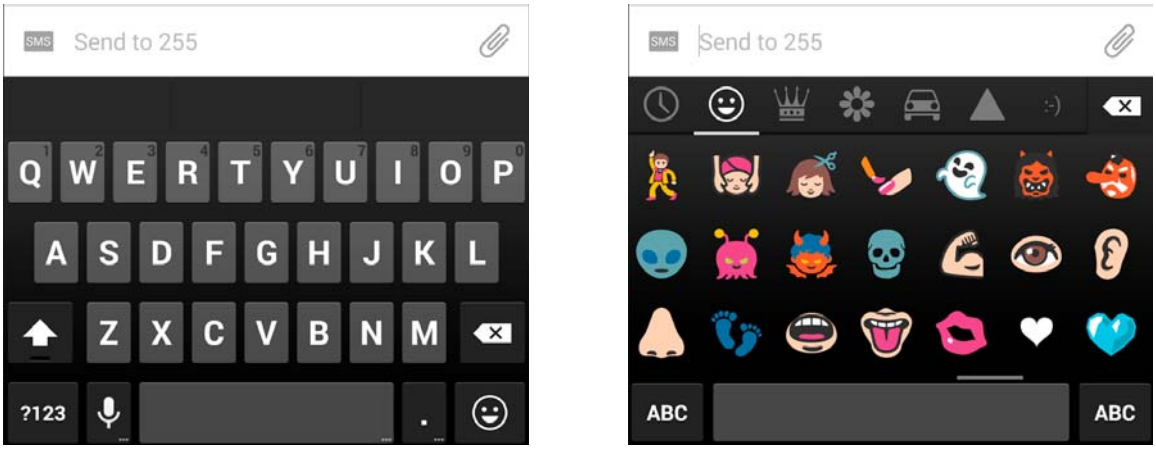
The emoji-ordering chart file shows an ordering for emoji characters that groups them together in a more natural fashion. This data has been incorporated into [CLDR].

This ordering groups characters presents a cleaner and more expected ordering for sorted lists of characters. The groupings include: faces, people, body-parts, emotion, clothing, animals, plants, food, places, transport, and so on. The ordering also groups more naturally for the purpose of selection in input palettes. However, for sorting, each character must occur in only one position, which is not a restriction for input palettes. See *Section 6 Input*.

6 Input

Emoji are not typically typed on a keyboard. Instead, they are generally picked from a palette, or recognized via a dictionary. The mobile keyboards typically have a ☺ button to select a palette of emoji, such as in the left image below. Clicking on the ☺ button reveals a palette, as in the right image.

Palette Input



The palettes need to be organized in a meaningful way for users. They typically provide a small number of broad categories, such as People, Nature, and so on. These categories typically have 100-200 emoji.

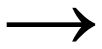
Many characters can be categorized in multiple ways: an orange is both a plant and a food. Unlike a sort order, an input palette can have multiple instances of a single character. It can thus extend the sort ordering to add characters in any groupings where people might reasonably be expected to look for them.

More advanced palettes will have long-press enabled, so that people can press-and-hold on an emoji and have a set of related emoji pop up. This allows for faster navigation, with less scrolling through the palette.

Annotations for emoji characters are much more finely grained keywords. They can be used for searching characters, and are often easier than palettes for entering emoji characters. For example, when someone types “hourglass” on their mobile phone, they could see and pick from either of the matching emoji characters ⌛ or ⏳. That is often much easier than scrolling through the palette and visually inspecting the screen. Input mechanisms may also map *emoticons* to emoji as keyboard shortcuts: typing :-) can result in 😄.

In some input systems, a word or phrase bracketed by colons is used to explicitly pick emoji characters. Thus typing in “I saw an *:ambulance:*” is converted to “I saw an 🚑”. For completeness, such systems might support all of the full Unicode names, such as *:first quarter moon with face:* for 🌛. Spaces within the phrase may be represented by _, as in the following:

“my *:alarm_clock:* didn’t work”



“my ⏱ didn’t work”.

However, in general the full Unicode names are not especially suitable for that sort of use; they were designed to be unique identifiers, and tend to be overly long or confusing.

7 Searching

Searching includes both searching for emoji characters in queries, and finding emoji characters in the target. These are most useful when they include the annotations as synonyms or hints. For example, when someone searches for ⛽ on yelp.com, they see matches for “gas station”. Conversely, searching for “gas pump” in a search engine could find pages containing ⛽. Similarly, searching for “gas pump” in an email program can bring up all the emails containing ⛽.

There is no requirement for uniqueness in both palette categories and annotations: an emoji should show up wherever users would expect it. A gas pump ⛽ might show up under “object” and “travel”; a heart 💔 under “heart” and “emotion”, a 😍 under “animal”, “cat”, and “heart”.

Annotations are language-specific: searching on yelp.de, someone would expect a search for ⛽ to result in matches for “Tankstelle”. Thus annotations need to be in multiple languages to be useful across languages. They should also include regional annotations within a given language, like “petrol station”, which people would expect search for ⛽ to result in on yelp.co.uk. An English annotation cannot simply be translated into different languages, since different words may have different associations in different languages. The emoji 🌵 may be associated with Mexican or Southwestern restaurants in the US, but not be associated with them in, say, Greece.

There is one further kind of annotation, called a *TTS name*, for text-to-speech processing. For accessibility when reading text, it is useful to have a short, descriptive name for an emoji character. A Unicode character name can often serve as a basis for this, but its requirements for name uniqueness often ends up with names that are overly long, such as *black right-pointing double triangle with vertical bar* for ⏯. TTS names are also outside the current scope of this document.

8 Longer Term Solutions

The longer-term goal for implementations should be to support embedded graphics, in addition to the emoji characters. Embedded graphics allow arbitrary emoji symbols, and are not dependent on additional Unicode encoding. Some examples of this are found in Skype and LINE—see the emoji press page for more examples.

However, to be as effective and simple to use as emoji characters, a full solution requires significant infrastructure changes to allow simple, reliable input and transport of images (stickers) in texting, chat, mobile phones, email programs, virtual and mobile keyboards, and so on. (Even so, such images will never interchange in environments

that only support plain text, such as email addresses.) Until that time, many implementations will need to use Unicode emoji instead.

For example, mobile keyboards need to be enhanced. Enabling embedded graphics would involve adding an additional custom mechanism for users to add in their own graphics or purchase additional sets, such as a + sign to add an image to the palette above. This would prompt the user to paste or otherwise select a graphic, and add annotations for dictionary selection.

With such an enhanced mobile keyboard, the user could then select those graphics in the same way as selecting the Unicode emoji. If users started adding many custom graphics, the mobile keyboard might even be enhanced to allow ordering or organization of those graphics so that they can be quickly accessed. The extra graphics would need to be disabled if the target of the mobile keyboard (such as an email header line) would only accept text.

Other features required to make embedded graphics work well include the ability of images to scale with font size, inclusion of embedded images in more transport protocols, switching services and applications to use protocols that do permit inclusion of embedded images (eg, MMS versus SMS for text messages). There will always, however, be places where embedded graphics can't be used—such as email headers, SMS messages, or file names. There are also privacy aspects to implementations of embedded graphics: if the graphic itself is not packaged with the text, but instead is just a reference to an image on a server, then that server could track usage.

Annex A: Data Files

The main data file is [emoji-data]. The format for that file is described in its header.

See [emoji-charts] for a collection of charts that have been generated from the emoji data file that may be useful in helping to understand it and the related [CLDR] emoji data (annotations and ordering). These charts are not versioned, and are purely illustrative; the data to use for implementation is in [emoji-data].

Annex B: Flags

26 REGIONAL INDICATOR symbols are used in pairs to represent country flags. Only valid sequences should be used, where:

- The valid region sequences are specified by Unicode region subtags as defined in [CLDR], excluding those that are designated private-use or deprecated in [CLDR]. (An overseas territory of a country may share the same flag as for the country itself.)

Emoji are generally presented with a square aspect ratio, which presents a problem for flags. The flag for Qatar is over 150% wider than tall; for Switzerland it is square; for Nepal it is over 20% taller than wide. To avoid a ransom-note effect, implementations may want to use a fixed ratio across all flags, such as 150%, with a blank band on the top and bottom. (The average width for flags is between 150% and 165%.) Narrower flags, such as the Swiss flag, may also have white bands on the side.

Flags should have a visible edge. One option is to use a 1 pixel gray line chosen to be contrasting with the adjacent field color.

The code point order of flags is by region code, which will not be intuitive for viewers, since that rarely matches the order of countries in the viewer's language. English speakers are surprised that the flag for Germany comes before the flag for Djibouti. An alternative is to present the sorted order according to the localized country name, using [CLDR] data.

For an open-source set of flag images (png and svg), see region-flags.

Annex C: Selection Factors

In the past, most emoji characters have been selected primarily on the basis of compatibility. The scope is being broadened to include other factors, as listed below.

To submit a proposal for a new emoji character, fill out the form for Submitting Character Proposals. To that form, also add an annex that lists each of the selection factors below, and for each one provides evidence as to what degree each proposed character would satisfy that factor.

None of these factors are completely determinant. For example, the word for an object may be extremely common on the internet, but the object not necessarily a good candidate due to other factors.

a. Compatibility. *Are these needed for compatibility with high-use emoji in existing systems, such as Gmail?*
   - For example, FACE WITH ROLLING EYES.
   - Compatibility is a strong selection factor. There are many cases where characters are or have been added for compatibility alone, such as 🆕 SQUARED NEW, or 👷 CONSTRUCTION WORKER. In such cases, many of the other factors don't apply.

b. Expected usage level. *Is there a high expected frequency of use?*
   - There are various possible measures of this that can be presented as evidence, such as:
     - whether closely-related characters show up above the median in emojitracker.com
     - the frequency of related words in web pages
     - the frequency in image search (eg, google or bing)
     - whether the object is commonly encountered in daily life
     - multiple usages, such as SHARK for not only the animal, but also for a *huckster*, in *jumping the shark*, *card shark*, *loan shark*, etc.

c. Image distinctiveness. *Is there a clearly recognizable image of physical objects that could serve as a paradigm, that would be distinct enough from other emoji?*
   - For example, CASSOULET or STEW probably couldn't be easily distinguished from 🍲 POT OF FOOD.
   - Simple words ("NEW") or abstract symbols ("卌") would not qualify as emoji.

- Note that objects often may represent activities or modifiers, such as 😢 CRYING FACE for *crying* or 🏃 RUNNER for *running*.

d. Disparity. *Does the proposed pictograph fill in a gap in existing types of emoji?*
   - For example, in Unicode 7.0 we have 🐯 TIGER, but not LION; ⛪ CHURCH but not MOSQUE.

e. Frequently requested. *Is it often requested of the Unicode Consortium, or of Unicode member companies?*
   - For example, HOT DOG or UNICORN.
   - Petitions are only considered as possible indications of potential frequency of usage, among the other selection factors. Citations of petition results should provide evidence as how reliable the petition mechanism is (in terms of preventing duplicates or robovotes) and account to what extent the results could be skewed by commercial promotion of the petition.

f. Generality. *Is the proposed character overly specific?*
   - For example, 🍣 SUSHI represents sushi in general, even though a common image will be of a specific type, such as Maguro. Adding SABA, HAMACHI, SAKE, AMAEBI and others would be overly specific.

g. Open-ended. *Is it just one of many, with no special reason to favor it over others of that type?*
   - For example, there are thousands of people, including occupations (DOCTOR, DENTIST, JANITOR, POLITICIAN, etc.): is there a special reason to favor particular ones of them?

h. Representable already. *Can the concept be represented by another emoji or sequence?*
   - For example, a *crying baby* can already be represented by 😢👶 CRYING FACE + BABY
   - A building associated with a particular religion might be represented by a PLACE OF WORSHIP emoji followed by a one of the many religious symbols in Unicode.
   - Halloween could be represented by either just 🎃 JACK-O-LANTERN, or a sequence of 🎃👻 JACK-O-LANTERN + GHOST.

i. Logos, Brands, UI icons, signage, specific people, deities. *Are the images unsuitable for encoding as characters?*
   - These are strong factors for **ex**clusion.
   - They include:
     - Images such as company logos, or those showing company brands as part or all of the image.
     - UI icons such as Material Design Icons, Winjs Icons, or Font Awesome Icons, which are often discarded or modified to meet evolving UI needs.
     - Signage such as  . See also Slate's *The Big Red Word vs. the Little Green Man*.
     - Specific people, whether historic or living.

- Deities.

Note that symbols used in signage or user interfaces may be encoded in Unicode for reasons unconnected with their use as emoji.

Annex D: Standard Additions ~~for Unicode 9.0~~

New candidates for the next version of Unicode are provided in the emoji charts, at Emoji Candidates. Those for recent versions of Unicode are in Emoji Accepted.

Annex E: Existing Use of ZWJ Sequences

The contents of this chart has been moved to the emoji charts, in Emoji ZWJ Sequences.

Review note: We probably want to change the name of the chart to Emoji Sequences, to also list keycaps, etc.

Acknowledgments

Mark Davis and Peter Edberg created the initial versions of this document, and maintain the text.

Thanks to Shervin Afshar, Julie Allen, Jeremy Burge, Michele Coady, Chenjintao (陈锦涛), Chenshiwei, Peter Constable, Craig Cummings, Andrew Glass, Paul Hunt, Hiroyuki Komatsu, Norbert Lindenberg, Ken Lunde, Gwyneth Marshall, Rick McGowan, Katsuhiko Momoi, Katsuhiro Ogata, Katrina Parrott, Michelle Perham, Addison Phillips, Roozbeh Pournader, Judy Safran-Aasen, Markus Scherer, Richard Tunnicliffe, and Ken Whistler for feedback on and contributions to this document, including earlier versions. ~~Thanks to Michael Everson and Microsoft for 8.0 chart images.~~

Thanks to Adobe / Paul Hunt, Apple, Michael Everson, Google, Microsoft, and iDiversicons for supplying images for illustration.

Rights to Emoji Images

The colored images used in this document and associated charts [emoji-charts] are for illustration only. They do not appear in the Unicode Standard, which has only black and white images. They are either made available by the respective vendors for use in this document, or are believed to be available for non-commercial reuse.

The Unicode Consortium is not a designer or purveyor of emoji images, nor is it the owner of any of the color images used in the document, nor does it negotiate licenses for their use. Inquiries for permission to use vendor images should be directed to those vendors, not to the Unicode Consortium.

References

[CLDR] CLDR – Unicode Common Locale Data Repository
http://cldr.unicode.org/

For the latest version of the associated specification (LDML), see:
http://www.unicode.org/reports/tr35/

[emoji-charts] The illustrative charts of emoji
http://unicode.org/emoji/charts/index.html

[emoji-data] The associated data file for emoji characters
For the latest version, see:
http://unicode.org/Public/emoji/latest/emoji-data.txt
For the 1.0 version, see:
http://unicode.org/Public/emoji/1.0/emoji-data.txt

[JSources] Review note: adjust in final release
The UCD sources for the JCarrier symbols
For the latest version, see:
http://unicode.org/Public/UCD/latest/ucd/EmojiSources.txt
For the 8.0 version, see:
http://unicode.org/Public/8.0.0/ucd/EmojiSources.txt

[Unicode] The Unicode Standard
For the latest version, see:
http://unicode.org/versions/latest/

[VSData] Review note: adjust in final release
A data file listing permissible variation sequences
For the latest version, see:
http://unicode.org/Public/UCD/latest/ucd/StandardizedVariants.txt
For the 8.0 version, see:
http://unicode.org/Public/8.0.0/ucd/StandardizedVariants.txt

[VSChart] Review note: adjust in final release
A chart of permissible variation sequences
For the latest version, see:
http://unicode.org/Public/UCD/latest/ucd/StandardizedVariants.htm
For the 8.0 version, see:
http://unicode.org/Public/8.0.0/ucd/StandardizedVariants.html

Modifications

The following summarizes modifications from the previous revisions of this document.

Revision 4

- Proposed update.
- General
  - Wording changes to prepare for Unicode 9.0.
- Section 1.4 Definitions
  - 1.4.5 Emoji Sequences Added subsection with definitions for:
    - ***ED-14. emoji core sequence***
    - ***ED-15. emoji sequence***
    - ***ED-16. emoji zwj sequence***
  - Added note on the possibility of removing the distinction between the levels.
- Section 2 Design Guidelines
  - Updated the text about combining marks
- Section 2.2.2 Implementations
  - Removed faces from the Secondary Set (Characters Subject to Emoji Modifiers) and added two other characters.
  - Added review note about intention to merge primary and secondary sets.
  - Added clarification and example for how to handle emoji modifier sequences where the base image doesn't have any visible skin.
- Section 2.3 Emoji ZWJ Sequences
  - Moved from Multi-Person Groupings to its own section.
- Annex D: Standard Additions for Unicode 8.0
  - Changed the title to be version-neutral
  - Changed contents to just point to the new charts
    - Emoji Accepted
    - Emoji Candidates
- Annex E: Existing Use of ZWJ Sequences
  - Moved the contents to the chart Emoji ZWJ Sequences.
- Acknowledgments
  - Added more contributors

Revision 3

- First approved published version.

Revision 2

- **Draft Unicode Technical Report**
- General
  - Updated Table of Contents to add links to tables
  - Added references for charts and CLDR, and updated references to them and to [emoji-data].

    - Changed links from Public/UNIDATA to Public/UCD/latest/ucd/
    - For clarity, replaced "minimal" with "primary" and "optional" with "secondary".
- Section 1 Introduction
    - Modified the introductory text slightly, incorporating material from the old section 9 Media.
    - Added single letter abbreviations to Table: Major Sources.
    - Added note on use of colored images.
    - Added new Section 1.4 Definitions. Added introductory text (replacing review note). This incorporates definitions from the old Annex A: Terminology (which is now replaced by a different Annex A) as well as the definitions from the old section 2.2.1 Implementations, and adds several new definitions.
- Section 2 Design Guidelines
    - *Section 2.2 Diversity*
        - Added several examples of non-realistic skin tones for generic human figures; clarified the guidelines on hair color.
        - Emphasized that an emoji modifier must occur immediately following a emoji character (optionally with an intervening variation selector) for it to have any effect on that emoji character.
    - *Section 2.2.1 Multi-Person Groupings*
        - New section from material formerly at the end of section 2.2; old section 2.2.1 Implementations now becomes section 2.2.2.
        - Expanded the discussion on gender and skin tone variations in emoji for multi-person groupings.
        - Added material on use of ZWJ in such sequences to request a single glyph if available.
        - Restructured to provide both display options, and move other text and example to Annex E.
    - *Section 2.2.2 Implementations*
        - Renumbered, was previously section 2.2.1.
        - Updated the table Characters Subject to Emoji Modifiers to add Unicode 8.0 characters as secondary. Includes faces (aside from ROBOT FACE) and SIGN OF THE HORNS.
        - Removed U+1F46F WOMAN WITH BUNNY EARS from primary, since it is typically represented as two people.
        - Removed the lists of codepoints in the table Characters Subject to Emoji Modifiers, since those were only for review. (The information is in the data file.)
        - Added sentence about ordering in the first table.
        - Recast list of conditions as new *Table: Emoji Modifiers and Variation Selectors*.
    - *Section 2.2.3 Emoji Modifiers in Text*
        - New section from material formerly at the end of old section 2.2.1 Implementations.
        - Illustrated that there is no particular order among the diversity

modifiers in *Table: Minipalettes*

- Section 3 Which Characters are Emoji
  - Changed the lists of characters to use block images for speed.
  - Added a figure showing the Japanese carrier emoji (feasible as a block).
  - Related the sets to the emoji levels, with new subsections 3.1 and 3.2 for each level.
  - Added a figure for standard additions from Unicode 8.0.
  - Adjusted the text around the lists (especially the 8.0 characters) to reflect the release status.
  - Removed HELM SYMBOL.
  - Adjusted the character counts accordingly, and dropping the three space characters — "a total of 1,281 level 1&2 emoji characters in Unicode 8.0."
  - Separated the material formerly at the end of section 3 into a new subsection 3.3. Methodology.
- Section 4 Presentation Style
  - Changed "text" and "presentation" to formal and informal, added more explanation.
  - Removed the table Emoji Environments, and modified table Emoji vs Text Display to dereify table rows for "Formal", "Informal", and "Mixed".
- Section 5 Ordering and Grouping, and Section 6 Input
  - Clarified the relation of grouping and ordering (replacing review note).
- Section 9 Media
  - Moved the text to early in the introduction, and removed the section.
- Section 10 Data Files becomes Annex A: Data Files, see next item.
- Annex A: Data Files
  - Replaces the old Annex A: Terminology, whose material has been moved to a new Section 1.4 Definitions.
  - Moved the html charts out of Annex A to a separate page for emoji charts, leaving only the data file, and a pointer to the charts page.
- Annex B: Flags
  - Some simplification of the text.
- Annex C: Selection Factors.
  - Added pointer to submission form, and a bit more explanation.
  - Added to factors: Historic or living people, Deities; petitions.
- Annex D: Standard Additions for Unicode 8.0
  - Removed DHYANI BUDDHA, changed name for BADMINTON RACQUET AND SHUTTLECOCK
  - Changed the titles and introductory text to reflect that these are no longer candidates.
  - Added code points to the table
  - Added double-links to the code points
  - Moved colored samples into the table, and extended them.
- Annex E: ZWJ Sequences In Use

    - Added
- Acknowledgments
    - Added additional members of the emoji subcommittee who've contributed to this document, and thanks to vendors for colored images.
- Rights to Emoji Images
    - Added information about the use of colored images.

Revision 1

- **Proposed Draft Unicode Technical Report**
- Added new selection factor Logos, Brands, UI icons, and signage, and added double-links to the clauses
- Moved Media list to Articles on Emoji, and reordered to most-recent-first.
- Restricted the list of recommended emoji in the data files, based on the emoji subcommittee review.
    - Correspondingly reduced the list of optional characters in *Table: Characters Subject to Emoji Modifiers*
    - Added explicit lists of emoji characters to *Section 3 Which Characters are Emoji* for easier review (than looking at the data files). Moved the numbers of characters to that section from the introduction.
- Added recommended breaking behavior to *Section 2.2.1 Implementations*.
- Added text-style.html chart for easier review of the default style for emoji characters.
- Added links to the Feedback section in relevant review notes, to make it easier for people to add feedback.
- Added Annex D: Emoji Candidates for Unicode 8.0.
- Added double-linked captions to tables.
- Added months to the dates in the table of Major Sources.
- Added more notes to Annex B: Flags, and on their ordering in Data File Descriptions
- Added text on the interaction between emoji modifiers and variation selectors in *Section 2.2.1 Implementations*
- Removed multiple-person emoji from the minimal set in Characters Subject to Emoji Modifiers
- Minor edits
- Added Annex C: Selection Factors
