Author	Jules Bertholet
Title	Proposal to change the East_Asian_Width property of various characters, and the derivation of Line_Break=Ambiguous
Date	2024-06-18

Proposal

Update the East_Asian_Width property values of the following characters to Ambiguous:
- All characters with the Extended_Pictographic property that currently have East_Asian_Width=Neutral, except for:
  - © U+00A9 COPYRIGHT SIGN,
  - ‼ U+203C DOUBLE EXCLAMATION MARK,
  - ⁉ U+2049 EXCLAMATION QUESTION MARK,
  - ⅈ U+2139 INFORMATION SOURCE, and
  - characters in the "Enclosed Alphanumeric Supplement" block.
- ⁂ U+2042 ASTERISM.
- ⁑ U+2051 TWO ASTERISKS ALIGNED VERTICALLY.
- The dentistry notation symbols in U+23BE..U+23CC.
- ⏎ U+23CE RETURN SYMBOL.
- ⦾ U+29BE CIRCLED WHITE BULLET and ⦿ U+29BF CIRCLED BULLET.
- ⧺ U+29FA DOUBLE PLUS and ⧻ U+29FB TRIPLE PLUS.
- Characters in the "Box Drawings" block that are not already Ambiguous: U+254C..U+254F and U+2574..U+257F.
- Box drawing characters from other blocks: U+1FBA0..U+1FBAF, and (new in Unicode 16) U+1CC1B..U+1CC20, U+1CE09..U+1CE0A, U+1CE16..U+1CE19.
- Various circled digits: U+24EA, U+2780..U+2793, U+1F10B..U+1F10C.
- All vulgar fractions (with Decomposition_Type=Fraction) that are not already Ambiguous: U+2150..U+2152, U+2155..U+215A, and U+215F.

Also, change the derivation of Line_Break=Ambiguous to the following, ensuring that it is a subset of East_Asian_Width=Ambiguous:
- Would otherwise be Line_Break=Alphabetic, AND
- Is East_Asian_Width=Ambiguous, AND
- Does not have a General_Category of Letter or Modifier_Symbol, AND
- Is not ® U+00AE REGISTERED SIGN.

**Proposal to change the East_Asian_Width property of various characters,
and the derivation of Line_Break=Ambiguous**

Characters with default emoji presentation that are currently Line_Break=Ambiguous should become Line_Break=Ideographic, for consistency with most other emojis.)

Rationale

UAX 11 states that characters with an East_Asian_Width of Neutral are those that "do not occur in East Asian typography." In contrast, the set of Ambiguous characters is "all characters that can be sometimes wide and sometimes narrow." The proposed changes align the assignments of these properties to their definitions:

- Pictographic characters (with the Extended_Photographic property) are generally displayed as wide by monospaced CJK fonts that support them; but despite this fact, most (though not all) are currently assigned East_Asian_Width=Neutral. This leads to several surprising inconsistencies. For example, the black and white spade, heart, and club card suit symbols are Ambiguous, but those for the diamond suit are Neutral. In addition, the musical sharp (♯) and flat (♭) signs are Ambiguous, but the natural sign (♮) is Neutral. Assigning a width of Ambiguous to all currently-Neutral Extended_Photographic characters, with a few exceptions reflecting common practice, resolves these issues.
  - The omitted characters are ones that are not actually pictographic in their default presentation, and have the Extended_Pictographic property only because they can become emoji when followed by U+FE0F.
- U+2051, U+23BE..U+23CC, U+23CE, U+29BE, U+29BF, U+29FA, and U+29FB originate from the Japanese JIS X 0213 standard (L2/00-0098), and display as wide in common CJK monospace fonts supporting them; therefore, they should be Ambiguous.
  - U+2042 ⁂ is similar to U+2051 ‑, so if the latter is to have Ambiguous width, so should the former. (This also matches the behavior of common fonts.)
- Certain box-drawing characters are Ambiguous, while others are currently Neutral. Box-drawing characters are meant to align with one another, so it's critical that they all have the same width. (The U+23BE..U+23CC dentistry symbols mentioned above are designed to interact with the box-drawing characters, and therefore must also have the same width as them.)
- Some circled digits and vulgar fractions are Ambiguous, while others are currently Neutral. Fonts will display these uniformly, so their properties should be made uniform also.

In an attempt to capture modern practice, the set of characters with Line_Break=Ambiguous has diverged both from UAX14 and from its stated derivation in UAX 14. Some of these divergences address the inconsistencies listed above; but, in the process, new inconsisten-

cies are introduced. In addition, the divergence between the two properties makes it more diffi-
cult for implementers to tailor both at once. The proposed new derivation resolves these issues,
by ensuring that Line_Break=Ambiguous is a strict subset of East_Asian_Width=Ambiguous
that reflects common modern practice.

Future Work

The set of arrows and mathematical symbols with East_Asian_Width=Ambiguous re-
mains inconsistent with common practice; a future proposal should investigate making further
changes to these.

Font samples

From left to right: a wide CJK ideograph, an Extended_Pictographic character, U+2042
ASTERISM, U+2051 TWO ASTERISKS ALIGNED VERTICALLY, a dentistry notation sym-
bol, U+23CE RETURN SYMBOL, U+29FB TRIPLE PLUS, ⊙ U+29BF CIRCLED BULLET,
a box-drawing character, and a circled digit.

Noto Sans Mono CJK HK	不｜☁｜⁂｜⁑｜⏛｜⏎｜⧻｜⊙｜--｜⓪
Source Han Mono K	不｜☁｜⁂｜⁑｜⏛｜⏎｜⧻｜⊙｜--｜⓪