

Combining character

8 languages

[Article](#) [Talk](#)

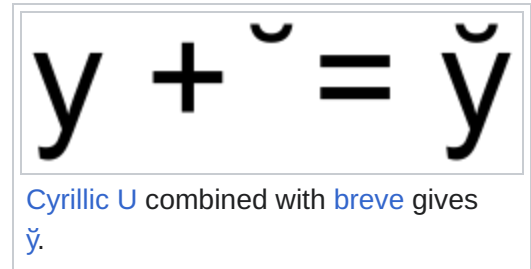
[Read](#) [Edit](#) [View history](#) [Tools](#)

From Wikipedia, the free encyclopedia

Not to be confused with [Spacing Modifier Letters](#).

In [digital typography](#), **combining characters** are [characters](#) that are intended to modify other characters. The most common combining characters in the Latin script are the **combining diacritical marks** (including **combining accents**).

[Unicode](#) also contains many [precomposed characters](#), so that in many cases it is possible to use both combining diacritics and precomposed characters, at the user's or application's choice. This leads to a requirement to perform [Unicode normalization](#) before comparing two Unicode strings and to carefully design encoding converters to correctly map all of the valid ways to represent a character in Unicode to a legacy encoding to avoid data loss.^[1]



This article contains [special characters](#). Without proper [rendering support](#), you may see [question marks](#), [boxes](#), or other [symbols](#).

In Unicode, the main block of combining diacritics for European languages and the [International Phonetic Alphabet](#) is U+0300–U+036F. Combining diacritical marks are also present in many other blocks of Unicode characters. In Unicode, diacritics are always added after the main character (in contrast to some older combining character sets such as [ANSEL](#)), and it is possible to add several diacritics to the same character, including stacked diacritics above and below, though some systems may not render these well.

Unicode ranges [\[edit\]](#)

Main articles: [Combining Diacritical Marks \(Unicode block\)](#), [Combining Diacritical Marks Extended \(Unicode block\)](#), [Combining Diacritical Marks Supplement \(Unicode block\)](#), [Combining Diacritical Marks for Symbols \(Unicode block\)](#), [Combining Half Marks \(Unicode block\)](#), and [Cyrillic Extended-A](#)

The following blocks are dedicated specifically to combining characters:

- Combining Diacritical Marks (0300–036F), since version 1.0, with modifications in subsequent versions down to 4.1
- Combining Diacritical Marks Extended (1AB0–1AFF), version 7.0
- Combining Diacritical Marks Supplement (1DC0–1DFF), versions 4.1 to 5.2
- Combining Diacritical Marks for Symbols (20D0–20FF), since version 1.0, with modifications in subsequent versions down to 5.1

- Cyrillic Extended-A (2DE0–2DFF), version 5.1
- Combining Half Marks (FE20–FE2F), versions 1.0, with modifications in subsequent versions down to 8.0

Combining characters are not limited to these blocks; for instance, the combining [dakuten](#) (U+3099) and combining [handakuten](#) (U+309A) are in the [Hiragana block](#), the [Devanagari](#) block contains combining vowel signs and other marks for use with that script, and so forth. Combining characters are assigned the [Unicode major category](#) "M" ("Mark").

Combining Diacritical Marks (0300–036F) [\[edit\]](#)

| Combining Diacritical Marks ^[1] | | | | | | | | | | | | | | | | |
|--|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| Official Unicode Consortium code chart (PDF) | | | | | | | | | | | | | | | | |
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| U+030x | ◌̐ | ◌̑ | ◌̒ | ◌̓ | ◌̔ | ◌̕ | ◌̖ | ◌̗ | ◌̘ | ◌̙ | ◌̚ | ◌̛ | ◌̜ | ◌̝ | ◌̞ | ◌̟ |
| U+031x | ◌̠ | ◌̡ | ◌̢ | ◌̣ | ◌̤ | ◌̥ | ◌̦ | ◌̧ | ◌̨ | ◌̩ | ◌̪ | ◌̫ | ◌̬ | ◌̭ | ◌̮ | ◌̯ |
| U+032x | ◌̰ | ◌̱ | ◌̲ | ◌̳ | ◌̴ | ◌̵ | ◌̶ | ◌̷ | ◌̸ | ◌̹ | ◌̺ | ◌̻ | ◌̼ | ◌̽ | ◌̾ | ◌̿ |
| U+033x | ◌̿ | ◌̺ | ◌̻ | ◌̼ | ◌̽ | ◌̾ | ◌̿ | ◌̺ | ◌̻ | ◌̼ | ◌̽ | ◌̾ | ◌̿ | ◌̺ | ◌̻ | ◌̼ |
| U+034x | ◌̐ | ◌̑ | ◌̒ | ◌̓ | ◌̔ | ◌̕ | ◌̖ | ◌̗ | ◌̘ | ◌̙ | ◌̚ | ◌̛ | ◌̜ | ◌̝ | ◌̞ | CGJ |
| U+035x | ◌̐ | ◌̑ | ◌̒ | ◌̓ | ◌̔ | ◌̕ | ◌̖ | ◌̗ | ◌̘ | ◌̙ | ◌̚ | ◌̛ | ◌̜ | ◌̝ | ◌̞ | ◌̟ |
| U+036x | ◌̐ | ◌̑ | ◌̒ | ◌̓ | ◌̔ | ◌̕ | ◌̖ | ◌̗ | ◌̘ | ◌̙ | ◌̚ | ◌̛ | ◌̜ | ◌̝ | ◌̞ | ◌̟ |

Notes

1.^ As of Unicode version 17.0

Codepoints U+032A and U+0346–034A are [IPA](#) symbols:

- U+032A ◌̠: [dental](#)
- U+0346 ◌̡: [dentolabial](#)
- U+0347 ◌̢: [alveolar](#)
- U+0348 ◌̣: strong [articulation](#)
- U+0349 ◌̤: weak articulation
- U+034A ◌̥: [denasal](#)

Codepoints U+034B–034E are IPA diacritics for [disordered speech](#):

- U+034B ◌̦: nasal escape
- U+034C ◌̧: velopharyngeal friction
- U+034D ◌̨: labial spreading
- U+034E ◌̩: whistled articulation

U+034F is the "[combining grapheme joiner](#)" (CGJ) and has no visible glyph.

| | | | | | | | | | | | | | | | | |
|--------|----------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| U+20Dx | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ |
| U+20Ex | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ | ◌◻ |
| U+20Fx | ◌ [*] | | | | | | | | | | | | | | | |

Notes

- 1.^ As of Unicode version 17.0
- 2.^ Grey areas indicate non-assigned code points

Combining Half Marks (FE20–FE2F) [edit]

| Combining Half Marks ^[1] | | | | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Official Unicode Consortium code chart (PDF) | | | | | | | | | | | | | | | | |
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| U+FE2x | ◌◌̇ | ◌◌̈ | ◌◌̉ | ◌◌̊ | ◌◌̋ | ◌◌̌ | ◌◌̍ | ◌◌̎ | ◌◌̏ | ◌◌̐ | ◌◌̑ | ◌◌̒ | ◌◌̓ | ◌◌̔ | ◌◌̕ | ◌◌̖ |

Notes

- 1.^ As of Unicode version 17.0

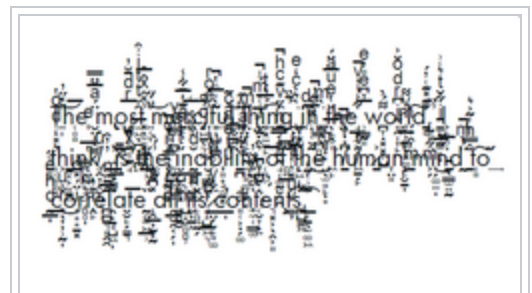
OpenType [edit]

OpenType has the *ccmp* "feature tag" to define glyphs that are compositions or decompositions involving combining characters, the *mark* tag to define the positioning of combining characters onto base glyph, and *mkmk* for the positionings of combining characters onto each other.

Zalgo text [edit]

Main article: [Zalgo text](#)

Combining characters have been used to create **Zalgo text**, which is text that appears "corrupted" or "creepy" due to an overuse of combining characters. This causes the text to extend vertically, overlapping other text.^[2] This is mostly used in horror contexts on the **Internet**. It is typically very challenging for most **software** to render, so the combining marks are often reduced or completely stripped off.



A sentence of Zalgo text

See also [edit]

- [Dotted circle](#)
- [Dead key](#)
- [Spacing Modifier Letters](#) which shouldn't combine (although they do erroneously on some implementations where a developer has confused "combining" with "modifier")

Notes [edit]

- [^] For example, when converting between [windows-1258](#) and [VISCII](#), the former uses combining diacritics whilst the latter has a large selection of precomposed characters so a converter using a simple mapping between code values and Unicode code points will corrupt text when converting between them.
- [^] Korpela, Jukka K. "How does Zalgo text work?" . *Stack Overflow*. Retrieved 11 April 2019.

External links [edit]

- [Combining diacritics chart](#) (in Adobe [PDF](#) format)
- [Combining diacritics supplement chart](#) (in Adobe [PDF](#) format)
- [Combining marks](#) test page facing combined and precomposed letters
- [Alan Wood's Unicode Resources](#)
- [DecodeUnicode.org combining diacritical marks reference](#)

V · T · E

Unicode

[\[show\]](#)

Category: [Unicode special code points](#)

This page was last edited on 30 March 2026, at 14:20 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike 4.0 License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

[Privacy policy](#) [About Wikipedia](#) [Disclaimers](#) [Contact Wikipedia](#) [Legal & safety contacts](#) [Code of Conduct](#) [Developers](#) [Statistics](#)

[Cookie statement](#) [Mobile view](#)

