

# Suggested improvements to the code chart annotations of "C0 Controls and Basic Latin" and "C1 Controls and Latin-1 Supplement"

Eduardo Marín Silva

16/04/2021

**Introduction.** This is the first in a series of documents that make suggestions to improve the code charts annotations of the first two blocks. It improves the completeness and readability of them, as well as remove certain mistakes.

**Additions.** Many punctuation and mathematical characters gained missing cross-references to other related characters in either its glyph or its semantics, as well as many script specific punctuation characters were added for their counterparts, such as the Slavonic asterisk (A673 ☆), as well as Armenian, Ethiopic, Canadian syllabics, Hebrew, Arabic and Kana. While that sounds like a lot, the fact that none of those has a large correspondence with the set in question, keeps the increase in clutter to a minimum. Another option would be to remove ALL (non-letter) script specific analogues, including the ones already there. The current inclusions and exclusions seem arbitrary. For comparison both versions of the charts are included.

With respect to the letters, many confusables or historically related characters were added (plus some modifier letters to the math and punctuation symbols), unless they were part of the basic Russian or Greek alphabet; this included the remaining epigraphic letters and Roman numerals. Many more missing symbols that were part of the "Letterlike Symbols" block, were added as well to their corresponding letter, completing that set so that those exclusions aren't arbitrary anymore.

Whenever a character was related to more than one character in these blocks (for instance 214A ₧ to L and P), it appears in both characters. This is consistent with other references in the current code charts such as the greek question mark (037E ;) , which is included both under the question mark (?) (semantic analogue) and the semicolon (;) (its homoglyph) or the latin letter dental click (01C0 l), appearing both under the solidus (/) and the vertical bar(|). Since these blocks are the most "important" it pays to be redundant.

Some characters received new explanatory bullet notes, but just a few had new informative aliases added: the number sign (#) has "hashtag", "hex" and "primorial", the ampersand (&) has "and", the apostrophe (') has "single quote" the solidus (002F /) has "forward slash" and "diagonal", the pound sign (£) has "lira" (as opposed to the more specific Italian and Turkish lira) and the plus-minus sign (±) has "biquintile" which is an astrological aspect.

**Removals.** The document doesn't have many removals, and of those, most were simply replaced with improved counterparts or consolidated into a single entry. The number sign for instance, had it's informative aliases broken up into different lines in order to better explain with parentheses, the origin of such aliases; so even though it seems we removed all the aliases, they are all still included as well as new ones. This isn't the case, for instance, for the hyphen-minus (002D -), where the alias "hyphen or minus sign" is not really all that useful, since that is already in both its formal name and the bullet note below; furthermore its not a common practice to read that character that way.

For the asterisk, the informative alias "star" was kept since its common practice to read the symbol that way, but the wrong claim that its used on phone keypads was removed. For the full stop (.) an incorrect recommendation practice was replaced with the correct one. For the low line ( \_ ) and the tilde ( ~ ), the unnecessary clarification "this is a spacing character" was removed; leaving it there breaks with consistency and can be confusing. The low line also had a reference to combining macron below (0331 ◌̣) removed, since I argued that they are not truly analogues, but its inclusion is up for debate. And for the superscript 3 (<sup>3</sup>), the reference to superscript 1 (<sup>1</sup>) was removed since it already is present in the previous character superscript 2 (<sup>2</sup>).

**Notation.** In order to indicate an addition a **yellow highlight** is used; in order to indicate deletion a **red highlight** is used, followed by a replacement in yellow when applicable; when minor corrections are applied a **blue highlight** is used.

Additionally the text ((that is formatted like this between double parentheses)), is used by the author to enclose the explanation for different decisions, right next to the change in question; they are not meant to be actually included in the chart, they are just meant for the reviewers of this document.

Finally, including every character in the block, even if many of them don't need modification, isn't very efficient but not indicating omissions can be confusing for the reviewers, so I use a vertical ellipsis ( : ) preceded by two tabs to indicate mayor omissions, and the midline horizontal ellipsis ( … ) preceded by one tab to indicate minor omissions. The header of the characters in question is always included (even when it itself doesn't need a change), for the convenience of the reviewers.

**Possible objection.** I anticipate that the committee will argue that all of these additions are excessive and will make the code charts harder to read by being so large. This argument would apply if that wasn't the case to begin with, but the truth is, it already is; specially for punctuation characters like the quotation mark, with 3 bullet notes spanning 6 lines, plus 6 character references, giving a total of 12 lines. Or considering the apostrophe with 2 informative aliases plus 4 bullet notes spanning 6 lines, plus 7 character references, giving a total of 15 lines.

As a compromise, I have proposed removing certain characters references, whose inclusion seems like a "stretch".

**Charts.** Here is a small table of contents, delimiting the different charts. Do note that there isn't two versions for the second block:

1. pages **3-11**: C0 Controls and Basic Latin (**BROAD VERSION**)
2. pages **12-20**: C0 Controls and Basic Latin (**NARROW VERSION**)
3. pages **21-25**: C1 Controls and Latin-1 Supplement













## SUGGESTED MODIFICATIONS: C0 Controls and Basic Latin (BROAD VERSION)

⋮











### ASCII punctuation and symbols

Based on ISO/IEC 646.











#### 0020 SPACE

- sometimes considered a control code
- other space characters: 2000  – 200A 
- 00A0  no-break space
- 1361  ethiopic wordspace
- 200B  zero width space
- 202F  narrow no-break space
- 2060  word joiner
- 2420  symbol for space
- 2422  blank symbol
- 2423  open box
- 3000  ideographic space
- FEFF  zero width no-break space

#### 0021 ! EXCLAMATION MARK

- = factorial
- = bang
- sometimes used (in suffix position) to indicate negation
- 00A1  inverted exclamation mark
- 00AC  not sign
- 01C3  latin letter retroflex click
- 055C  armenian exclamation mark
- 203C  double exclamation mark
- 203D  interrobang
- 26A0  warning sign
- 2757  heavy exclamation mark symbol
- 2762  heavy exclamation mark ornament
- A71D  modifier letter raised exclamation mark

#### 0022 " QUOTATION MARK

- neutral (vertical), used as opening or closing quotation mark
- preferred characters in English for paired quotation marks are 201C “ & 201D ”
- 05F4 ” is preferred for gershayim when writing Hebrew
- 02BA  modifier letter double prime
- 02DD  double acute accent
- 02EE  modifier letter double apostrophe
- 030B  combining double acute accent *((has less of a relation than the next mark in this list))*
- 030E  combining double vertical line above
- 05F4  hebrew punctuation gershayim
- 201C  left double quotation mark
- 201D  right double quotation mark
- 2033  double prime
- 3003  ditto mark

#### 0023 # NUMBER SIGN

- = pound sign, hash, crosshatch, octothorpe *((broke up informative aliases to be clearer))*
- = pound sign (weight)
- = hashtag, hash
- = octothorpe (original name from Bell Labs; has many variant spellings)

= hex (Singapore and Malaysia)

= crosshatch (fill pattern)

- for denoting "sharp" 266F # is preferred

- also used for denoting (in postfix position) the primorial operation or the cardinality of a set

→ 2114 ℔ l b bar symbol

→ 2216 № numero sign

→ 22D5 ∥ equal and parallel to

→ 2317 ▣ viewdata square

→ 266F # music sharp sign

→ 29E3 # equals sign and slanted parallel

## 0024 \$ DOLLAR SIGN

= milréis, escudo

- used for many peso currencies in Latin America and elsewhere

- often used as a generic currency sign, instead of 00A4 Ⱶ

- other currency symbol characters start at 20A0 ₧

→ 00A2 ¢ cent sign

→ 00A4 Ⱶ currency sign

→ 20B1 ₱ peso sign

→ 1F10F ₵ circled dollar sign with overlaid backslash

→ 1F4B0 💵 money bag

→ 1F4B2 💵 heavy dollar sign

→ 1F4B5 💵 banknote with dollar sign

## 0026 & AMPERSAND

= and

- historically derived from a ligature of the letters 'e' and 't' when they were used for the word "et" which is latin for "and"

- sometimes used as an operator for denoting the 'logical AND'

→ 204A 7 tironian sign et

→ 214B 8 turned ampersand

→ 2227 ^ logical and

→ 1F674 & heavy ampersand ornament

## 0027 ' APOSTROPHE

= apostrophe-quote (1.0)

= single quote

= APL quote

- neutral (vertical) glyph with mixed usage

- 2019 ’ is preferred for apostrophe

- serif fonts may use this character with a glyph like 2019 ’

- preferred characters in English for paired quotation marks are 2018 ‘ & 2019 ’

- 05F3 ’ is preferred for gersh when writing Hebrew

→ 02B9 ’ modifier letter prime

→ 02BC ’ modifier letter apostrophe

→ 02C8 ’ modifier letter vertical line

→ 0301 ◌ combining acute accent

→ 030D ◌ combining vertical line above *((replaces the previous character with a more similar one))*

→ 05F3 ’ hebrew punctuation gersh

→ 2018 ‘ left single quotation mark

→ 2019 ’ right single quotation mark

→ 2032 ’ prime *((removed to avoid confusion))*

→ A78C ' latin small letter saltillo

## 002A \* ASTERISK

= star **(on phone keypads)** *((this is incorrect, this character is not used in phone keypads, or when transcribing phone numbers in most contexts, but rather a midline glyph like 2217 \*, or an enlarged one like 2731 \* or 1F7B6 \* is used. The character however is normally read as "star" so the alias is still correct))*

- can have five or six spokes

→ 066D \* arabic five pointed star

→ 2020 † dagger

→ 2042 \*\* asterism

→ 204E \* low asterisk

→ 2051 \* two asterisks aligned vertically

→ 20F0 ⚡ combining asterisk above

→ 2217 \* asterisk operator

→ 26B9 ✕ sextile

→ 2731 \* heavy asterisk

→ A673 ☆ slavonic asterisk

→ 1F7B6 \* medium six spoked asterisk

## 002B + PLUS SIGN

- some cultures prefer a '+' glyph due to religious sensibilities

- sometimes used for the 'logical AND'

→ 02D6 ⋅ modifier letter plus sign

→ 2064 ± invisible plus

→ 2212 − minus sign

→ 2795 ⚡ heavy plus sign

→ FB29 ׀ hebrew letter alternative plus sign

→ 1F7A2 ⚡ light greek cross

## 002C , COMMA

= decimal separator, thousands separator

- some locales switch the functions of the comma and the full stop in their number formatting, and others prefer to use a narrow space instead or even the apostrophe

→ 055D ֊ armenian comma

→ 060C ٫ arabic comma

→ 066C ٫ arabic thousands separator

→ 1363 ፡ ethiopic comma

→ 201A ٫ single low-9 quotation mark

→ 2063 ٫ invisible separator

→ 275F ٫ heavy low single comma quotation mark ornament

→ 2E34 ٫ raised comma

→ 2E41 ٫ reversed comma

→ 2E4C ٫ medieval comma

→ 3001 ٫ ideographic comma

## 002D - HYPHEN-MINUS

= hyphen or minus sign *((not a useful informative alias, broken up below))*

= hyphen, dash

= minus sign

- used for either hyphen or minus sign

- its function is broad, so it can be used as either the punctuation hyphen or as a minus sign operator *((clearer than the previous version above))*

→ 00AD ̣ soft hyphen

→ 02D7 - modifier letter minus sign

- 2010 - hyphen
- 2011 ̄ non-breaking hyphen
- 2012 – figure dash
- 2013 – en dash
- 2027 · hyphenation point
- 2043 ▪ hyphen bullet
- 2212 – minus sign
- 10191 – roman uncia sign

## 002E . FULL STOP

= period, dot, decimal point

• may be rendered as a raised decimal point in old style numbers *((this should not be an encouraged practice, rather the practice described below))*

• in some contexts (like old style number formatting), a raised decimal point is preferred, for those cases 00B7 · should be used

- 00B7 · middle dot
- 0589 : armenian full stop
- 066B , arabic decimal separator
- 06D4 . arabic full stop
- 1362 :: ethiopic full stop
- 16EE x canadian syllabics full stop
- 2024 . one dot leader
- 2026 ... horizontal ellipsis
- 2E33 · raised dot
- 2E3C x stenographic full stop
- 3002 。 ideographic full stop

## 002F / SOLIDUS

= slash, forward slash, virgule

= diagonal (Spanish)

- 01C0 | latin letter dental click
- 0338 ø combining long solidus overlay
- 2044 / fraction slash
- 2215 / division slash
- 27CB ⁄ mathematical rising diagonal
- 29F8 ⁄ big solidus
- 2AFD // double solidus operator
- 2E4A ⁄ dotted solidus

⋮

## ASCII punctuation and symbols

### 003A : COLON

- also used to denote division or scale; for that mathematical use 2236 : is preferred (except in applications that require ASCII characters)
- in Finnish and Swedish, the colon can be used as the equivalent of an apostrophe, without spaces or word breaks
- 02D0 : modifier letter triangular colon
- 02F8 : modifier letter raised colon
- 0589 : armenian full stop
- 05C3 : hebrew punctuation sof pasuq
- 1361 : ethiopic wordspace
- 1365 ÷ ethiopic colon
- 205A : two dot punctuation

- 205D : tricolon
- 2236 : ratio
- A789 : modifier letter colon
- FE30 : presentation form for vertical two dot leader

### 003B ; SEMICOLON

- this, and not 037E ;, is the preferred character for 'Greek question mark'
- 037E ; greek question mark
- 061B ؛ arabic semicolon
- 1364 ፤ ethiopic semicolon
- 204F Ⴗ reversed semicolon
- 2E35 Ⴟ turned semicolon

### 003C < LESS-THAN SIGN

- 02C2 < modifier letter left arrowhead
- 2039 ‹ single left-pointing angle quotation mark
- 2264 ≤ less-than or equal to
- 2329 ‹ left-pointing angle bracket ((this character is deprecated))
- 27E8 ‹ mathematical left angle bracket
- 3008 ‹ left angle bracket

### 003D = EQUALS SIGN

- other related characters: 2241 ≈ – 2263 ≡
- 1400 = canadian syllabics hyphen
- 2248 ≈ almost equal to
- 2260 ≠ not equal to
- 2261 ≡ identical to
- 2E40 = double hyphen
- 30A0 = katakana-hiragana double hyphen
- A78A = modifier letter short equals sign
- 10190 = roman sextans sign

### 003E > GREATER-THAN SIGN

- 02C3 > modifier letter left arrowhead
- 203A › single right-pointing angle quotation mark
- 2265 ≥ greater-than or equal to
- 232A › right-pointing angle bracket ((this character is deprecated))
- 27E9 › mathematical right angle bracket
- 3009 › right angle bracket

### 003F ? QUESTION MARK

- 00BF ¿ inverted question mark
- 037E ; greek question mark
- 055E Ԑ armenian question mark
- 061F ؟ arabic question mark
- 1367 ፤ ethiopic question mark
- 203D ʔ interrobang
- 2047 ?? double question mark
- 2048 ?! question exclamation mark ((replace with the double question mark))
- 2049 !? exclamation question mark ((ditto))
- 2753 ？ black question mark ornament
- 2BD1 𐤒 uncertainty sign
- 2E2E ʇ reversed question mark
- FFFD 𐀀 replacement character

...

## Uppercase Latin alphabet

...

### 0043 C LATIN CAPITAL LETTER C

- 2102 ¢ double-struck capital c
- 2103 °C degree celsius
- 2104 ₠ centre line symbol
- 212D Ć black-letter capital c
- 216D Ċ roman numeral one hundred
- 2183 Ɔ roman numeral reversed one hundred

### 0044 D LATIN CAPITAL LETTER D

- 2145 Ⓓ double-struck italic capital d
- 216E D̄ roman numeral five hundred

...

### 0046 F LATIN CAPITAL LETTER F

- 2109 °F degree fahrenheit
- 2131 ₣ script capital f
- 2132 ƒ turned capital f
- A7FB ƒ̄ latin epigraphic letter reversed f

...

### 0049 I LATIN CAPITAL LETTER I

- Turkish and Azerbaijani use 0131 ı for lowercase
- 0130 İ latin capital letter i with dot above
- 01C0 Ꝁ latin letter dental click
- 0399 Ι greek capital letter iota
- 0406 І cyrillic capital letter byelorussian-ukrainian i
- 04C0 І cyrillic letter palochka
- 2110 ℑ script capital i
- 2111 Ĩ black-letter capital i
- 2160 I roman numeral one
- A7F7 Ꝁ̄ latin epigraphic letter sideways i
- A7FE Ī latin epigraphic letter i longa

### 004A J LATIN CAPITAL LETTER J

- 037F Ј greek capital letter yot

...

### 004C L LATIN CAPITAL LETTER L

- 2104 ₠ centre line symbol
- 2112 ℒ script capital l
- 214A ₧ property line
- 216C L̄ roman numeral fifty

### 004D M LATIN CAPITAL LETTER M

- 216F M̄ roman numeral one thousand
- A7FD Ⅿ latin epigraphic letter inverted m
- A7FF 𐌚 latin epigraphic letter archaic m

### 004E N LATIN CAPITAL LETTER N

- 2115 ₧ double-struck capital n
- 2216 № numero sign

...



## 0050 P LATIN CAPITAL LETTER P

- 2119 ₧ double-struck capital p
- 214A ₧ property line
- A7FC Ɔ latin epigraphic letter reversed p

## 0051 Q LATIN CAPITAL LETTER Q

- 051A Q cyrillic capital letter qa
- 211A Ɔ double-struck capital q
- 213A Ɔ rotated capital q

## 0052 R LATIN CAPITAL LETTER R

- 211B ℞ script capital r
- 211C ʀ black-letter capital r
- 211D ℝ double-struck capital r
- 211E ℞ prescription take
- 211F ʀ response

...

## 0056 V LATIN CAPITAL LETTER V

- 2123 ʒ versicle
- 2164 V roman numeral five

## 0057 W LATIN CAPITAL LETTER W

- 051C W cyrillic capital letter we

## 0058 X LATIN CAPITAL LETTER X

- 2169 X roman numeral ten

## 0059 Y LATIN CAPITAL LETTER Y

- 00A5 ¥ yen sign
- 04AE Y cyrillic capital letter straight u

## 005A Z LATIN CAPITAL LETTER Z

- 01B5 Z latin capital letter z with stroke
- 2124 Z double-struck capital z
- 2128 Ʒ black-letter capital z

## ASCII punctuation and symbols

...

## 005C \ REVERSE SOLIDUS

- = backslash
- 20E5 Ɔ combining reverse solidus overlay
- 2216 \ set minus
- 27CD \ mathematical falling diagonal
- 29F9 \ big reverse solidus

...

## 005E ^ CIRCUMFLEX ACCENT

- this is a spacing character
- often used to indicate exponentiation when superscript characters aren't available

...

## 005F \_ LOW LINE

- = spacing underscore (1.0)
- this is a spacing character *((this is already obvious; removed to avoid confusion))*
- joins with adjacent instances of itself
- not considered a decomposable character
- 02CD \_ modifier letter low macron
- 0331 ̲ combining macron below *((not really an analogue, since it doesn't join with adjacent instances of itself))*

- 0332 ̲ combining low line
- 0640 ۰ arabic tatweel
- 2017 ͡ double low line
- 203E ͤ overline
- FE2D ͆ combining conjoining macron below

...

## Lowercase Latin alphabet

...

### 0064 d LATIN SMALL LETTER D

- 0501 d cyrillic small letter komi de
- 2146 d double-struck italic small d

### 0065 e LATIN SMALL LETTER E

- 212E e estimated symbol
- 212F e script small e
- 2147 e double-struck italic small e

...

### 0069 i LATIN SMALL LETTER I

- Turkish and Azerbaijani use 0130 İ for uppercase
- 0131 ı latin small letter dotless i
- 2071 <sup>i</sup> superscript latin small letter i
- 2139 i information source
- 2148 i double-struck italic small i
- 1D6A4 i mathematical italic small dotless i *((this reference is inconsistent with the rest of the annotations on the letters))*

### 006A j LATIN SMALL LETTER J

- 0237 j latin small letter dotless j
- 03F3 j greek letter yot
- 2149 j double-struck italic small j
- 1D6A5 j mathematical italic small dotless j *((ditto))*

...

### 006C l LATIN SMALL LETTER L

- 01C0 l latin letter dental click
- 1231 l cyrillic small letter palochka
- 2113 l script small l
- 1D4C1 l mathematical script small l *((ditto))*

...

### 0070 p LATIN SMALL LETTER P

- 2118 p script capital p
- 214C p per sign

### 0071 q LATIN SMALL LETTER Q

- 051B q cyrillic small letter qa

...

### 0073 s LATIN SMALL LETTER S

- 017F ſ latin small letter long s

...

### 0077 w LATIN SMALL LETTER W

- 051D w cyrillic small letter we

### 0078 x LATIN SMALL LETTER X

- 00D7 × multiplication sign

...

## ASCII punctuation and symbols

...

### 007C | VERTICAL LINE

= vertical bar, pipe

- used in pairs to indicate the absolute value of a term or expression, the norm of a function, the determinant of a matrix or the cardinality of a set
- can also be used as an unpaired separator or fence
- for cases where a relation or an operator is intended (such as the 'bitwise OR' or the 'negated AND') 2223 | is preferred (except in applications that require ASCII characters)

→ 00A6 | broken bar

→ 01C0 | latin letter dental click

→ 05C0 | hebrew punctuation paseq

→ 0964 | devanagari danda

→ 2016 || double vertical line

→ 20D2 ϕ combining long vertical line overlay

→ 2223 | divides

→ 23D0 | vertical line extension

→ 2502 | box drawings light vertical

→ 2758 | light vertical bar

...

### 007E ~ TILDE

- this is a spacing character *((this is already obvious; removed to avoid confusion))*

- sometimes used (in suffix position) to indicate negation

→ 00AC ¬ not sign

→ 02DC ~ small tilde

→ 0303 ¯ combining tilde

→ 0334 ⅈ combining tilde overlay

→ 2053 ~ swung dash

→ 223C ~ tilde operator

→ 2E1B ≈ tilde with ring above

→ 2E2F ' vertical tilde

→ FF5E ∼ fullwidth tilde

## SUGGESTED MODIFICATIONS: C0 Controls and Basic Latin (NARROW VERSION)

### ASCII punctuation and symbols

Based on ISO/IEC 646.

#### 0020 **SPACE**

- sometimes considered a control code
- other space characters: 2000 **EQ** – 200A **LS**
- 00A0 **uhs** no-break space
- 200B **zws** zero width space
- 202F **unbs** narrow no-break space
- 2060 **wj** word joiner
- 2420 **sp** symbol for space
- 2422 **b** blank symbol
- 2423 **□** open box
- 3000 **IS** ideographic space *((removed for being script specific))*
- FEFF **BOM** zero width no-break space

#### 0021 **! EXCLAMATION MARK**

- = factorial
- = bang
- sometimes used (in suffix position) to indicate negation
- 00A1 **¡** inverted exclamation mark
- 00AC **¬** not sign
- 01C3 **!** latin letter retroflex click
- 203C **!!** double exclamation mark
- 203D **?** interrobang
- 26A0 **⚠** warning sign
- 2757 **!** heavy exclamation mark symbol
- 2762 **⚡** heavy exclamation mark ornament
- A71D **!** modifier letter raised exclamation mark

#### 0022 **" QUOTATION MARK**

- neutral (vertical), used as opening or closing quotation mark
- preferred characters in English for paired quotation marks are 201C “ & 201D ”
- 05F4 ” is preferred for gershayim when writing Hebrew
- 02BA **”** modifier letter double prime
- 02DD **”** double acute accent
- 02EE **”** modifier letter double apostrophe
- 030B **”** combining double acute accent *((has less of a relation than the next mark in this list))*
- 030E **”** combining double vertical line above
- 05F4 **”** hebrew punctuation gershayim *((removed for being script specific))*
- 201C **“** left double quotation mark
- 201D **”** right double quotation mark
- 2033 **”** double prime
- 3003 **”** ditto mark

#### 0023 **# NUMBER SIGN**

- = pound sign, hash, crosshatch, octothorpe *((broke up informative aliases to be clearer))*
- = pound sign (weight)
- = hashtag, hash
- = octothorpe (original name from Bell Labs) (has many variant spellings)
- = hex (Singapore and Malaysia)
- = crosshatch (fill pattern)

- for denoting "sharp" 266F # is preferred
- also used for denoting (in postfix position) the primorial operation or the cardinality of a set
- 2114 ℔ lb bar symbol
- 2216 № numero sign
- 22D5 ∥ equal and parallel to
- 2317 ⇈ viewdata square
- 266F # music sharp sign
- 29E3 # equals sign and slanted parallel

#### 0024 \$ DOLLAR SIGN

- = milréis, escudo
- used for many peso currencies in Latin America and elsewhere
- often used as a generic currency sign, instead of 00A4 ₭
- other currency symbol characters start at 20A0 ₧
- 00A2 ¢ cent sign
- 00A4 ₭ currency sign
- 20B1 ₱ peso sign
- 1F10F ₧ circled dollar sign with overlaid backslash
- 1F4B0 💰 money bag
- 1F4B2 \$ heavy dollar sign
- 1F4B5 💵 banknote with dollar sign

#### 0026 & AMPERSAND

- = and
- historically derived from a ligature of the letters 'e' and 't' when they were used for the word "et" which is latin for "and"
- sometimes used as an operator for denoting the 'logical AND'
- 204A 7 tironian sign et
- 214B 8 turned ampersand
- 2227 ^ logical and
- 1F674 & heavy ampersand ornament

#### 0027 ' APOSTROPHE

- = apostrophe-quote (1.0)
- = single quote
- = APL quote
- neutral (vertical) glyph with mixed usage
- 2019 ’ is preferred for apostrophe
- serif fonts may use this character with a glyph like 2019 ’
- preferred characters in English for paired quotation marks are 2018 ‘ & 2019 ’
- 05F3 ׳ is preferred for geresh when writing Hebrew
- 02B9 ’ modifier letter prime
- 02BC ’ modifier letter apostrophe
- 02C8 ’ modifier letter vertical line
- 0301 ◌ combining acute accent
- 030D ◌ combining vertical line above *((replaces the previous character with a more similar one))*
- 05F3 ׳ hebrew punctuation geresh
- 2018 ‘ left single quotation mark
- 2019 ’ right single quotation mark
- 2032 ’ prime *((removed to avoid confusion))*
- A78C ' latin small letter saltillo

## 002A \* ASTERISK

= star **(on phone keypads)** *((this is incorrect, this character is not used in phone keypads, or when transcribing phone numbers in most contexts, but rather a midline glyph like 2217 \*, or an enlarged one like 2731 \* or 1F7B6 \* is used. The character however is normally read as "star" so the alias is still correct))*

- can have five or six spokes

→ 066D \* arabic five pointed star

→ 2020 † dagger

→ 2042 \*\* asterism

→ 204E \* low asterisk

→ 2051 \* two asterisks aligned vertically

→ 20F0 ⌘ combining asterisk above

→ 2217 \* asterisk operator

→ 26B9 ✕ sextile

→ 2731 \* heavy asterisk

→ 1F7B6 \* medium six spoked asterisk

## 002B + PLUS SIGN

- some cultures prefer a 'ⴰ' glyph due to religious sensibilities

- sometimes used for the 'logical AND'

→ 02D6 ⋅ modifier letter plus sign

→ 2064 ± invisible plus

→ 2212 − minus sign

→ 2795 ⴰ heavy plus sign

→ 1F7A2 ⴰ light greek cross

## 002C , COMMA

= decimal separator, thousands separator

- some locales switch the functions of the comma and the full stop in their number formatting, and others prefer to use a narrow space instead or even the apostrophe

→ 060C , arabic comma *((removed for being script specific))*

→ 201A , single low-9 quotation mark

→ 2063 \_ invisible separator

→ 275F , heavy low single comma quotation mark ornament

→ 2E34 , raised comma

→ 2E41 , reversed comma

→ 2E4C ɀ medieval comma

→ 3001 ٫ ideographic comma *((removed for being script specific))*

## 002D - HYPHEN-MINUS

= hyphen or minus sign *((not a useful informative alias, broken up below))*

= hyphen, dash

= minus sign

- used for either hyphen or minus sign

- its function is broad, so it can be used as either the punctuation hyphen or as a minus sign operator *((clearer than the previous version above))*

→ 00AD ̣ soft hyphen

→ 02D7 - modifier letter minus sign

→ 2010 - hyphen

→ 2011 ̮ non-breaking hyphen

→ 2012 – figure dash

→ 2013 – en dash

→ 2027 · hyphenation point

→ 2043 • hyphen bullet

- 2212 – minus sign
- 10191 – roman uncia sign

## 002E . FULL STOP

= period, dot, decimal point

• may be rendered as a raised decimal point in old style numbers ((this should not be an encouraged practice, rather the practice described below))

• in some contexts (like old style number formatting), a raised decimal point is preferred, for those cases 00B7 · should be used

→ 00B7 · middle dot

→ 06D4 . arabic full stop ((removed for being script specific))

→ 2024 . one dot leader

→ 2026 ... horizontal ellipsis

→ 2E33 · raised dot

→ 2E3C . stenographic full stop

→ 3002 . ideographic full stop ((ditto))

## 002F / SOLIDUS

= slash, forward slash, virgule

= diagonal (Spanish)

→ 01C0 | latin letter dental click

→ 0338 ø combining long solidus overlay

→ 2044 / fraction slash

→ 2215 / division slash

→ 27CB / mathematical rising diagonal

→ 29F8 / big solidus

→ 2AFD // double solidus operator

→ 2E4A / dotted solidus

⋮

## ASCII punctuation and symbols

### 003A : COLON

• also used to denote division or scale; for that mathematical use 2236 : is preferred (except in applications that require ASCII characters)

• in Finnish and Swedish, the colon can be used as the equivalent of an apostrophe, without spaces or word breaks

→ 02D0 : modifier letter triangular colon

→ 02F8 : modifier letter raised colon

→ 0589 : armenian full stop ((removed for being script specific))

→ 05C3 : hebrew punctuation sof pasuk ((ditto))

→ 205A : two dot punctuation

→ 205D : tricolon

→ 2236 : ratio

→ A789 : modifier letter colon

→ FE30 : presentation form for vertical two dot leader

### 003B ; SEMICOLON

• this, and not 037E ;, is the preferred character for 'Greek question mark'

→ 037E ; greek question mark

→ 061B ; arabic semicolon ((removed for being script specific))

→ 204F ; reversed semicolon

→ 2E35 : turned semicolon

### 003C < LESS-THAN SIGN

- 02C2 < modifier letter left arrowhead
- 2039 ‹ single left-pointing angle quotation mark
- 2264 ≤ less-than or equal to
- 2329 ‹ left-pointing angle bracket ((this character is deprecated))
- 27E8 ‹ mathematical left angle bracket
- 3008 ‹ left angle bracket

### 003D = EQUALS SIGN

- other related characters: 2241 ≈ – 2263 ≡
- 2248 ≈ almost equal to
- 2260 ≠ not equal to
- 2261 ≡ identical to
- 2E40 = double hyphen
- A78A = modifier letter short equals sign
- 10190 = roman sextans sign

### 003E > GREATER-THAN SIGN

- 02C3 > modifier letter left arrowhead
- 203A › single right-pointing angle quotation mark
- 2265 ≥ greater-than or equal to
- 232A › right-pointing angle bracket ((this character is deprecated))
- 27E9 › mathematical right angle bracket
- 3009 › right angle bracket

### 003F ? QUESTION MARK

- 00BF ¿ inverted question mark
- 037E ; greek question mark
- 061F ؟ arabic question mark
- 203D ? interrobang
- 2047 ?? double question mark
- 2048 ?! question exclamation mark ((replace with the double question mark))
- 2049 !? exclamation question mark ((ditto))
- 2753 ? black question mark ornament
- 2BD1 ⚡ uncertainty sign
- 2E2E ؟ reversed question mark
- FFFD ⬮ replacement character

...

## Uppercase Latin alphabet

...

### 0043 C LATIN CAPITAL LETTER C

- 2102 ℄ double-struck capital c
- 2103 °C degree celsius
- 2104 ₠ centre line symbol
- 212D Ć black-letter capital c
- 216D Ċ roman numeral one hundred
- 2183 ↀ roman numeral reversed one hundred

### 0044 D LATIN CAPITAL LETTER D

- 2145 Ⓓ double-struck italic capital d
- 216E ↀ roman numeral five hundred

...



#### 0046 F LATIN CAPITAL LETTER F

- 2109 °F degree fahrenheit
- 2131 ℱ script capital f
- 2132 ƒ turned capital f
- A7FB Ɔ latin epigraphic letter reversed f

...

#### 0049 I LATIN CAPITAL LETTER I

- Turkish and Azerbaijani use 0131 ı for lowercase
- 0130 İ latin capital letter i with dot above
- 01C0 Ĭ latin letter dental click
- 0399 Ι greek capital letter iota
- 0406 І cyrillic capital letter byelorussian-ukrainian i
- 04C0 І cyrillic letter palochka
- 2110 ℑ script capital i
- 2111 ɇ black-letter capital i
- 2160 I roman numeral one
- A7F7 ⅃ latin epigraphic letter sideways i
- A7FE ⅃ latin epigraphic letter i longa

#### 004A J LATIN CAPITAL LETTER J

- 037F Ј greek capital letter yot

...

#### 004C L LATIN CAPITAL LETTER L

- 2104 ₧ centre line symbol
- 2112 ℒ script capital l
- 214A ₧ property line
- 216C Ⅴ roman numeral fifty

#### 004D M LATIN CAPITAL LETTER M

- 216F Ⅰ roman numeral one thousand
- A7FD ⅃ latin epigraphic letter inverted m
- A7FF ⅃ latin epigraphic letter archaic m

#### 004E N LATIN CAPITAL LETTER N

- 2115 Ⅱ double-struck capital n
- 2216 № numero sign

...

#### 0050 P LATIN CAPITAL LETTER P

- 2119 Ⅱ double-struck capital p
- 214A ₧ property line
- A7FC Ɔ latin epigraphic letter reversed p

#### 0051 Q LATIN CAPITAL LETTER Q

- 051A Q cyrillic capital letter qa
- 211A Ⅱ double-struck capital q
- 213A Ⅱ rotated capital q

#### 0052 R LATIN CAPITAL LETTER R

- 211B ℞ script capital r
- 211C Ƀ black-letter capital r
- 211D Ⅱ double-struck capital r
- 211E Ⅱ prescription take
- 211F Ⅱ response

...

## 0056 V LATIN CAPITAL LETTER V

→ 2123  $\mathfrak{V}$  versicle

→ 2164 V roman numeral five

## 0057 W LATIN CAPITAL LETTER W

→ 051C W cyrillic capital letter we

## 0058 X LATIN CAPITAL LETTER X

→ 2169 X roman numeral ten

## 0059 Y LATIN CAPITAL LETTER Y

→ 00A5 ¥ yen sign

→ 04AE Y cyrillic capital letter straight u

## 005A Z LATIN CAPITAL LETTER Z

→ 01B5 Z latin capital letter z with stroke

→ 2124 Z double-struck capital z

→ 2128 3 black-letter capital z

## ASCII punctuation and symbols

...

## 005C \ REVERSE SOLIDUS

= backslash

→ 20E5 ¤ combining reverse solidus overlay

→ 2216 \ set minus

→ 27CD ↘ mathematical falling diagonal

→ 29F9 ↵ big reverse solidus

...

## 005E ^ CIRCUMFLEX ACCENT

- this is a spacing character

- often used to indicate exponentiation when superscript characters aren't available

...

## 005F \_ LOW LINE

= spacing underscore (1.0)

- this is a spacing character *((this is already obvious; removed to avoid confusion))*

- joins with adjacent instances of itself

- not considered a decomposable character

→ 02CD ¯ modifier letter low macron

→ 0331 ◌◌ combining macron below *((not really an analogue, since it doesn't join with adjacent instances of itself))*

→ 0332 ◌◌ combining low line

→ 2017 = double low line

→ 203E ¯ overline

→ FE2D ◌◌ combining conjoining macron below

...

## Lowercase Latin alphabet

...

## 0064 d LATIN SMALL LETTER D

→ 0501 d cyrillic small letter komi de

→ 2146 d double-struck italic small d

## 0065 e LATIN SMALL LETTER E

→ 212E e estimated symbol

→ 212F e script small e

→ 2147 e double-struck italic small e

...

## 0069 i LATIN SMALL LETTER I

- Turkish and Azerbaijani use 0130 İ for uppercase
- 0131 ĭ latin small letter dotless i
- 2071 <sup>ı</sup> superscript latin small letter i
- 2139 ĩ information source
- 2148 *ı* double-struck italic small i
- 1D6A4 *ı* mathematical italic small dotless i ((this reference is inconsistent with the rest of the annotations on the letters))

## 006A j LATIN SMALL LETTER J

- 0237 j latin small letter dotless j
- 03F3 j greek letter yot
- 2149 *j* double-struck italic small j
- 1D6A5 *j* mathematical italic small dotless j ((ditto))

...

## 006C l LATIN SMALL LETTER L

- 01C0 l latin letter dental click
- 1231 l cyrillic small letter palochka
- 2113 *ℓ* script small l
- 1D4C1 *ℓ* mathematical script small l ((ditto))

...

## 0070 p LATIN SMALL LETTER P

- 2118 *ⵑ* script capital p
- 214C *ꝑ* per sign

## 0071 q LATIN SMALL LETTER Q

- 051B q cyrillic small letter qa

...

## 0073 s LATIN SMALL LETTER S

- 017F ſ latin small letter long s

...

## 0077 w LATIN SMALL LETTER W

- 051D w cyrillic small letter we

## 0078 x LATIN SMALL LETTER X

- 00D7 × multiplication sign

...

## ASCII punctuation and symbols

...

## 007C | VERTICAL LINE

- = vertical bar, pipe
- used in pairs to indicate the absolute value of a term or expression, the norm of a function, the determinant of a matrix or the cardinality of a set
- can also be used as an unpaired separator or fence
- for cases where a relation or an operator is intended (such as the 'bitwise OR' or the 'negated AND') 2223 | is preferred (except in applications that require ASCII characters)
- 00A6 ¦ broken bar
- 01C0 l latin letter dental click
- 05C0 | hebrew punctuation paseq ((removed for being script specific))
- 2016 || double vertical line
- 20D2 ⋮ combining long vertical line overlay
- 2223 | divides
- 23D0 | vertical line extension

→ 2502 | box drawings light vertical

→ 2758 | light vertical bar

...

#### 007E ~ TILDE

- this is a spacing character ((this is already obvious; removed to avoid confusion))

- sometimes used (in suffix position) to indicate negation

→ 00AC ¬ not sign

→ 02DC ~ small tilde

→ 0303 ¨ combining tilde

→ 0334 ⌘ combining tilde overlay

→ 2053 ~ swung dash

→ 223C ~ tilde operator

→ 2E1B ¸ tilde with ring above

→ 2E2F ˇ vertical tilde

→ FF5E ∼ fullwidth tilde









## SUGGESTED MODIFICATIONS: C1 Controls and Latin-1 Supplement

⋮



### Latin-1 punctuation and symbols

Based on ISO/IEC 8859-1 (aka Latin-1) from here.





#### 00A0 NO-BREAK SPACE

- commonly abbreviated as 
- 0020  space
- 2007  figure space
- 202F  narrow no-break space
- 2060  word joiner
- 237D  shouldered open box
- FEFF  zero width no-break space
- ≈ <noBreak> 0020 









#### 00A1 INVERTED EXCLAMATION MARK

- Spanish, Asturian, Galician
- 0021 ! exclamation mark
- A71E  modifier letter raised inverted exclamation mark
- A71F  modifier letter low inverted exclamation mark



#### 00A2 CENT SIGN

- 0024 \$ dollar sign
- 023C  latin small letter c with stroke
- 20A5  mill sign
- 20B5  cedi sign
- 1D135  musical symbol cut time








#### 00A3 POUND SIGN

- = pound sterling, irish punt,  lira,  Italian lira,  Turkish lira, etc.
- may have one or two crossbars, but the glyph with a single crossbar is preferred to distinguish it from 20A4 
- 20A4  lira sign
- 20BA  turkish lira sign
- 10192  roman semuncia sign
- 1F4B7  banknote with pound sign

#### 00A4 CURRENCY SIGN

- intended to be a generic currency sign, or as a fallback when the currency in question is unknown
- other currency symbol characters: 20A0  – 20BF 
- 0024 \$ dollar sign

#### 00A5 YEN SIGN

- = yuan sign (renmimbi)
- Japan, People's Republic of China
- glyph may have one or two crossbars, but the official symbol in Japan and the PRC  has two (as of 2020, it isn't used in Hong Kong, Macau, Singapore or the ROK)
- the sign is formally used today, by the PCR and Japan in international contexts. However in their native contexts, the use of a single ideograph to act as the currency sign may be preferred
- exactly what ideograph is used depends on the exact context, but the most common are 5143  for the yuan and 5186  for the yen
- 5143 
- 5186 
- 1F4B4  banknote with yen sign
- 1F4B9  chart with upwards trend and yen sign

## 00A6 | **BROKEN BAR**

= broken vertical bar (1.0)

= parted rule (in typography)

- originally intended to be an allograph of 007C |, in order to avoid confusability in certain contexts, but disunified to follow legacy practice

→ 007C | vertical line

## 00A7 § **SECTION SIGN**

- paragraph sign in some European usage

- sometimes used as a paragraph separator in Europe *((clearer than the version above and in line with other similar annotations))*

→ 2E39 § top half section sign

...

## 00AA <sup>a</sup> **FEMININE ORDINAL INDICATOR**

- Spanish

→ 1D43 <sup>a</sup> modifier letter small a

≈ <super> 0061 a

...

## 00AC ¬ **NOT SIGN**

= angled dash (in typography)

→ 0021 ! exclamation mark

→ 007E ~ tilde

→ 0305 ¯ combining overline

→ 2310 ¬ reversed not sign

→ 2319 ⇄ turned not sign

→ FE41 ⌊ presentation form for vertical left corner bracket

## 00AD sh **SOFT HYPHEN**

= discretionary hyphen

- commonly abbreviated as |SHY|

- this character allows for hyphenation; i.e. it indicates a mid-word line breaking opportunity, that only upon breaking may show a visible hyphen (the exact way is dependent on the locale)

→ 2010 - hyphen

## 00AE ® **REGISTERED SIGN**

= registered trade mark sign (1.0)

→ 24C7 ® circled latin capital letter r

→ 2122 ™ trade mark sign

→ 1F16B <sup>MD</sup> raised md sign

→ 1F16C <sup>MR</sup> raised mr sign

## 00AF ¯ **MACRON**

= overline, APL overbar

- this is a spacing character

- unlike 203E ¯, repeated instances of this character do not join with each other

→ 02C9 ¯ modifier letter macron

→ 0304 ¯ combining macron

→ 0305 ¯ combining overline *((not really an analogue, since it joins with adjacent instances of itself))*

→ 203E ¯ overline

→ FE26 ¯ combining conjoining macron

≈ 0020 <sup>SP</sup> 0304 ¯

## 00B0 ° DEGREE SIGN

- this is a spacing character *((this is already obvious; removed to avoid confusion))*

→ 02DA ° ring above

→ 030A ° combining ring above

→ 1D52 ° modifier letter small o

→ 2070 ° superscript zero *((shape is never circular))*

→ 2103 °C degree celsius

→ 2109 °F degree fahrenheit

→ 2218 ° ring operator *((not related historically or semantically))*

## 00B1 ± PLUS-MINUS SIGN

= biquintile (astrological aspect; 144°)

→ 2213 ∓ minus-or-plus sign

## 00B2 ² SUPERSCRIFT TWO

= squared

- other superscript digit characters: 2070 ⁰ –2079 ⁹

→ 00B9 ¹ superscript one

≈ <super> 0032 ²

## 00B3 ³ SUPERSCRIFT THREE

= cubed

→ 00B9 ¹ superscript one *((redundant given the reference of the character above))*

≈ <super> 0033 ³

## 00B4 ´ ACUTE ACCENT

- this is a spacing character

→ 02B9 ´ modifier letter prime

→ 02CA ´ modifier letter acute accent

→ 0301 ´ combining acute accent

→ 0384 ´ greek tonos

→ 1FFD ´ greek oxia

→ 2032 ´ prime

≈ 0020 [SP] 0301 ´

...

## 00B9 ¹ SUPERSCRIFT ONE

→ 00B2 ² superscript two

→ 00B3 ³ superscript three

→ 2070 ⁰ superscript zero

≈ <super> 0031 ¹

## 00B7 · MIDDLE DOT

= midpoint, interpunct, centered dot (in typography) *((added extra aliases and removed unnecessary clarification))*

= Georgian comma

= Greek middle dot semicolon (ano teleia) *((modified informative alias and italicized the non English term))*

- also used as a raised decimal point or to denote multiplication; for multiplication 22C5 · is preferred

- for when an operator (like multiplication) is intended, 22C5 · is preferred (except for applications that restrict the admissible code-points to ASCII and Latin-1)

→ 002E . full stop

→ 02D9 ´ dot above

→ 0387 ´ greek ano teleia

→ 16EB · runic single punctuation *((including this would imply including other ancient dot like characters, and the list is long already))*

→ 2022 • bullet *((not related as much as the other characters))*

→ 2024 . one dot leader

→ 2027 • hyphenation point

→ 2219 • bullet operator *((this is related to the bullet, not the middle dot))*

→ 22C5 • dot operator

→ 2E31 • word separator middle dot

→ 2E33 • raised dot

→ 30FB • katakana middle dot

→ A78F • latin letter sinological dot

## 00BA ° MASCULINE ORDINAL INDICATOR

• Spanish

→ 1D52 ° modifier letter small o

→ 2216 № numero sign

≈ <super> 006F o

⋮

## Letters

...

## 00D0 Ð LATIN CAPITAL LETTER ETH

• lowercase is 00F0 ð

→ 0110 Ð latin capital letter d with stroke

→ 0189 Ð latin capital letter african d

→ A7C7 Ð latin capital letter d with short stroke overlay

...

## Mathematical operator

### 00D7 × MULTIPLICATION SIGN

= z notation Cartesian product

→ 0078 x latin small letter x

→ 2062 ∞ invisible times

→ 22C5 • dot operator

→ 2613 X saltire

→ 2715 X multiplication x

→ 274C ✕ cross mark

→ 2A2F × vector or cross product

→ 1F5D9 ✕ cancellation x

→ 1F7A9 × light saltire

## Letters

...

## 00DE Þ LATIN CAPITAL LETTER THORN

→ 03F7 Þ greek capital letter sho

→ 16A6 Þ runic letter thurisaz thurs thorn

## 00DF ß LATIN SMALL LETTER SHARP S

= Eszett

• German (except for Swiss High German)

• not used in Swiss High German *((consolidated in the note above))*

• uppercase is "SS" (formal case mapping) or 1E9E ß in certain contexts

• typographically the glyph for this character can be based on a ligature of 017F ſ with either 0073 s or with an old-style glyph for 007A z (the latter similar in appearance to 0292 ȝ). Both forms exist interchangeably today.

→ 017F ſ latin small letter long s

→ 0292 ȝ latin small letter ezh



- 03B2 β greek small letter beta
- 1E9E ß latin capital letter sharp s

...

#### 00F0 ð LATIN SMALL LETTER ETH

- Icelandic, Faroese, Old English, IPA
- uppercase is 00D0 Ð
- 0111 đ latin small letter d with stroke
- 03B4 δ greek small letter delta
- 2202 ∂ partial differential

...

#### Mathematical operator

##### 00F7 ÷ DIVISION SIGN

- = obelus
- occasionally used as an alternate, more visually distinct version of 2212 – or 2011 ⁂ in some contexts
- historically used as a punctuation mark to denote questionable passages in manuscripts
- 070B ܐ syriac harklean obelus
- 2011 ⁂ non-breaking hyphen
- 2052 % commercial minus sign
- 2212 – minus sign
- 2215 / division slash
- 2223 | divides
- 2236 : ratio
- 2797 ÷ heavy division sign
- 27CC ⁝ long division
- 2E13 ⁄ dotted obelos

#### Letters

...

#### 00FE þ LATIN SMALL LETTER THORN

- Icelandic, Old English, Old Norse and various phonetic notations
- Runic letter borrowed into the Latin script
- used for variations of dental affricates (/th/ in modern English)
- historically, incorrectly transcribed with 0079 y in the USA
- 03F8 ϣ greek small letter sho
- 16A6 𐝒 runic letter thurisaz thurs thorn