



Junicode

the font for medievalists



*specimens and user manual*

for version 2





# Contents

1	About Junicode	1
2	Specimens	3
3	Getting Started with Junicode	8
4	Feature Reference	12
4.1	Introduction . . . . .	12
4.2	Case-Related Features . . . . .	15
4.2.1	<code>smcp</code> – Small Capitals . . . . .	15
4.2.2	<code>c2sc</code> – Small Capitals from Capitals . . . . .	15
4.2.3	<code>pcap</code> – Petite Capitals . . . . .	15
4.2.4	<code>case</code> – Case-Sensitive Forms . . . . .	16
4.3	Numbers and Sequencing . . . . .	16
4.3.1	<code>frac</code> – Fractions . . . . .	16
4.3.2	<code>numr</code> – Numerators . . . . .	16
4.3.3	<code>dnom</code> – Denominators . . . . .	16
4.3.4	<code>nalt</code> – Alternate Annotation Forms . . . . .	16
4.3.5	<code>tnum</code> – Tabular Figures . . . . .	17
4.3.6	<code>onum</code> – Oldstyle Figures . . . . .	17
4.3.7	<code>pnum</code> – Proportional Figures . . . . .	17
4.3.8	<code>lnum</code> – Lining Figures . . . . .	17
4.3.9	<code>zero</code> – Slashed Zero . . . . .	17
4.4	Superscripts and Subscripts . . . . .	18
4.4.1	<code>sups</code> – Superscripts . . . . .	18

4.4.2	<b>subs</b> – Subscripts	18
4.5	Ornaments	18
4.5.1	<b>ornm</b> – Ornaments	18
4.6	Alphabetic Variants	19
4.6.1	<b>cv01–cv52</b> – Basic Latin Variants	19
4.6.2	<b>cv53–cv66</b> – Other Latin Letters	21
4.6.3	<b>ss01</b> – Alternate thorn and eth	22
4.6.4	<b>ss02</b> – Insular Letter-Forms	22
4.6.5	<b>ss04</b> – High Overline	23
4.6.6	<b>ss05</b> – Medium-High Overline	23
4.6.7	<b>ss06</b> – Enlarged Minuscles	23
4.6.8	<b>ss07</b> – Underdotted Text	23
4.6.9	<b>ss08</b> – Contextual Long s	23
4.6.10	<b>ss16</b> – Contextual r Rotunda	24
4.6.11	<b>cv68</b> – Variant of ? (U+0294, glottal stop)	24
4.7	Greek	24
4.7.1	<b>ss03</b> – Alternate Greek	25
4.8	Punctuation	25
4.8.1	<b>ss18</b> – Old-Style Punctuation Spacing	26
4.8.2	<b>cv69</b> – Variants of ꝛꝛ (U+204A / U+2E52, Tironian nota)	26
4.8.3	<b>cv70</b> – Variants of . (period)	26
4.8.4	<b>cv71</b> – Variant of · (U+00B7, middle dot)	26
4.8.5	<b>cv72</b> – Variants of , (comma)	26
4.8.6	<b>cv73</b> – Variants of ; (semicolon)	26
4.8.7	<b>cv74</b> – Variants of ⁙ (U+2E4E, <i>punctus elevatus</i> )	26
4.8.8	<b>cv75</b> – Variant of ! (exclamation mark)	26
4.8.9	<b>cv76</b> – Variants of ? (question mark)	27
4.8.10	<b>cv77</b> – Variant of ~ (ASCII tilde)	27
4.8.11	<b>cv78</b> – Variant of * (asterisk)	27
4.8.12	<b>cv79</b> – Variants of / (slash)	27
4.9	Abbreviations	27
4.9.1	<b>cv80</b> – Variant of ꝛ (U+A75D, rum abbreviation)	27
4.9.2	<b>cv81</b> – Variants of Ϳ (U+035B, combining zigzag above)	27
4.9.3	<b>cv82</b> – Variants of spacing ʹ (U+A770)	28

4.9.4	cv83 – Variants of 3 (U+A76B, “et” abbreviation)	28
4.10	Combining Marks	28
4.10.1	cv84 – MUFI combining marks (variants of U+0304)	28
4.10.2	cv67 – Spacing zigzag (variant of U+00AF, spacing macron)	29
4.10.3	ss10 – Character Entities for Combining Marks	29
4.10.4	ss20 – Low Diacritics	30
4.10.5	cv85 – Variant of � (U+1DD3, combining open a)	31
4.10.6	cv86 – Variant of � (U+1DD8, combining insular d)	31
4.10.7	cv87 – Variant of � (U+1DE3, combining r rotunda)	31
4.10.8	cv88 – Variant of combining dieresis (U+0308)	31
4.10.9	cv89 – Variant of � (U+0305, combining overline)	31
4.10.10	cv90 – Variants of �� (U+035E, combining double macron)	31
4.10.11	cv92 – Variant of breve below (U+032E)	31
4.11	Currency and Weights	31
4.11.1	cv93 – Variants of � (U+0044, generic currency sign)	31
4.11.2	cv94 – Variant of �� (U+2114)	32
4.11.3	cv95 – Variants of � (U+00A3, British pound sign)	32
4.11.4	cv96 – Variant of � (U+20B0, German penny sign)	32
4.11.5	cv97 – Variant of � (U+0192, florin)	32
4.11.6	cv98 – Variant of � (U+2125, Ounce sign)	32
4.12	Gothic	33
4.12.1	ss19 – Latin to Gothic Transliteration	33
4.13	Runic	33
4.13.1	ss12 – Early English Futhorc	33
4.13.2	ss13 – Elder Futhark	33
4.13.3	ss14 – Younger Futhark	33
4.13.4	ss15 – Long Branch to Short Twig	33
4.13.5	rtlm – Right to Left Mirrored Forms	34
4.14	Ligatures and Digraphs	34
4.14.1	hlig – Historic Ligatures	34
4.14.2	dlig – Discretionary Ligatures	35
4.14.3	ss17 – Rare Digraphs	35
4.15	Required Features	35

5	Non-MUFI Code Points	36
6	Entering characters with tags	37
7	Transcribing records	55
7.1	A preliminary note on transcription . . . . .	55
7.2	1. Common combining marks . . . . .	57
7.3	2. Spacing characters . . . . .	59
7.4	3. Other formatting . . . . .	60
7.5	4. On the web . . . . .	61
8	The Enlarge Axis	63
9	Junicode on the Web	65
9.1	Subsetting Junicode . . . . .	65
9.2	Junicode and CSS/HTML . . . . .	68

## 1. About Junicode

Junicode is modeled on the Pica Roman type purchased by Oxford University in 1692 and used to set the bulk of the Latin text of George Hickes, *Linguarum vett. septentrionalium thesaurus grammatico-criticus et archaeologicus* (Oxford, 1703–5). This massive two-volume folio is not only a major work of scholarship on the languages and literatures of northern Europe in the Middle Ages, but also a fine example of the work of the Oxford Press at this period: printed in multiple types (for every language had to have its proper type) and lavishly illustrated with engravings of manuscript pages, coins and artifacts.

Junicode also includes two other typefaces from the *Thesaurus*: Pica Saxon, used to set passages in the Old English language, and also a typeface reproducing the Gothic alphabet (“Gothic” here being not the late medieval style, but rather the earliest extensively attested Germanic language). These were commissioned by the literary scholar Franciscus Junius (1591–1677) and bequeathed by him to the University. Examples of all these typefaces can be found in *A Specimen of the Several Sorts of Letter Given to the University by Dr. John Fell, Sometime Lord Bishop of Oxford. To Which Is Added the Letter Given by Mr. F. Junius* (Oxford, 1693).

Junicode has two distinct Greek faces. The first, newly designed to

harmonize with the roman face, is upright and modern. The other, accompanying the italic face, is based on type designed by Alexander Wilson (1714–86) of Glasgow and used in numerous books published by the Foulis Press, most notably the great Glasgow Homer of 1756–58.

The Junicode project began around 1998, when the developer began to revise his older (early 1990s) “Junius” fonts for medievalists to take account of the Unicode standard, then relatively new. The font’s name, a contraction of “Junius Unicode,” was supposed to be a stopgap, serving until a more suitable name could be found, but the name “Junicode” is now so well known that it can’t be changed.<sup>1</sup> The project has been active for its entire history, responding to frequent requests from users and changes in font technology; the developer is now pushing towards the release of Junicode version 2, an extended font family of five weights and five widths, with both static and variable versions, which aims to promote best practices in the presentation of medieval texts, especially in the area of accessibility. This aspect of the font is explored in the Introduction to the Feature Reference.

*Users should be aware that Junicode 2 is a beta version in the roman face, complete but unstable and subject to change. In the italic it is an alpha, about ninety-five percent complete (as of March 2022) and more unstable than the roman. Until the official release of version 2 (probably in late 2022), versions of Junicode numbered 1.00N should be used for projects that require stability. The latest stable version can be downloaded [here](#).*

---

<sup>1</sup> An effort to change the name to “JuniusX” produced only confusion. If you find a font by the name JuniusX on a free font site, that is nothing more than an early version of Junicode 2.



## 2. Specimens

### Old and Middle English

Wē æthrynon mid ūrum ārum þā yðan þæs dēopan wāles; wē ġesāwon ēac þā muntas ymbe þære sealtan sǣ strande, and wē mid āðēnedum hrægle and ġesundfullum windum þær ġewicodon on þām ġemǣrum þære fæġerestān þēode. Þā yðan ġetācniað þisne dēopan cræft, and þā muntas ġetācniað ēac þā micelnyssa þisses cræftes. (Regular)

Siþen þe sege and þe assault watz sesed at Troye,  
þe borȝ brittēd and brent to brondez and askez,  
þe tulk þat þe trammes of tresoun þer wroȝt  
Watz tried for his tricherie, þe trewest on erthe:  
Hit watz Ennias þe athel, and his highe kynde,  
þat siþen deprecd prouinces, and patrounes bicomē  
Welneȝe of al þe wele in þe west iles. (SemiExpanded)

*Apply the OpenType feature ss02 (Stylistic Set 2) for insular letter-forms.*

Her cýnepulſ benam riȝebryht hſ rices ȝ peſſreaxna riotan ƿop unſryhtum deðū buton  
hamtúnſcipe ȝ he hæfde þa oþ he ofſlog þone aldormon þe hi lengeſt punode ȝ hiene þa  
cýnepulſ on andƿeð aðræfde ȝ h þær punaðe oþ þæt hine án ƿpán ofſtanȝ æt ƿrýſeteſflo-  
ðan ȝ he ƿpēc þone aldormon cumbpan ȝ ȝe cýnepulſ ofſt miclum ġeſeohtum ƿeaht uuþ  
bƿetpalū. (SemiCondensed)

### Old Irish

Fect n-oen do Ailill ȝ do Meidb iar n-dergud a rígleptha dóib i Cruachanráith Chonnacht, arrecaim

## 4 SPECIMENS

comrad chind-cherchailli eturru. Fírbriathar, a ingen, bar Ailill, is maith ben ben dagfir. Maith omm, bar ind ingen. Cid diatá latsu ón. Is de atá lim, bar Ailill, ar it ferr-su indiu indá in lá thucus-sa thu. (Condensed Medium)

*For insular letter-forms, apply the OpenType feature ss02 (Stylistic Set 2), making sure the language is set to Irish.*

ḃamaíth-je pemut, ap Meḃb. Íj maíth nach cualammap ʒ nach ʒetammap, ap Ailill, acht ḃo bithriu ap bantincup mnaa ʒ bḃba na cḃich ba neʒrom ḃuit oc bḃe-íth ḃo ʒlaít ʒ ḃo chpech ı ʒúatach úait. Nı ʒamlaıḃ báʒa, ap Meḃb, acht m'athair ı n-apḃrıʒı hḂpenn .ı. Eocho ʒeıblech mac ʒıḃ meic ʒıḃomain meic ʒıḃeoin meic ʒıḃʒunı meic Roʒein Rúaiḃ meic Rıʒéoin meic ḃlathachta meic ḃeothachta meic Ḃna Aʒnıʒ meic Oengurʒa Ḃurḃıʒ. ḃátar aice ʒe ingena ḃ'ingenaıḃ: Ḃeḃriu, Ḃthı ʒ Éle, Clothru, Muʒain, Meḃb, meʒı ba uarlıu ʒ ba upraıtu ḃıḃ. (Regular)

### Old Icelandic

*For Nordic shapes of þ and ð in an English context, specify the appropriate language (e.g. Icelandic or Norwegian); or apply the OpenType ss01 (Stylistic Set 1) feature.*

Um haustit sendi Mörðr Valgarðsson orð at Gunnarr myndi vera einn heimi, en lið alt myndi vera niðri í eyjum at lúka heyverkum. Riðu þeir Gizurr Hvíti ok Geirr Goði austr yfir ár, þegar þeir spurðu þat, ok austr yfir sanda til Hofs. Þá sendu þeir orð Starkaði undir Þríhyrningi; ok fundusk þeir þar allir er at Gunnari skyldu fara, ok réðu hversu at skyldi fara. (SemiExpanded Medium)

### Runic

*JunICODE has features for automated transliteration of Latin letters into various runic systems.*

ƿıɥk ƿıƿḄn ƿḄƿ ƿı ƿmrxmıḃmrix ƿƿrḃ xƿ:ʒrik xrrḃı ıƿr Ḅm ƿı xrmnı  
xıʒƿḄm Ḅrḃıƿ Ḃı  
rḃmƿƿııı ƿıḄ rminmƿƿıııı ıƿḂxmı xıbrḃıƿr ƿƿḂḄḄ Ḅıƿ ƿḄıƿ ƿı rḃmḃ  
kḃʒıı: ƿııı ııııı (Expanded)

### German

Ich sag üch aber / minen fründen / Fözchtēd üch nit vo2 denen die den lyb tōdend / vnd darnach nichts

habennd das fy mer thügind. Ich wil üch aber zeigē vo2 welchem ir üch fōzchten follend. Fōzchtend  
 üch vo2 dem / der / nach dem er tōdet hat / ouch macht hat zewerffen inn die hell: ja ich sag üch /  
 vo2 dem selben fōzchtēd üch. Koufft man nit fünf Sparen v̄m zween pfennig? (Condensed)

## Latin

*Junicond contains the most common Latin abbreviations, making it suitable for diplomatic editions of Latin texts.*

Adiuuanos dñ falutaris noster & pp̄i gl̄am nominis tui dn̄e libanos & ppitiuſ esto  
peccatis nostris pp̄ter nomen tuum. Ne forte dicant ingentib: ubi est dñ eorum &  
innotescat in nationib: corā oculis nr̄is. Posuerunt morticina feruorū ruorū efcaſ uolatilib:  
cēli carnes scōz̄ tuoꝝ bestiis tenice. Facti sum⁹ obpbrium uicinis nr̄is. (Light)

# Gothic

jabai auk hoas gasailviþ þuk þana habandan kunþi in galiuge stada anakumbjandan, niu miþwis-  
sei is siukis wisandins timrjada du galiugagudam gasaliþ matjan? fraqistniþ auk sa unmahteiga  
ana þeinamma witubnja broþar in þize Xristus gaswalt. swaþ þan frawaurkjandans wiþra  
broþrums, slahandans ize gahugd siuka, du Xristau frawaurkeiþ. (SemiCondensed Light)

*Use ss19 to produce Gothic letters automatically from transliterated text.*

ԳԼԵՒԻ ԼՈՒՔ ՕՒՏ ԴՏՏՈՒՄ ՓՈՒՔ ՓՆԱՂ ԿԼԵՒՆԴԱՆ ԿՈՆՓԻ ԻՆ ԴԼԻՆԴԵ ՏՂԺԺ  
 ԼՈՒԿՈՄԵԾԱՆԴԱՆ, ՆԻՆ ՄԻՓՎԻՏՏԵԻ ԻՏ ՏԻՆԿԻՏ ՎԻՏԱՆԴՈՒՆ ՏԻՄԿԵԾԺԺ ԺՆ ԴԼԻՆ-  
 ԴԼՆԴԱՄ ԴՏՏՈՒՄ ՄԺԴԵՆ? ԳԼԵՒԻ ԼՈՒՔ ՕՒՏ ԴՏՏՈՒՄ ՓՈՒՔ ՓՆԱՂ ԿԼԵՒՆԴԱՆ  
 ԿՈՆՓԻ ԻՆ ԴԼԻՆԴԵ ՏՂԺԺ ԼՈՒԿՈՄԵԾԱՆԴԱՆ, ՆԻՆ ՄԻՓՎԻՏՏԵԻ ԻՏ ՏԻՆԿԻՏ ՎԻՏԱՆ-  
 ԴՈՒՆ ՏԻՄԿԵԾԺԺ ԺՆ ԴԼԻՆԴԼՆԴԱՄ ԴՏՏՈՒՄ ՄԺԴԵՆ? (SemiExpanded Bold)

### Sanskrit Transliteration

mānam dvividham viṣayadvai vidyāśaktyaśaktiṭaḥ  
arthakriyāyām keśadirnārtho 'narthādhimoksataḥ

sadr̥ṣāsadr̥ṣatvācca viṣayāviṣayatvataḥ  
śabdasyānyanimittānām bhāve dhisadasattvataḥ (SemiCondensed Medium)

## International Phonetic Alphabet

hwan θat a:prɪl wɪθ ɪs ʃu:rəs so:tə θə drʊxt ɔf mɑrtʃ hɑθ pe:rsəd to: θə ro:te and  
ba:ðəd ɛvri væɪn ɪn swɪtʃ lɪku:r ɔf hwɪtʃ vɜrtɪu ɛndʒɛndrəd ɪs θə flu:r hwan zɛfɪrʊs e:k  
wɪθ hɪs swe:tə bræ:θ (Regular)

## Greek

βίβλος γενέσεως ἰησοῦ χριστοῦ υἱοῦ δαυίδ υἱοῦ ἀβραάμ. ἀβραάμ ἐγέννησεν τὸν ἰσαάκ,  
ἰσαάκ δὲ ἐγέννησεν τὸν ἱακώβ, ἱακώβ δὲ ἐγέννησεν τὸν ἰούδαν καὶ τοὺς ἀδελφοὺς αὐτοῦ,  
ἰούδας δὲ ἐγέννησεν τὸν φάρες καὶ τὸν ζάρα ἐκ τῆς θαμάρ, φάρες δὲ ἐγέννησεν τὸν  
ἔσρώμ, ἔσρώμ δὲ ἐγέννησεν τὸν ἄράμ, ἄράμ δὲ ἐγέννησεν τὸν ἀμιναδάβ, ἀμιναδάβ δὲ  
ἐγέννησεν τὸν ναασσών, ναασσών δὲ ἐγέννησεν τὸν σαλμών, σαλμών δὲ ἐγέννησεν τὸν  
βόες ἐκ τῆς ῥαχά (Regular)

βίβλος γενέσεως ἰησοῦ χριστοῦ υἱοῦ δαυίδ υἱοῦ ἀβραάμ. ἀβραάμ ἐγέννησεν τὸν ἰσαάκ,  
ἰσαάκ δὲ ἐγέννησεν τὸν ἱακώβ, ἱακώβ δὲ ἐγέννησεν τὸν ἰούδαν καὶ τοὺς ἀδελφοὺς αὐτοῦ,  
ἰούδας δὲ ἐγέννησεν τὸν φάρες καὶ τὸν ζάρα ἐκ τῆς θαμάρ, φάρες δὲ ἐγέννησεν τὸν ἔσρώμ,  
ἔσρώμ δὲ ἐγέννησεν τὸν ἄράμ, ἄράμ δὲ ἐγέννησεν τὸν ἀμιναδάβ, ἀμιναδάβ δὲ ἐγέννησεν  
τὸν ναασσών, ναασσών δὲ ἐγέννησεν τὸν σαλμών, σαλμών δὲ ἐγέννησεν τὸν βόες ἐκ τῆς  
ῥαχά (Italic)

## Lithuanian

*Lithuanian poses several typographical challenges. Make sure Contextual Alternates (calt) is turned on; for  
į, use i followed by combining dot accent (U+0307) and acute (U+0301).*

Visa žemė turėjo vieną kalbą ir tuos pačius žodžius. Kai žmonės kėlėsi  
iš rytų, jie rado slėnį Šinaro krašte ir ten įsikūrė. Vieni kitiems sakė:  
Eime, pasidirkime plytų ir jas išdekime. – Vietoj akmens jie naudojo  
plytas, o vietoj kalkių – bitumą. Eime, – jie sakė, – pasistatykime mi-  
estą ir bokštą su dangų siekiančia viršūne ir pasidarykime sau vardą, kad  
nebūtume išblaškyti po visą žemės veidą. (Expanded)

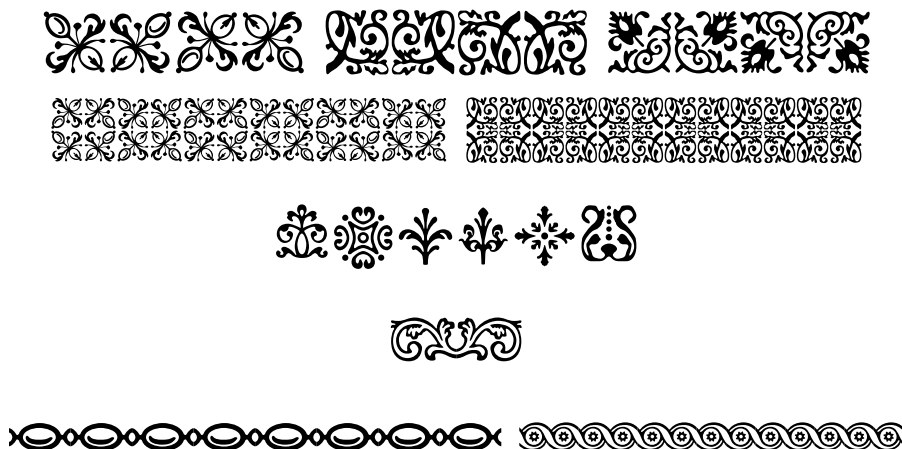
## Polish

*The default shape and position of ogonek in Junicon are suitable for modern Polish. For the medieval Latin e-caudata, consider using cv62.*

Mieszkańcy całej ziemi mieli jedną mowę, czyli jednakowe słowa. A gdy wędrowali ze wschodu, napotkali równinę w kraju Szinear i tam zamieszkali. I mówili jeden do drugiego: Chodźcie, wyrabiamy cegłę i wypalmy ją w ogniu. A gdy już mieli cegłę zamiast kamieni i smołę zamiast zaprawy murarskiej, rzekli: Chodźcie, zbudujemy sobie miasto i wieżę, której wierzchołek będzie sięgał nieba, i w ten sposób uczynimy sobie znak, abyśmy się nie rozproszyli po całej ziemi. (Condensed Medium)

## Fleurons

*Junicon contains a number of fleurons (floral ornaments) copied from a 1785 Caslon specimen book. Access these via the OpenType feature [ornm](#). Fleurons have only one weight and width, and they are the same in roman and italic.*



### 3. Getting Started with Junicode

When installing Junicode on your system, you must choose the kind of font that will work best for you and which faces you are most likely to use. You can choose between TrueType and Compact Font Format (or CFF), and between static and variable.<sup>1</sup> Most users will prefer TrueType fonts, as these will look best on computer screens, and static fonts, as these will work best with programs like Microsoft Word. You can find static TrueType fonts in the fonts/ttf folder.

Use the CFF version (fonts/otf) if you run into problems with TrueType (for example, a few Mac printer drivers mess up the placement of diacritics when rendering these fonts). Use the TrueType variable version if you're using Junicode in a web page, and the CFF variable version if you are using Lua $\TeX$  (see the document `LuaTeX_JunicodeVF.pdf` for instructions and pointers).

Junicode has five weights and five widths, which are combined in many ways for a total of twenty-two styles in both roman and italic. It is not necessary to install all of these; in fact, your life will be simplified (font menus easier to navigate) if you make a selection. You will probably want the traditional Regular, Bold, Italic, and Bold Italic fonts, but you should survey the styles displayed in the Specimen section of this booklet, choose the ones that look best to you, and install only those. A reasonable selection for many users will include the traditional four styles for main text, several

---

<sup>1</sup> TrueType is a font format developed by Apple Computer in the 1980s; it is still the most commonly used format. Compact Font Format (or CFF, often inaccurately called “OpenType”) was developed by Adobe Systems. Both TrueType and CFF fonts are capable of fully supporting the OpenType standard, developed in the late 1990s by Microsoft and Apple, which dramatically expanded the capabilities of computer fonts and continues to evolve. Static fonts are the ones users are most familiar with, each font file having a single set of outlines scalable to any size. By contrast, a single variable font file stores a set of outlines that can morph in various ways—for example, becoming bolder or lighter, narrower or wider.

SemiExpanded styles for footnotes, and SemiCondensed for titles.

With more than 4,700 characters, Junicode is a large font. Finding the things you want in a collection that size can be a challenge, and then entering them in your documents is another challenge. This document will help, but it presupposes a certain amount of knowledge—for example, how to install a font in Windows, Mac OS or Linux and how to install and use different kinds of software.

Medievalists will find the [\*MUFI Character Recommendation\*](#), version 4.0 (2015) an essential supplement to this document. The *Recommendation* lists thousands of characters identified by the Medieval Unicode Font Initiative as being of interest to medievalists. Junicode contains all of these characters. There are two versions of the *Recommendation*: you will probably find the “Alphabetical Order” version most helpful.

From the *MUFI Character Recommendation* or, alternatively, the *Junicode List of Encoded Characters* (packaged with the font) you can find out the **code points**<sup>2</sup> of the characters you need. These code points can be used to enter characters in your documents when they cannot be typed on the keyboard.

To enter any Unicode character in a Windows application, type its four-digit code, followed by Alt-X. To do the same in the Mac OS, first install and switch to the “Unicode Hex Input” keyboard, then type the code while holding down the Option key. In most Linux distributions you can enter a code by typing Shift-Control-U, then the code followed by Return or Enter.

**Combining marks** (diacritics and certain abbreviation signs) can pose special problems for medievalists. Unicode contains a great many **precomposed characters** consisting of a base letter plus one or more marks. If these are all you need you’re fortunate—especially if they can be typed on an international keyboard (not all can).<sup>3</sup> But medieval manuscripts frequently contain combinations of base + mark that are not used in modern written languages. For these, you’ll have to enter bases and marks separately.

To position a mark correctly over a base character, first enter the base, followed

---

<sup>2</sup> A Unicode code point is a numerical identifier for a character. It is generally expressed as a four-digit hexadecimal (or base-16) number with a prefix of “U+”. The letter capital “A,” for example, is U+0041 (65 in decimal notation), and lowercase “3” (Middle English yogh) is U+021D.

<sup>3</sup> Both Windows and the Mac OS come with international keyboards that make it easy to type special letters and diacritics. To find out how to enable these, search online for “Mac OS International Keyboard” or “Windows International Keyboard.”

by the mark or marks. The sequence **m** + ◌̈ (U+1DD9) will make **m̈**; **y** + ◌̄ (U+0304) + ◌̇ (U+0306) will make **ÿ**; **e** + ◌̇ (U+0323) + ◌̈ (U+1DE0) will make **ĕ**.

More than sixteen hundred characters in Junicode can only be accessed via OpenType features—that is, by way of the programming built into the font—and many others *should* be accessed that way for reasons explained in the Introduction to the Feature Reference section of this document.

For example, some programs (including Microsoft Word) produce small caps by scaling capitals down to approximately the height of lowercase letters. These always look too thin and light. But Junicode contains hundreds of TRUE SMALL CAPS designed to harmonize with the surrounding text. These can only be accessed via the OpenType **smcp** feature, which you can apply to a run of text much as you apply italic or bold styles: select some text and then apply the feature.

Unfortunately, not every program supports OpenType features, and some that do either support only a few or make them difficult to access. Programs that support Junicode’s features fully include the free word processor **LibreOffice Writer**, all major browsers (Firefox, Chrome, Safari and Edge), and the typesetting programs **LuaL<sup>A</sup>T<sub>E</sub>X** and **X<sub>L</sub>L<sup>A</sup>T<sub>E</sub>X**. Adobe InDesign supports OpenType features only partially, but provides access to all of Junicode’s characters through its own interface.

Microsoft Word, unfortunately, provides only limited support for OpenType features. It supports the **Required Features** discussed below, and also variant number forms and Stylistic Sets (though only one at a time). Many characters (for example, TRUE SMALL CAPS and those accessible only via Character Variant features) cannot be accessed at all. To activate Word’s OpenType support, you must open the “Font” dialog, click over to the “Advanced” tab, and check the “Kerning” box. (Oddly, the “Kerning” box enables all other OpenType features.) Then, in the same tab, select Standard Ligatures, Contextual Alternates, and any other features you want. OpenType features are best applied to character styles rather than directly to text: this will save you from having to perform this operation repeatedly.

It is also good to set the language properly for the text you’re working on. Programs like Word will automatically set the language to the default for your system. If you change to a language other than your own for a passage (or even a single word), you should set the language for that passage appropriately. This will unlock a number of capabilities. For example, in Old and Middle English, Word and other programs will use the English form of thorn and eth (þð) instead of the modern Icelandic (þð),



and in ancient Greek you will be able to type accents after vowels instead of looking up the codes for hundreds of polytonic vowel + accent combinations. But these and other capabilities are only available when you set the language correctly.

In LibreOffice and InDesign you can set the language with a drop-down menu in the “Character” dialog. In Word there is a separate “Language” dialog, accessible from the “Tools” menu.

If you have questions about any aspect of Junicode, post a query in the [Junicode discussion forum](#). If you notice a bug, please open an [issue](#) at the font’s [development site](#). If you need help with programming, subsetting or other tasks, contact the developer directly.<sup>4</sup>

---

<sup>4</sup> b dot tarde at gmail dot com.

## 4. Feature Reference

### 4.1. Introduction

The OpenType features of Junicode version 2 and its variable counterpart (hereafter referred to together as “Junicode”) have two purposes. One is to provide convenient access to the rich character set of the Medieval Unicode Font Initiative (MUFI) recommendation. The other is to enable best practices in the presentation of medieval text, promoting accessibility in electronic texts from PDFs to e-books to web pages.

Each character in the MUFI recommendation has a code point associated with it: either the one assigned by Unicode or, where the character is not recognized by Unicode, in the Private Use Area (PUA) of the Basic Multilingual Plane, a block of codes, running from U+E000 to U+F8FF, that are assigned no value by Unicode but instead are available for font designers to use in any way they please.

The problem with PUA code points is precisely their lack of any value. Consider, as a point of comparison, the letter **a** (U+0061). Your computer, your phone, and probably a good many other devices around the house store a good bit of information about this **a**: that it’s a letter in the Latin script, that it’s lowercase, and that the uppercase equivalent is **A** (U+0041). All this information is available to word processors, browsers, and other applications running on your computer.

Now suppose you’re preparing an electronic text containing what MUFI calls LATIN SMALL LETTER NECKLESS A (**a**). It is assigned to code point U+F215, which belongs to the PUA. Beyond that, your computer knows nothing about it: not that it is a variant of **a**, or that it is lowercase, or a letter in the Latin alphabet, or even a character in a language system. A screen reader cannot read, or even spell out, a word with U+F215 in it; a search engine will not recognize the word as containing the letter **a**.

Junicode offers the full range of MUFI characters—you can enter the PUA code points—but also a solution to the problems posed by those code points. Think of an electronic text (a web page, perhaps, or a PDF) as having two layers: an underlying text, stable and unchanging, and the displayed text, generated by software at the instant it is needed and discarded when it is no longer on the screen. For greatest accessibility the underlying text should contain the plain letter **a** (U+0061) along with markup indicating how it should be displayed. To generate the displayed text, a program called a “layout engine” will (simplifying a bit here) read the markup and apply the OpenType feature `cv02[5]`<sup>1</sup> to the underlying **a**, bypassing the PUA code point, with the result that readers see **ᵃ**—the “neckless a.” And yet the letter will still register as **a** with search engines, screen readers, and so on.

This is the Junicode model for text display, but it is not peculiar to Junicode: it is widely considered to be the best practice for displaying text using current font technology.

The full range of OpenType features listed in this document is supported by all major web browsers, LibreOffice, XeTeX, LuaTeX, and (presumably) other document processing applications. All characters listed here are available in Adobe InDesign, though that program supports only a selection of OpenType features. Microsoft Word, unfortunately, supports only Stylistic Sets, ligatures (all but the standard ones in peculiar and probably useless combinations), number variants, and the [Required Features](#). In terms of OpenType support, Word is the most primitive of the major text processing applications.

Many MUFI characters cannot be produced by using the OpenType variants of Junicode. These characters fall into three categories:

- Those with Unicode (non-PUA) code points. MUFI has done valuable work obtaining Unicode code points for medieval characters. All such characters (those with hexadecimal codes that *do not* begin with **E** or **F**) are presumed safe to

---

<sup>1</sup> Many OpenType features produce different outcomes depending on an index passed to an application’s layout engine along with the feature tag. Different applications have different ways of entering this index: consult your application’s documentation. Here, the index is recorded in brackets after the feature tag. Users of fontspec (with Xe<sub>La</sub>TeX or Lua<sub>Te</sub>X) should also be aware that fontspec indexes start at zero while OpenType indexes start at one. Therefore all index numbers listed in this document must be reduced by one for use with fontspec.

use in accessible and searchable text. However, some of these are covered by Junicode OpenType features for particular reasons.

- Precomposed characters—those consisting of base character + one or more diacritics. For greatest accessibility, these should be entered not as PUA code points, but rather as sequences consisting of base character + diacritics. For example, instead of MUFI U+E498 LATIN SMALL LETTER E WITH DOT BELOW AND ACUTE, use **e** + U+0323 COMBINING DOT BELOW + U+0301 COMBINING ACUTE ACCENT: **é** (when applying combining marks, start with any marks below the character and work downwards, then continue with any marks above the character and work upwards. For example, to make **ő**, place characters in this order: **o**, COMBINING OGONEK U+0328, COMBINING DOT BELOW U+0323, COMBINING MACRON U+0304, COMBINING ACUTE U+0301). Some MUFI characters have marks in unconventional locations, e.g. **ö** LATIN SMALL LETTER O WITH DOT ABOVE AND ACUTE, where the acute appears beside the dot instead of above. This and other characters like it should still be entered as a sequence of base character + marks (here **o**, COMBINING DOT ABOVE U+0307, COMBINING ACUTE U+0301). Junicode will position the marks correctly.
- Characters for which a base character (a Unicode character to which it can be linked) cannot be identified, or for which there may be an inconsistency in the MUFI recommendation. These include:
  - **fl** U+E8AF. This is a ligature of long **s** and **l** with stroke, but there are no base characters with this style of stroke.
  - **ŭŰ** U+EFD8 and U+EFD9. MUFI lists these as ligatures (corresponding to the historic ligatures **uW**), but they cannot be treated as ligatures in the font because a single diacritic is positioned over the glyphs as if they were digraphs like **aA**.
  - **pp̈** U+EBE7 and U+EBE6, for the same reason.
  - **ð** U+F159 LATIN ABBREVIATION SIGN SMALL DE. Neither a variant of **d** nor an eth (**ð**), this character may be a candidate for Unicode encoding.
- Characters for which OpenType programming is not yet available. These will be added as they are located and studied. [Check: U+EBF1, and smcp version.]

These characters should be avoided, even if you are otherwise using MUFI's PUA characters:

- U+F1C5 COMBINING CURL HIGH POSITION. Use U+1DCE COMBINING OGONEK ABOVE. The positioning problem mentioned in the MUFI recommendation is solved in Junicode (and, to be fair, many other fonts with OpenType programming).
- U+F1CA COMBINING DOT ABOVE HIGH POSITION. Use U+0307 COMBINING DOT ABOVE. It will be positioned correctly on any character.

## 4.2. Case-Related Features

### 4.2.1. **smcp** – Small Capitals

Converts lowercase letters to small caps; also several symbols and combining marks. All lower- and uppercase pairs (with exceptions noted below) have a small cap equivalent. Lowercase letters without matching caps may lack matching small caps. fghij → FGHIJ.

Note: Precomposed characters defined by MUFI in the Private Use Area have no small cap equivalents. Instead, compose characters using combining diacritics, as outlined in the introduction. For example, **smcp** applied to the sequence **t** + COMBINING OGONEK (U+0328) + COMBINING ACUTE (U+0301) will change **ť** to **ť̂**.

### 4.2.2. **c2sc** – Small Capitals from Capitals

Use with **smcp** for all-small-cap text. ABCDE → ABCDE.

### 4.2.3. **pcap** – Petite Capitals

Produces small caps in a smaller size than **smcp**. Use these when small caps have to be mixed with lowercase letters. The whole of the basic Latin alphabet is covered, plus several other letters. klmnop → KLMNOP.

#### 4.2.4. `case` – Case-Sensitive Forms

Produces combining marks that harmonize with capital letters:  $\text{Ř}$ ,  $\text{X̣}$ , etc. Use of this feature reduces the likelihood that a combining mark will collide with a glyph in the line above. Some applications turn this feature on automatically for runs of capitals, and precomposed characters (e.g.  $\text{É}$  U+00C9,  $\text{Ū}$  U+016A) already use case-appropriate combining marks.

### 4.3. Numbers and Sequencing

#### 4.3.1. `frac` – Fractions

Applied to a slash and surrounding numbers, produces fractions with diagonal slashes.  $6/9$  becomes  $\frac{6}{9}$ ,  $16/91$  becomes  $\frac{16}{91}$ .

#### 4.3.2. `numr` – Numerators

Changes numbers to those suitable for use on the left/upper side of fractions with diagonal stroke (U+2044). This can be used, with `dnom`, to manually construct fractions, but for most users `frac` will be a better solution.

#### 4.3.3. `dnom` – Denominators

Changes numbers to those suitable for use on the right/lower side of fractions with diagonal stroke (U+2044). This can be used, with `numr`, to manually construct fractions, but for most users `frac` will be a better solution.

#### 4.3.4. `nalt` – Alternate Annotation Forms

Produces letters and numbers circled, in parenthesis, or followed by periods, as follows:

`nalt[1]`, circled letters or numbers:  $\textcircled{a}$   $\textcircled{b}$  ...  $\textcircled{z}$ ;  $\textcircled{0}$   $\textcircled{1}$   $\textcircled{2}$  ...  $\textcircled{20}$ .

`nalt[2]`, letter or numbers in parentheses: (a) ... (z); (0) (1) ... (20).

`nalt[3]`, double-circled numbers:  $\textcircled{\textcircled{0}}$   $\textcircled{\textcircled{1}}$  ...  $\textcircled{\textcircled{10}}$ .

`nalt[4]`, white numbers in black circles:  $\textcircled{0}$   $\textcircled{1}$   $\textcircled{2}$   $\textcircled{3}$  ...  $\textcircled{20}$ .

`nalt[5]`, numbers followed by period:  $\textcircled{0}$   $\textcircled{1}$  ...  $\textcircled{20}$ .

For enclosed figures 10 and higher, **rlig** (Required Ligatures) must also be enabled (as it should be by default: see [Required Features](#) below).

#### 4.3.5. **tnum** – Tabular Figures

Fixed-width figures: 0123456789 (default or with **lnum**), 0123456789 (with **onum**).

#### 4.3.6. **onum** – Oldstyle Figures

Figures that harmonize with lowercase characters: 0123456789 (default or with **tnum**), 0123456789 (with **pnum**). When combined with **pnum**, this feature also affects subscripts and superscripts.

#### 4.3.7. **pnum** – Proportional Figures

Proportionally spaced figures: 0123456789 (default or with **lnum**), 0123456789 (with **onum**). When combined with **onum**, this feature also affects subscripts and superscripts. Most applications (including MS Word) with any support of OpenType features will support this feature and **lnum** in such a way that you don't have to enter them manually.

#### 4.3.8. **lnum** – Lining Figures

Figures in a uniform height, harmonizing with uppercase letters: 0123456789 (default or with **tnum**), 0123456789 (with **pnum**).

#### 4.3.9. **zero** – Slashed Zero

Produces slashed zero in all number styles: 0 0 0 0. Includes superscripts, subscripts, and fractions formed with **frac**: 0<sup>0</sup> 0<sub>0</sub> 10/30.

## 4.4. Superscripts and Subscripts

### 4.4.1. `sups` – Superscripts

Produces superscript numbers and letters. Only affects lining tabular and oldstyle proportional figures. All lowercase letters of the basic Latin alphabet are covered, and most uppercase letters: <sup>0123 4567 abcde ABDEG</sup>. Wherever superscripts are needed (e.g. for footnote numbers), use `sups` instead of the raised and scaled characters generated by some programs. With `sups`: <sup>4567</sup>. Scaled: <sup>4567</sup>.






























### 4.4.2. `subs` – Subscripts

Produces subscript numbers. Only affects lining tabular (the default numbers) and oldstyle proportional figures (use `pnum` and `onum`): <sub>8901 2345</sub>.

## 4.5. Ornaments

### 4.5.1. `ornm` – Ornaments

Produces ornaments (fleurons) in either of two ways: as an indexed variant of the bullet character (U+2022) or as variants of a-z, A-C (all fleurons are available by either method):

a, 1		i, 9		q, 17		y, 25	
b, 2		j, 10		r, 18		z, 26	
c, 3		k, 11		s, 19		A, 27	
d, 4		l, 12		t, 20		B, 28	
e, 5		m, 13		u, 21		C, 29	
f, 6		n, 14		v, 22			
g, 7		o, 15		w, 23			
h, 8		p, 16		x, 24			

The method with letters of the alphabet is easier, but the method with bullets will produce a more satisfactory result when text is displayed in an environment where Unicode is not available or `ornm` is not implemented.



## 4.6. Alphabetic Variants

### 4.6.1. cv01-cv52 – Basic Latin Variants

These features also affect small cap (**smcp**) and underdotted (**ss07**) forms, where available. Variants in **magenta** are also available via **ss06** “Enlarged Minuscles.” Use the **cvNN** features instead of **ss06** when you want to substitute an enlarged minuscule for a capital (or, less likely, a lowercase) letter everywhere in a text.

Variant of	cvNN	Variants
A	cv01	1=Ⓐ, 2=Ⓐ̄, 3=Ⓐ̇, 4= <b>a</b>
a	cv02	1=ⓐ, 2=ⓐ̄, 3=ⓐ̇, 4=ⓐ̂, 5=ⓐ̃, 6= <b>a</b>
B	cv03	1= <b>b</b> , 2=ⓑ
b	cv04	1= <b>b</b>
C	cv05	1=Ⓒ, 2= <b>c</b>
c	cv06	1=Ⓒ̇, 2=Ⓒ̂
D	cv07	1=Ⓓ, 2= <b>d</b> , 3=Ⓓ̇
d	cv08	1=ⓓ, 2=ⓓ̄, 3=ⓓ̇, 4= <b>d</b> , 5=ⓓ̂ (for 1, see also sso2)
E	cv09	1=Ⓔ, 2=Ⓔ̄, 3= <b>e</b> , 4=Ⓔ̂
e	cv10	1=ⓔ, 2=ⓔ̄, 3=ⓔ̇, 4= <b>e</b>
F	cv11	1=Ⓕ, 2= <b>f</b> , 3=Ⓕ̇
f	cv12	1=ⓕ, 2=ⓕ̄, 3=ⓕ̇, 4=ⓕ̂, 5=ⓕ̃, 6=ⓕ̄̇, 7= <b>f</b> , 8=ⓕ̂̇
G	cv13	1=Ⓖ, 2=Ⓖ̄, 3=Ⓖ̇, 4= <b>g</b>
g	cv14	1=ⓖ, 2=ⓖ̄, 3=ⓖ̇, 4=ⓖ̂, 5=ⓖ̃, 6=ⓖ̄̇, 7=ⓖ̂̇, 8= <b>g</b>
H	cv15	1=Ⓗ, 2= <b>h</b>
h	cv16	1=ⓗ, 2=ⓗ̄, 3= <b>h</b>
I	cv17	1=Ⓘ, 2=Ⓘ̄, 3= <b>i</b>

## 20 FEATURE REFERENCE

Variant of	cvNN	Variants
i	cv18	1= $\dot{\text{i}}$ , 2= $\dot{\text{i}}$ , 3= $\dot{\text{i}}$ , 4= $\dot{\text{i}}\dot{\text{j}}$ , 5= $\dot{\text{j}}$ , 6= $\dot{\text{i}}^*$
J	cv19	1= $\dot{\text{j}}$ , 2= $\dot{\text{j}}$
j	cv20	1= $\dot{\text{j}}$ , 2= $\dot{\text{j}}$ , 3= $\dot{\text{j}}$ , 4= $\dot{\text{j}}$
K	cv21	1= $\dot{\text{k}}$
k	cv22	1= $\dot{\text{k}}$ , 2= $\dot{\text{k}}$ , 3= $\dot{\text{k}}$ , 4= $\dot{\text{k}}$ , 5= $\dot{\text{k}}$
L	cv23	1= $\dot{\text{l}}$
l	cv24	1= $\dot{\text{l}}$ , 2= $\dot{\text{l}}$ , 3= $\dot{\text{l}}$ , 4= $\dot{\text{l}}$ , 5= $\dot{\text{l}}$ (roman only)
M	cv25	1= $\dot{\text{m}}$ , 2= $\dot{\text{m}}$ , 3= $\dot{\text{m}}$ , 4= $\dot{\text{m}}$
m	cv26	1= $\dot{\text{m}}$ , 2= $\dot{\text{m}}$ , 3= $\dot{\text{m}}$ , 4= $\dot{\text{m}}$
N	cv27	1= $\dot{\text{n}}$ , 2= $\dot{\text{n}}$
n	cv28	1= $\dot{\text{n}}$ , 2= $\dot{\text{n}}$ , 3= $\dot{\text{n}}$ , 4= $\dot{\text{n}}$ , 5= $\dot{\text{n}}$
O	cv29	1= $\dot{\text{o}}$ , 2= $\dot{\text{o}}$
o	cv30	1= $\dot{\text{o}}$ , 2= $\dot{\text{o}}$
P	cv31	1= $\dot{\text{p}}$ , 2= $\dot{\text{p}}$
p	cv32	1= $\dot{\text{p}}$
Q	cv33	1= $\dot{\text{q}}$ , 2= $\dot{\text{q}}$ , 3= $\dot{\text{q}}$ , 4= $\dot{\text{q}}$
q	cv34	1= $\dot{\text{q}}$ , 2= $\dot{\text{q}}$
R	cv35	1= $\dot{\text{r}}$ , 2= $\dot{\text{r}}$ , 3= $\dot{\text{r}}$
r	cv36	1= $\dot{\text{r}}$ , 2= $\dot{\text{r}}$ , 3= $\dot{\text{r}}$ , 4= $\dot{\text{r}}$
S	cv37	1= $\dot{\text{s}}$ , 2= $\dot{\text{s}}$ , 3= $\dot{\text{s}}$ , 4= $\dot{\text{s}}$
s	cv38	1= $\dot{\text{s}}$ , 2= $\dot{\text{s}}$ , 3= $\dot{\text{s}}$ , 4= $\dot{\text{s}}$ , 5= $\dot{\text{s}}$ , 6= $\dot{\text{s}}$ , 7= $\dot{\text{s}}$
T	cv39	1= $\dot{\text{t}}$ , 2= $\dot{\text{t}}$
t	cv40	1= $\dot{\text{t}}$ , 2= $\dot{\text{t}}$ , 3= $\dot{\text{t}}$

Variant of	cvNN	Variants
U	cv41	I=U
u	cv42	I=U
V	cv43	I=V
v	cv44	I=V, 2=V̇, 3=V̈, 4=V̋, 5=V̌
W	cv45	I=W
w	cv46	I=W
X	cv47	I=X
x	cv48	I=X, 2=Ẋ, 3=Ẍ, 4=X̋, 5=X̌
Y	cv49	I=Ẏ, 2=Ÿ
y	cv50	I=Ẏ, 2=Ÿ, 3=Y̋
Z	cv51	I=Ż, 2=Z̈
z	cv52	I=Ż, 2=Z̈, 3=Z̋

\* **cv18[4]** changes ii to ij at the end of a word; **cv18[5]** changes i to j at the end of a word whether another i precedes or not. These variants are chiefly useful for roman numbers, but also for Latin words ending in -ii. The j produced by this feature is searchable as i.

#### 4.6.2. cv53-cv66 – Other Latin Letters

Some features affect both upper- and lowercase forms. **cv62** also affects combining **e** with ogonek, accessible via **ss10** with the entity reference **&\_eogo;**. features.

Variant of	cvNN	Variants
Ą (U+0104)	cv53	I=Ą, 2=Ą̇, 3=Ą̈
ą (U+0105)	cv54	I=ą, 2=ą̇

## 2.2 FEATURE REFERENCE

Variant of	cvNN	Variants
Ƕ (U+A733)	cv55	1=Ƕ, 2=Ƕ, 3=Ƕ
Æ (U+00C6)	cv56	1=Æ, 2=Æ
æ (U+00E6)	cv57	1=æ, 2=æ, 3=æ, 4=æ
Ɔ (U+A734)	cv58	1=Ɔ, 2=Ɔ, 3=Ɔ
Ɔ (U+A735)	cv59	1=Ɔ, 2=Ɔ, 3=Ɔ
ʒ (U+A739)	cv60	1=ʒ
ḏ (U+0111)	cv61	1=ḏ
Ǝ, Ǝ ... (U+0118, U+0119)	cv62	1=Ǝ, Ǝ ...; 2=Ǝ, Ǝ ...
Ȝ, ȝ (U+021C, U+021D)	cv63	1=Ȝ, 2=ȝ
ø (U+A7C1)	cv65	1=ø, 2=ø, 3=ø, 4=ø
þ, þ (U+A765)	cv66	1=þ, þ

### 4.6.3. ss01 – Alternate thorn and eth

Produces Nordic thorn and eth (þðÞ) when the language is English, and English thorn and eth (þðþ) with any other language, reversing the font’s ordinary usage. This also affects small caps, crossed thorn (þ þ—see also [cv66](#)), combining mark eth (U+1DD9, ̥ ̥), and enlarged thorn and eth (see [ss06](#)). This feature depends on [loca](#) (Localized Forms), which in most applications will always be enabled.

### 4.6.4. ss02 – Insular Letter-Forms

Produces insular letter-forms, e.g. ȝȝȝȝȝ. The result is different, depending on whether the language is English or Irish (make sure the language for your document is set properly). In English text, capitals are not affected (except W), as these do not commonly have insular shapes in early manuscripts; instead, enter the Unicode code points or use the Character Variant ([cvNN](#)) features. In English text, sso2 imitates the typography of the Old English passages of Hickes’s *Thesaurus*, not the usage of Old English or Anglo-Latin manuscripts. In Irish texts, it imitates the distribution of insular characters but cannot imitate the style of particular scribal hands or typefaces.

#### 4.6.5. **ss04 – High Overline**

Produces a high overline over letters used as roman numbers:  $\overline{\text{cdijlmvx}} \overline{\text{CDIJLMVXO}}$ .

#### 4.6.6. **ss05 – Medium-High Overline**

Produces a medium-high overline over (or through the ascenders of) letters used as roman numbers, and some others as well:  $\overline{\text{bcdhijklmfvxp}}$ .

#### 4.6.7. **ss06 – Enlarged Minuscles**

Letters that are lowercase in form but uppercase in function, and between upper- and lowercase in size. They are often used to begin sentences in medieval manuscripts: abcdefg. The feature covers the whole of the basic Latin alphabet and a number of other letters that occur at the beginnings of sentences: consult the MUFI recommendation for details. Uppercase letters are also covered by this feature so that enlarged minuscles can, if you like, be searched as capitals.

If you are using the variable version of the font (JunicodeVF), consider using the [Enlarged axis](#) instead, for reasons of flexibility and accessibility.

#### 4.6.8. **ss07 – Underdotted Text**

Produces underdotted text (indicating deletion in medieval manuscripts) for many letters (including the whole of the basic Latin alphabet and a number of other letters), e.g.  $\text{a}\grave{\text{b}}\grave{\text{c}}\grave{\text{d}}\grave{\text{e}}\grave{\text{f}}\grave{\text{g}} \text{H}\grave{\text{I}}\grave{\text{J}}\grave{\text{K}}\grave{\text{L}}\grave{\text{M}}$ . This also affects small caps, e.g.  $\text{ABCDEF} \rightarrow \text{A}\grave{\text{B}}\grave{\text{C}}\grave{\text{D}}\grave{\text{E}}\grave{\text{F}}$ . For letters without corresponding underdotted forms (e.g. U+A751, p), use U+0323, combining dot below ( $\text{p}\grave{\text{}}$ ).

#### 4.6.9. **ss08 – Contextual Long s**

In English and French text only, varies **s** and **f** according to rules followed by many early printers:  $\text{fports, effence, ftoomy, difheveled, transfusions, flynefs, cliffside}$ . For this feature to work properly, **calt** “Contextual Alternates” must also be enabled (as it should be by default: see [Required Features](#) below). This feature does not work in Lua<sub>T</sub><sub>E</sub>X, except in harf mode.

#### 4.6.10. `ss16` – Contextual r Rotunda

Converts `r` to `ʀ` (lowercase only) following the most common rules of medieval manuscripts: `pziest`, `firmer`, `frost`, `oznament`. For this feature to work properly, `calt` “Contextual Alternates” must also be enabled (as it should be by default: see [Required Features](#) below). This feature does not work in LuaTeX, except in harf mode.

#### 4.6.11. `cv68` – Variant of ʔ (U+0294, glottal stop)

`ʔ=ʔ`.

### 4.7. Greek

Unicode has two distinct styles of Greek. In the roman face, it is upright and modern, especially designed to harmonize with Unicode’s Latin letters. In the italic, it is slanted and old-style, being based on the eighteenth-century Greek type designed by Alexander Wilson and used by the Foulis Press in Glasgow. Both Greek styles include the full polytonic and monotonic character sets:  $\alpha\beta\gamma\delta\epsilon\zeta$   $\alpha\beta\gamma\delta\epsilon\zeta$ .

To set Greek properly (especially polytonic text) requires that both `locl` and `ccmp` be active, as they should be by default in most text processing applications (but in MS Word they must be explicitly enabled by checking the “kerning” box on the “Advanced” tab of the Font Dialog).

Modern monotonic Greek should be set using only characters from the Unicode “Greek and Coptic” range (U+0370–U+03FF). When monotonic text is set in all caps, Unicode suppresses accents automatically (except in single-letter words, for which you must substitute unaccented forms manually). This substitution is not performed on text containing visually identical letters from the “Greek Extended” range (U+0F00–U+1FFF). Thus when setting polytonic Greek, one should use (for example)  $\text{Α}$  (U+1FBF), not  $\text{Α}$  (U+0386), though they look the same.

You can set polytonic Greek either by entering code points from the Greek Extended range or by entering sequences of base characters and diacritics. When using the latter method, you must first make sure the language for the text in question (whether a single word, a short passage, or a complete document) is set to Greek, and then enter characters in canonical order (that is, the sequence defined by Unicode as equivalent to the composite character). The order is as follows: 1. base character;

2. diacritics positioned either above or in front of the base character, working from left to right or bottom to top; 3. the *ypogegrammeni* (U+0345), or for capitals, if you prefer, the *prosgegrammeni* (U+1FBE).<sup>2</sup>

For example, the sequence ω (U+03C9) ◊ (U+0313) ◊ (U+0301) ◊ (U+0345) produces ◊. Substitute capital Ω (U+03A9) and the result is "Ω◊. Note that in a number of applications the layout engine will perform these substitutions before Junicode's own programming is invoked. If either the layout engine or Junicode fails to produce your preferred result, try placing U+034F COMBINING GRAPHEME JOINER somewhere in the sequence of combining marks—for example, before the *ypogegrammeni* to make "Ω◊.

#### 4.7.1. ss03 – Alternate Greek

Provides alternate shapes of β γ θ π φ χ ω: β γ θ π φ χ ω. These are chiefly useful in linguistics, as they harmonize with IPA characters.

### 4.8. Punctuation

MUFI encodes nearly twenty marks of punctuation in the PUA. In Junicode these can be accessed in either of two ways: all are indexed variants of . (period), and all are associated with the Unicode marks of punctuation they most resemble (but it should not be inferred that the medieval marks are semantically identical with the Unicode marks, or that there is an etymological relationship between them). The first method will be easier for most to use, but the second is more likely to yield acceptable fallbacks in environments where Junicode is not available.

Marks with Unicode encoding are not included here, as they can safely be entered directly. In MUFI 4.0 several marks have PUA encodings, but have since that release been assigned Unicode code points: *paragraphus* (¶ U+2E4D), medieval comma (Ꞥ U+2E4C), *punctus elevatus* (⋈ U+2E4E), *virgula suspensiva* (⁄ U+2E4A), triple dagger (‡ U+2E4B).

<sup>2</sup> Some applications will automatically reorder sequences of letters and accents, sparing you the trouble of remembering the canonical order.

#### 4.8.1. **ss18** – Old-Style Punctuation Spacing

Colons, semicolons, parentheses, quotation marks and several other glyphs are spaced as in early printed books.

#### 4.8.2. **cv69** – Variants of ꝛꝛ (U+204A / U+2E52, Tironian nota)

1=ꝛꝛ, 2=ꝛꝛ̇.

#### 4.8.3. **cv70** – Variants of . (period)

1=·, 2=., 3=ʹ, 4=;̇, 5=.,̇, 6=;̇, 7=;̇, 8=;̇, 9=ʹ, 10=ʹ, 11=ʹ, 12=ʹ, 13=ʹ, 14=ʹ, 15=ʹ, 16=ʹ, 17=ʹ, 18=ʹ, 19=ʹ, 20=ʹ. This feature provides access to all non-Unicode MUFI punctuation marks. Some of them are available via other features (see below).

#### 4.8.4. **cv71** – Variant of · (U+00B7, middle dot)

1=· (*distinctio*).

#### 4.8.5. **cv72** – Variants of , (comma)

1=,, 2=ʹ.

#### 4.8.6. **cv73** – Variants of ; (semicolon)

1=; (*punctus versus*), 2=.,, 3=;̇, 4=;̇, 5=;̇.

#### 4.8.7. **cv74** – Variants of ⁜ (U+2E4E, *punctus elevatus*)

1=⁜, 2=⁜, 3=⁜, 4=⁜ (*punctus flexus*).

#### 4.8.8. **cv75** – Variant of ! (exclamation mark)

1=⁜ (*punctus exclamativus*).



#### 4.8.9. cv76 – Variants of ? (question mark)

1=¿, 2=⸐, 3=⸑. Shapes of the *punctus interrogativus*.

#### 4.8.10. cv77 – Variant of ~ (ASCII tilde)

1=˜ (same as MUFI U+F1F9, “wavy line”).

#### 4.8.11. cv78 – Variant of \* (asterisk)

1=⋔. MUFI defines U+F1EC as a *signe de renvoi*. Manuscripts employ a number of shapes (of which this is one) for this purpose. Junicon defines it as a variant of the asterisk—the most common modern *signe de renvoi*.

#### 4.8.12. cv79 – Variants of / (slash)

1=⸀, 2=⸁. The first of these is Unicode, U+2E4E.

### 4.9. Abbreviations

#### 4.9.1. cv80 – Variant of ꝛ (U+A75D, rum abbreviation)

1=ꝛ.

#### 4.9.2. cv81 – Variants of ͆ (U+035B, combining zigzag above)

1=͆, 2=͇, 3=͈. Positioning of the zigzag can differ from that of other combining marks, e.g. ͆, ͇, ͈. If **calt** “Contextual Alternates” is enabled (as it is by default in most apps), variant forms of **cv81[2]** will be used with several letters, e.g. ͇, ͈, ͉. Enable **case** for forms that harmonize with capitals (͇ ͈ ͉ ͊), **smcp** for forms that harmonize with small caps (͇ ͈ ͉ ͊).

### 4.9.3. cv82 – Variants of spacing ʹ (U+A770)

1=ʹ, 2=ᵹ. **cv82**[1] produces the baseline *-us* abbreviation (same as MUFI U+F1A6). MUFI also has an uppercase baseline *-us* abbreviation (U+F1A5), but as there is no uppercase version of U+A770 to pair it with, it is indexed separately here.

### 4.9.4. cv83 – Variants of ʒ (U+A76B, “et” abbreviation)

1=;̣, 2=;̇. [1] is identical in shape to a semicolon, but as it is semantically the same as U+A76B, it is preferable to use that character with this feature. [2] produces a subscript version of the character, a common variant in printed books.

## 4.10. Combining Marks

### 4.10.1. cv84 – MUFI combining marks (variants of U+0304)

MUFI encodes a number of combining marks in the PUA (with code points between E000 and F8FF), but when these characters are entered directly, they can interfere with searching and accessibility, and some important applications fail to position them correctly over their base characters. To avoid these problems, enter U+0304 (̄, COMBINING MACRON) and apply **cv84**, with the appropriate index, to both mark and base character. This collection of marks does not include any Unicode-encoded marks (from the “Combining Diacritical Marks” ranges), as these can safely be entered directly. It does include three marks (**cv84**[30], [31] and [32]) that lack MUFI code points but are used to form MUFI characters, and three more ([2], [33], and [34]) that have no code points in Unicode or MUFI.

This feature may often appear to have no effect. When this happens it is because an application replaced a sequence like **a U+0304** with a precomposed character like **ā** (U+0101) before Junicode’s OpenType programming had a chance to work. This process is called normalization, and it usually has the effect of simplifying text processing tasks, but can sometimes prevent the proper functioning of OpenType features. To disable it, place the character U+034F COMBINING GRAPHEME JOINER (don’t waste time puzzling over the name) between the base character and the combining mark (or the first combining mark). For example, to produce the combination **ū**, enter **u U+034F U+0304**. (Without U+034F, you would get **ū**).

These marks can sometimes be produced by other **cvNN** features, which may be preferable to **cv84** as providing more suitable fallbacks for applications that do not support Character Variant (**cvNN**) features.

For some marks with PUA code points, users may find it easier to use **entities** than this feature.

These marks are not affected by most other features. This is to preserve flexibility, given the rule that the feature that produces them must be applied to both the mark and the base character. For example, if you had to apply **smcp** “Small Caps” to **U+1DE8** <sup>b</sup> to get **cv84[11]** <sup>b</sup>, it would be impossible to produce the sequence **nāa** (or the reverse case **NĀA**) with the diacritic properly positioned.

1= <sup>2</sup>	10= <sup>8</sup>	19= <sup>m</sup>	28= <sup>œ</sup>
2= <sup>3</sup>	11= <sup>8</sup>	20= <sup>ö</sup>	29= <sup>œ</sup>
3= <sup>4</sup>	12= <sup>9</sup>	21= <sup>9</sup>	30= <sup>5</sup>
4= <sup>7</sup>	13= <sup>ë</sup>	22= <sup>9</sup>	31= <sup>6</sup>
5= <sup>9</sup>	14= <sup>ë</sup>	23= <sup>9</sup>	32= <sup>7</sup>
6= <sup>1</sup>	15= <sup>1</sup>	24= <sup>2</sup>	33= <sup>8</sup>
7= <sup>œ</sup>	16= <sup>j</sup>	25= <sup>7</sup>	34= <sup>8</sup>
8= <sup>œ</sup>	17= <sup>j</sup>	26= <sup>y</sup>	
9= <sup>8</sup>	18= <sup>8</sup>	27= <sup>b</sup>	

#### 4.10.2. **cv67** – Spacing zigzag (variant of U+00AF, spacing macron)

A spacing version of <sup>5</sup> (U+035B, combining zigzag) appears in John Hutchins, *The History and Antiquities of the County of Dorset* (London, 1774). It is not in Unicode or MUFI. In the future this feature may be used, as necessary, for other spacing marks of abbreviation.

#### 4.10.3. **ss10** – Character Entities for Combining Marks

Instead of **cv84** for representing non-Unicode combining marks, some users may wish to use XML/HTML-style entities. When **ss10** is turned on (preferably for the entire text), these entities appear as combining marks and are correctly positioned over base characters. For example, the letter **e** followed by **&\_eogo;** will yield <sup>e</sup>. An advantage of entities is that they are (unlike the PUA code points or the indexes of

**cv84**) mnemonic and thus easy to use. A disadvantage is that searches cannot ignore combining marks entered by this method as they can using the **cv84** method. (Every method of entering non-Unicode combining marks has disadvantages: users should weigh these, choose a method, and stick with it.)

If you use any of these entities in a work intended for print publication, you should call your publisher's attention to them, since they will likely have their own method of representing them.

<code>&amp;x_ansc;</code> →	<code>&amp;x_idotl;</code> →	<code>&amp;x_orr;</code> →
<code>&amp;x_an;</code> →	<code>&amp;x_j;</code> →	<code>&amp;x_oru;</code> →
<code>&amp;x_ar;</code> →	<code>&amp;x_jdotl;</code> →	<code>&amp;x_q;</code> →
<code>&amp;x_arsc;</code> →	<code>&amp;x_ksc;</code> →	<code>&amp;x_ru;</code> →
<code>&amp;x_bsc;</code> →	<code>&amp;x_munc;</code> →	<code>&amp;x_tsc;</code> →
<code>&amp;x_dsc;</code> →	<code>&amp;x_oogo;</code> →	<code>&amp;x_y;</code> →
<code>&amp;x_eogo;</code> →	<code>&amp;x_osh;</code> →	<code>&amp;x_thorn;</code> →
<code>&amp;x_emac;</code> →	<code>&amp;x_omac;</code> →	

For another function of Stylistic Set 10, see [Chapter 6, Entering Characters with Tags](#).

#### 4.10.4. ss20 – Low Diacritics

The MUFI recommendation includes a number of precomposed characters with base letters b, h, k, þ, ð and ð and combining marks (U+0363), (U+0364), (U+0304/**cv84**[15]), (U+0366), (U+036C), (U+1DE2), (U+036D), (U+036E), (U+1DE6) and (U+0304/**cv84**[19]). Instead of being positioned above ascender height as usual (e.g. ), the MUFI glyphs have the marks positioned above the x-height (e.g. ). Using the MUFI code points for these precomposed glyphs can interfere with searching and drastically reduce accessibility. Users of Junicode should instead use a sequence of base character + combining mark, and apply **ss20** to the two glyphs. A variant shape of eth () that accommodates the combining mark will be substituted for the normal base character (but this is not necessary for the other base characters). Examples: , , , , , .

**ss20** affects only the diacritics and base characters listed here; other combinations (e.g. , ) are not affected. It will therefore probably be safe to apply this feature to the whole text if it is needed anywhere.

4.10.5. **cv85** – Variant of ˆ (U+1DD3, combining open a)

1=ˆ.

4.10.6. **cv86** – Variant of ˆ (U+1DD8, combining insular d)

1=ˆ.

4.10.7. **cv87** – Variant of ˆ (U+1DE3, combining r rotunda)

1=ˆ.

4.10.8. **cv88** – Variant of combining dieresis (U+0308)

1=¨. This also affects precomposed characters on which this variant dieresis may occur, e.g. á.

4.10.9. **cv89** – Variant of ¯ (U+0305, combining overline)

1=¯.

4.10.10. **cv90** – Variants of ¯ (U+035E, combining double macron)

1=¯, 2=¯.

4.10.11. **cv92** – Variant of breve below (U+032E)

1=◌◌◌. Position the mark after the middle of three glyphs, and apply **cv92** to both the mark and (at least) the middle glyph. This mark is not available via **cv84**.

## 4.11. Currency and Weights

4.11.1. **cv93** – Variants of ₪ (U+0044, generic currency sign)

1=Π	8=gr	15=ℓ	22=ⱥr
2=✕	9=Φ	16=£	23=ⱥ <sup>8</sup>
3=¥	10=⋈	17=ⱥ	24=ℓ
4=℔	11=℔	18=ⱥ	25=ℓ
5=₪	12=℔	19=mⱥ	26=ℓ
6=ⱥ	13=ⱥ	20=ⱥ	27=ⱥ
7=ƒ	14=ℓ	21=ⱥ	

All of MUFI's currency and weight symbols (those that do not have Unicode code points) are gathered here, but some are also variants of other currency signs (see below).

#### 4.II.2. cv94 – Variant of ℔ (U+2114)

1=℔. Same as MUFI U+F2EB (French Libra sign).

#### 4.II.3. cv95 – Variants of £ (U+00A3, British pound sign)

1=℔, 2=℔, 3=ⱥ, 4=ℓ, 5=ℓ, 6=£. Same as MUFI U+F2EA, F2EB, F2EC, F2ED, F2EE, F2EF, pound signs from various locales.

#### 4.II.4. cv96 – Variant of ⱥ (U+20B0, German penny sign)

1=ⱥr. Same as MUFI U+F2F5.

#### 4.II.5. cv97 – Variant of ƒ (U+0192, florin)

1=ƒ. Same as MUFI U+F2E8.

#### 4.II.6. cv98 – Variant of ⱥ (U+2125, Ounce sign)

1=ⱥ. Same as MUFI U+F2FD, Script ounce sign.

## 4.12. Gothic

### 4.12.1. ss19 – Latin to Gothic Transliteration

Produces Gothic letters from Latin: Warþ þan in dagans jainans → 𐍄𐍂𐍆 𐍆𐍅𐍅 𐍇𐍅𐍅𐍅 𐍇𐍅𐍅𐍅 𐍇𐍅𐍅𐍅. In web pages, the letters will be searchable as their Latin equivalents.

## 4.13. Runic

### 4.13.1. ss12 – Early English Futhorc

Changes Latin letters to their equivalents in the early English futhorc. Because of the variability of the runic alphabet, this method of transliteration may not produce the result you want. In that case, it may be necessary to manually edit the result. fisc flodu ahof → 𐍆𐍆𐍅 𐍆𐍆𐍅𐍅 𐍆𐍆𐍅𐍅.

### 4.13.2. ss13 – Elder Futhark

Changes Latin letters to their equivalents in the Elder Futhark. Because of the variability of the runic alphabet, this method of transliteration may not produce the result you want. In that case, it may be necessary to manually edit the result. ABCDEFG → 𐍆𐍅𐍅𐍅𐍅𐍅𐍅.

### 4.13.3. ss14 – Younger Futhark

Changes Latin letters to their equivalents in the Younger Futhark. Because of the variability of the runic alphabet, this method of transliteration may not produce the result you want. In that case, it may be necessary to manually edit the result. ABCDEFG → 𐍆𐍅𐍅𐍅𐍅𐍅.

### 4.13.4. ss15 – Long Branch to Short Twig

In combination with ss14, converts long branch (the default for the Younger Futhark) to short twig runes: 𐍆𐍅𐍅𐍅𐍅𐍅 → 𐍆𐍅𐍅𐍅𐍅.

### 4.13.5. `rtlm` – Right to Left Mirrored Forms

Produces mirrored runes, e.g.  $\text{ᚱᚷᚢᚱᚱᚱᚱᚱᚱ} \rightarrow \text{ᚱᚱᚱᚱᚱᚱᚱᚱ}$ . This feature cannot change the direction of text.

## 4.14. Ligatures and Digraphs

### 4.14.1. `hlig` – Historic Ligatures

Produces ligatures for combinations that should not ordinarily be rendered as ligatures in modern text.<sup>3</sup> Most of these are from the MUFI recommendation, ranges B.1.1(b) and B.1.4. This feature does not produce digraphs (which have a phonetic value), for which see [ss17](#). The ligatures:

$\text{af} \rightarrow \text{ᚠ}$	$\text{ck} \rightarrow \text{ᚠᚵ}$	$\text{hr} \rightarrow \text{ᚱ}$	$\text{PP} \rightarrow \text{ᚱᚱ}$	$\text{fch} \rightarrow \text{ᚠᚵᚱ}$
$\text{ar} \rightarrow \text{ᚠᚱ}$	$\text{ðð} \rightarrow \text{ᚢᚢ}$	$\text{hf} \rightarrow \text{ᚠᚱ}$	$\text{pp} \rightarrow \text{ᚱᚱ}$	$\text{ftr} \rightarrow \text{ᚠᚱᚱ}$
$\text{ag} \rightarrow \text{ᚠᚱ}$	$\text{ey} \rightarrow \text{ᚠᚱ}$	$\text{hf} \rightarrow \text{ᚠᚱ}$	$\text{pp} \rightarrow \text{ᚱᚱ}$	$\text{fp} \rightarrow \text{ᚠᚱ}$
$\text{al} \rightarrow \text{ᚠᚱ}$	$\text{fä} \rightarrow \text{ᚠᚱ}$	$\text{kr} \rightarrow \text{ᚠᚱ}$	$\text{Ps} \rightarrow \text{ᚱᚱ}$	$\text{ττ} \rightarrow \text{ᚠᚱ}$
$\text{an} \rightarrow \text{ᚠᚱ}$	$\text{gd} \rightarrow \text{ᚠᚱ}$	$\text{kf} \rightarrow \text{ᚠᚱ}$	$\text{ps} \rightarrow \text{ᚱᚱ}$	$\text{UE} \rightarrow \text{ᚠᚱ}$
$\text{aN} \rightarrow \text{ᚠᚱ}$	$\text{gð} \rightarrow \text{ᚠᚱ}$	$\text{kf} \rightarrow \text{ᚠᚱ}$	$\text{Psi} \rightarrow \text{ᚱᚱ}$	$\text{ue} \rightarrow \text{ᚠᚱ}$
$\text{ap} \rightarrow \text{ᚠᚱ}$	$\text{gð} \rightarrow \text{ᚠᚱ}$	$\text{ll} \rightarrow \text{ᚠᚱ}$	$\text{psi} \rightarrow \text{ᚱᚱ}$	$\text{UU} \rightarrow \text{ᚠᚱ}$
$\text{ar} \rightarrow \text{ᚠᚱ}$	$\text{gg} \rightarrow \text{ᚠᚱ}$	$\text{nf} \rightarrow \text{ᚠᚱ}$	$\text{qp} \rightarrow \text{ᚠᚱ}$	$\text{uu} \rightarrow \text{ᚠᚱ}$
$\text{aR} \rightarrow \text{ᚠᚱ}$	$\text{gg} \rightarrow \text{ᚠᚱ}$	$\text{oc} \rightarrow \text{ᚠᚱ}$	$\text{q3/q3} \rightarrow \text{ᚠᚱ}$	$\text{pp} \rightarrow \text{ᚠᚱ}$
$\text{ap} \rightarrow \text{ᚠᚱ}$	$\text{go} \rightarrow \text{ᚠᚱ}$	$\text{O2} \rightarrow \text{ᚠᚱ}$	$\text{qq} \rightarrow \text{ᚠᚱ}$	$\text{pr} \rightarrow \text{ᚠᚱ}$
$\text{bb} \rightarrow \text{ᚠᚱ}$	$\text{gp} \rightarrow \text{ᚠᚱ}$	$\text{o2} \rightarrow \text{ᚠᚱ}$	$\text{Q2} \rightarrow \text{ᚠᚱ}$	$\text{ps} \rightarrow \text{ᚠᚱ}$
$\text{bg} \rightarrow \text{ᚠᚱ}$	$\text{gr} \rightarrow \text{ᚠᚱ}$	$\text{O2} \rightarrow \text{ᚠᚱ}$	$\text{q2} \rightarrow \text{ᚠᚱ}$	$\text{pf} \rightarrow \text{ᚠᚱ}$
$\text{ch} \rightarrow \text{ᚠᚱ}$	$\text{Hr} \rightarrow \text{ᚠᚱ}$	$\text{o2} \rightarrow \text{ᚠᚱ}$	$\text{fä} \rightarrow \text{ᚠᚱ}$	$\text{pb} \rightarrow \text{ᚠᚱ}$

Note: For the ligature  $\text{ᚠ}$  to work properly, U+017F  $\text{ᚠ}$  must be entered directly, not by applying an OpenType feature to  $\text{s}$ .

<sup>3</sup> Some fonts define `hlig` differently, as including all ligatures in which at least one element is an archaic character, e.g. those involving long s ( $\text{ſ}$ ). In Junicod, however, a historic ligature is defined not by the characters it is composed of, but rather by the join between them. If two characters (though modern) should not be joined except in certain historic contexts, they form a historic ligature. If they should be joined in all contexts (even if archaic), the ligature is not historic and should be defined in `liga`.



#### 4.14.2. `dlig` – Discretionary Ligatures

Produces lesser-used ligatures, but also roman numbers, e.g. ii, II, xi, XI. The lesser-used ligatures: `ct`, `fp`, `str`, `st`, `tr`, `tt`, `ty`.

#### 4.14.3. `ss17` – Rare Digraphs

By “digraph” we mean conjoined letters that represent a phonetic value: the most common examples for western languages are `æ` and `œ` (though these, because they are so common, are not included in this feature). Use of this feature in web pages enables easier searches: for example, producing `pau` from `pau` allows the word to be searched as “pau.” The digraphs covered by this feature are `aa`, `æ`, `ai`, `æ`, `ay`, `oo`, `wy`, plus capital and small cap equivalents and digraph + diacritic combinations anticipated in the MUFI recommendation. To produce such a digraph + diacritic combination, either type a letter + diacritic combination as the second element of the digraph or type the diacritic after the second element. For example, `a` + `ú` yields `á`, and so does `a` + `u` + U+0301 (combining acute accent). To produce a digraph + diacritic combination not covered by MUFI (e.g. `ð`), you may have to place U+034F COMBINING GRAPHEME JOINER (see [cv84](#) above) between the second element of the digraph and the combining mark. Without U+034F: `ao`. With U+034F: `ð`.

### 4.15. Required Features

Required features, which provide some of the font’s most basic functionality—ligatures, support for other features, kerning, and more—include `ccmp` (Glyph Composition/Decomposition), `calt` (Contextual Alternates), `liga` (Standard Ligatures), `loca` (Localized Forms), `rlig` (Required Ligatures), `kern` (Horizontal Kerning), and `mark/mkmm` (Mark Positioning). In MS Word these features have to be explicitly enabled on the Advanced tab of the Font dialog (Ctrl-D or Cmd-D: enable Kerning, Standard Ligatures, and Contextual Alternates, and the others will be enabled automatically), but in most other applications they are enabled by default.





































## 5. Non-MUFI Code Points


Characters in Junicode that do not have Unicode code points should be accessed via OpenType features whenever possible. MUFI/PUA code points should be used only in applications that do not support OpenType, or that support it only partially (for example, MS Word). For certain characters that lack either Unicode or MUFI code points, code points in the Supplementary Private Use Area-A (plane 15) are available.


U+F0000 <i>A</i>	U+F0008 <i>ḁ</i>	U+F0010 <i>m̄</i>	U+F0018 <i>v̄</i>
U+F0001 <i>ȧ</i>	U+F0009 <i>Ḃ</i>	U+F0011 <i>o</i>	U+F0019 <i>x̄</i>
U+F0002 <i>Ȧ</i>	U+F000A <i>ḃ</i>	U+F0012 <i>ø</i>	U+F001A <i>x̄</i>
U+F0003 <i>ċ</i>	U+F000B <i>f</i>	U+F0013 <i>o</i>	U+F001B <i>ʔ</i>
U+F0004 <i>ċ</i>	U+F000C <i>ī</i>	U+F0014 <i>o</i>	U+F001C <i>ʔ</i>
U+F0005 <i>ḋ</i>	U+F000D <i>j</i>	U+F0015 <i>q</i>	U+F001D <i>ʔ</i>
U+F0006 <i>Ḍ</i>	U+F000E <i>j̄</i>	U+F0016 <i>ʒ</i>	U+F001E <i>ʔ</i>
U+F0007 <i>dʼ</i>	U+F000F <i>P</i>	U+F0017 <i>v̄</i>	U+F001F <i>ʔ</i>


## 6. Entering characters with tags



Any character in Junicode that can be rendered using a Character Variant (**c**vNN) feature can also be rendered using a sequence consisting of a base character and two Unicode tags—that is, characters from the Unicode tag range. This range duplicates the ASCII character set (which consists, roughly, of things that can be typed on a U.S. English keyboard), but the characters it contains are normally invisible. They are used as modifiers for a preceding character (in other fonts, usually a flag symbol). Junicode contains a partial collection of tag characters:

 U+E0061	 U+E0062	 U+E0063
 U+E0064	 U+E0065	 U+E0066
 U+E0067	 U+E0068	 U+E0069
 U+E006A	 U+E006B	 U+E006C
 U+E006D	 U+E006E	 U+E006F
 U+E0070	 U+E0071	 U+E0072
 U+E0073	 U+E0074	 U+E0075
 U+E0076	 U+E0077	 U+E0078
 U+E0079	 U+E007A	 U+E0030
 U+E0031	 U+E0032	 U+E0033
 U+E0034	 U+E0035	 U+E0036
 U+E0037	 U+E0038	 U+E0039

When creating web pages or XML documents, you can enter these using character entities (e.g. **&#xE0073**; for ); along the same lines, you can use the **\char**

command (e.g. `\char"0E0073`) when composing TeX documents. But such entities and commands are not generally available in editors and word processors, and entering the tags themselves can be tricky because they are not only invisible, but also ignored by the application (the cursor skips over them). Instead of attempting to enter the code points for tags directly, use character entities consisting of **an ampersand, two underscores, the tag character, and a semicolon**. These will yield visible tags. For example, typing `&__6;` will yield the tag .

To use the tag method of entering characters, first type the base character from the table below, then the two tags. For example, to make a square C () , enter one of these sequences:

Web page:	<code>C&amp;#xE0073;&amp;#xE0071;</code>
TeX:	<code>C\char"0E0073\char"0E0071</code>
Word processor:	<code>C&amp;__s;&amp;__q;</code> (appears as C   )

Then apply the OpenType feature **ss10** (Stylistic Set 10) to the passage or passages containing the tags or, if tags occur throughout, to the whole document. The tags will disappear and the preceding characters (the base characters) will be transformed.

For variants of the combining macron (U+0304) and perhaps other combining marks, it will often be necessary to place the COMBINING GRAPHEME JOINER (U+034F) between the macron and the base character that precedes it. This will prevent normalization (the shaping engine changing sequences of character + combining mark into precomposed characters), which interferes with the operation of tags.

Most of the two-tag sequences documented here are designed to be mnemonic. For example, the sequence in the example above stands for “square.” However, two-tag sequences are not capable of describing characters in any detail, and in some cases, where the number of variants is large (especially for period, combining macron, and currency), the tags are not descriptive at all, but rather an index (the same numbers used in the corresponding **cvNN** features).

These tags are compatible with Junicode’s other OpenType features, including the **cvNN** features, and can be mixed with them. They will not interfere with the placement of combining marks. Use a **cvNN** feature when a variant should appear throughout the text, repeatedly in a particular passage (for example, a block quotation), or in a style. A tag sequence may be preferable for isolated forms or to override an OpenType feature.















































In the list below, records for each character are color-coded as follows:

green	MUFI characters with PUA code points
yellow	Other characters with PUA code points
blue	Characters with Unicode code points
red	Characters without code points















































Tags or **cvNN** features are usually to be preferred to PUA code points, which should be used only where accessibility and searchability are not issues (mainly in printed texts). Unicode code points can safely be entered directly. Junicode makes a few of them accessible via **cvNN** features and tags because it may often be desirable to associate these characters with their bases rather than the Unicode code points. For example, the insular T (**Ƨ**) is sure to be searchable as T if entered with the sequence **T**in, but if entered as **U+A786** it may or may not be searchable as T, depending on the application.

Characters without code points can only be entered via tags or OpenType features.
























Base	Sequence	Result	Description / Code point
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">1</span>	·	Distinctio / F1F8
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">2</span>	.	Comma positura / F1E2
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">3</span>	’	High comma positura / F1E3
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">4</span>	;	Punctus versus / F1EA
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">5</span>	.,	Punctus with comma positura / F1E4
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">6</span>	∴	Colon with middle comma positura / F1E5
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">7</span>	∴	Two dots over comma positura / F1F2
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">8</span>	∴	Three dots over comma positura / F1E6
period	<span style="border: 1px solid black; padding: 0 2px;">0</span> <span style="border: 1px solid black; padding: 0 2px;">9</span>	↗	Punctus elevatus diagonal stroke / F1F0
period	<span style="border: 1px solid black; padding: 0 2px;">1</span> <span style="border: 1px solid black; padding: 0 2px;">0</span>	ˆ	Punctus elevatus with high back / F1FA

Base	Sequence	Result	Description / Code point
period	.  	./	Punctus elevatus with onset / F1FB
period	.  	?	Punctus flexus / F1F5
period	.  	!	Punctus exclamativus / F1E7
period	.  	?	Punctus interrogativus / F160
period	.  	˘	Punctus interrogativus horizontal tilde / F1E8
period	.  	∞	Punctus interrogativus lemniskate / F1F1
period	.  	~	Wavy line / F1F9
period	.  	∴	Signe de renvoi / F1EC
period	.  	/	Virgula suspensiva / F1F4
period	.  	/	Short virgula / F1F7
U+0304	◌  	◌̈	Combining curly bar above / F1CC
U+0304	◌  	◌̆	Combining diagonal macron
U+0304	◌  	◌̇	Combining bar above with dot / F1C0
U+0304	◌  	◌̈́	Combining angular zigzag / F1C7
U+0304	◌  	◌̈́	Combining curly zigzag / F1C8
U+0304	◌  	◌̈́	Combining vertical tilde / 033E
U+0304	◌  	◌̈́	Combining ligature an / F036
U+0304	◌  	◌̈́	Combining ligature aN / F03A
U+0304	◌  	◌̈́	Combining ligature ar / F038
U+0304	◌  	◌̈́	Combining ligature aR / F130
U+0304	◌  	◌̈́	Combining B / F013
U+0304	◌  	◌̈́	Combining D / F016
U+0304	◌  	◌̈́	Combining e with macron / F136
























Base	Sequence	Result	Description / Code point
U+0304	̄ [1 TAG] [4 TAG]	ė	Combining e with ogonek / F135
U+0304	̄ [1 TAG] [5 TAG]	ï	Combining dotless i / F02F
U+0304	̄ [1 TAG] [6 TAG]	ĵ	Combining j / F030
U+0304	̄ [1 TAG] [7 TAG]	ĵ	Combining dotless j / F031
U+0304	̄ [1 TAG] [8 TAG]	ķ	Combining K / F01C
U+0304	̄ [1 TAG] [9 TAG]	ṁ	Combining uncial m / F01F
U+0304	̄ [2 TAG] [0 TAG]	ō	Combining o with macron / F13F
U+0304	̄ [2 TAG] [1 TAG]	ȯ	Combining o with ogonek / F13E
U+0304	̄ [2 TAG] [2 TAG]	ő	Combining o with stroke / F032
U+0304	̄ [2 TAG] [3 TAG]	q̇	Combining q / F033
U+0304	̄ [2 TAG] [4 TAG]	ꝛ	Combining rum abbreviation / F040
U+0304	̄ [2 TAG] [5 TAG]	ṛ	Combining T / F02A
U+0304	̄ [2 TAG] [6 TAG]	ȳ	Combining y / F02B
U+0304	̄ [2 TAG] [7 TAG]	þ	Combining thorn / F03D
U+0304	̄ [2 TAG] [8 TAG]	ꝝ	Combining ligature o r rotunda / F03E
U+0304	̄ [2 TAG] [9 TAG]	Ꝟ	Combining ligature letter o rum / F03F
U+0304	̄ [3 TAG] [0 TAG]	ö̇	Combining diagonal dieresis
U+0304	̄ [3 TAG] [1 TAG]	ž	Combining dot and acute
U+0304	̄ [3 TAG] [2 TAG]	q̣	Combining ogonek and dot
U+0304	̄ [3 TAG] [3 TAG]	ō̇	Combining narrow macron
U+0304	̄ [3 TAG] [4 TAG]	ō̋	Combining macron with serifs
A	A [1 TAG] [0 TAG]	Ɽ	Insular A / F201
A	A [8 TAG] [9 TAG]	ⱦ	Square A / F13A

Base	Sequence	Result	Description / Code point
A	A  	Ȧ	A with diagonal stroke / E8DA
A	A  	a	Enlarged minuscule A
a	a  	ɑ	Insular a / F200
a	a  	Ɑ	Uncial a / F214
a	a  	Ɫ	Open a / F202
a	a  	ⱦ	Closed a / F203
a	a  	Ɽ	Neckless a / F215
a	a  	ⱥ	Enlarged minuscule a / EEE0
B	B  	b	Enlarged minuscule B
B	B  	Ɱ	Insular B
b	b  	ⱦ	Enlarged minuscule b / EEE1
C	C  	ⱪ	Square C / F106
C	C  	c	Enlarged minuscule C
c	c  	ⱪ	c with curl / F198
c	c  	c	Enlarged minuscule c / EEE2
D	D  	ⱦ	Enlarged minuscule insular D
D	D  	Ɱ	Insular D / A779
D	D  	d	Enlarged minuscule D
d	d  	ⱦ	Insular d default form / A77A
d	d  	Ɱ	Insular d second form
d	d  	ⱦ	Enlarged minuscule insular d / EEE4
d	d  	ⱪ	d with curl / F193
d	d  	d	Enlarged minuscule d / EEE3






Base	Sequence	Result	Description / Code point
E	E 	Ʒ	Uncial closed E / F217
E	E 	Ʒ	Uncial E / F10A
E	E 	Ʒ	Insular E
E	E 	Ʒ	Enlarged minuscule E
e	e 	Ʒ	Uncial e / F218
e	e 	Ʒ	e with bar / F219
e	e 	Ʒ	High e with bar / F21A
e	e 	Ʒ	Enlarged minuscule e / EEE6
F	F 	Ʒ	Insular F / A77B
F	F 	Ʒ	Enlarged minuscule F / EEE7
F	F 	Ʒ	Enlarged minuscule insular F
f	f 	Ʒ	Insular f / A77C
f	f 	Ʒ	Insular f with dotted hooks / F21C
f	f 	Ʒ	Semi-closed insular f / EBD5
f	f 	Ʒ	Closed insular f / EBD6
f	f 	Ʒ	Enlarged minuscule insular f / EEFF
f	f 	Ʒ	Narrow f / F000B
f	f 	Ʒ	f with curl / F194
f	f 	Ʒ	Enlarged minuscule f / EEE7
G	G 	Ʒ	Insular G / A77D
G	G 	Ʒ	Ormulum G with bar / A7D0
G	G 	Ʒ	Square G / F10E
G	G 	Ʒ	Enlarged minuscule G

Base	Sequence	Result	Description / Code point
g	$g \begin{smallmatrix} \boxed{i} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	ȝ	Insular g / 1D79
g	$g \begin{smallmatrix} \boxed{o} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{r} \\ \text{TAG} \end{smallmatrix}$	ȝ	Ormulum g with bar / A7D1
g	$g \begin{smallmatrix} \boxed{s} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{c} \\ \text{TAG} \end{smallmatrix}$	g	Script g / 0261
g	$g \begin{smallmatrix} \boxed{c} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{u} \\ \text{TAG} \end{smallmatrix}$	gʁ	g with curl / F196
g	$g \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	g	Enlarged minuscule g / EEE8
g	$g \begin{smallmatrix} \boxed{c} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{1} \\ \text{TAG} \end{smallmatrix}$	g	Closed g with separate loops / F21D
g	$g \begin{smallmatrix} \boxed{c} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{2} \\ \text{TAG} \end{smallmatrix}$	g	Closed g with large lower loop / F21E
g	$g \begin{smallmatrix} \boxed{c} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{3} \\ \text{TAG} \end{smallmatrix}$	g	Closed g with small lower loop / F21F
H	$H \begin{smallmatrix} \boxed{u} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	ḥ	Uncial H / F110
H	$H \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	h	Enlarged minuscule H
h	$h \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix}$	h	h with descender / F23A
h	$h \begin{smallmatrix} \boxed{a} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{u} \\ \text{TAG} \end{smallmatrix}$	hr	h-shaped <i>autem</i> abbreviation / E8A3
h	$h \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	h	Enlarged minuscule h / EEE9
I	$I \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{a} \\ \text{TAG} \end{smallmatrix}$	I	I with dot above / 0130
I	$I \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix}$	I	I with descender / A7FE
I	$I \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	i	Enlarged minuscule I
i	$i \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{l} \\ \text{TAG} \end{smallmatrix}$	ı	Dotless i / 0131
i	$i \begin{smallmatrix} \boxed{l} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{o} \\ \text{TAG} \end{smallmatrix}$	ı	Long I / F220
i	$i \begin{smallmatrix} \boxed{b} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix}$	ḡ	i with double bar / E8A1
i	$i \begin{smallmatrix} \boxed{f} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{1} \\ \text{TAG} \end{smallmatrix}$	ij	i final form (when i precedes)
i	$i \begin{smallmatrix} \boxed{f} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{2} \\ \text{TAG} \end{smallmatrix}$	j	i final form
i	$i \begin{smallmatrix} \boxed{e} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{n} \\ \text{TAG} \end{smallmatrix}$	i	Enlarged minuscule i / EEEA
J	$J \begin{smallmatrix} \boxed{d} \\ \text{TAG} \end{smallmatrix} \begin{smallmatrix} \boxed{a} \\ \text{TAG} \end{smallmatrix}$	Ĵ	J with dot above / E15C

Base	Sequence	Result	Description / Code point
J	J 	j	Enlarged minuscule J
j	j 	j	Dotless j / 0237
j	j 	ĵ	j with double bar / E8A2
j	j 	j	j with dot accent / F000D
j	j 	j	Enlarged minuscule j / EEEB
K	K 	k	Enlarged minuscule K
k	k 	k	Uncial k / F208
k	k 	k	Closed k, one / F221
k	k 	ķ	Closed k, two / F209
k	k 	k̤	k with curl / F195
k	k 	k	Enlarged minuscule k / EEEC
L	L 	l	Enlarged minuscule L
l	l 	l	Descending l / F222
l	l 	l	Enlarged minuscule l / EEED
l	l 	l	l with high stroke / A749
l	l 	l	l with high stroke ending with flourish / F000F
l	l 	l	Insular l (roman only) / F000F
M	M 	Ⓜ	Uncial M with descender / F224
M	M 	Ⓜ	Uncial M / F11A
M	M 	ℳ	Epigraphic M / A7FF
M	M 	m	Enlarged minuscule M
m	m 	Ⓜ	Uncial m with descender / F23D
m	m 	m	Uncial m / F23C

Base	Sequence	Result	Description / Code point
m	m <sub>ENG</sub> <sub>ENG</sub>	ṁ	m with descender / F223
m	m <sub>ENG</sub> <sub>ENG</sub>	m	Enlarged minuscule m / EEEE
N	N <sub>ENG</sub> <sub>ENG</sub>	Ṛ	N with descender / F229
N	N <sub>ENG</sub> <sub>ENG</sub>	n	Enlarged minuscule n / EEEF
n	n <sub>ENG</sub> <sub>ENG</sub>	ṅ	n with descender / F228
n	n <sub>ENG</sub> <sub>ENG</sub>	ñ	n with bar / E7B2
n	n <sub>ENG</sub> <sub>ENG</sub>	ṇ	n with curl / F19A
n	n <sub>ENG</sub> <sub>ENG</sub>	n	Enlarged minuscule n / EEEF
n	n <sub>ENG</sub> <sub>ENG</sub>	Ṛ	Small cap n with descender / F22A
O	O <sub>ENG</sub> <sub>ENG</sub>	o	Enlarged minuscule O
o	o <sub>ENG</sub> <sub>ENG</sub>	o	Enlarged minuscule o / EEf0
P	P <sub>ENG</sub> <sub>ENG</sub>	Ɔ	Reversed P / A7FC
P	P <sub>ENG</sub> <sub>ENG</sub>	p	Enlarged minuscule P
p	p <sub>ENG</sub> <sub>ENG</sub>	p	Enlarged minuscule p / EEf1
Q	Q <sub>ENG</sub> <sub>ENG</sub>	q	Q with stem / F22C
Q	Q <sub>ENG</sub> <sub>ENG</sub>	Q <sub>1</sub>	Q with long tail, one
Q	Q <sub>ENG</sub> <sub>ENG</sub>	Q <sub>2</sub>	Q with long tail, two
Q	Q <sub>ENG</sub> <sub>ENG</sub>	q	Enlarged minuscule Q
q	q <sub>ENG</sub> <sub>ENG</sub>	q̇	q with diagonal stroke / E8B4
q	q <sub>ENG</sub> <sub>ENG</sub>	q	Enlarged minuscule q / EEf2
R	R <sub>ENG</sub> <sub>ENG</sub>	Ṛ	Insular R / A783
R	R <sub>ENG</sub> <sub>ENG</sub>	Ꝛ	R rotunda / A75A
R	R <sub>ENG</sub> <sub>ENG</sub>	r	Enlarged minuscule R

Base	Sequence	Result	Description / Code point
r	$r \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͥ} \end{smallmatrix}$	ɹ	Insular r / A782
r	$r \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ʀ	r rotunda / A75B
r	$r \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͣ} \end{smallmatrix}$	ɽ	r with curl / F19B
r	$r \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ŕ	Enlarged minuscule r / EE3
S	$S \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͥ} \end{smallmatrix}$	ſ	Insular S / A784
S	$S \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͧ} \end{smallmatrix}$	Ꝣ	Sigmoid S / A7D8
S	$S \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	Ꝥ	Enlarged minuscule S
S	$S \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ꝥ	S with diagonal stroke (Sanctus abbrev)
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͥ} \end{smallmatrix}$	ɣ	Insular s / A785
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͧ} \end{smallmatrix}$	Ꝣ	Sigmoid s / A7D9
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ƒ	Long s / 017F
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule s / EE4
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ꝥ	Long s with descender / F127
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ꝥ	Long s with flourish / E8B7
s	$s \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͦ} \end{smallmatrix}$	ꝥ	Long s with diagonal stroke / E8B8
T	$T \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͥ} \end{smallmatrix}$	Ƨ	Insular T / A786
T	$T \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule T
t	$t \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͥ} \end{smallmatrix}$	ɿ	Insular t / A787
t	$t \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͣ} \end{smallmatrix}$	ɿ	t with curl / F199
t	$t \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule t / EE5
U	$U \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule U
u	$u \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule u / EE7
V	$V \begin{smallmatrix} \text{͡} \\ \text{ͥ} \end{smallmatrix} \begin{smallmatrix} \text{͢} \\ \text{ͤ} \end{smallmatrix}$	ꝥ	Enlarged minuscule V















































Base	Sequence	Result	Description / Code point
v	v  	ṽ	Enlarged minuscule v / EEf8
v	v  	ṽ	v with diagonal stroke / E8BA
v	v  	ṽ	v with bar / E74E
v	v  	V	Enlarged minuscule v / EEf8
v	v  	ṽ	v with two bars / E8BC
W	W  	W	Enlarged minuscule W
w	w  	W	Enlarged minuscule w / EEf9
X	X  	x	Enlarged minuscule X
x	x  	X	Enlarged minuscule x / EEfA
x	x  	ṽ	x with long left leg / AB57
x	x  	ṽ	x with diagonal stroke (upper left) / E8BD
x	x  	ṽ	x with diagonal stroke (lower right) / E8BE
x	x  	ṽ	x with two diagonal strokes (lower right) / E8CE
Y	Y  	Y	Y with diagonal stroke (Hymnus abbrev) / E8DB
Y	Y  	y	Enlarged minuscule Y
y	y  	y	y with main right stroke / F233
y	y  	ṽ	y with bar / E77B
y	y  	y	Enlarged minuscule y / EEfB
Z	Z  	Ź	Visigothic Z / A762
Z	Z  	z	Enlarged minuscule Z
z	z  	Ź	Visigothic z / A763
z	z  	ḥ	Middle High German z / F238
z	z  	Z	Enlarged minuscule z / EEfC

Base	Sequence	Result	Description / Code point
U+0104	A <sub>Ꝛ</sub> <span>Ꝛ</span> <span>1</span>	Ⱥ	A with short diagonal stroke / F0000
U+0104	A <sub>Ꝛ</sub> <span>Ꝛ</span> <span>2</span>	Ȼ	A with long diagonal stroke / F001E
U+0104	A <sub>Ꝛ</sub> <span>Ꝛ</span> <span>1</span>	Ⱥ	A with flourish / F0002
U+0105	a <sub>Ꝛ</sub> <span>Ꝛ</span> <span>1</span>	Ȣ	a with short diagonal stroke / F0001
U+0105	a <sub>Ꝛ</sub> <span>Ꝛ</span> <span>2</span>	ȣ	a with long diagonal stroke / F001F
U+A733	aa <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɑ	Uncial aa
U+A733	aa <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɑ	Closed aa / EFA0
U+A733	aa <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɑ	Enlarged minuscule aa / EFDF
U+00C6	Æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɱ	Neckless Æ / EFAE
U+00C6	Æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɱ	Enlarged minuscule Æ
U+00E6	æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɐ	Neckless æ / EFA1
U+00E6	æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɐ	Enlarged minuscule æ / EAF1
U+00E6	æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɐ	Open æ / F204
U+00E6	æ <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɐ	Uncial æ
U+A734	AO <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Neckless AO / F205
U+A734	AO <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Enlarged minuscule AO
U+A734	AO <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Enlarged minuscule AO with smaller O
U+A735	ao	Ɒ	Enlarged minuscule ao / EFDE
U+A735	ao <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Enlarged minuscule ao with smaller o / EAF2
U+A735	ao <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Neckless ao / F206
U+A735	ao <sub>Ꝛ</sub> <span>Ꝛ</span>	Ɒ	Uncial ao
U+A739	av <sub>Ꝛ</sub> <span>Ꝛ</span>	ⱱ	Neckless av / EFA2
U+0111	d <sub>Ꝛ</sub> <span>Ꝛ</span>	ȡ	d with flourish / F0007

Base	Sequence	Result	Description / Code point
U+0118	E <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	Ě	E with centered ogonek
U+0118	E <sub>1</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	Ě̌	E with diagonal stroke / F0009
U+0118	Ě <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	Ě́	E with dot, acute, and centered ogonek
U+0118	Ě <sub>1</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	Ě̌́	E with dot, acute, and diagonal stroke
U+0119	e <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	ě	e with centered ogonek
U+0119	e <sub>1</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	ě̌	e with diagonal stroke / F000A
U+0119	ě	ě́	e with dot, acute, and centered ogonek
U+0119	ě <sub>1</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	ě̌́	e with dot, acute, and diagonal stroke
&_eogo;	◌&_eogo; <sub>2</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	◌̊	Combining e with centered ogonek
&_eogo;	◌&_eogo; <sub>3</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	◌̋	Combining e with diagonal stroke
U+021C	3 <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	Ȝ	Yogh with flat top
U+021C	3 <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	ȝ	Yogh with insular shape
U+021D	3 <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	ȝ	yogh with flat top
U+021D	3 <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	ȝ	yogh with insular shape
U+A7C1	<sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub> <sup>1</sup>	o	Old Polish o with broken slash / F0011
U+A7C1	<sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	ø	Old Polish o with short slash / F0012
U+A7C1	<sub>1</sub> <sub>TAG</sub> <sub>3</sub> <sub>TAG</sub>	o	Old Polish o with lower left slash / F0013
U+A7C1	<sub>1</sub> <sub>TAG</sub> <sub>4</sub> <sub>TAG</sub>	o	Old Polish o with upper right slash / F0014
U+A765	þ <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	þ̊	thorn with stroke with different slant / F149
U+A765/ENG	þ <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	þ̊	thorn with stroke with different slant
U+0241	ʔ <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	ʔ̊	Alternate glottal stop / F001B
U+204A	ɀ <sub>1</sub> <sub>TAG</sub> <sub>1</sub> <sub>TAG</sub>	ɀ	Tironian et sign later form / F001D
U+204A	ɀ <sub>1</sub> <sub>TAG</sub> <sub>2</sub> <sub>TAG</sub>	ɀ̄	Tironian et sign later form with bar





Base	Sequence	Result	Description / Code point
slash	/  	⁄	Virgula suspensiva / F1F4
slash	/  	/	Short virgula / F1F7
U+A75D	ŕ  	ꝛ	Alternate rum abbreviation / F0016
U+035B	◌  	◌̃	Combining angular zigzag / F1C7
U+035B	◌  	◌̈́	Combining curly zigzag / F1C8
U+035B	◌  	◌̈́	Combining vertical zigzag (tilde) / 033E
U+A770	ʹ  	ᵹ	Baseline spacing us abbreviation / F1A6
U+A770	ʹ  	ᵶ	Uppercase us abbreviation / F1A5
U+A76B	₃  	;	Semicolon-like et abbreviation / F1AC
U+A76B	₃  	₳	Subscript et abbreviation
U+1DD3	◌  	◌̄	Combining flattened a with macron / F1C1
U+1DD8	◌  	◌̈́	Alternate combining insular d / F0005
U+A75B	◌  	◌̈́	Combining ur abbreviation lemniskate / F1C2
U+00E4	ä  	ä	a with diagonal dieresis / E8D5
U+00F6	ö  	ö	o with diagonal dieresis / E8D7
U+0308	◌  	◌̈́	Combining diagonal dieresis
U+0305	◌  	◌̇	Combining horizontal stroke with dot / F1C0
U+032E	◌   ◌	◌̈́◌̈́◌̈́	Breve below three letters / F1FC
U+00A4	◻  	ℙ	Latin as libralis sign / F2E0
U+00A4	◻  	℥	Latin small capital letter x with bar / F2E2
U+00A4	◻  	℥	Latin small capital letter y with bar / F2E3
U+00A4	◻  	℥	Latin small capital letter d with slash / F2E4
U+00A4	◻  	℥	Pharmaceutical dram sign / F2E6

Base	Sequence	Result	Description / Code point
U+00A4	☐ <sub>TAG</sub> 0 6	₣	Ecu sign / F2E7
U+00A4	☐ <sub>TAG</sub> 0 7	₧	Floren sign with loop / F2E8
U+00A4	☐ <sub>TAG</sub> 0 8	₥	Groschen sign / F2E9
U+00A4	☐ <sub>TAG</sub> 0 9	⊕	Helbing sign / F2FB
U+00A4	☐ <sub>TAG</sub> 1 0	₰	Krone sign / F2FA
U+00A4	☐ <sub>TAG</sub> 1 1	₧	Dutch libra sign / F2EA
U+00A4	☐ <sub>TAG</sub> 1 2	₥	French libra sign / F2EB
U+00A4	☐ <sub>TAG</sub> 1 3	₥	Italian libra sign / F2EC
U+00A4	☐ <sub>TAG</sub> 1 4	₥	Flemish libra sign / F2ED
U+00A4	☐ <sub>TAG</sub> 1 5	₥	Lira nuova sign / F2EE
U+00A4	☐ <sub>TAG</sub> 1 6	₥	Lira sterlina sign / F2EF
U+00A4	☐ <sub>TAG</sub> 1 7	₥	Old mark sign / F2F0
U+00A4	☐ <sub>TAG</sub> 1 8	₥	Old flourish mark sign / F2F1
U+00A4	☐ <sub>TAG</sub> 1 9	₥	Marked small letter m sign / F2F2
U+00A4	☐ <sub>TAG</sub> 2 0	₥	Flourished small letter m sign / F2F3
U+00A4	☐ <sub>TAG</sub> 2 1	₥	Pharmaceutical obelus sign / F2F4
U+00A4	☐ <sub>TAG</sub> 2 2	₥	Penning sign / F2F5
U+00A4	☐ <sub>TAG</sub> 2 3	₥	Old Reichstaler sign / F2F6
U+00A4	☐ <sub>TAG</sub> 2 4	₥	German schilling sign / F2F7
U+00A4	☐ <sub>TAG</sub> 2 5	₥	German script schilling sign / F2F8
U+00A4	☐ <sub>TAG</sub> 2 6	₥	Scudi sign / F2F9
U+00A4	☐ <sub>TAG</sub> 2 7	₥	Script ounce sign / F2FD
U+2114	₥ <sub>TAG</sub> 3 1	₥	French libra sign / F2EB

Base	Sequence	Result	Description / Code point
U+00A3	£	ⷂ	Dutch libra sign / F2EA
U+00A3	£	ⷃ	French libra sign / F2EB
U+00A3	£	ⷄ	Italian libra sign / F2EC
U+00A3	£	ⷅ	Flemish libra sign / F2ED
U+00A3	£	ⷆ	Lira nuova sign / F2EE
U+00A3	£	⷇	Lira sterlina sign / F2EF
U+20B1	ſ	ⷈ	Penning sign / F2F5
U+0192	f	ⷉ	Floren sign with loop / F2E8
U+2125	₯	ⷊ	Script ounce sign / F2FD

---

<sup>1</sup>Unicode entities (like &\_\_a;) following U+A7C1 (Old Polish o) are not properly displayed or resolved for technical reasons related to U+A7C1 having very recently been added to Unicode. However, tags entered directly (as here) or via HTML/XML entity references and TeX commands will still work, and the Unicode entity will work once applications are updated with the latest additions to the Unicode standard.

## 7. Transcribing records

This chapter provides guidance for persons transcribing records (laws and other public documents) in the style of Charles Trice Martin’s *The Record Interpreter, Statutes of the Realm*, and similar guides and editions. Unlike most editions of early texts, these retain (or recommend retaining) the capitalization, punctuation, and abbreviations of their manuscript sources.

### 7.1. A preliminary note on transcription

Here are a few observations, based on a long career as a scholarly editor of medieval and eighteenth-century texts.

Before embarking on the task of transcribing an old document, ask yourself what value you want to add to the document as it already exists, because different kinds of transcription add different kinds of value. The kind of transcription that adds the least is that which aims at the exact *visual* reproduction of a document. A transcript is not a facsimile: it needs to do something that a photograph can’t do.

Converting a document from visual image to Unicode-encoded text adds a good bit of value all by itself, but only if done with due regard for the semantics of Unicode characters. Every Unicode character has a meaning, and that meaning is a help to readers. Using the wrong character is a hinderance to readers, even if it *looks* right.

For example, in transcribing a Middle English text, you may decide that the Unicode `ezh` (3, U+0292) looks more like the yogh in your source than the Unicode `yogh` (3, U+021D) and therefore decide to use it for yogh. But the `ezh` is not a yogh! It is a character in the International Phonetic Alphabet and a letter in the alphabets of several minor languages—but not a letter in the Middle English language. If you use it where the yogh is called for, it will make your text less accessible and less search-

able. Indexing, concordance and bibliographical programs may be misled by it; screen readers will misinterpret it. To solve one problem (that of visual representation), you may well have introduced a host of far more serious problems.

Fortunately, Junicode offers a solution for this particular problem. The OpenType feature **cv63** substitutes for the yogh a character that *looks* like the *ezh* but is semantically a yogh and therefore will be handled correctly by applications. But neither Junicode nor any other font can solve every problem of this kind. Sometimes you will have to call to mind the important principle stated above: *A transcript is not a facsimile*. It is much more important that it should have the same *meaning* as the original than that it should have the same *look*.

This document concerns the transcription of texts in Latin (and to some extent, other archaic languages, e.g. Old and Middle English, Old French). It is long-standing custom, when transcribing certain kinds of documents, to retain marks of abbreviation—for example, the **ꝑ** you may find in a manuscript or printed edition representing the word *propterea*. This is okay—and Junicode can help with the task. But when dealing with the abbreviations, punctuation, and diacritics of an old text it is more important than ever that you use semantically correct characters for your transcription, as this will help readers who already face significant challenges.

For example, the abbreviation **ꝑ** as printed here consists of an underlying sequence of Unicode characters: **ꝑ** (U+A753, the common abbreviation for *pro*) + **ꝑ** + **◌̣** (U+0363, the combining small *a*). The OpenType feature **hlig** (Historical Ligatures) has been applied to this sequence, changing its appearance but not its underlying value. That underlying value is intelligible to computer applications in the sense that they can recognize each character.

This doesn't mean, though, that computer programs can correctly interpret **ꝑ** as *propterea*. Many (probably *most*) Latin abbreviations are ambiguous: this one, for example, can mean *propterea* or *propria*. Some abbreviations (most notoriously **◌̣** U+035B) can mean many things, depending on context. It takes a human being with a knowledge of Latin to interpret them correctly.

So another way you can add value in your transcript is by interpreting abbreviations like **ꝑ** and supplying expansions of them. Fortunately, systems for representing texts often offer ways to handle this task gracefully. For example, in a TEI (Text Encoding Initiative) text, you would use this construction:

```
<choice>
```

```
<abbr rend="hlig">ꝑꝑ</abbr>
<expan>propterea</expan>
</choice>
```

This kind of structure can be approximated in HTML, with supporting CSS and scripting to allow readers to choose between a “diplomatic” version, with unexpanded abbreviations, and a “reading” version, with expanded abbreviations and perhaps other amenities, such as modern punctuation and capitalization.

There are other ways to add value to a transcript—for example, by correcting errors, annotating the content, or writing textual notes. Each of these operations takes your transcript farther from the facsimile and closer to the edition.

## 7.2. 1. Common combining marks

A **combining mark** is a character that combines with another character (called the **base**) to form a character with accent (e.g. é) or an abbreviation (e.g. ꝑ for *prae*). Unicode and the Medieval Unicode Font Initiative (MUFI) offer code points for many precomposed combinations of base + combining mark, but it is also possible to place any mark over any base character by entering first the base and then the combining mark. It is also possible to place a combining mark over another combining mark. For example, to produce ꝑ̄, enter this sequence: q (U+0071) + U+0363 + U+0304.

Junicode 2 contains many variants of combining marks: for example the curly zigzag ǝ is a variant of Unicode’s angular zigzag ǝ (U+035B), produced by applying the OpenType feature `cv81[2]` to **both the base character and the combining mark**. Sometimes the combination of base + combining mark + OpenType feature will not produce the desired effect. When this happens, place U+034F (a special invisible combining mark, included in Unicode for exactly this purpose) between the base and the (visible) mark.

a. For a straight stroke over any letter, use the COMBINING MACRON (U+0304):

ōnis *omnis*; ōmis *omnis*; dāpna *dampna*; dāmpa *dampna*.

The combining macron can also be applied above superscripts and combining marks. Apply the OpenType feature `cv84[33]` for a narrower macron:

antiquā *antiquam*; q̄ *quam*.

For the superscript *a*, use the OpenType feature `sup`s (see r. below).

b. For a straight stroke through a tall letter, use the COMBINING SHORT STROKE OVERLAY (U+0335): **f d l**. But Unicode also has precomposed versions of **d**, **l** and other characters with stroke, e.g. **d** (U+0111), **l** (U+019A).

c. For **~** above any character, use the COMBINING TILDE (U+0303):

*ã ac, apud; ã alias.*  
*dñs dominus; cařina carmina; fčis factis.*  
*põita posita.*

d. For **~** through a vertical stroke, use the TILDE OVERLAY (U+0334): **l d** (U+0303) would be positioned above the letter, e.g. **l̃, d̃**. For the ligatures **fl**, **bb**, and **ff**, type the sequence for **l**, etc. twice.

e. For the tilde positioned above two letters, use COMBINING DOUBLE TILDE (U+0360) between the letters. It is automatically repositioned to clear tall characters: **cō tō dō oī**. The same is true of DOUBLE BREVE (U+035D) **cö dö**, DOUBLE MACRON (U+035E) **cō dō**, DOUBLE INVERTED BREVE (U+0361) **cô dô**, and DOUBLE CIRCUMFLEX (U+1DCD) **ô ô**.

f. The figure used to represent *er* (and other similar combinations) is a common medieval abbreviation which takes many forms. The semantically correct Unicode character is the COMBINING ZIGZAG (U+035B), but the best match in Junicode 2 for the figure as it appears in the *Record Interpreter* and the *Statutes* is a gothic variant of this, which MUFI encodes as U+F1C8 (the curly form zigzag). However, because for technical reasons many applications will not position the MUFI character correctly over the base, that code point should be avoided. The best way to access this variant is to apply `cv81[2]` to U+035B, as here:

*dēbe debere; iñ inter; P̃rū ferrum; gñō generatio; p̃; prae; ser̃ve servire.*

The curly form of the combining zigzag may be attached to any letter, and it may change shape depending on the letter it is attached to (including caps, for which use the `case` feature, and small caps: **ÄBČß**).

g. All letters a–z, and several others too, have combining forms. You may access these via their code points, when they are standard Unicode, via the `cv84` feature, or



via Junicode’s special entity references. For details, see [4.10.3, Character Entities for Combining Marks](#).

q̇ quo; q̇ qui; quatt̃ quattuor.

### 7.3. 2. Spacing characters

h. The symbol for *is*, *es* and a number of other abbreviations is the IS-SIGN (U+A76D):

forꝑ *foris*; omꝑ *omnes*; 9tꝑ *competentes*; infꝑ *infortunium*.

This character will sometimes ligature with the preceding letter. The italic version differs from the roman stylistically (*for<sup>℞</sup> om<sup>℞</sup> 9t<sup>℞</sup> inf<sup>℞</sup>*), but it will be intelligible to informed readers.

i. There are two characters for *-us* in Unicode: SPACING US U+A770 (do not confuse this with CON U+A76F) and COMBINING US U+1DD2. The *Record Interpreter* and *Statutes* appear to use only the spacing character:

īꝑiꝑ *ipsius*; ūsꝑ *uersus*; pꝑtea *postea*; pꝑ *post*.

j. The three-like sign is the ET SIGN (◌3, U+A76B, also used for *us*—do not confuse this with Middle English yogh: ȝ, U+021D):

quibꝓ *quibus*; licꝓ *licet*; sꝓ *sed*.

k. For *-rum* the Unicode RUM ROTUNDA (U+A75D) is like the one in MUFI/Junicode. The one in the *Record Interpreter* and *Statutes* is a late stylized version of this. Use U+A75D and apply OpenType feature **cv8o** to obtain the correct shape:

āĩaz̃ *animarum*; coꝝpere *corrumper*; beatoꝝ *beatorum*.

l. For *cum*, *con*, etc. use SMALL LETTER CON (U+A76F):

9ꝑutus *computus*; 9a *contra*; 9nouit *cognouit*.

m. For *per* (or sometimes *par* and other similar sequences), use P WITH STROKE U+A751:

psōa *persona*; 9ꝑet *comparet*.

- n. For *pro*, use P WITH FLOURISH U+A753:

ꝑceres *proceres*.

- o. For *prae*, *præ*, *pre*, there is no separate character; use a variant of the ZIGZAG (f. above) with **p**:

ꝑsēs *praesens*.

- p. For **q** with stroke through the descender, there are two Unicode points: U+A757 for a straight stroke, and U+A759 for a diagonal stroke (the *Record Interpreter* appears to use only the former, and neither is listed among the *Statutes* abbreviations):

Ꝛ *quod*; ꝛ *quid*; ꝛꝛ *quibus*.

- q. For *quae*, *que*, use **q** followed by ET (U+A76B) with or without **hlig**: **Ꝟ ꝟ**. For the semicolon-like ET sign (q;), use **cv83[1]**; for the subscripted version (which can also form a ligature via **hlig**), use **cv83[2]**: **Ꝟ ꝟ**.

- r. All of the letters a–z are available in superscript form. Access with the **sup**s OpenType feature:

Ꝟ<sup>o</sup>s *quos*; ꝛ<sup>l</sup>o *circulo*; ꝛ<sup>i</sup> *capituli*.

The basic Latin letters a–z have anchors that allow you to position combining marks over them (see a. above)

- s. Tironian ET sign ꝛ U+204A, cap ꝛ U+2E52. With **cv69[1]** ꝛꝛ; with **cv69[2]** ꝛꝛ.

- t. For *est*, use ꝛ U+223B HOMOTHETIC. Use of a mathematical sign for this purpose is not ideal, but Unicode offers no better solution.

- u. For *tz* (Old French), use **z** U+01B6 Z WITH STROKE.

- v. For an abbreviation for *Rex*, use **R** U+211E or **R̄** U+211F.

- w. At least one edition uses a spacing version of the COMBINING ZIGZAG (f. above). Neither Unicode nor MUFI has a matching character: with Junicode, apply **cv67** to the spacing MACRON (U+00AF): ̄.

## 7.4. 3. Other formatting

- x. For underdotted text, use Stylistic Set 7, Underdotted. For letters that lack an underdotted form, use U+0323 COMBINING DOT BELOW.

## 7.5. 4. On the web

Because Junicode is a very large font, web pages should use a subsetted version to speed loading (see [Chapter 9, Junicode on the Web](#), for instructions). The variable version of the font is better for web use than the static fonts, since one variable font file can do the work of many static font files.

All major web browsers (Firefox, Chrome, Safari, Edge) are capable of accessing all of Junicode's characters via OpenType features, use of which promotes accessibility and searchability. When building a web page, study which features will be needed and write them into the appropriate element or class definition of the page's CSS style sheet. For example, if you use the curly form of the zigzag (U+035B) anywhere, you are likely to want it everywhere, and so it should be included in the CSS styling for the `<body>` element:

```
body {
  font-family: Junicode;
  font-feature-settings: "cv81" 2;
}
```

But the `hlig` feature, if applied to the whole text, will produce many unwanted effects, so it should be included in a class definition to be used in a `<span>` applied just to the target sequence:

```
.que {
  font-feature-settings: "hlig" on;
}
filio<span class="que">q&#xA76B;</span>
```

The illustrations here use the low-level CSS `font-feature-settings` property. There are higher-level properties for some OpenType features, but as these are not (yet) universally supported by browsers, and some implementations are buggy, it is best to stick with `font-feature-settings` for now.

For the purposes addressed in this document, the `font-feature-settings` for the `<body>` element should probably be as follows:

```
font-feature-settings: 'cv69' 2, 'cv80' 1, 'cv81' 2;
```

And the following classes should be defined:

```
.super {  
  font-feature-settings: 'supr' on, 'cv84' 39;  
}
```

```
.que {  
  font-feature-settings: 'hlig' on;  
}
```

```
.deleted {  
  font-feature-settings: 'ss07' on;  
}
```

## 8. The Enlarge Axis

The character recommendation of the Medieval Unicode Font Initiative (MUFI) includes a class of characters called “Enlarged Minuscules,” for representing characters that are lowercase in shape but intermediate between lowercase and uppercase in size: these are often used to begin sentences in medieval manuscripts. MUFI encodes these characters in the Private Use Area, posing accessibility and searchability problems, as explained in the introduction to the “Feature Reference” chapter of this manual.

Junicond provides a solution to these problems via the OpenType feature Stylistic Set 6 (ss06, “Enlarged minuscules”). This feature also works in JunicondVF, the variable version of Junicond, which in addition offers a far more flexible way of representing enlarged minuscules—the Enlarge axis.

An “axis” is an aspect of a font that can be varied along a numerical range. A family of traditional fonts like Times New Roman has a weight axis with a font file on either end: Regular and Bold. Other font families have more weights along this axis: for example, Light, Medium, ExtraBold. Most variable fonts also have a weight axis, but all weights are contained in a single file, and users are not restricted to just a few weights, but can select any weight between the extremes.

Because almost every font family has at least two weights, Weight is the most familiar axis. But several other axes are frequently found in both variable fonts and extended font families. Junicond has Weight and Width axes (Width varying from 75 Condensed to 125 Expanded, with 100 Regular in the middle), and the variable font also has an Enlarge axis, which can vary the size of many lowercase letters from that of the font’s capitals to that of the lowercase letters: Just as the size of these sentence-

Ōñs Ōñs Ōñs Ōñs òñs

initial letters varies widely in manuscripts, so it can vary on web pages and in print (though few applications for producing printed documents currently support variable fonts). Notice that the letters are not simply scaled: the proportions change and the weight remains consistent (a lowercase letter scaled up would look too heavy, but a letter scaled via the Enlarged axis will have its original weight at the lower end of the axis and the same weight as a capital at the top).

The Enlarge axis runs from 0 to 100. You can choose any number in that range: to match the effect of sso6 precisely, choose 47.

To use the axis in a web page, declare a CSS class specifying the value for the axis. For example, the second of the examples in the figure above has the axis set to 75:

```
.SentenceInitial {
  font-variation-settings: "wght" 400, "wdth" 100, "ENLA" 75;
}
```

In the text, enclose the first letter of a sentence in a `<span>` with the class “SentenceInitial” (the entity is for insular d):

```
<span class="SentenceInitial">&#xA77A;</span>ñs
```

The result will be an abbreviation that begins with an “Enlarged Minuscule” insular d, precisely matching the look of the second example in the figure above.

## 9. Junicode on the Web

If you are using Junicode on a web page, you should prefer the variable fonts (those with “VF” in the family name and filename) to the static fonts. One variable font file can do the work of many traditional font files. For example, the [Test of High-Level CSS Properties](#) web page displays Junicode in regular, bold and semicondensed styles. It used to be that your user would have to download three font files, one for each style, but one variable font will display all three.

But you may be thinking, *That font is big!* It’s true: even the compressed webfont (.woff2) is nearly a megabyte in size—enough to seriously slow down the loading of a web page.

To solve this problem, you’ll need to subset the font—that is, produce a copy that contains only what you need. The subsetting font that downloads with the property test web page is approximately 275k in size—almost thirty percent of the size of the full webfont. It’s still a pretty big download, but that’s because the page displays a lot of the font’s features. If I were displaying, say, a diplomatic transcript of a Latin text, the font would be much smaller.

So the first section of this chapter will talk about how to subset the Junicode font.

### 9.1. Subsetting Junicode

First the legalities. It is perfectly all right to create a modified version of Junicode via subsetting, compress it into a webfont (almost certainly in woff2 format), and host it on your web server. This is because “Junicode” is not a “Reserved Font Name” (which complicates web use of many fonts licensed under the Open Font License). If you are nervous about the legal requirements of the Open Font License, you can change the font name to something arbitrary with the `--obfuscate-names` option of the

pyftsubset program, and you can embed the Open Font License, or a link to it, in your CSS. These steps should settle any ambiguity about whether you are in compliance with the license.

Generating a subsetted version of Junicode should be one of the last tasks you perform before deploying your web page(s). Until then, it is recommended that you work with the unmodified font. After subsetting, review your pages thoroughly to make sure everything is displayed properly. If you have forgotten to include a glyph in your subsetted font, you will see little boxes where characters should be or, perhaps, the correct characters displayed in the wrong typeface. If you have omitted features, you will see default instead of transformed characters.

There are many subsetting programs, some online and very easy to use. But for maximum control (and thus the smallest fonts), you should choose pyftsubset, a part of the [fontTools](#) library, which runs under Python 3.7 or higher. This is a command-line tool which takes a long list of arguments; you should create a shell script to run it.

Here is the script used to create the subsetted font for the property test web page mentioned above:

```
#!/bin/zsh
pyftsubset JunicodeTwoBetaVF-Roman.ttf \
--flavor=woff2 --output-file=JunicodeVFsubset.woff2 \
--recommended-glyphs \
--text="jq" --text-file=Junicode-2-feature-test.html \
--layout-features+=liga,ss01,ss02,ss03,ss04,ss05,ss06,ss07,ss08,\
ss10,ss12,ss13,ss14,ss15,ss16,ss17,ss18,ss19,ss20,cv01,cv02,cv05,\
cv06,cv07,cv08,cv09,cv10,dlig,hlig,onum,pnum,pcap,smcp,c2sc,subs,\
supers,zero \
--layout-features-=rlig
```

For those unfamiliar with shell scripts, the first line specifies the shell the script is to run under (in this case the default shell for Mac OS, but bash is another possible choice), and the backslashes mean that the command continues on the next line. The rest of the file is a list of arguments passed to pyftsubset. Let's walk through them.

First after the program name comes the name of the unsubsetted, uncompressed font file. After that, the `--flavor` argument tells the program that you want a web-



font in woff2 format, and `--output` is the name of the font file you want the program to save.

Having taken care of this preliminary business, we tell `pyftsubset` what we want the font to contain.

`--recommended-glyphs` includes a few characters that every font should have, according to the OpenType specification—though in fact modern browsers don’t care. It’s best, however, to conform to the specification, since it’s impossible to say with absolute certainty that no program will ever reject the font because of the absence of these few characters.

`--text-file` is the name of a file to treat as a list of characters that *must* be included in the font. In this case I have simply used the `html` file for the web page for this purpose. If your site contains multiple web pages, your job will be more complicated. You must make sure the text file contains all the characters used on the site—either that or supplement the text file with a `--text` argument (which here adds two lower-case letters that don’t appear in the web page—just in case). The text file will contain only encoded characters—you don’t have to worry here about unencoded characters produced by OpenType features.

`--layout-features+` tells the program which OpenType features you want to retain in the font. All others, except for the [Required Features](#), are discarded. All of the characters referenced in these features will also be included in the output file, as long as those characters are variants of characters in your text file. For example, the `smcp` (Small Caps) feature has many more small caps than there are letters of the alphabet, but most of them are not included in the subsetted font. The program’s parsimony with characters keeps the font file as small as possible. Note that some features are included automatically: `ccmp`, `locl`, `calt`, `liga`, `rlig`, `kern`, `mark`, and `mkmk`.

`--layout-features-` tells the program which OpenType features to omit. Normally, `rlig` (Required Ligatures) is automatically included in fonts by `pyftsubset`, but as it has no relevance to this web page, it can be omitted.

These are the most useful arguments, but there are many more. Type `pyftsubset --help` for a complete list. Once you have written your script, run it (in Mac OS or Linux you need to make the file executable by typing `chmod +x mysubsetsript` on the command line).

Before you put your subsetted font to work, check it carefully in a program like [FontGoggles](#), which lets you preview the font and test all its OpenType features. If

you find errors, revise your script and run it again.

## 9.2. Junicode and CSS/HTML

This section assumes a basic knowledge of HTML (Hypertext Markup Language, used to construct web pages) and CSS (Cascading Style Sheets, used to format them). If you want to learn about these subjects, the number of good books and online tutorials is so great that it makes no sense to try to list them. Just make sure that the instructional materials you choose are of recent vintage, because the relevant standards are always changing.

In the CSS for your web page, the `@font-face` at-rule for a variable font is a little different from the one for a static font in that the range of possible values for each axis can be declared:

```
@font-face {
    font-family: "Junicode Two Beta VF";
    src: url("../webfiles/JunicodeVFsubset.woff2");
    font-weight: 300 700;
    font-stretch: 75% 125%;
    font-style: normal;
}
```

These ranges are not strictly necessary, but they will prevent your supplying invalid values for `font-weight` and `font-stretch` (that is, width) in other CSS rules.

Once you have declared the font, you can invoke it in setting up classes. For example:

```
body {
    font-family: "Junicode Two Beta VF";
    font-size: 28px;
    font-weight: normal; /* that is, 400 */
    font-stretch: 112.5%; /* that is, semiexpanded */
}
h1 {
    font-family: "Junicode Two Beta VF";
    font-size: 125%;
```

```

    font-weight: 600; /* that is, semibold */
    font-stretch: 112.5%; /* that is, semiexpanded */
}
.annotation {
    font-size: 90%;
    font-weight: 300; /* that is, light */
    font-stretch: 87.5%; /* that is, semicondensed */
}

```

These classes should be tested in all browsers. If any fail to display text properly, you can use `font-variation-settings` instead of the high-level `font-weight` and `font-stretch`:

```

body {
    font-family: "Junicode Two Beta VF";
    font-size: 28px;
    font-variation-settings: "wght" 400, "wdth" 112.5;
}
h1 {
    font-family: "Junicode Two Beta VF";
    font-size: 125%;
    font-variation-settings: "wght" 600, "wdth" 112.5;
}
.annotation {
    font-size: 90%;
    font-variation-settings: "wght" 300, "wdth" 87.5;
}

```

To accommodate older browsers, you should make a selection of Junicode static fonts, subset them, and include them in your CSS. For example, if you need normal and bold weights of Junicode roman, your `@font-face` at-rule may look like this:

```

@font-face {
    font-family: "Junicode Two Beta VF";
    src: url("../webfiles/JunicodeVFsubset.woff2");
    font-weight: 300 700;
    font-stretch: 75% 125%;
}

```

```

        font-style: normal;
    }
    @font-face {
        font-family: "Junicode Two Beta";
        src: url("../webfiles/JunicodeTwoBeta-Regular.woff2");
        font-weight: 400;
        font-style: normal;
    }
    @font-face {
        font-family: "Junicode Two Beta";
        src: url("../webfiles/JunicodeTwoBeta-Bold.woff2");
        font-weight: 700;
        font-style: normal;
    }

```

Now use `@supports` in your CSS rules to determine which font gets downloaded:

```

body {
    font-family: "Junicode Two Beta", serif;
}
@supports (font-variation-settings: normal) {
    body {
        font-family: "Junicode Two Beta VF", serif;
    }
}
b {
    font-weight: 700;
}
@supports (font-variation-settings: "wght" 700) {
    b {
        font-variation-settings: "wght" 700;
    }
}

```

The variable version of Junicode will be downloaded only if the browser supports it, and the static version will be downloaded only if needed.

*This document was set in 12pt Junicode SemiExpanded  
using the X<sub>Y</sub>L<sub>A</sub>T<sub>E</sub>X typesetting system with fontspec for font management.  
The source for the document, Feature\_Reference.tex, is available at  
<https://github.com/psb1558/Junicode-font>.*