Proposal for a Latin Script Root Zone LGR

LGR Version 4.0
Date: 2019-10-11
Document version: 5.0
Authors: Latin Generation Panel

Table of Content

1. General Information

The purpose of this document is to give an overview of the proposed LGR in the XML format and the rationale behind the design decisions taken.
It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used, and information on the contributors.
The formal specification of the LGR can be found in the accompanying XML document:
proposal-lgr-latin-20180910.xml
The test labels of the LGR can be found in the accompanying file:
TO BE DEVELOPED

2. Script for Which the LGR is Proposed

The Latin script has the following specifications:

- ISO 15924 code: Latn

- ISO 15924 no.: 215
- ISO 15924 English Name: Latin

Native name of the script:

- It is written differently in different languages.

A partial list of script names in different languages is given below:

- Latin (English, French),
- Latein (German),
- Latino (Italian, Portuguese),
- Latín (Spanish)
- Latinica (Croatian, Serbian)
- Kịch bản latin (Vietnamese)
- Umbhalo we-latin (Zulu)

Maximal Starting Repertoire (MSR) version: MSR-4

As per the *Procedure to Develop and Maintain the Label Generation Rules for the DNS Root Zone in Respect of IDNA Labels* (referred to simply as [Procedure] in the following), only code points included in the latest version of the Maximal Starting Repertoire (currently version 4 and referred to simply as [MSR] in the following) were considered.

The set of code points in the Latin script, as specified by [MSR], contains 346 selected code points, i.e. 326 letters and 20 Combining Diacritical Marks. Code points are from the following Unicode ranges as listed in table 1 below. [MSR] excludes the Unicode ranges listed in table 2 below.

Table 1. Unicode ranges included in [MSR].

Latin Script	Range of Unicode code points
Controls and Basic Latin	U+0061 – U+007A
Controls and Latin-1 Supplement	U+00DF - U+00F6 U+00F8 - U+00FF
Latin Extended-A	U+0101 – U+017F
Latin Extended-B	U+0180 – U+024F
IPA Extensions	U+0250 – U+02AF
Combining Diacritical Marks	U+0300 – U+036F
Combining Diacritical Marks Supplement	U+1DC0 – U+1DFF
Latin Extended Additional	U+1E00 – U+1EFF
Latin Extended-C	U+2C60 – U+2C7F

Table 2. Unicode ranges excluded from [MSR].

Latin Script	Range of Unicode code points
Latin Extended-D; technical use (phonetic)/obsolete/punctuation	U+A720 – U+A7FF

Latin Ligatures; compatibility characters not PVALID in IDNA 2008	U+FB00 – U+FB0F
Full-width Latin Letters; compatibility characters not PVALID in IDNA 2008	U+FF00 – U+FF5E

When a single, precomposed code point is equivalent to the combination of letter code point and a diacritic mark code point, only the precomposed code point may be used as per [IDNA 2008]. Furthermore, only lower case letters are considered in creating the repertoire, as upper case ones may not be used in IDNs following [IDNA 2008].

3. Background on Script and Principal Languages Using It

The Latin script[1] is a major writing system of the world today, and the most widely used in terms of number of languages and number of speakers, with circa 70% of the world's readers and writers making use of this script[2] [Wikipedia-Latin script].

3.1 Principal Languages Using Latin Script

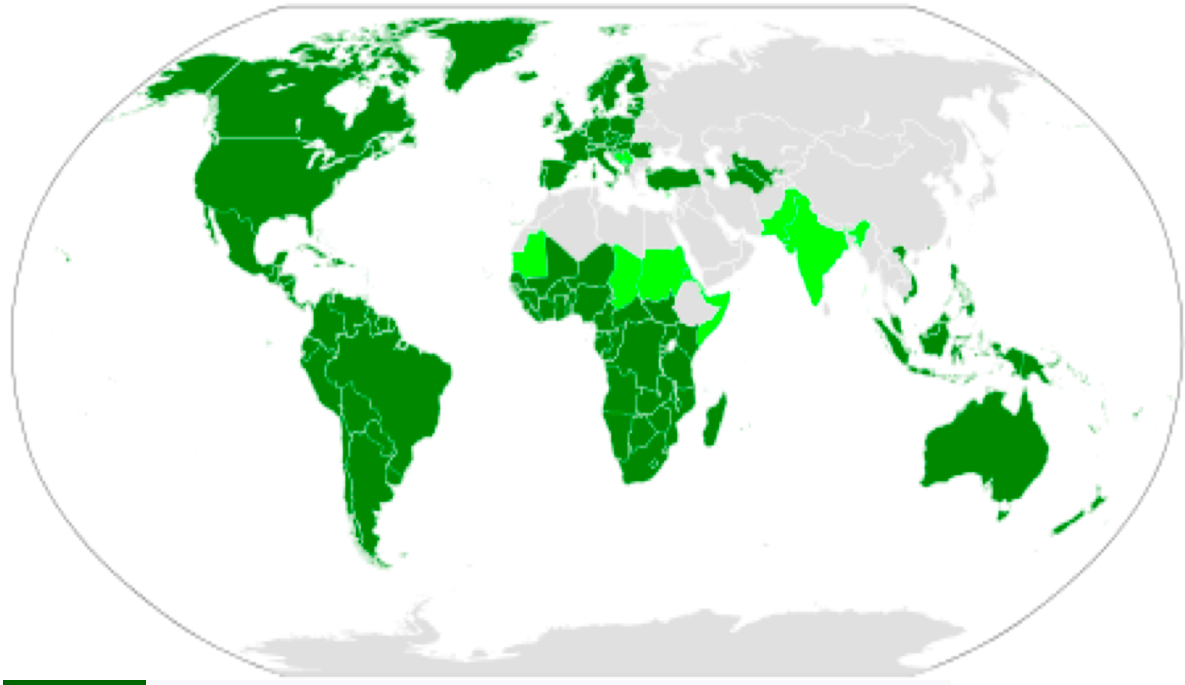
The list of languages taken into consideration contains relevant data for 455 languages using Latin script. The table with languages using Latin script was derived using data from http://www.omniglot.com/writing/langalph.htm and https://www.ethnologue.com/browse/names and was attached to "Proposal for Generation Panel for Latin Script Label Generation Ruleset for the Root Zone".

3.2 Geographic Territories or Countries With Significant User Communities

Per Wikipedia the distribution of the Latin script on the world map is:

[1] *Script* is used here to indicate the whole writing system including basic letters, ligatures and diacritics. See also RFC 6365 and ISO 15924.

[2] However, several orthographies on the basis of different scripts are frequently used simultaneously, both historically and contemporarily.



Dark green marks countries where the Latin script is the sole main script.
Light green marks countries where Latin co-exists with other scripts.
Grey marks areas, in which supposedly Latin-script is not used or used only unofficially for second language.

3.3 Related Scripts

Latin GP has agreed that following scripts are directly related to Latin script, as all are ultimately derived from Phoenician:

- Cyrillic
- Greek
- Armenian

4. Overall Development Process and Methodology

The work has been done according to the work plan given in "Proposal for the Generation Panel (GP) for the Latin Script Label Generation Ruleset (LGR) for the Root Zone".
The panel formed two working groups:

- Repertoire WG
- Variant WG

which worked in parallel.

First task for each group was to define the Principles for developing Repertoire and the Principles for developing Variants. Principles were sent to Integration panel for comments and suggestions and were also offered for public unofficial comment. Comments from Integration panel were encompassed in final version of Principles.

During the Repertoire definition phase, the Panel reviewed and processed 181 languages with EGIDS level 1 through 4, and 29 languages with EGIDS Level 5, which have more than 1, 000, 000 speakers. The processed languages are listed in Appendix B.

The Latin Generation Panel used [MSR] as the starting point and after processing 210 languages Latin GP found:

- 193 MSR Unicode code points verified
- 22 Code Point Sequences (defined below) detected
- 6 New code points added to MSR

The panel also found some languages that use letters matching code points outside [MSR]. In some cases, they were rejected and in some cases the panel made successful requests for inclusion in [MSR]. This is described in more detail in Appendix A.
The second phase of Latin GP work was mainly devoted to defining in-script and cross-script Variants.

5. Repertoire

Based on the discussions within the GP, the principles for inclusion and exclusion of code points in the Repertoire are as follows.

5.1 Definitions

Language: The present document and its principles deal with any language making use of Latin script[3] today. Languages are restricted to natural human languages in active use. Both the socio-political situation (such as the political or legal status of a language in a country or community) and the socio-linguistic roles of languages in society (such as the absolute or relative frequency of use) are explicitly not considered for the current purposes. Super- or sub-units of languages, such as dialect, regiolect (a dialect spoken in a particular geographical region), or language clusters, are all considered equivalent to language. However, notions such as official language, national language, standard language and vernacular, are not considered at all in determining whether something is a language.
Letter Code Point is a Unicode code point with General Category property value of Lx (Lu, Ll, Lt, Lm, Lo), as defined in the Unicode Character Database.
Mark Code Point is a Unicode code point with General Category property value of Mx (Mn, Mc, Me), as defined in the Unicode Character Database.
Code Point Sequence is a sequence of two or more Code Points (e.g. Letter Code Point followed by one or more Mark Code Point(s)).
Established contemporary use of a letter means it is in active use by a community today. Such use may be demonstrated by, for example, educational resources, published material, media, or other materials and sources. This does not depend on their material or non-material form, such as handwritten or typed manuscripts or digitally produced text. There may be multiple sources for acquiring such evidence, including (but not limited to) the following:

- Members of Language communities,
- Members of the Latin GP,
- Other experts
- Language tables submitted by ccTLD in the context of IDNA 2008 in the IANA repository, and
- Published standards (e.g. by a language authority or any other national or international body).

[3] Latin script is also known as Roman script in academic literature.

5.2 Principles for Developing Repertoire

5.2.1 Inclusion Principles

If a Code Point is included and delegated as part of the label, the Code Point cannot be retracted in future revisions of the LGR. All applicable criteria must be met to include a Code Point.

1. Only languages which have a rating of levels of 0-4 under the Expanded Graded Intergenerational Disruption Scale (EGIDS) are considered as supporting the inclusion of a Code Point. Languages with EGIDS 5 may be included in special cases where there is additional evidence that it is in widespread use, notwithstanding its formal EGIDS rating.
2. Code Points may only be included if they have established contemporary use in one or more of the languages considered.
3. If the Code Point in question is a Mark Code Point, then it can only be included in its context. That is, a Mark Code Point is included as part of a sequence consisting of a Lower Letter (Ll) or Other Letter (Lo) and the subsequent mark or marks. (See Section 5.3.1)
4. Any combination of Code Points is defined by its sequence. To be included, a sequence must be supported by some included language in the same way as a separate Code Point of type Ll or Lo.
5. If a character can be represented by multiple Code Point Sequences, each Code Point Sequence must be separately justified to be included.
6. A Code Point Sequence can only be included if there is no pre-composed alternative available unless there is specific evidence that a language eligible for inclusion under Criterion 1 makes alternate use of such a sequence.
7. If the Code Point in question is a Modifier letter (Lm), then it can only be included together with its context. That is a sequence of Lm plus Ll or Lo (or the other way around), unless there is strong evidence that the Lm can be used in any context, or that such a sequence or order cannot be defined.

5.2.2 Exclusion Principles

A Code Point is excluded if at least one of these exclusion principles is met. If a Code Point can neither be included nor excluded on the basis of these principles, the Code Point is automatically excluded from the proposed LGR for Latin Script, per RFC 6912.

1. The Code Point is DISALLOWED or UNASSIGNED by IDNA 2008 protocol.
2. The Code Point presents a security or stability issue which cannot be resolved at any other stage of the analysis (e.g., stage of determining Code Points, variants, Contextual Rules or Whole Label Evaluation Rules).
3. The Code Point is either deprecated or not recommended for use in Unicode Standard -- unless it meets all of the applicable inclusion criteria, with no alternative Code Point or Code Point sequence.
4. The Code Point is used exclusively in a subset of textual genres, such as technical or religious texts, and is not otherwise used as described in Section 2 above.
5. The Code Point is predominantly used in one of the following functions, apart from any other uses in orthography**:**
    a. Formatting character or mark
    b. Numerical digit
    c. Punctuation mark

d. Honorific mark or symbol
e. Mathematical symbol

5.3 Code Points Included

The table below lists the code points proposed for inclusion in the root zone LGR for the Latin script. The table also lists examples of languages using the code point and their EGIDS rating. All references for specific code points found during language processing are included.
This table is sorted by Unicode column.
The table with the same data, sorted by glyph, can be found in Appendix C.
Description of References supporting inclusion of code point is in section 9.1

Table 3. Code Points Included in the Repertoire of Latin Script LGR.

#	Unicode	Glyph	Unicode name	Languages using the code point (EGIDS)	Reference supporting inclusion (URL etc.)
1.	0061	a	LATIN SMALL LETTER A	Basic Latin	[0]
2.	0061 + 0331	a̲	LATIN SMALL LETTER A + COMBINING MACRON BELOW	Nuer (4)	[146], [129]
3.	0062	b	LATIN SMALL LETTER B	Basic Latin	[0]
4.	0063	c	LATIN SMALL LETTER C	Basic Latin	[0]
5.	0064	d	LATIN SMALL LETTER D	Basic Latin	[0]
6.	0065	e	LATIN SMALL LETTER E	Basic Latin	[0]
7.	0065 + 0331	e̲	LATIN SMALL LETTER E + COMBINING MACRON BELOW	Nuer (4)	[146]
8.	0066	f	LATIN SMALL LETTER F	Basic Latin	[0]
9.	0067	g	LATIN SMALL LETTER G	Basic Latin	[0]

10.	0067 + 0303	g̃	LATIN SMALL LETTER G + COMBINING TILDE	Guarani (1)	[142], [143]
11.	0068	h	LATIN SMALL LETTER H	Basic Latin	[0]
12.	0069	i	LATIN SMALL LETTER I	Basic Latin	[0]
13.	0069 + 0331	i̲	LATIN SMALL LETTER I + COMBINING MACRON BELOW	Nuer (4)	[146]
14.	006A	j	LATIN SMALL LETTER J	Basic Latin	[0]
15.	006B	k	LATIN SMALL LETTER K	Basic Latin	[0]
16.	006C	l	LATIN SMALL LETTER L	Basic Latin	[0]
17.	006D	m	LATIN SMALL LETTER M	Basic Latin	[0]
18.	006D + 0327	m̧	LATIN SMALL LETTER M + COMBINING CEDILLA	Marshallese (1)	[213], [136], [214]
19.	006E	n	LATIN SMALL LETTER N	Basic Latin	[0]
20.	006E + 0304	n̄	LATIN SMALL LETTER N + COMBINING MACRON	Raga (Hano) (3) Marshallese (1)	[200], [213], [136]
21.	006E + 0308	n̈	LATIN SMALL LETTER N + COMBINING DIAERESIS	Malagasy (1)	[230]
22.	006F	o	LATIN SMALL LETTER O	Basic Latin	[0]

23.	006F + 0327	o̧	LATIN SMALL LETTER O + COMBINING CEDILLA	Marshallese (1)	[136]
24.	006F + 0331	o̱	LATIN SMALL LETTER O + COMBINING MACRON BELOW	Nuer (4)	[146], [129]
25.	0070	p	LATIN SMALL LETTER P	Basic Latin	[0]
26.	0071	q	LATIN SMALL LETTER Q	Basic Latin	[0]
27.	0072	r	LATIN SMALL LETTER R	Basic Latin	[0]
28.	0072 + 0303	r̃	LATIN SMALL LETTER R WITH COMBINING TILDE	Hausa (2)	[147]
29.	0073	s	LATIN SMALL LETTER S	Basic Latin	[0]
30.	0074	t	LATIN SMALL LETTER T	Basic Latin	[0]
31.	0075	u	LATIN SMALL LETTER U	Basic Latin	[0]
32.	0076	v	LATIN SMALL LETTER V	Basic Latin	[0]
33.	0077	w	LATIN SMALL LETTER W	Basic Latin	[0]
34.	0078	x	LATIN SMALL LETTER X	Basic Latin	[0]
35.	0079	y	LATIN SMALL LETTER Y	Basic Latin	[0]
36.	007A	z	LATIN SMALL LETTER Z	Basic Latin	[0]
37.	00DF	ß	LATIN SMALL LETTER SHARP S	German (1)	[119]

38.	00E0	à	LATIN SMALL LETTER A WITH GRAVE	Italian (1) Galician (2) Wolof (4)	[130], [131], [106], [132]
39.	00E1	á	LATIN SMALL LETTER A WITH ACUTE	Spanish (1) French (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Chuukese (2) Galician (2) Lule Sámi (2) Northern Sámi (2)	[100], [101], [102], [103], [104], [105], [106], [107], [108], [114]
40.	00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	Vietnamese (1) Romanian (1) Skolt Sami (2) Kirundi (1) French (1) Galician (2) West Frisian (2) Friulian (4) Xavante (4)	[109], [110], [113], [104], [114], [106], [115], [116], [117]
41.	00E3	ã	LATIN SMALL LETTER A WITH TILDE	Umbundu (3) Guarani (1) Nauruan (3) Khoekhoe (4)	[141], [142], [143], [144], [145]
42.	00E4	ä	LATIN SMALL LETTER A WITH DIAERESIS	German (1) Finnish (1) Turkmen (1) Estonian (1) Swedish (1) Lule Sámi (2) Yapese (2) Dinka (4) Kaqchikel (4) Bashkir (4) Alsatian (5) Nuer (4)	[119], [120], [121], [122], [123], [107], [124], [125], [126], [127], [128], [129]
43.	00E5	å	LATIN SMALL LETTER A WITH RING ABOVE	Danish (1) Finnish (1) Chamorro (1) Swedish (1) Lule Sámi (2)	[139], [120], [140], [123], [107]

44.	00E6	æ	LATIN SMALL LETTER AE	Danish (1) Icelandic (1) Faroese (2)	[139], [102], [103]
45.	00E7	ç	LATIN SMALL LETTER C WITH CEDILLA	Turkish (1) Turkmen (1) Kurdish (2) French (1) Azerbaijani (1) Basque (1) Galician (2) Friulian (4) Bashkir(4)	[157], [121], [158], [114], [159], [160], [161], [106], [116], [127]
46.	00E8	è	LATIN SMALL LETTER E WITH GRAVE	French (1) Italian (1) Afrikaans (1) Kirundi (1) Haitian Creole (1) French (1)	[114], [130], [175], [104], [182], [183]
47.	00E9	é	LATIN SMALL LETTER E WITH ACUTE	French (1) Italian (1) Spanish (1) Czech (1) Icelandic (1) Kirundi (1) Chuukese (2) Galician (2) Wolof (4) XAVANTE (4) West Frisian (2)	[114], [130], [100], [101], [102], [104], [105], [106], [132], [117], [115]
48.	00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX	French (1) Tswana (1) Afrikaans (1) Vietnamese (1) Kurdish (2) Kirundi (1) West Frisian (2) Friulian (4)	[114], [173], [174], [175], [109], [158], [104], [115], [116]
49.	00EB	ë	LATIN SMALL LETTER E WITH DIAERESIS	Afrikaans (1) Kirundi (1) Albanian (1) French (1) Chuukese (2) Uyghur (2)	[175], [104],[176], [177], [114], [178], [179], [124], [132], [180], [126], [115], [129]

				Yapese (2) Wolof (4) Drehu (4) Kaqchikel (4) West Frisian (2) Nuer (4)	
50.	00EC	ì	LATIN SMALL LETTER I WITH GRAVE	Italian (1) Kirundi (1)	[130], [206], [208]
51.	00ED	í	LATIN SMALL LETTER I WITH ACUTE	Spanish (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Galician (2) Bashkir(4)	[100], [101], [102], [103], [104], [106], [127]
52.	00EE	î	LATIN SMALL LETTER I WITH CIRCUMFLEX	Afrikaans (1) Romanian (1) Kurdish (2) Kirundi (1) French (1) Friulian (4)	[175], [110], [158], [104], [114], [116]
53.	00EF	ï	LATIN SMALL LETTER I WITH DIAERESIS	Afrikaans (1) French (1) Kaqchikel (4) Dinka (4) West Frisian (2)	[175], [114], [126], [125], [115]
54.	00F0	ð	LATIN SMALL LETTER ETH	Faroese (2) Icelandic (1)	[103], [102]
55.	00F1	ñ	LATIN SMALL LETTER N WITH TILDE	Spanish (1) Pulaar (3) Chamorro (1) Filipino (1) Guarani (1) Chavacano (4) Basque (1) Galician (2) Iloco (3) Quechua (3) Cape Verdean Creole (4) Waray-Waray (3) Wolof (4)	[221], [250],[222], [142], [143], [223], [160], [106], [224], [225], [226], [227], [228], [132], [144], [229], [127], [136], [197], [205]

				Nauruan (3) Lozi (4) Bashkir (4) Marshallese (1) Mandinka (5) Igbo (2)	
56.	00F2	ò	LATIN SMALL LETTER O WITH GRAVE	Italian (1) Haitian Creole (1)	[130], [182], [183]
57.	00F3	ó	LATIN SMALL LETTER O WITH ACUTE	Spanish (1) Polish (1) Czech (1) Icelandic (1) Kirundi (1) Chuukese (2) Galician (2) Wolof (4)	[100], [152], [101], [102], [104], [105], [106], [132]
58.	00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX	Tswana (1) Afrikaans (1) Vietnamese (1) Kirundi (1) French (1) Northern Sotho (1) West Frisian (2) Galician (2) Friulian (4) Xavante(4)	[173], [174], [175], [109], [104], [114], [230], [115], [106], [116], [117]
59.	00F5	õ	LATIN SMALL LETTER O WITH TILDE	Estonian (1) Skolt Sami (2) Umbundu (3) Guarani (1) Nauruan (3) Xavante (4) Khoekhoe (4)	[122], [113], [141], [142], [143], [144], [117], [235]
60.	00F6	ö	LATIN SMALL LETTER O WITH DIAERESIS	German (1) Finnish (1) Afrikaans (1) Turkish (1) Swedish (1) Uygur (2) Yapese (2) Drehu (4) Kaqchikel (4)	[119], [120], [175], [157], [123], [179], [124], [180], [126], [125], [127], [231], [232], [115], [129]

				Dinka (4) Bashkir (4) Chechen (2) 1992 Version West Frisian (2) Nuer (4)	
61.	00F8	ø	LATIN SMALL LETTER O WITH STROKE	Danish (1) Faroese (2)	[139], [103]
62.	00F9	ù	LATIN SMALL LETTER U WITH GRAVE	Italian (1) French (1) Papiamento (1)	[130], [206], [245], [246], [253]
63.	00FA	ú	LATIN SMALL LETTER U WITH ACUTE	Spanish (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Chuukese (2) West Frisian (2) Galician (2)	[100], [101], [102], [103], [104], [105], [115], [106]
64.	00FB	û	LATIN SMALL LETTER U WITH CIRCUMFLEX	Afrikaans (1) Kurdish (2) Kirundi (1) French (1) Miskito (2) West Frisian (2) Friulian (4) Zazaki (4)	[175], [158], [104], [114], [243], [115], [116], [244]
65.	00FC	ü	LATIN SMALL LETTER U WITH DIAERESIS	German (1) Spanish (1) Afrikaans (1) Turkish (1) Swedish (1) French (1) Azeri (1) Basque (1) Galician (2) Uygur (2) Kaqchikel (4) Bashkir (4)	[119], [100], [175], [157], [123], [114], [159], [161], [106], [179], [126], [127], [231]

66.	00FD	ý	LATIN SMALL LETTER Y WITH ACUTE	Turkmen (1) Czech (1) Icelandic (1) Faroese (2) Guarani (1)	[121], [101], [102], [103], [142], [143]
67.	00FE	þ	LATIN SMALL LETTER THORN	Icelandic (1)	[102]
68.	00FF	ÿ	LATIN SMALL LETTER Y WITH DIAERESIS	French (1)	[114], [253], [257]
69.	0101	ā	LATIN SMALL LETTER A WITH MACRON	Latvian (1) Tongan (1) Hawaiian (2) Marshallese (1)	[133], [134], [135], [136]
70.	0103	ă	LATIN SMALL LETTER A WITH BREVE	Vietnamese (1) Romanian (1)	[109], [110]
71.	0105	ą	LATIN SMALL LETTER A WITH OGONEK	Polish (1) Lithuanian (1)	[137], [138]
72.	0107	ć	LATIN SMALL LETTER C WITH ACUTE	Croatian (1) Serbian (1) Polish (1)	[150], [151], [152]
73.	0109	ĉ	LATIN SMALL LETTER C WITH CIRCUMFLEX	Esperanto (3)	[255]
74.	010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE	Maltese (1)	[163]
75.	010D	č	LATIN SMALL LETTER C WITH CARON	Croatian (1) Serbian (1) Latvian (1) Slovak (1) Northern Sámi (2) Lithuanian (1)	[150], [151], [133], [153], [108], [154]
76.	010F	ď	LATIN SMALL LETTER D WITH CARON	Czech (1) Slovak (1)	[101], [153]

77.	0111	đ	LATIN SMALL LETTER D WITH STROKE	Croatian (1) Serbian (1) Vietnamese (1) Northern Sámi (2)	[150], [151], [109], [108], [168]
78.	0113	ē	LATIN SMALL LETTER E WITH MACRON	Latvian (1) Hawaiian (2) Tongan (1) Minangkabau (5)	[133], [135], [134], [184]
79.	0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE	Lithuanian (1)	[138], [154]
80.	0119	ę	LATIN SMALL LETTER E WITH OGONEK	Polish (1) Palauan (2) Lithuanian (1)	[152], [185], [138], [154]
81.	011B	ě	LATIN SMALL LETTER E WITH CARON	Czech (1) Kirundi (1) Sorbian (4)	[101], [104], [172]
82.	011D	ĝ	LATIN SMALL LETTER G WITH CIRCUMFLEX	Esperanto (3)	[255]cute
83.	011F	ğ	LATIN SMALL LETTER G WITH BREVE	Turkish (1) Tatar (2) Azeri (1) Bashkir (4) Zaza (5)	[157], [201], [159], [127], [202]
84.	0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE	Maltese (1)	[163]
85.	0123	ģ	LATIN SMALL LETTER G WITH CEDILLA	Latvian (1) Brahui (5)	[133], [168]
86.	0125	ĥ	LATIN SMALL LETTER H WITH CIRCUMFLEX	Esperanto (3)	[255]
87.	0127	ħ	LATIN SMALL LETTER H WITH STROKE	Maltese (1)	[163]

88.	0129	ĩ	LATIN SMALL LETTER I WITH TILDE	Guarani (1) Cubeo (3) Khoekhoe (4) Kikuyu (5)	[142], [143], [186], [145], [209]
89.	012B	ī	LATIN SMALL LETTER I WITH MACRON	Latvian (1) Lithuanian (1) Hawaiian (2) Tongan (1)	[133], [138], [135], [134]
90.	012F	į	LATIN SMALL LETTER I WITH OGONEK	Lithuanian (1)	[154]
91.	0131	ı	LATIN SMALL LETTER DOTLESS I	Turkish (1) Tatar (2) Azeri (1)	[157], [203], [201], [159]
92.	0135	ĵ	LATIN SMALL LETTER J WITH CIRCUMFLEX	Esperanto (3)	[255]
93.	0137	ķ	LATIN SMALL LETTER K WITH CEDILLA	Latvian (1)	[133]
94.	013A	Í	LATIN SMALL LETTER L WITH ACUTE	Slovak (1)	[153]
95.	013C	ļ	LATIN SMALL LETTER L WITH CEDILLA	Latvian (1) Marshallese (1) Brahui (5)	[133], [213], [214], [168]
96.	013E	ľ	LATIN SMALL LETTER L WITH CARON	Slovak (1)	[153]
97.	0142	ł	LATIN SMALL LETTER L WITH STROKE	Polish (1)	[152]
98.	0144	ń	LATIN SMALL LETTER N WITH ACUTE	Polish (1) Lule Sámi (2) Sorbian (4) Brahui (5)	[152], [107], [172], [168]
99.	0146	ņ	LATIN SMALL LETTER N WITH CEDILLA	Latvian (1) Marshallese (1)	[133], [136]

100.	0148	ň	LATIN SMALL LETTER N WITH CARON	Turkmen (1) Czech (1) Slovak (1)	[121], [101], [153]
101.	014B	ŋ	LATIN SMALL LETTER ENG	Inari Sami (2) Dagaare Burkina Faso (4) Dagbani (Dagomba) (4) Northern Sami (2) Ewondo (3) Luganda (3) Wolof (4) Adzera (4) Nuer (4) Ga (4) Dinka (4) Duala (3) Ewe (3) Soga (5) Alur (5) Mandinka (5) Acholi (5) Bambara (4) Nuer (4)	[188], [148], [189], [108], [190], [191], [132], [192], [146], [193], [125], [194], [170], [195], [196], [197], [198], [199], [129]
102.	014D	ō	LATIN SMALL LETTER O WITH MACRON	Hawaiian (2) Marshallese (1) Tongan (1)	[135], [136], [134]
103.	0151	ő	LATIN SMALL LETTER O WITH DOUBLE ACUTE	Hungarian (1)	[233], [234]
104.	0153	œ	LATIN SMALL LIGATURE OE	French (1)	[114], [253]
105.	0155	ŕ	LATIN SMALL LETTER R WITH ACUTE	Slovak (1) Brahui (5)	[153], [168]
106.	0159	ř	LATIN SMALL LETTER R WITH CARON	Czech (1) Sorbian (4)	[101], [172]
107.	015B	ś	LATIN SMALL LETTER S WITH ACUTE	Polish (1) Montenegrin (1)	[152], [258]

108.	015D	ŝ	LATIN SMALL LETTER S WITH CIRCUMFLEX	Esperanto (3)	[255]
109.	015F	ş	LATIN SMALL LETTER S WITH CEDILLA	Turkish (1) Turkmen (1) Kurdish (2) Tatar (2) Azeri (1) Bashkir (4) Brahui (5) Zaza (5)	[157], [121], [158], [201], [159], [127], [168], [202]
110.	0161	š	LATIN SMALL LETTER S WITH CARON	Tswana (1) Croatian (1) Serbian (1) Latvian (1) Northern Sotho (1) Northern Sami (2) Lithuanian (1)	[174], [150], [151], [133], [230], [108], [154]
111.	0165	ť	LATIN SMALL LETTER T WITH CARON	Czech (1) Slovak (1)	[101], [153]
112.	0167	ŧ	LATIN SMALL LETTER T WITH STROKE	Northern Sami (2) Brahui (5)	[108], [168]
113.	0169	ũ	LATIN SMALL LETTER U WITH TILDE	Umbundu (3) Guarani (1) Nauruan (3) Khoekhoe (4) Kikuyu (5)	[141], [142], [143], [144], [145], [209]
114.	016B	ū	LATIN SMALL LETTER U WITH MACRON	Latvian (1) Hawaiian (2) Lithuanian (1) Marshallese (1) Tongan (1)	[133], [135], [138], [154], [136], [134]
115.	016D	ŭ	LATIN SMALL LETTER U WITH BREVE	Esperanto (3)	[255]
116.	016F	ů	LATIN SMALL LETTER U WITH RING ABOVE	Czech (1)	[101]

117.	0171	ű	LATIN SMALL LETTER U WITH DOUBLE ACUTE	Hungarian (1)	[233], [234]
118.	0173	ų	LATIN SMALL LETTER U WITH OGONEK	Lithuanian (1)	[154], [138]
119.	0175	ŵ	LATIN SMALL LETTER W WITH CIRCUMFLEX	Chichewa (3) Welsh (2)	[247], [256]
120.	0177	ŷ	LATIN SMALL LETTER Y WITH CIRCUMFLEX	Welsh (2)	[256]
121.	017A	ź	LATIN SMALL LETTER Z WITH ACUTE	Polish (1) Brahui (5) Sorbian (4) Montenegrin (1)	[152], [252], [168], [172], [258]
122.	017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE	Polish (1) Maltese (1)	[152], [163]
123.	017E	ž	LATIN SMALL LETTER Z WITH CARON	Lithuanian (1) Croatian (1) Serbian (1) Turkmen (1) Latvian (1) Slovak (1) Northern Sami (2) Chechen (2) 1925 Version	[154], [150], [151], [121], [133], [153], [108], [232]
124.	0192	ƒ	LATIN SMALL LETTER F WITH HOOK	Ewe (3)	[170]
125.	0199	ƙ	LATIN SMALL LETTER K WITH HOOK	Hausa (2)	[147]
126.	01A1	ơ	LATIN SMALL LETTER O WITH HORN	Vietnamese (1)	[109]

127.	01B0	ư	LATIN SMALL LETTER U WITH HORN	Vietnamese (1)	[109]
128.	01B4	ƴ	LATIN SMALL LETTER Y WITH HOOK	Dagaare-Burkina Faso (4) Fula (3)	[148], [251], [149]
129.	01CE	ǎ	LATIN SMALL LETTER A WITH CARON	Kirundi (1)	[104] **https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx** Jean Paul Nkurunziza (personal communication)
130.	01D0	ǐ	LATIN SMALL LETTER I WITH CARON	Kirundi (1)	[104]
131.	01D2	ǒ	LATIN SMALL LETTER O WITH CARON	Kirundi (1)	[104]
132.	01D4	ǔ	LATIN SMALL LETTER U WITH CARON	Kirundi (1)	[104]
133.	01DD	ə	LATIN SMALL LETTER TURNED E	Kanuri (3)	[240]
134.	01E7	ǧ	LATIN SMALL LETTER G WITH CARON	Skolt Sami (2)	[113]
135.	01E9	ǩ	LATIN SMALL LETTER K WITH CARON	Skolt Sami (2)	[113]
136.	01EF	ǯ	LATIN SMALL LETTER EZH WITH CARON	Skolt Sami (2)	[113]
137.	0219	ș	LATIN SMALL LETTER S WITH COMMA BELOW	Romanian (1)	[110]

138.	021B	ț	LATIN SMALL LETTER T WITH COMMA BELOW	Romanian (1)	[110]
139.	024D	ɍ	LATIN SMALL LETTER R WITH STROKE	Kanuri (3)	[240]
140.	0253	ɓ	LATIN SMALL LETTER B WITH HOOK	Hausa (2) Dagaare-Burkina Faso (4) Pulaar, (3)	[147], [148], [250]
141.	0254	ɔ	LATIN SMALL LETTER OPEN O	Dagaare - Burkina Faso (4) Dagbani (Dagomba) (4) Lingala (2) Akan (3) Ewondo (3) Fon (3) Nuer (4) Ga (4) Duala (3) Ewe (3) Nuer (4)	[148], [189], [236], [237], [190], [169], [146], [193], [194], [170], [129]
142.	0254 + 0308	ɔ̈	LATIN SMALL LETTER OPEN O + COMBINING DIAERESIS	Dinka (4)	[125]
143.	0254 + 0331	ɔ̱	LATIN SMALL LETTER OPEN O + COMBINING MACRON BELOW	Nuer (4)	[129], [146]
144.	0256	ɖ	LATIN SMALL LETTER D WITH TAIL	Fon (3) Ewe (3)	[169], [170]
145.	0257	ɗ	LATIN SMALL LETTER D WITH HOOK	Hausa (2) Pulaar (3)	[147], [166], [250]
146.	0259	ə	LATIN SMALL LETTER SCHWA	Azeri, Azerbaijani (1) Ewondo (3) Ewe (3) Bugis (3)	[159], [190], [170], [241]

147.	025B	ɛ	LATIN SMALL LETTER OPEN E	Dagaare - Burkina Faso (4) Lingala (2) Akan (3) Ewondo (3) Dagbani (Dagomba) (4) Fon (3) Mossi (3) Ga (4) Ewe (3) Duala (3) Bambara (4) Nuer (4)	[148], [236], [237], [190], [189], [169], [212], [238], [193], [170], [194], [199], [129]
148.	025B + 0308	ɛ̈	LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS	Nuer (4) Dinka (4)	[129], [146], [239], [125]
149.	025B + 0331	ɛ̱	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW	Nuer (4)	[129], [146], [239]
150.	025B + 0331 + 0308	ɛ̱̈	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS	Nuer (4)	[146], [239]
151.	0263	ɣ	LATIN SMALL LETTER GAMMA	Dagbani (Dagomba) (4) Nuer (4) Dinka (4) Ewe (3) Nuer (4)	[189], [146], [125], [170], [129]
152.	0268	ɨ	LATIN SMALL LETTER I WITH STROKE	Cubeo (3) Dagbani (Dagomba) (4) HIxkaryána (4) Maasai (5)	[186], [189], [210], [211]
153.	0268 + 0303	ɨ̃	LATIN SMALL LETTER I WITH STROKE +	Cubeo (3)	[186]

			COMBINING TILDE		
154.	0269	ɩ	LATIN SMALL LETTER IOTA	Dagaare - Burkina Faso (4) Mossi (3)	[148], [212]
155.	0272	ɲ	LATIN SMALL LETTER N WITH LEFT HOOK	Susu (4) Zarma (4) Bambara (4)	[218], [219], [199]
156.	0289	ʉ	LATIN SMALL LETTER U BAR	Cubeo (3) Maasai (5)	[186], [187], [211]
157.	0289 + 0303	ʉ̃	LATIN SMALL LETTER U BAR + COMBINING TILDE	Cubeo (3)	[186], [187]
158.	028B	ʋ	LATIN SMALL LETTER V WITH HOOK	Dagaare - Burkina Faso (4) Mossi (3) Ewe (3)	[148], [212], [238], [170]
159.	0292	ʒ	LATIN SMALL LETTER EZH	Skolt Sami (2) Dagbani (Dagomba) (4)	[113], [189]
160.	1E13	ḓ	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
161.	1E21	ḡ	LATIN SMALL LETTER G + MACRON	Raga (Hano) (3)	[200]
162.	1E37	ḷ	LATIN SMALL LETTER L WITH DOT BELOW	Marshallese (1)	[213], [214], [215], [216]
163.	1E3D	ḽ	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
164.	1E43	ṃ	LATIN SMALL LETTER M WITH DOT BELOW	Marshallese (1)	[213], [136], [215], [216]

165.	1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE	Venda (1)	[164], [257]
166.	1E47	ṇ	LATIN SMALL LETTER N WITH DOT BELOW	Marshallese (1)	[136], [215], [216]
167.	1E49	n̲	LATIN SMALL LETTER N WITH LINE BELOW	Pitjantjatjara (4)	[220]
168.	1E4B	ṉ̭	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
169.	1E63	ṣ	LATIN SMALL LETTER S WITH DOT BELOW	Yoruba (2)	[181]
170.	1E6D	ṭ	LATIN SMALL LETTER T WITH DOT BELOW	Mizo (4)	[242]
171.	1E71	ṯ̭	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
172.	1E8D	ẍ	LATIN SMALL LETTER X WITH DIAERESIS	Mam (4)	[248], [249]
173.	1EA1	ạ	LATIN SMALL LETTER A WITH DOT BELOW	Vietnamese (1)	[109]
174.	1EA3	ả	LATIN SMALL LETTER A WITH HOOK ABOVE	Vietnamese (1)	[109]
175.	1EA5	ấ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]
176.	1EA7	ầ	LATIN SMALL LETTER A WITH	Vietnamese (1)	[109]

			CIRCUMFLEX AND GRAVE		
177.	1EA9	ẩ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
178.	1EAB	ẫ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]
179.	1EAD	ậ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]
180.	1EAF	ắ	LATIN SMALL LETTER A WITH BREVE AND ACUTE	Vietnamese (1)	[109]
181.	1EB1	ằ	LATIN SMALL LETTER A WITH BREVE AND GRAVE	Vietnamese (1)	[109]
182.	1EB3	ẳ	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	Vietnamese (1)	[109]
183.	1EB5	ẵ	LATIN SMALL LETTER A WITH BREVE AND TILDE	Vietnamese (1)	[109]
184.	1EB7	ặ	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	Vietnamese (1)	[109]
185.	1EB9	ẹ	LATIN SMALL LETTER E WITH DOT BELOW	Yoruba (2)	[181]

186.	1EB9 + 0300	ệ	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT	Yoruba (2)	[254]
187.	1EB9 + 0301	ệ	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT	Yoruba (2)	[254]
188.	1EBB	ẻ	LATIN SMALL LETTER E WITH HOOK ABOVE	Vietnamese (1)	[109]
189.	1EBD	ẽ	LATIN SMALL LETTER E WITH TILDE	Umbundu (3) Guarani (1) Cubeo (3) Xavante (4)	[141], [142], [143], [186], [187], [117]
190.	1EBF	ế	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]
191.	1EC1	ề	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	Vietnamese (1)	[109]
192.	1EC3	ể	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
193.	1EC5	ễ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]
194.	1EC7	ệ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]

195.	1EC9	ỉ	LATIN SMALL LETTER I WITH HOOK ABOVE	Vietnamese (1)	[109]
196.	1ECB	ị	LATIN SMALL LETTER I WITH DOT BELOW	Igbo (2)	[205]
197.	1ECD	ọ	LATIN SMALL LETTER O WITH DOT BELOW	Igbo (2) Yoruba (2) Marshallese (1)	[204], [205], [181], [136], [215], [216]
198.	1ECD + 0300	ọ̀	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT	Yoruba (2)	[254]
199.	1ECD + 0301	ọ́	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT	Yoruba (2)	[254]
200.	1ECF	ỏ	LATIN SMALL LETTER O WITH HOOK ABOVE	Vietnamese (1)	[109]
201.	1ED1	ố	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]
202.	1ED3	ồ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	Vietnamese (1)	[109]
203.	1ED5	ổ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
204.	1ED7	ỗ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]

205.	1ED9	ộ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]
206.	1EDB	ớ	LATIN SMALL LETTER O WITH HORN AND ACUTE	Vietnamese (1)	[109]
207.	1EDD	ờ	LATIN SMALL LETTER O WITH HORN AND GRAVE	Vietnamese (1)	[109]
208.	1EDF	ở	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE	Vietnamese (1)	[109]
209.	1EE1	ỡ	LATIN SMALL LETTER O WITH HORN AND TILDE	Vietnamese (1)	[109]
210.	1EE3	ợ	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	Vietnamese (1)	[109]
211.	1EE5	ụ	LATIN SMALL LETTER U WITH DOT BELOW	Vietnamese (1) Igbo (2)	[109], [204], [205]
212.	1EE7	ủ	LATIN SMALL LETTER U WITH HOOK ABOVE	Vietnamese (1)	[109]
213.	1EE9	ứ	LATIN SMALL LETTER U WITH HORN AND ACUTE	Vietnamese (1)	[109]
214.	1EEB	ừ	LATIN SMALL LETTER U WITH HORN AND GRAVE	Vietnamese (1)	[109]

215.	1EED	ử	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	Vietnamese (1)	[109]
216.	1EEF	ữ	LATIN SMALL LETTER U WITH HORN AND TILDE	Vietnamese (1)	[109]
217.	1EF1	ự	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	Vietnamese (1)	[109]
218.	1EF3	ỳ	LATIN SMALL LETTER Y WITH GRAVE	Vietnamese (1)	[109]
219.	1EF5	ỵ	LATIN SMALL LETTER Y WITH DOT BELOW	Vietnamese (1)	[109]
220.	1EF7	ỷ	LATIN SMALL LETTER Y WITH HOOK ABOVE	Vietnamese (1)	[109]
221.	1EF9	ỹ	LATIN SMALL LETTER Y WITH TILDE	Vietnamese (1) Guarani (1)	[109] [142]

5.3.1 Combining Marks

There are six Unicode code points included in the Latin repertoire which are non-space Combining Marks and which are presented below in Table 4. They are not listed individually in the repertoire, since they cannot be used independently. Also, they cannot be arbitrarily combined with just any other code points from the repertoire. They are used only in specific combinations that are included as sequences in the repertoire above. (See Section 5.2.1, Inclusion Principle #3.)

Table 4. Combining Marks Included in the Repertoire of Latin Script LGR.

Unicode	Glyph	Unicode name
0300	◌̀	COMBINING GRAVE ACCENT
0301	◌́	COMBINING ACUTE ACCENT
0303	◌̃	COMBINING TILDE
0304	◌̄	COMBINING MACRON

0308	¨	COMBINING DIAERESIS
0327	¸	COMBINING CEDILLA

5.4 Code Points Excluded

The Internet Architecture Board (IAB) has mandated that punctuation marks cannot be used in domain names. This includes punctuation marks themselves, code points that look like punctuation marks, and letters which, although they are single letters in a particular language's alphabet, *look like* punctuation marks. Accordingly, the following letters from various languages using the Latin script have been excluded from the repertoire.

Table 5. Punctuation Marks or Punctuation Mark Look-Alikes

Unicode	Glyph	Unicode Name	Language	Reference
02BB	ʻ	MODIFIER LETTER TURNED COMMA	Hawaiian (2)	https://www.omniglot.com/writing/hawaiian.htm
02BC	ʼ	MODIFIER LETTER APOSTROPHE	Chamorro - (1) Dagaare-Burkina Faso (4) Dagbani (Dagomba) (4) Dholuo (5) Garo (2) Hausa (2) Mossi (3) Tartar (2) Tausūg (3) Tongan (1) Uzbek (1)	https://www.omniglot.com/writing/chamorro.htm http://www.omniglot.com/writing/dagaare.htm http://www.omniglot.com/charts/dagbani.pdf http://www.omniglot.com/writing/dholuo.php https://www.omniglot.com/writing/garo.htm http://www.omniglot.com/writing/hausa.htm https://www.omniglot.com/writing/mossi.htm http://www.omniglot.com/writing/tatar.htm https://www.omniglot.com/writing/tausug.htm http://www.omniglot.com/writing/tongan.htm http://www.omniglot.com/writing/uzbek.htm
A78C	'	LATIN SMALL LETTER SALTILLO	Central Sinama (4) Guarani (1) Kaqchikel (4) Oromo (Afaan) (5)	https://www.omniglot.com/writing/centralsinama.htm http://sinama.org/bahasa-sinama/sama-alphabet/

			Pangasinan (3)	http://www.omniglot.com/writing/guarani.htm https://en.wikipedia.org/wiki/Guarani_alphabet https://www.omniglot.com/writing/kaqchikel.htm https://www.omniglot.com/writing/oromo.htm https://www.omniglot.com/writing/pangasinan.htm
01C3	!	LATIN LETTER RETROFLEX CLICK	Khoekhoe (4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm

Table 6. Letters Combined With Punctuation Marks or Punctuation Mark Look-Alikes.

Unicode	Glyph	Unicode Name	Language	Reference
0063 + 0068 + A78C	ch'	LATIN SMALL LETTER C +LATIN SMALL LETTER H + LATIN SMALL LETTER SALTILLO	Quechua (3)	https://www.omniglot.com/writing/quechua.htm
0067 + 02BC	g’	LATIN SMALL LETTER G + MODIFIER LETTER APOSTROPHE	Uzbek (1)	https://en.wikipedia.org/wiki/Uzbek_alphabet#Distinct_characters
02BC + 0068	’h	LATIN MODIFIER LETTER APOSTROPHE WITH LATIN SMALL LETTER H	Dagaare - Burkina Faso (4)	http://www.omniglot.com/writing/dagaare.htm
006B + A78C	k'	LATIN SMALL LETTER K + LATIN SMALL LETTER SALTILLO	Quechua (3)	https://www.omniglot.com/writing/quechua.htm
02BC + 006C	’l	LATIN MODIFIER LETTER APOSTROPHE WITH LATIN SMALL LETTER L	Dagaare - Burkina Faso (4)	http://www.omniglot.com/writing/dagaare.htm

006C + 02BC	l’	LATIN SMALL LETTER L + MODIFIER LETTER APOSTROPHE	Garo (2)	http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm
006D + 02BC	m’	LATIN SMALL LETTER M + MODIFIER LETTER APOSTROPHE	Garo (2)	http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm
006E + 02BC	n’	LATIN SMALL LETTER N + MODIFIER LETTER APOSTROPHE	Garo (2)	http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm
006E + 0067 + 02BC	ng’	LATIN SMALL LETTER N + LATIN SMALL LETTER G + MODIFIER LETTER APOSTROPHE	Garo (2)	http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm
014B + 02BC	ŋ’	LATIN SMALL LETTER ENG WITH MODIFIER LETTER APOSTROPHE	Adzera (4)	http://www.omniglot.com/writing/adzera.htm
006F + 02BC	o’	LATIN SMALL LETTER O + MODIFIER LETTER APOSTROPHE	Uzbek (1)	https://en.wikipedia.org/wiki/Uzbek_alphabet#Distinct_characters
0070 + A78C	p'	LATIN SMALL LETTER O + LATIN SMALL LETTER SALTILLO	Quechua (3)	https://www.omniglot.com/writing/quechua.htm
0071 + A78C	q'	LATIN SMALL LETTER Q + LATIN SMALL LETTER SALTILLO	Quechua (3)	https://www.omniglot.com/writing/quechua.htm
0074 + A78C	t'	LATIN SMALL LETTER T + LATIN SMALL LETTER SALTILLO	Quechua (3)	https://www.omniglot.com/writing/quechua.htm
02BC + 0077	’w	LATIN MODIFIER LETTER APOSTROPHE WITH LATIN SMALL LETTER W	Dagaare - Burkina Faso (4)	http://www.omniglot.com/writing/dagaare.htm

5.4.1 Other Excluded Letters

The Integration Panel has declined to include three letters, proposed by Latin GP for inclusion in [MSR], because of unspecified “security concerns”. These letters are marked as homoglyphs of punctuation.

Complete explanation could be found in https://www.icann.org/en/system/files/files/msr-3-overview-28mar18-en.pdf - Section 5.7.5 (pg. 24).

Table 7. Homoglyphs of Punctuation Marks Excluded from the Repertoire of Latin Script LGR.

Unicode	Glyph	Unicode Name	Language	Reference
01C0	ǀ	LATIN LETTER DENTAL CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm
01C1	ǁ	LATIN LETTER LATERAL CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm
01C2	ǂ	LATIN LETTER ALVEOLAR CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm

A fourth letter that the Latin GP proposed for inclusion and which was declined by the Integration Panel is the Middle Dot (00B7). This character is an integral part of the Catalan language. The reasoning for exclusion is the fact that the status of this code point under IDNA 2008 is CONTEXTO and “code points permitted by IDNA2008 under the CONTEXTO and CONTEXTJ rules are automatically excluded” according to the RZ-LGR Procedure Section B.3.4.2.

Table 8. CONTEXTO and CONTEXTJ Code Points Excluded from the Repertoire of Latin Script LGR.

Unicode	Glyph	Unicode Name	Language	Reference
00B7	·	MIDDLE DOT	Catalan(2)	https://en.wikipedia.org/wiki/Interpunct#Catalan http://www.omniglot.com/writing/catalan.htm

6. Variants

This section discusses the definition of variants for the Latin script, the discovery methodology, and the proposed candidates.

In accordance with the Procedure, an IDN variant for the Latin Root Zone LGR is going to be an alternate code point (or sequence of code points) that could be substituted for a code point (or sequence of code points) in a candidate label to create a variant label that is considered the “same”.

6.1 Principles for Developing Variants

For the Latin Root Zone LGR the meaning of “same” will slightly vary. Latin GP determined that there are two dimensions for sameness for the Latin script:

- visual
- non-visual

In addition to the above, Latin GP has reviewed other cases which may or may not fall under those categories, such as IDNA2003 compatibility and HTML underlining.

For the XML, a matrix will be developed, which will indicate for any codepoint, why it is considered a variant. The following matrix is an example but it is still under discussion and has not found consensus as of yet.

Table 9. Variants Principles Matrix.

Index #	Principle	Reason	Disposition	Example
1	Visual variant (homoglyph)	Security	Blocked	
2	Visual variant (glyph nearly identical)	Security	Blocked	
3	Visual variant (generally acceptable font design)	Security	Blocked	
4	Non-visual variant	Security	Blocked	
5	Symmetry property {a:b}	Security	Blocked	
6	Transitivity property {a:b; b:c}	Security	Blocked	
7	URL underlining	Security	Blocked	
8	IDNA2003 Compatibility	Security	Blocked	
9	Function (alternate orthography)	Usability	Allocatable	

6.1.1 Distinguishing Visual From Non-Visual Variants

Latin GP has analyzed variants on the basis of both visual and non-visual aspects. While the criteria for visual similarity are fairly consistent across both in-script and cross-script variants, the non-visual variation was less clear-cut.

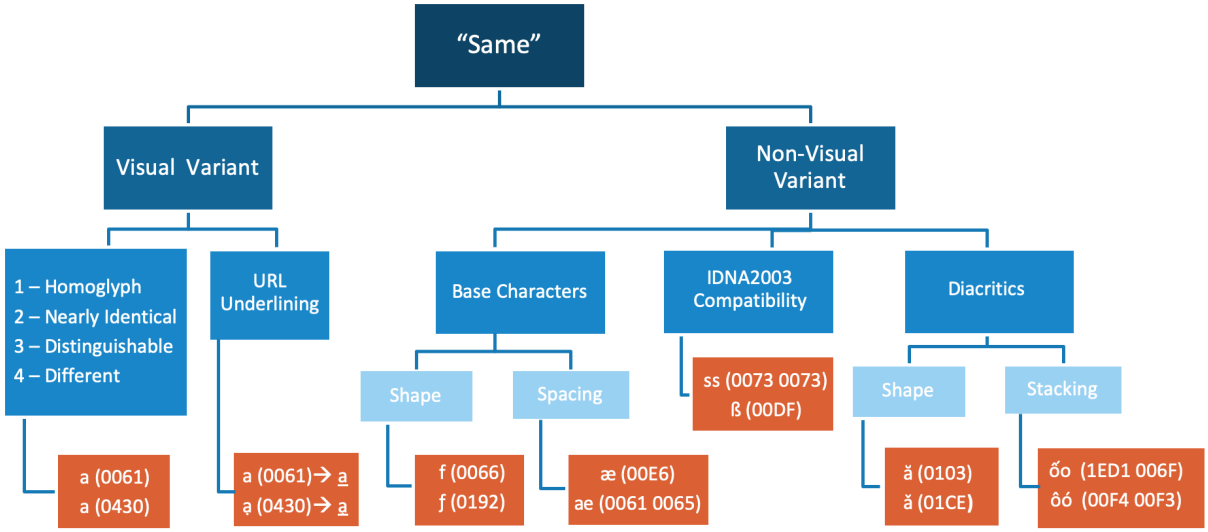
With non-visual variants the issue is essentially two-fold:

- Either readers (of domain name labels) may consider two glyphs conceptually identical despite being able to visually tell them apart, or
- readers may identify glyphs wrongly with other letters or sequences of letters in certain contexts.

Both issues relate to the psycholinguistic process of reading and writing, which is based not only on graphic aspects, but also on other aspects such as linguistic, contextual and cognitive factors. However, the second issue also overlaps strongly with visual similarity. While such capacities are generally individual to single readers, Latin GP had to identify certain key areas where such non-visual similarity may be confusable across significant parts of the script-using community and across individual readers.

GP has identified several aspects, which may play into as to why two or more code points may be considered “same”, as summarized in the following diagram:

Diagram 1: The Sub-Types of “Same” in Latin Script



Section 6.1.2 below discusses first the types of visual similarity (on the left-hand branch of the diagram).

6.1.2 Visual Variants

Per [MSR],
“the kinds of variants to be defined in the Root Zone LGR are limited to homoglyphs, which are characters essentially identical appearance by design, instead of merely similar appearance” (22 March 2017, IP Feedback to Latin GP Proposal, Document Version 1).

However, based on discussions within the GP and by the GP with IP, the panel came to the conclusion that the GP found that homoglyphs are not a categorial but a gradual distinction. Accordingly, Latin GP devised a four-point scale to determine whether a given pair of candidate characters tended to fall into the “essentially identical appearance by design” group, i.e. a clear-cut case of a homoglyph, or rather into the a “merely similar appearance” group.

This scale was found to be useful by the GP, because it places similar interpretations next to one another: While both categories Homoglyphs and Different visa-a-vis one another are not only self-explanatory but were also judged very coherently across different members of the GP, the debates usually revolved around the difference between a Homoglyph and Nearly Identical case, a Nearly Identical Case versus a Distinguishable case, and - to a lesser degree - a Distinguishable case versus a Different case. Accordingly, such a scale allowed the GP to express such gradual distinctions. The elements of that scale are presented together with a concise definition below in Table 10:

Table 10. Scale for Classifying Degree of Visual Identity

Score	Category

1	Homoglyphs A pair of code points in this category have essentially identical appearance by design.
2	Nearly Identical A pair of code points is considered Nearly Identical when the visual confusion can be attributed to font design.
3	Distinguishable A pair of code points is considered Distinguishable when any of the code point's glyphs have recognizably different features from the other code point.
4	Different When the two glyphs in the pair are sufficiently different.

Over time, a rough consensus evolved as summarized by the concise definitions of the items of this scale above in Table 10. The GP decided that a Latin code point will be deemed a visual variant with another code point when the two code points or sequence of code points are either

- homoglyphs (i.e. visual score = 1), or
- nearly identical (i.e. visual score = 2).

Nonetheless, numerous debates about the precise rating between different pairs of variant candidates according to this scale took place, which eventually were resolved only by means of explicit vote by each active member, to establish majority decisions. However, during this very long process the GP came to the understanding, that visual appearance, was not the only aspect which led to users considering code points as variants. For pragmatic reasons, this other category, which found no explicit mention in MSR, but which by consensus of the Panel was understood to be included under "characters essentially identical appearance by design", was simply termed 'Non-Visual Variant', as rendered on the right-hand branch of in Diagram 1 above, and as discussed in the following sections.

6.1.3 Non-Visual Variants

6.1.3.1 Shape of Base Characters

Historically, the classical Latin or Roman alphabet consisted of only 23 letters. Most new letters developed since are based on already existing letters and are therefore derived letters, or they were inspired by or adopted from other scripts, that is borrowed letters. Derived letters were usually modified by extending certain lines (e.g. k vs. ƙ or f vs. ƒ) or by dropping elements (e.g. i vs. ɪ). In handwriting practices, where a cursive writing style dominates connecting most letters to the right in order to speed up handwriting, the same kinds of changes to letters are made in order to make those connections; that is lines are extended and elements are dropped. Accordingly, Latin GP hypothesized that some hand-written forms may end up taking similar or the same shapes as some derived letters, and that readers may consider such unknown derived letters as hand-written variations of familiar letters, such as e.g. v vs. ʋ.

Also, some letters have traditionally different shapes in hand-written and printed forms such as a vs. ɑ (with the latter shaping being the traditional form encountered in handwriting). Many such differences

also overlap with the difference between upper and lower case, such as e.g. e vs. ɛ, with the latter glyph being a common upper-case form in handwriting to the former glyph and letter.

6.1.3.2 Spacing of Base Characters

Several letters have been derived by putting more closely together sequences of two or more letters, and the result of such modifications of spacing in between letters are called ligatures. This strategy to develop new letters was already employed in antiquity, with e.g. w being derived out of a sequence of two v, i.e. vv (https://en.wikipedia.org/wiki/History_of_the_Latin_script).

While the origins are still somehow recognizable in the case of w, in other cases the ligatures are not recognizable anymore as combinations of their original letters, such as ß which was formed on the hand-written basis of s and z (https://en.wikipedia.org/wiki/%C3%9F). In such cases where letters are recognizable as being composed of two or more letters, confusion could arise among readers and depending on the spacing in between those glyphs in a font (which depends on typographic factors such as e.g. kerning), ligatures may become indistinguishable from a sequence of letters of which the same ligature was originally composed.

6.1.3.3 IDNA 2003 Compatibility

In Section 5.5 of Maximal Starting Repertoire — MSR-4 Overview and Rationale, Integration Panel highlighted risks due to IDNA compatibility issues:
“In IDNA2003, case folding is applied which creates compatibility issues between IDNA2008 and IDNA2003 for several code points. This arguably makes the affected code points candidates for summary exclusion from the MSR on grounds of Longevity (§2.1).”

Of those code points, two belong to the Latin-script repertoire, namely 00DF LATIN SMALL LETTER SHARP S and 0131 LATIN SMALL LETTER DOTLESS I. The solutions based on a point of view of IDNA compatibility are presented in sections 6.7.2 and 6.7.3, while the considerations involving those code points and leading to those solutions are discussed in further detail in Appendix D.5.

However those two code points were also considered under other aspects, including cross-script variants between Latin and Greek script (cf. section 6.3.3), Generic Glyphs across scripts (cf. section 6.3.4) and in-script Variants based on the shaping of base characters (cf. Appendix D.1).

6.1.3.4 Diacritics

6.1.3.4.1 Shaping of Diacritics

Diacritics are modifiers surrounding basic letter shapes. While in some cases diacritics are considered part and parcel of a letter shape, such as e.g. the dot on top of i, generally they are recognized as distinct graphic elements of the script employed to form new letters, such as é based on e featuring an acute accent on top, and the majority of derived letters of Latin script were developed using this strategy. Over time however, novel diacritics became employed which were based on other diacritics, such as e.g. on ű, which features a base character u with a double acute (˝), a diacritic which is in turn based on the single acute (´). Many novel diacritics are very limited in use and occur in only a few languages, as they were developed to express less common distinctive linguistic features of languages written in Latin script, such as Tone, and often such are only familiar to users of such languages. Essentially there are three types of potential issues with such modifiers:

First, certain diacritics may be considered conceptually the same as others by significant parts of the user community, such as dot below or a comma below.

Secondly, in some cases certain diacritics are not kept apart from one another in handwriting traditions, such as e.g. a caron often being written in the same way as a breve, or a dot above (even where they are considered part of a basic letter shape) being written in the same way as an acute. Furthermore, in cursive hand-writing writers make use of particular strategies to write letters more quickly, modifying them in ways in which the diacritics become visually identical or confusable with others, such as a diaeresis being replaced by two vertical strokes, which could be mistaken for a double acute in italic fonts, or a tilde being written 'simply' as a simple horizontal stroke above, i.e. a macron.

Lastly, since a number of these diacritics are used only in a very limited part of the script using community, this may lead to confusion with significant parts of the script-using community or even the majority. For example, the horn (as e.g. used in combination with the basic letter shape o on 01A1 ơ LATIN SMALL LETTER O WITH HORN) could be conceptually mistaken by some readers for a misplaced acute (´) or even an apostrophe (') -- for those users unaware that punctuation marks are excluded from use in IDN-labels because of the LDH principle. By consequence, diacritics considered conceptually different in both print and Unicode may in handwriting be considered as being interchangeable or even the same, or may become visually confusable or identical to other diacritics for readers.

6.1.3.4.2 Stacking of Diacritics

Diacritics are also combined with one another, such as ấ featuring both a circumflex and an acute. Such combinations are for the most part comparatively recent innovations, which again were often developed for linguistically distinctive features absent from European languages and therefore not traditionally represented in Latin script, such as Tone. Such novel elements of the script were often encoded in later revisions of Unicode and glyphs have been developed only for a very limited number of fonts.

By consequence, many fonts either use fallback rendering, replacing missing glyphs by taking them from any other font featuring the missing glyph and available to the user's client, or such glyphs are not represented correctly at all by fonts, with overlapping and misplacement of diacritics occurring frequently. Therefore, glyphs featuring base characters with several diacritics may become visually identical or confusable to readers with sequences of glyphs featuring the same diacritics on two separate code points or may even become effectively invisible in context by crossing over into adjacent glyphs.

6.2 Methodology For Developing Cross-Script Variants

Latin GP has analyzed variant relationships across related scripts, such as Cyrillic, Armenian and Greek. In addition, cases where a character shape is so generic that it occurs in multiple unrelated languages were examined. To wit, a straight vertical line (LATIN SMALL LETTER L), a circle (LATIN SMALL LETTER O), and a crescent (LATIN SMALL LETTER C and LATIN SMALL LETTER OPEN O).

To test this, Latin GP selected three fonts to represent Latin script, which it deemed to be widespread enough to be representative, i.e. Arial, Courier New, and Times New Roman, to compare glyphs across scripts. In the case of Armenian script, it was noted that there were varying glyph shapes, depending on

the application used for rendering strings, which made the initial analysis much more difficult[4]. The GP consulted the Armenian Proposal to identify which glyphs the Armenian GP had chosen for representation in its Proposal [ARMENIAN] and considered those as standard for purposes of comparison with Latin script. To demonstrate the glyphs as seen and considered by Latin GP, we use screenshots in parts of this document to ensure that the reader sees the same shapes.

6.3 Cross-Script Variants

6.3.1 Armenian Script

Latin GP proposes the following cross-script variants with the Armenian script.
The two tables below display the same information; the second table, however, is a screenshot taken from Microsoft Excel to demonstrate the glyph shapes as seen by the GP during the cross-script variant analysis

Table 11. Armenian Cross-Script Variants

Source Unicode Name	Source Code Point	Source Glyph	Variant Relation-ship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Rationale
LATIN SMALL LETTER G	0067	g	↔	g	0581	ARMENIAN SMALL LETTER CO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER H	0068	h	↔	h	0570	ARMENIAN SMALL LETTER HO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER N	006E	n	↔	n	0578	ARMENIAN SMALL LETTER VO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER O	006F	o	↔	o	0585	ARMENIAN SMALL LETTER OH	Blocked	Homoglyph
LATIN SMALL LETTER Q	0071	q	↔	q	0566	ARMENIAN SMALL LETTER ZA	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER U	0075	u	↔	u	057D	ARMENIAN SMALL LETTER SEH	Blocked	Glyphs nearly identical due to font design

[4] Google Sheets, the tool used for cross-script analysis, did not offer variety of font designs for Armenian letters, which made it difficult for the Latin GP to replicate Armenian GP's results. Thus, an alternate application such as Microsoft Excel, which did offer more variety of font styles as seen in the snapshot, was used.

LATIN SMALL LETTER IOTA	0269	ɩ	↔	ʟ	0582	ARMENIAN SMALL LETTER YIWN	Blocked	Glyphs nearly identical due to font design

Screenshot taken from Microsoft Excel. The three glyphs for each code point are set in Times New Roman, Arial, and Courier, respectively:

Latin			Armenian			Disposition	Rationale
Unicode Name	Unicode	Glyph	Glyph	Unicode	Unicode Name		
LATIN SMALL LETTER O	006F	o o o	o o o	0585	ARMENIAN SMALL LETTER OH	Blocked	Homoglyph
LATIN SMALL LETTER Q	0071	q q q	q q q	0566	ARMENIAN SMALL LETTER ZA	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER H	0068	h h h	h h h	0570	ARMENIAN SMALL LETTER HO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER N	006E	n n n	n n n	0578	ARMENIAN SMALL LETTER VO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER U	0075	u u u	u u u	057D	ARMENIAN SMALL LETTER SEH	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER G	0067	g g g	g g g	0581	ARMENIAN SMALL LETTER CO	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER IOTA	0269	ɩ ɩ ɩ	ʟ ʟ ʟ	0582	ARMENIAN SMALL LETTER YIWN	Blocked	Glyphs nearly identical due to font design

6.3.2 Cyrillic Script

The Latin GP proposes the following cross-script variants with Cyrillic script:

Table 12: Cyrillic Cross-Script Variants

Source Unicode Name	Source Code Point	Source Glyph	Variant Relationship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Rationale
LATIN SMALL LETTER R	0072	r	↔	r	0433	CYRILLIC SMALL LETTER GHE	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER Y	0079	y	↔	ү	04AF	CYRILLIC SMALL LETTER STRAIGHT U	Blocked	Glyphs nearly identical due to font design. See [C1] below.
LATIN SMALL LETTER C WITH CEDILLA	00E7	ç	↔	ç	04AB	CYRILLIC SMALL LETTER ES WITH DESCENDER	Blocked	Glyphs nearly identical due to font design

LATIN SMALL LETTER Y WITH DIAERESIS	00FF	ÿ	↔	ӳ	04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER R WITH ACUTE	0155	ŕ	↔	ѓ	0453	CYRILLIC SMALL LETTER GJE	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER R WITH STROKE	024D	ɍ	↔	ғ	0493	CYRILLIC SMALL LETTER GHE WITH STROKE	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER U WITH DOT BELOW	1EE5	ụ	↔	џ	045F	CYRILLIC SMALL LETTER DZHE	Blocked	Glyphs nearly identical due to font design. See [C2] below.
LATIN SMALL LETTER A	0061	a	↔	а	0430	CYRILLIC SMALL LETTER A	Blocked	Homoglyph
LATIN SMALL LETTER C	0063	c	↔	с	0441	CYRILLIC SMALL LETTER ES	Blocked	Homoglyph
LATIN SMALL LETTER E	0065	e	↔	е	0435	CYRILLIC SMALL LETTER IE	Blocked	Homoglyph
LATIN SMALL LETTER H	0068	h	↔	һ	04BB	CYRILLIC SMALL LETTER SHHA	Blocked	Homoglyph
LATIN SMALL LETTER I	0069	i	↔	і	0456	CYRILLIC SMALL LETTER BELARUSIAN-UKRAINIAN I	Blocked	Homoglyph
LATIN SMALL LETTER J	006A	j	↔	ј	0458	CYRILLIC SMALL LETTER JE	Blocked	Homoglyph
LATIN SMALL LETTER L	006C	l	↔	ӏ	04CF	CYRILLIC SMALL LETTER PALOCHKA	Blocked	Homoglyph

LATIN SMALL LETTER O	006F	o	↔	o	043E	CYRILLIC SMALL LETTER O	Blocked	Homoglyph
LATIN SMALL LETTER P	0070	p	↔	p	0440	CYRILLIC SMALL LETTER ER	Blocked	Homoglyph
LATIN SMALL LETTER S	0073	s	↔	s	0455	CYRILLIC SMALL LETTER DZE	Blocked	Homoglyph
LATIN SMALL LETTER X	0078	x	↔	x	0445	CYRILLIC SMALL LETTER HA	Blocked	Homoglyph
LATIN SMALL LETTER Y	0079	y	↔	y	0443	CYRILLIC SMALL LETTER U	Blocked	Homoglyph. See [C1] below.
LATIN SMALL LETTER A WITH DIAERESIS	00E4	ä	↔	ä	04D3	CYRILLIC SMALL LETTER A WITH DIAERESIS	Blocked	Homoglyph
LATIN SMALL LETTER AE	00E6	æ	↔	æ	04D5	CYRILLIC SMALL LIGATURE A IE	Blocked	Homoglyph
LATIN SMALL LETTER E WITH DIAERESIS	00EB	ë	↔	ë	0451	CYRILLIC SMALL LETTER IO	Blocked	Homoglyph
LATIN SMALL LETTER I WITH DIAERESIS	00EF	ï	↔	ï	0457	CYRILLIC SMALL LETTER YI	Blocked	Homoglyph
LATIN SMALL LETTER O WITH DIAERESIS	00F6	ö	↔	ö	04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS	Blocked	Homoglyph
LATIN SMALL LETTER A WITH BREVE	0103	ă	↔	ă	04D1	CYRILLIC SMALL LETTER A WITH BREVE	Blocked	Homoglyph

LATIN SMALL LETTER H WITH STROKE	0127	ħ	↔	ħ	045B	CYRILLIC SMALL LETTER TSHE	Blocked	Homoglyph
LATIN SMALL LETTER TURNED E	01DD	ə	↔	ə	04D9	CYRILLIC SMALL LETTER SCHWA	Blocked	Homoglyph
LATIN SMALL LETTER SCHWA	0259	ə	↔	ə	04D9	CYRILLIC SMALL LETTER SCHWA	Blocked	Homoglyph
LATIN SMALL LETTER EZH	0292	ʒ	↔	ʒ	04E1	CYRILLIC SMALL LETTER ABKHASIAN DZE	Blocked	Homoglyph

[C1] Cyrillic GP has already classified 0079 and 0443 as variants [CYRILLIC]. In addition to that, Latin GP considers 04AF to be sufficiently similar to 0079 to warrant a variant relationship between the two characters. By consequence, this finding leads towards an in-script variant in Cyrillic script between 04AF and 0443, due to the requirement of transitivity.

[C2] In Arial and Courier New, the glyphs of 1EE5 and 045F look nearly identical. The screenshot below presents the glyphs in those two fonts in the second and third rows, respectively (The first row presents the glyphs in Times New Roman).

LATIN SMALL LETTER U WITH DOT BELOW	1EE5	ụ ụ ụ	ц ц ц	045F	CYRILLIC SMALL LETTER DZHE

6.3.3 Greek Script

The Latin GP proposes the following cross-script variants with Greek script:

Table 13: Greek Cross-Script Variants

Source Unicode Name	Source Code Point	Source Glyph	Variant Relationship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Rationale
LATIN SMALL LETTER O	006F	o	↔	ο	03BF	GREEK SMALL LETTER OMICRON	Blocked	Homoglyph
LATIN SMALL LETTER I WITH ACUTE	00ED	í	↔	ί	03AF	GREEK SMALL LETTER IOTA WITH TONOS	Blocked	Homoglyph

LATIN SMALL LETTER I WITH DIAERESIS	00EF	ï	↔	ΐ	03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA	Blocked	Homoglyph
LATIN SMALL LETTER O WITH ACUTE	00F3	ó	↔	ό	03CC	GREEK SMALL LETTER OMICRON WITH TONOS	Blocked	Homoglyph
LATIN SMALL LETTER DOTLESS I	0131	ı	↔	ι	03B9	GREEK SMALL LETTER IOTA	Blocked	Homoglyph
LATIN SMALL LETTER OPEN E	025B	ɛ	↔	ε	03B5	GREEK SMALL LETTER EPSILON	Blocked	Homoglyph
LATIN SMALL LETTER IOTA	0269	ɩ	↔	ι	03B9	GREEK SMALL LETTER IOTA	Blocked	Homoglyph
LATIN SMALL LETTER V	0076	v	↔	ν	03BD	GREEK SMALL LETTER NU	Blocked	Glyphs nearly identical due to font design.
LATIN SMALL LETTER A	0061	a	↔	α	03B1	GREEK SMALL LETTER ALPHA	Blocked	Glyphs nearly identical due to font design. See **[G1]** below.
LATIN SMALL LETTER P	0070	p	↔	ρ	03C1	GREEK SMALL LETTER RHO	Blocked	Glyphs nearly identical due to font design. See **[G2]** below.
LATIN SMALL LETTER U	0075	u	↔	υ	03C5	GREEK SMALL LETTER UPSILON	Blocked	Glyphs nearly identical due to font design. See **[G3]** below.
LATIN SMALL LETTER Y	0079	y	↔	γ	03B3	GREEK SMALL LETTER GAMMA	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER SHARP S	00DF	ß	↔	β	03B2	GREEK SMALL LETTER BETA	Blocked	Glyphs nearly identical due to font design. See **[G4]** below.
LATIN SMALL LETTER A WITH ACUTE	00E1	á	↔	ά	03AC	GREEK SMALL LETTER ALPHA WITH TONOS	Blocked	Glyphs nearly identical due to font design

LATIN SMALL LETTER U WITH ACUTE	00FA	ú	↔	ύ	03CD	GREEK SMALL LETTER UPSILON WITH TONOS	Blocked	Glyphs nearly identical due to font design. See **[G3]** below.
LATIN SMALL LETTER U WITH DIAERESIS	00FC	ü	↔	ϋ	03CB	GREEK SMALL LETTER UPSILON WITH DIALYTIKA	Blocked	Glyphs nearly identical due to font design
LATIN SMALL LETTER O WITH HORN	01A1	ơ	↔	σ	03C3	GREEK SMALL LETTER SIGMA	Blocked	Glyphs nearly identical due to font design. See **[G5]** below.
LATIN SMALL LETTER V WITH HOOK	028B	ʋ	↔	υ	03C5	GREEK SMALL LETTER UPSILON	Blocked	Glyphs nearly identical due to font design. See **[G3]** below.

[G1] Latin-script users consider 0061 LATIN SMALL LETTER A and GREEK SMALL LETTER ALPHA 03B1 as variants on non-visual grounds:

0061 is regularly represented using a glyph (nearly) identical with 03B1 in handwriting, which is why significant parts of the Latin script-using community may consider them equivalent, despite being able to visually tell the difference between the two glyphs. For example, 0061 is considered the block- or print-letter shape to the hand-written shape of 03B1 in large parts of the script-using community, and a shape similar to 03B1 is used in standard primers and repertoire of handwriting as taught to school children, such as e.g. the Grundschrift (https://en.wikipedia.org/wiki/Grundschrift)[5] demonstrated in Figure G02:

Figure G02. Repertoire of Standard Handwriting repertoire as official in the German state Hamburg, taken from https://en.wikipedia.org/wiki/Grundschrift#/media/File:Hamburger_Druckschrift_ab_2011.jpg

[5] Grundschrift is the current standard repertoire by law for the German state of Hamburg and is being endorsed for use across all German states. Similar glyphs are also used in other repertoires of didactic hand-writing repertoires of German-speaking countries such as the Swiss Basisschrift - https://www.basisschrift.ch/aufbau-und-didaktik).

This variation between glyphs is however not limited to the German speaking user community or didactic hand-writing repertoires: Similar shapes to both 0061 and 03B1 are featured prominently in the graphic design of logos of international brand names in, which constantly reiterates the inter-changeability to the minds of readers:

- US TV-station ABC (http://logos.wikia.com/wiki/ABC_(United_States)?file=Abc_2013_logo_dark_grey.svg),
- Beats by Dr. Dre (https://cdn.dealspotr.com/zc-images/merchants/beats-by-dre.jpg),
- Macys (https://en.wikipedia.org/wiki/Macy%27s),
- Adidas (https://en.wikipedia.org/wiki/Adidas)
- German TV station ARD-Alpha (https://de.wikipedia.org/wiki/ARD-alpha),
- Former US airline AirTran (http://logos.wikia.com/wiki/AirTran_Airways?file=AirTran_A.svg)

The variation in between the two character shapes occurs also within the same logos

- e.g. http://logos.wikia.com/wiki/Save-A-Lot)

This inter-changeability is also historically established and has been used for decades in the typography employed in movies (cf. the initial “a” Paramount movie openers (http://logos.wikia.comwiki/Paramount_Cartoon_Studios).

While IP has noted that logos should not be used as evidence since they use ad-hoc font styles (as noted during the conference call with IP in October 2018), the large number of well-known logos across language communities together with the independent evidence from font renderings constitutes sufficient evidence for Latin GP to be considered as valid evidence in favor of a variant relationship.

In summary, Latin GP concluded that users of Latin script may not be able to differentiate 03B1 from 0061 based on non-visual grounds[6], and therefore 03B1 should be in a variant relationship with 0061.

[G2] LATIN SMALL LETTER P (0070) and GREEK SMALL LETTER RHO (03C1) are visually nearly identical in isolation in several widespread fonts (such as Times New Roman and Courier New, presented in the first and third row, respectively, of the screenshot below).

Figure G02: 0070 vs. 03C1

[6] Cf. also the discussion of the in-script variant in between 00E6 LATIN SMALL LETTER AE and 0153 LATIN SMALL LIGATURE OE (D.2.1. and D.2.2.).

LATIN SMALL LETTER P	0070	p	ρ	03C1	GREEK SMALL LETTER RHO
		p	ρ		
		p	ρ		

In such cases, the two code-points are visually only distinguishable in context because of their relative positioning towards the baseline, since 0070 crosses below the baseline but 03C1 does not. Given that there are several variant candidates among the cross-script variants, numerous plausible labels could be made up, such as .ρορ or .ραγ ,which most Latin-script users would be hard-pressed to distinguish in context.

Furthermore, designers from the Latin-script using community have exploited the visual similarity[7] between these two code-points and have created logos for globally used brand-names, which employ glyphs baring more resemblance to Greek 03C1 rather than Latin 0070, such as Pepsi (cf. https://perma.cc/6GTA-98C9?type=image). Again, this use in logo designs is neither limited to the Pepsi-brand logo nor the English-using community - cf.

- http://logos.wikia.com/wiki/Logopedia:Theme/Logos_with_the_letter_P?file=Publix_logo.png,
- http://logos.wikia.com/wiki/File:150px-Android_P_logo.png
- http://logos.wikia.com/wiki/File:Vpf.png,
- http://logos.wikia.com/wiki/Category:Red_PAT -

,and it is featured in historically established logos – cf.

- http://logos.wikia.com/wiki/File:Pba_83_on_city_2_Vintage_Sports.jpg.

By consequence, Latin-script users tend to recognize glyphs resembling Greek 03C1 as non-visual variants of 0070, even where they are able to visually distinguish the two shapes and irrespective of the fact, that for Greek users, 03C1 is clearly distinctive from Latin 0070, therefore constituting a variant on non-visual grounds.

[G3] 0075-03C5: The two glyphs look “nearly identical” in Arial font (as shown in the second row in the image below).

Figure G03.1: 0075 vs. 03C5

LATIN SMALL LETTER U	0075	u	υ	03C5	GREEK SMALL LETTER UPSILON
		u	υ		
		u	υ		

028B-03C5: Also these two glyphs look “nearly identical” in Times New Roman font (as shown in the first row in the image below).

Figure G03.2: 028B vs. 03C5

[7] This similarity is not accidental but based on the historic relationship between the two characters, since p probably developed on the basis of Rho (together with Cyrillic Er (P)) (cf. [259]).

LATIN SMALL LETTER V WITH HOOK	028B	ʋ ʊ ᴜ	ᴠ ᴜ ᵁ	03C5	GREEK SMALL LETTER UPSILON

The same analysis applies to 00FA and 03CD, which are essentially the same characters with the addition of a modifying diacritic on top (an Acute in the case of Latin and a Tonos in the case of Greek script).

Since the former two variant sets feature one and the same code point from Greek script but two different code points from Latin script, this therefore imposes an in-script variant relationship between 0075 and 028B due to transitivity. The two code points Latin U (0075) and Latin V with Hook (028B) are however both used in a distinguishing manner in the orthography of Mossi – a language of Burkina Faso[8]. Latin GP foresees no issues and accepts the imposed variant relationship between the two code points, given that the variant relationship between U and V with Hook will still permit users from the Mossi community to employ both code points in labels, and since there won't be any particular security risk for the Mossi community, such as spoofing, as the variant set will have a the disposition of "blocked".

[G4] The Greek script code point 03B2 β (Letter Beta) is visually nearly identical due to font design to Latin script code point 00DF ß (Letter Sharp S). While those differences may be argued to be sufficiently different from a point of view of Greek script users, particularly the German users from the German language community may consider these code points confusable, since the typical rendering of the Greek variant is one of the forms taught to elementary school pupils as a hand-written form of the Latin-script code point 00DF ß across the German-speaking part of the script-using community, as demonstrated by Figure G04 below.

Figure G04: A handwritten form of the German Lexeme Grüße 'greetings', taken from https://de.wikipedia.org/wiki/%C3%9F#/media/File:Gruesse-Schneidler-Legende.png (Cf. e.g. https://de.wikipedia.org/wiki/%C3%9F#/media/File:Gruesse-Schneidler-Legende.png, where the German lexeme Grüße is spelled with such a hand-written form).

Therefore, adult script-users may also consider them in their minds to be the same, despite them being able to see the visual differences between the glyphs. Given that there are several Greek code points in a variant relationship to Latin code points, which are used by the German orthography, there are numerous plausible labels which could be made up, such as Greek. voβ, which may be identified with the German surname Voß. Additionally, German orthography commonly replaces 00DF ß by a sequence of two ss, and the same variation is also encountered in personal names, i.e. both Voß and Voss are used, which gives further scope to this potential confusion among readers (This issue is further complicated by the issue of IDNA compatibility – c.f. section 6.7.2). Accordingly, there is a concrete risk

[8] The official language of Burkina Faso is French - cf. //en.wikipedia.org/wiki/Burkina_Faso)

for the safety and stability of the zone, which should be dealt with at the level of the LGR definition by a variant relationship between those two code points (and others), despite them not being homoglyphs in a strict sense in a number of fonts.

[G5] In Courier New (represented by the third row in the screenshot of Figure G04 below) the glyphs are deemed nearly identical due to font design:

Figure G05: 01A1 vs. 03C3

LATIN SMALL LETTER O WITH HORN	01A1	ơ	σ	03C3	GREEK SMALL LETTER SIGMA
		ơ	σ		
		ơ	σ		

6.3.4 Generic Glyphs

In MSR, IP did also highlight the risk of "a number of homoglyphs of code points that cross scripts", providing examples of "circle glyph" from seven scripts:

"Because simple glyph shapes like this give effectively no hint of script identity, the IP encourages the Generation Panels to consider cross-script variants in such cases even for otherwise unrelated scripts. Among related scripts, there may be pairs of code points that are identical or nearly identical despite having more complex shapes. Where these can be used to form a label that is a homograph of a label in another script, they should be investigated for variant status." [MSR, page 22-23]

Most scripts have used similar graphic elements to distinguish basic letter shapes. Accordingly, there are a few shapes which are sufficiently generic that they occur in both related and unrelated scripts[9], such as the "circle glyph" referenced by IP. For Latin script, next to such a circle shape (Latin Small Letter O 006F and Latin Small Letter Open O 0254) this includes a single straight line (Latin Small Letter Dotless I 0131) or a crescent (Latin Small Letter C 0053). While these examples are independent code points in Latin script, in other scripts they may occur as combining mark code points.

Latin GP has identified the following variant relationships based on an analysis of generic glyphs of scripts included in [MSR], while all shortlisted variant candidates are presented in Appendix E.

Source Unicode Name	Source Code Point	Source Glyph	Variant Relationship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Rationale

6.4 Methodology for Developing In-Script Variants

In the case of visual variants, the following cases will be proposed as in-script variant:

[9] Only very few script creations occurred in complete isolation (cf. [DANIELS], inter alia), and most scripts have inspired one another through linguistic and cultural contact in terms of features expressed and graphic elements employed, irrespective of whether such scripts were related historically in a linguistic sense or not.

- Homoglyphs (i.e. visual score = 1): when any given pair of code points or code point sequences are visually identical as represented in a common use font (e.g., Arial, Times New Roman or Courier New) by Internet applications, such as internet browsers.

In the case of non-visual variants, the methodology is different depending on the type of suspected variance:
To test the hypotheses regarding the influence of handwriting on font design and the conception of readers, Latin GP looked at both handwriting samples as well as font design. The Latin GP looked comprehensively at font design when evaluating possible variants. In addition, in some cases, we looked at how handwriting typically renders letters in order to understand other ways that users might be accustomed to visualizing particular cases. This was not done systematically, just an aid to guide our review in particular cases.. In the case of shaping of base characters and diacritics, it was assumed that if such handwriting practices would cross-over into the printed forms, there should be fonts in which such potential variant pairs would turn out to be identical or nearly identical in appearance by a significant number of fonts:

While in the case of cross-script variants, the GP initially examined glyphs only in three widely used fonts, namely Arial, Courier New, and Times New Roman, in the case of in-script variants the GP choose to compare glyphs across a wide number of fonts to see if a significant minority of fonts gave way to a variant relationship between several code points. The reason for this is that there is no stability for the fonts employed by software which render strings. Not only are different fonts used across different types of software as well as across different platforms, but most clients offer the option to change the fonts, while some protocols allow the server to freely specify a different font just as well.

Therefore, the only way to predict what will be a plausible case for a variant relationship, is to look for trends in the rendering of certain glyphs, and see if even a significant minority of fonts renders the same glyph in a distinctly different manner. Since fonts designers are free to play with shapes and graphic elements, which make out glyphs recognizable by most users as one specific letter, there will always be ‘extreme’ cases, which may not be representative of the typical rendering of a character. However, if several fonts make use of the same graphical features in rendering of a glyph, such a shared feature may already give way to a similarity, which can pose a risk to stability and which may have to be dealt with at LGR-level.

In some cases the panel identified, potential variant cases, where a significant minority of glyphs shared some features, which suggested a variant relationship to other code points, however it was decided that it did not rise to the level of variant status based on a vote among members actively participating in that discussion, and in such cases the GP decided that such cases should be amended to Latin In-Script Confusables shortlist (cf. Appendix E), which should highlight such potential risks to any party looking to implement the LGR.

The GP used the website https://wordmark.it/ to compare strings across such a large number of different fonts. In order to attain results which were less dependent on pre-installed fonts on specific platforms and user interfaces, renderings were compared using Google Fonts, a font library employed by many APIs, instead of system fonts as rendered by that website.

Where shaping of base characters or diacritics was assumed to give way to variant candidates, strings containing the two code points, such as ff or vice versa, i.e. 0066 + 0192, or strings containing code points featuring the two diacritics, such as ăă or vice versa, i.e. 0103 + 01CE, were compared.

Meanwhile where spacing of base characters or stacking of diacritics were assumed to give way to variant candidates, strings containing the ligature plus the separate elements of the ligature, such as e.g. œoe or vice versa, i.e. 0153 + 006F + 0065 were compared, or strings containing code points featuring the stacked diacritics followed by the base character which the stacked diacritics modifies as well as sequences of code points featuring those diacritics separately (where available), such as e.g. ő◌ô◌ó, i.e. 1ED1 + 006F + 00F4 + 00F3.

This analysis was conducted for all code points featured in the suggested repertoire, as well as relevant candidates from other scripts. Code points not included in the repertoire as well as historical cases, such as w being a ligature of the sequence vv, were excluded, since such a derivation is part of the basic set of modern Latin script and therefore part of ASCII and as such out of scope for a variant analysis, since no IDN variant rules may occur which would impose variant relationships on non-IDN labels.

Variance based on compatibility to with old revisions of IDNA is discussed separately below in section 6.7.2.

6.5 In-Script Latin Variants

In the following, the variant sets confirmed by Latin GP are presented together with the relevant data and rationale. The full list of potential variant candidates shortlisted and analyzed by the GP including such cases which were not confirmed, is presented further below in Appendix D.

Table 14. In-Script Latin Variants

Source Unicode Name	Source Code Point	Source Glyph	Variant Relationship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Rationale
LATIN SMALL LETTER TURNED E	01DD	ǝ	↔	ə	0259	LATIN SMALL LETTER SCHWA	Blocked	In-script variant due to transitivity relationship of 04D9 Cyrillic Small Letter Schwa
LATIN SMALL LETTER DOTLESS I	0131	ı	↔	ɩ	0269	LATIN SMALL LETTER IOTA	Blocked	In-script variant due to transitivity relationship of 03B9 Greek Small Letter Iota
LATIN SMALL LETTER U	0075	u	↔	ʊ	028B	LATIN SMALL LETTER V WITH HOOK	Blocked	In-script variant due to transitivity relationship

								of 03C5 Greek Small Letter Upsilon
Source Unicode Name	Source Code Point	Source Glyph	Variant Relation- ship	Target Glyph	Target Code Point	Target Unicode Name	Disposition	Principle(s)
			↔					

6.7 Other Considerations for Variant Analysis

Apart from cross-script variants and in-script variants, Latin GP has also considered three other potential security risks, which could affect the safety and stability of the root zone, namely the effect of URL underlining, full compliance with IDNA 2003 but not IDNA 2008, as well as generic shapes of glyphs across related and unrelated scripts in [MSR]. The results of that analysis is summarized in the present section, with details of the analysis presented in Appendix D.

6.7.1 URL Underlining

In their communique by email from August 29, 2018, Integration Panel highlighted recent security risks based on the underlining of labels in URLs, which may obfuscate modifiers below or near the baseline, and asked the GP to take such risks into particular consideration:

“There are recent and widely published examples of phishing attacks using Latin IDNs in which the key features involved were diacritics below the letter. […] Of all diacritics, diacritics below can be difficult to distinguish or be prone to clipping -- there is less space below the baseline than between the typical lowercase glyph and the top of the line. […] The IP would like to encourage the LatinGP (and any other GP facing cases like this) to explicitly examine this example and other cases like it, where code points can become indistinguishable in common usage scenarios for IDNs, and formally conclude whether and how to take these into account when designing their LGR.“

In many user interfaces and software clients for different protocols making use of IDNs, IDN labels are linkified by converting them into protocol-specific hyperlinks and are usually highlighted by underlining the URL, and - in many instances - by color coding (visited and unvisited) hyperlinks. Often such URLs are further abbreviated by showing only the domain name label, in an attempt to present very simplified clickable links to internet users. Both the linkification and simplification as well as the underlining have consequences for the safety and stability of the root zone. While linkification and underlining cannot be predicted at all and is therefore a general and uncontrollable risk, the visual highlighting by means of underlining may obfuscate parts of such IDN-labels, where parts of letters or diacritics to such letters encoded by the code points of that label cross below the baseline and may therefore become entirely or partially obscured by the underline.

Accordingly, the GP decided to redeploy the same methodology and framework used for analysis of cross-script variants (see section 6.2 above) to identify which sets of code points were confusingly similar or visually the same due to this underlining. The same three fonts, namely Arial, Courier New,

and Times New Roman were used to compare strings, and it was decided that a visual score of 1-2, that is homoglyphs or code points nearly identical, would constitute variants.

While shortlisting relevant code points (the glyphs of which crossed into or below the baseline) were comparatively easy to identify and shortlist for analysis, it wasn't always clear which code points to compare them to and in several cases new or extended potential variant sets evolved after the data had been prepared and initially analyzed, since the obfuscation of certain 'extensions' of the letters led to a wider then expected similarity (which relates to the fact that most letters were developed based on others as discussed above in section 6.1.3). Generally, any code point included in the repertoire and represented by a glyph which features a modifier below the baseline was compared with the code point representing the same glyph without any modifier below the baseline, such as e.g ą, a̱, or ạ vs a. In the end, this analysis proofed to be even more difficult than e.g. the cross-script variant analysis and in many instances the final verdict on potential variant sets was arrived at only by means of majority vote. Any set of code points positively identified as variants was automatically assigned the disposition of Blocked.

The tables below present the variant candidate sets positively confirmed by the GP after such an analysis. All the candidate sets analyzed, including those which could not be confirmed are presented together with the data in Appendix D.6.

Table 15. In-Script Variants Due to Underlining

Group	Underlining							
Target			Source			Variant Candidate [Yes/No]	Disposition [Allocatable /	Rationale
Code Point	Glyph	Name	Code Point	Glyph	Name		Blocked]	
0061	a	LATIN SMALL LETTER A	0105	ą	LATIN SMALL LETTER A WITH OGONEK	YES	Blocked	Glyphs nearly identical due to underlining
0061	a	LATIN SMALL LETTER A	0061 + 0331	a̱	LATIN SMALL LETTER A + COMBINING MACRON BELOW	YES	Blocked	Glyphs nearly identical due to underlining

0061	a	LATIN SMALL LETTER A	1EA1	ạ	LATIN SMALL LETTER A WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0103	ă	LATIN SMALL LETTER A WITH BREVE	1EA7	ặ	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	1EAD	ậ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0062	b	LATIN SMALL LETTER B	00FE	þ	LATIN SMALL LETTER THORN	YES	Blocked	Glyphs nearly identical due to underlining
0064	d	LATIN SMALL LETTER D	1E13	ḓ	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0065	e	LATIN SMALL LETTER E	1EB9	ẹ	LATIN SMALL LETTER E WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining

0065	e	LATIN SMALL LETTER E	0065 + 0331	ḛ	LATIN SMALL LETTER E + COMBINING MACRON BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0065	e	LATIN SMALL LETTER E	0119	ę	LATIN SMALL LETTER E WITH OGONEK	YES	Blocked	Glyphs nearly identical due to underlining
0065	e	LATIN SMALL LETTER E	0019	ę	LATIN SMALL LETTER E WITH OGONEK	YES	Blocked	Glyphs nearly identical due to underlining
00E9	é	LATIN SMALL LETTER E WITH ACUTE	1EB9 + 0301	ẹ́	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT	YES	Blocked	Glyphs nearly identical due to underlining
00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX	1EC7	ệ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
00E8	è	LATIN SMALL LETTER E WITH GRAVE	1EB9 + 0300	ẹ̀	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT	YES	Blocked	Glyphs nearly identical due to underlining

025B	ɛ	LATIN SMALL LETTER OPEN E	ɛ̲	025B + 0331	LATIN SMALL LETTER OPEN E WITH COMBINING MACRON BELOW	YES	Blocked	Glyphs nearly identical due to underlining
025B + 0308	ɛ̈	LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS	025B + 0331 + 0308	ɛ̲̈	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS	YES	Blocked	Glyphs nearly identical due to underlining
0069	i	LATIN SMALL LETTER I	0069 + 0331	i̲	LATIN SMALL LETTER I + COMBINING MACRON BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0069	i	LATIN SMALL LETTER I	1ECB	ị	LATIN SMALL LETTER I WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006A	j	LATIN SMALL LETTER J	012F	į	LATIN SMALL LETTER I WITH OGONEK	YES	Blocked	Glyphs nearly identical due to underlining
006B	k	LATIN SMALL LETTER K	0137	ķ	LATIN SMALL LETTER K WITH CEDILLA	YES	Blocked	Glyphs nearly identical due to underlining

006C	l	LATIN SMALL LETTER L	013C	ļ	LATIN SMALL LETTER L WITH CEDILLA	YES	Blocked	Glyphs nearly identical due to underlining
006C	l	LATIN SMALL LETTER L	1E37	ḷ	LATIN SMALL LETTER L WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006C	l	LATIN SMALL LETTER L	1E3D	ḽ	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006D	m	LATIN SMALL LETTER M	1E43	ṃ	LATIN SMALL LETTER M WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006E	n	LATIN SMALL LETTER N	1E47	ṇ	LATIN SMALL LETTER N WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006E	n	LATIN SMALL LETTER N	1E49	n̲	LATIN SMALL LETTER N WITH LINE BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006E	n	LATIN SMALL LETTER N	014B	ŋ	LATIN SMALL LETTER ENG	YES	Blocked	Glyphs nearly identical due to underlining

0146	ņ	LATIN SMALL LETTER N WITH CEDILLA	1E4B	n̰	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW	YES	Blocked	Glyphs nearly identical due to underlining
006F	o	LATIN SMALL LETTER O	006F + 0331	o̲	LATIN SMALL LETTER O + COMBINING MACRON BELOW	YES	Blocked	Glyphs nearly identical due to underlining
00F3	ó	LATIN SMALL LETTER O WITH ACUTE	1ECD + 0301	ọ́	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT	YES	Blocked	Glyphs nearly identical due to underlining
00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX BELOW	1ED9	ộ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
00F2	ò	LATIN SMALL LETTER O WITH GRAVE	1ECD + 0300	ọ̀	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT	YES	Blocked	Glyphs nearly identical due to underlining
01A1	ơ	LATIN SMALL LETTER O WITH HORN	1EE3	ợ	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining

00F4	ô	LATIN SMALL LETTER OPEN O WITH CIRCUMFLEX	1ED9	ộ	LATIN SMALL LETTER OPEN O WITH CIRCUMFLEX AND DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0073	s	LATIN SMALL LETTER S	015F	ş	LATIN SMALL LETTER S WITH CEDILLA	YES	Blocked	Glyphs nearly identical due to underlining
015F	ş	LATIN SMALL LETTER S WITH CEDILLA	0219	ș	LATIN SMALL LETTER S WITH COMMA BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0074	t	LATIN SMALL LETTER T	021B	ț	LATIN SMALL LETTER T WITH CEDILLA	YES	Blocked	Glyphs nearly identical due to underlining
0074	t	LATIN SMALL LETTER T	1E71	ṯ	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0074	t	LATIN SMALL LETTER T	1E6D	ṭ	LATIN SMALL LETTER T WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining

021B	ț	LATIN SMALL LETTER T WITH COMMA BELOW	1E71	ṱ	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0075	u	LATIN SMALL LETTER U	1EE5	ụ	LATIN SMALL LETTER U WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
0079	y	LATIN SMALL LETTER Y	1EF5	ỵ	LATIN SMALL LETTER Y WITH DOT BELOW	YES	Blocked	Glyphs nearly identical due to underlining
1E3D	ḽ	Latin Small Letter L with Circumflex Below	013C	ļ	Latin Small Letter L with Cedilla	YES	Blocked	Glyphs nearly identical due to underlining
006E	n	Latin Small Letter N	0146	ņ	Latin Small Letter N with Cedilla	YES	Blocked	Glyphs nearly identical due to underlining
006F	o	Latin Small Letter O	1ECD	ọ	Latin Small Letter O with Dot Below	YES	Blocked	Glyphs nearly identical due to underlining
0254	ɔ	Latin Small Letter Open O	0254 + 0331	ɔ̲	Latin Small Letter Open O + Combining Macron Below	YES	Blocked	Glyphs nearly identical due to underlining

0073	s	Latin Small Letter S	1E63	ṣ	Latin Small Letter S with Dot Below	YES	Blocked	Glyphs nearly identical due to underlining	
0075	u	Latin Small Letter U	0173	ų	Latin Small Letter U with Ogonek	YES	Blocked	Glyphs nearly identical due to underlining	
01B0	ư	Latin Small Letter U with Horn	1EF1	ự	Latin Small Letter U with Horn and Dot Below	YES	Blocked	Glyphs nearly identical due to underlining	

6.7.2 IDNA2003 Compatibility

The Latin GP has analyzed and discussed the pros and cons of a different solutions to mitigate risks arising from IDNA 2003 compatibility issues, as discussed in detail in Appendix D.5.

In the case of Latin Small Letter Sharp S (00DF), the LGR proposes a solution including the code point with a variant relationship with the sequence of letters ‘ss’ (0073 0073), as follows:
Table 16. In-Script Variants for Latin Small Letter Sharp S (00DF)

Source Code Point	Variant Relationship	Target Code Point	Disposition
00DF Latin Small Letter Sharp S	→	0073 0073 Latin Small Letter S + Latin Small Letter S	Allocatable
0073 0073 Latin Small Letter S + Latin Small Letter S	→	00DF Latin Small Letter Sharp S	Blocked

The GP has not yet reached final consensus on a solution to the case of Latin Small Letter Dotless I (0131). The preliminary detailed analysis is presented in Appendix D.5.2.

7 Whole Label Evaluation Rules (WLE) and contextual rules

In LGR contextual rules or restrictions can be defined in several ways. One technique is called Whole Label Evaluation Rules (WLE).

For Latin LGR no WLEs are planned, but the analysis is yet to be conducted. The only code points that need contextual restrictions are the non-space marks (see section 5.3.1). The restriction of those is that they are only allowed, in the Latin LGR, after specific letter code points. That restriction is achieved by not listing the marks as individual code points in the LGR, but only as part of the permitted sequence of a letter code point and the non-space mark (in one instance, the sequence of a letter code point plus two ordered non-space marks).

8. Contributors

Michael Bauland
Chris Dillon (Chair of Latin GP until 2016 chair)
Mats Dufberg
Hazem Hezzah
Bill Jouris
Meikal Mumin
Jean Paul Nkurunziza
Dennis Tan Tanaka
Mirjana Tasić (Chair of Latin GP from 2016)

9 References

9.1 References used in developing Repertoire

[0] The Unicode Consortium, Unicode® 11.0.0, http://www.unicode.org/versions/Unicode11.0.0/, 5 September 2018
[100], ICANN, Second Level Reference Label Generation Rules for Spanish, https://www.icann.org/sites/default/files/packages/lgr/lgr-second-level-spanish-30aug16-en.html, 31 August 2018
[101], Omniglot, Czech (čeština), http://www.omniglot.com/writing/czech.htm, 31 August 2018
[102], Omniglot, Icelandic (Íslenska), http://www.omniglot.com/writing/icelandic.htm, 31 August 2018
[103], Omniglot, Faroese (føroyskt mál), http://www.omniglot.com/writing/faroese.htm, 31 August 2018
[104], Wikipedia, Burundi Bwacu, https://en.wikipedia.org/wiki/Burundi_Bwacu#Kirundi_.28with_tonal_diacritics_.E2.80.94_utw.C3.A2tuzo.29, 31 August 2018
[105], Omniglot, Chuukese (Chuuk), http://www.omniglot.com/writing/chuukese.htm, 31 August 2018
[106], SCRIPTSOURCE, Galician written with Latin script, http://www.webcitation.org/6siTI8ieQ, 31 August 2018
[107], Omniglot, Lule Sámi (julevsámegiella), http://www.omniglot.com/writing/lulesami.htm, 31 August 2018
[108], Wikipedia, Northern Sami, https://en.wikipedia.org/wiki/Northern_Sami, 4 September 2018
[109], Omniglot, Vietnamese (tiếng việt / 喃越), http://www.omniglot.com/writing/vietnamese.htm, 4 September 2018
[110], Omniglot, Romanian (limba română), http://www.omniglot.com/writing/romanian.htm, 4 September 2018

[113], Omniglot, Skolt Sámi (Sää´mǩiõll / Nuõrttsää’m),
http://www.omniglot.com/writing/skoltsami.htm, 4 September 2018
[114], Omniglot, French (français), http://omniglot.com/writing/french.htm, 4 September 2018
[115], Omniglot, West Frisian (Frysk), http://www.omniglot.com/writing/westfrisian.htm, 4 September 2018
[116], Omniglot, Friulian (furlan/marilenghe), http://www.omniglot.com/writing/friulian.htm, 4 September 2018
[117], Anteriormente Summer Institute of Linguistics, Pequeno dicionário: Xavante-Português, Português-Xavante, http://www.silbrasil.org.br/resources/archives/17019, 4 September 2018
[119], Omniglot, German (Deutsch), http://www.omniglot.com/writing/german.htm, 4 September 2018
[120], Omniglot, Finnish (suomi), http://www.omniglot.com/writing/finnish.htm, 4 September 2018
[121], Omniglot, Turkmen (Türkmen dili / Түркмен дили),
http://www.omniglot.com/writing/turkmen.htm, 4 September 2018
[122], Omniglot, Estonian (eesti keel), http://www.omniglot.com/writing/estonian.htm, 4 September 2018
[123], Omniglot, Swedish (svenska), http://www.omniglot.com/writing/swedish.htm, 4 September 2018
[124], Omniglot, Yapese (Waab), http://www.omniglot.com/writing/yapese.htm, 4 September 2018
[125], Omniglot, Dinka (Thuɔŋjäŋ), https://www.omniglot.com/writing/dinka.php, 4 September 2018
[126], Omniglot, Kaqchikel (Kaqchikel Ch'ab'äl), http://www.omniglot.com/writing/kaqchikel.htm, 4 September 2018
[127], Omniglot, Bashkir/Bashkort (Башҡорт теле / Başqort tele),
http://www.omniglot.com/writing/bashkir.htm, 4 September 2018
[128], Omniglot, Alsatian (Ëlsässisch), https://www.omniglot.com/writing/alsatian.htm, 4 September 2018
[129], Wikipedia, Nuer language, https://en.wikipedia.org/wiki/Nuer_language, 4 September 2018
[130], Omniglot, Italian (italiano), http://www.omniglot.com/writing/italian.htm, 4 September 2018
[131], Wikipedia, Italian orthography, https://en.wikipedia.org/wiki/Italian_orthography, 4 September 2018
[132], Omniglot, Wolof (Wollof), http://www.omniglot.com/writing/wolof.htm, 4 September 2018
[133], Omniglot, Latvian (latviešu valoda), http://www.omniglot.com/writing/latvian.htm, 4 September 2018
[134], Omniglot, Tongan (Faka-Tonga), http://www.omniglot.com/writing/tongan.htm, 4 September 2018
[135], Omniglot, Hawaiian (‘Ōlelo Hawai‘i), http://www.omniglot.com/writing/hawaiian.htm, 4 September 2018
[136], Omniglot, Marshallese (kajin ṃajeḷ), http://www.omniglot.com/writing/marshallese.php, 4 September 2018
[137], Omniglot, Polish (polski), http://www.omniglot.com/writing/polish.htm, 4 September 2018
[138], Omniglot, Lithuanian (lietuvių kalba), http://www.omniglot.com/writing/lithuanian.htm, 4 September 2018
[139], Omniglot, Danish (dansk), http://www.omniglot.com/writing/danish.htm, 4 September 2018
[140], Omniglot, Chamorro (chamoru), http://www.omniglot.com/writing/chamorro.htm, 4 September 2018
[141], Omniglot, Umbundu (Úmbúndú), http://www.omniglot.com/writing/umbundu.htm, 4 September 2018
[142], Omniglot, Guaraní (Avañe'ẽ), http://www.omniglot.com/writing/guarani.htm, 4 September 2018
[143], Wikipedia, Guarani alphabet, https://en.wikipedia.org/wiki/Guarani_alphabet, 4 September 2018

[144], Omniglot, Nauruan (Ekaiairũ Naoero), http://www.omniglot.com/writing/nauruan.htm, 4 September 2018
[145], Omniglot, Khoekhoe (Khoekhoegowab), https://www.omniglot.com/writing/khoekhoe.htm, 4 September 2018
[146], Omniglot, Nuer (Naath), https://www.omniglot.com/writing/nuer.htm, 4 September 2018
[147], Omniglot, Hausa (Harshen Hausa / هَرْشَن هَوْس), http://www.omniglot.com/writing/hausa.htm , 4 September 2018
[148], Omniglot, Dagaare, http://www.omniglot.com/writing/dagaare.htm, 4 September 2018
[149], Omniglot, Fula (Fulfulde, Pulaar, Pular'Fulaare), http://www.omniglot.com/writing/fula.htm, 4 September 2018
[150], Omniglot, Croatian (Hrvatski), http://www.omniglot.com/writing/croatian.htm, 4 September 2018
[151], Omniglot, Serbian (српски / srpski), http://www.omniglot.com/writing/serbian.htm, 4 September 2018
[152], Wikipedia, Polish language, https://en.wikipedia.org/wiki/Polish_language, 4 September 2018
[153], Omniglot, Slovak (slovenčina), http://www.omniglot.com/writing/slovak.htm, 4 September 2018
[154], Evertype Publishing, Lithuanian lietuvių kalba Version 1.1, http://www.evertype.com/alphabets/lithuanian.pdf, 4 September 2018
[157], Omniglot, Turkish (Türkçe), http://www.omniglot.com/writing/turkish.htm, 4 September 2018
[158], Omniglot, Kurdish (Kurdî / کوردی), http://www.omniglot.com/writing/kurdish.htm, 4 September 2018
[159], Omniglot, Azerbaijani (آذربایجانجا دیلي / Азәрбајҹан дили / Azərbaycan dili), http://www.omniglot.com/writing/azeri.htm, 4 September 2018
[160], Omniglot, Basque (euskara), http://www.omniglot.com/writing/basque.htm, 4 September 2018
[161], Wikipedia, Basque language, https://en.wikipedia.org/wiki/Basque_language#Writing_system, 4 September 2018
[163], Omniglot, Maltese (Malti), http://www.omniglot.com/writing/maltese.htm, 4 September 2018
[164], Omniglot, Venda (Tshivenḓa / Luvenḓa), http://www.omniglot.com/writing/venda.htm, 4 September 2018
[166], Wikipedia, Hausa language, https://en.wikipedia.org/wiki/Hausa_language , 4 September 2018
[167], Christian Chanard and Rhonda L. Hartell. 2014 , Pulaar sound inventory (AA), http://phoible.org/inventories/view/809#tsource, 4 September 2018
[168], Omniglot, Brahui (Bráhuí / براوی), https://www.omniglot.com/writing/brahui.htm, 4 September 2018
[169], Wikipedia, Fon language, https://en.wikipedia.org/wiki/Fon_language , 4 September 2018
[170], Omniglot, Ewe (Eʋegbe), http://www.omniglot.com/writing/ewe.htm, 4 September 2018
[172], Omniglot, Sorbian (hornjoserbsce/dolnoserbski), https://www.omniglot.com/writing/sorbian.htm, 4 September 2018
[173], Peace corps, Botswana, An Introduction to Setswana Language, http://files.peacecorps.gov/multimedia/audio/languagelessons/botswana/Bw_Setswana_Language_Lessons.pdf, 4 September 2018
[174], Omniglot, Tswana (Setswana), http://omniglot.com/writing/tswana.php, 4 September 2018
[175], Wikipedia, Afrikaans, https://en.wikipedia.org/wiki/Afrikaans, 4 September 2018
[176], Omniglot, Albanian (shqip / gjuha shqipe), http://www.omniglot.com/writing/albanian.htm, 4 September 2018
[177], Wikipedia, Albanian alphabet, https://en.wikipedia.org/wiki/Albanian_alphabet, 4 September 2018

[178], Compiled by Jay Hinner, So you want to learn chuukese?, http://www.jesuitvolunteers.org/wp-content/uploads/2015/08/So_you_want_to_learn_chuukese_-_only_for_Chuuk_JVs.pdf, 4 September 2018
[179], Wikipedia, Uyghur Latin alphabet, https://en.wikipedia.org/wiki/Uyghur_Latin_alphabet, 4 September 2018
[180], Omniglot, Drehu (De'u), http://www.omniglot.com/writing/drehu.php, 4 September 2018
[181], Omniglot, Yoruba (Èdè Yorùbá), http://www.omniglot.com/writing/yoruba.htm, 4 September 2018
[182], Omniglot, Haitian Creole (Kreyòl ayisyen), http://www.omniglot.com/writing/haitiancreole.htm, 4 September 2018
[183], Wikipedia, Haitian Creole, https://en.wikipedia.org/wiki/Haitian_Creole#Orthography, 4 September 2018
[184], Omniglot, Minangkabau (Baso Minangkabau / باسو مينغكاباو), http://www.omniglot.com/writing/minangkabau.htm, 4 September 2018
[185], Omniglot, Palauan (a tekoi er a Belau), http://www.omniglot.com/writing/palauan.htm, 4 September 2018
[186], Omniglot, Cubeo (pãmié), http://www.omniglot.com/writing/cubeo.htm, 4 September 2018
[187], Editorial Alberto Lleras Camargo, Diccionario Ilustrado Bilingüe cubeo-español español-cubeo, https://www.sil.org/system/files/reapdata/10/58/27/105827858436939923317665060690738956 20/40337_01.pdf, 4 September 2018
[188], Omniglot, Inari Saami (Anarâškielâ), http://www.omniglot.com/writing/inarisami.htm, 4 September 2018
[189], Omniglot, Compiled by Wolfram Siegel, DAGBANI, http://www.omniglot.com/charts/dagbani.pdf, 4 September 2018
[190], Omniglot, Ewondo, http://www.omniglot.com/writing/ewondo.php, 4 September 2018
[191], Omniglot, Luganda (Oluganda), http://www.omniglot.com/writing/ganda.php, 4 September 2018
[192], Omniglot, Adzera, http://www.omniglot.com/writing/adzera.htm, 4 September 2018
[193], Omniglot, Ga (Gã), http://www.omniglot.com/writing/ga.htm, 4 September 2018
[194], Omniglot, Duala (Duálá), http://www.omniglot.com/writing/duala.php, 4 September 2018
[195], Omniglot, Soga (Lusoga), http://www.omniglot.com/writing/soga.htm, 4 September 2018
[196], Omniglot, Alur (Lur), http://www.omniglot.com/writing/alur.htm, 4 September 2018
[197], Omniglot, Mandinka (Mandi'nka kango / لغة مندنكا), http://www.omniglot.com/writing/mandinka.htm, 4 September 2018
[198], Omniglot, Acholi (Lwo), https://www.omniglot.com/writing/acholi.htm, 4 September 2018
[199], Omniglot, Bambara (Bamanankan), http://www.omniglot.com/writing/bambara.htm, 4 September 2018
[200], Omniglot, Raga (Hano), http://www.omniglot.com/writing/raga.htm, 4 September 2018
[201], Omniglot, Tatar (tatarça / татарча / تاتارچا), http://www.omniglot.com/writing/tatar.htm, 4 September 2018
[202], Omniglot, Zaza (Zazaki / زازاکی), https://www.omniglot.com/writing/zazaki.htm, 4 September 2018
[203], Wikipedia, Turkish alphabet, https://en.wikipedia.org/wiki/Turkish_alphabet, 4 September 2018
[204], School of English, Adam Michiewicz University, Poznań, Poland, Poznań Studies in Contemporary Linguistics 43(1),2007, pp. 169-180, A Demographic Igbo Orthography, https://www.degruyter.com/downloadpdf/j/psicl.2007.43.issue-1/v10010-007-0009-0/v10010-007-0009-0.pdf, 4 September 2018
[205], Omniglot, Igbo (Asụsụ Igbo), http://www.omniglot.com/writing/igbo.htm, 4 September 2018

[206], ItalianPod101, Italian Accents and Proper Italian Pronunciation, https://www.italianpod101.com/italian-accents, 4 September 2018
[208], Reverso Dictionary, venerdì translation | Italian-English dictionary, http://dictionary.reverso.net/italian-english/venerd%C3%AC, 4 September 2018
[209], Omniglot, Kikuyu (Gĩkũyũ), http://www.omniglot.com/writing/kikuyu.htm, 4 September 2018
[210], Omniglot, Hixkaryána, http://www.omniglot.com/writing/hixkaryana.htm, 4 September 2018
[211], Omniglot, Maasai (ɔl Maa), http://www.omniglot.com/writing/maasai.htm, 4 September 2018
[212], Omniglot, Mossi (Mòoré), http://www.omniglot.com/writing/mossi.htm, 4 September 2018
[213], Omniglot, Jenesis. The Bible in Marshallese, 2009., Contributed by Wolfgang Kuhl, http://www.omniglot.com/babel/marshallese.htm, 4 September 2018
[214], Wikipedia, Cedilla, https://en.wikipedia.org/wiki/Cedilla#Marshallese, 4 September 2018
[215], Wikipedia, Marshallese language, https://en.wikipedia.org/wiki/Marshallese_language#Display_issues, 4 September 2018
[216], Trussel, Marshallese-English Online Dictionary, http://www.trussel2.com/MOD/, 4 September 2018
[218], Omniglot, Susu (Sosoxi), https://www.omniglot.com/writing/susu.htm, 4 September 2018
[219], Omniglot, Zarma (Zarmaciine), https://www.omniglot.com/writing/zarma.htm, 4 September 2018
[220], Omniglot, Pitjantjatjara, https://www.omniglot.com/writing/pitjantjatjara.htm, 4 September 2018
[221], Omniglot, Spanish (español/castellano), http://www.omniglot.com/writing/spanish.htm, 4 September 2018
[222], Omniglot, Filipino (wikang Filipino), http://www.omniglot.com/writing/filipino.htm, 4 September 2018
[223], Omniglot, Chavacano, http://www.omniglot.com/writing/chavacano.php, 4 September 2018
[224], Wikipedia, Ilocano language, https://en.wikipedia.org/wiki/Ilocano_language#Modern_alphabet, 4 September 2018
[225], Omniglot, Quechua (Runasimi), http://www.omniglot.com/writing/quechua.htm, 4 September 2018
[226], Wikipedia, Quechua alphabet, https://en.wikipedia.org/wiki/Quechua_alphabet, 4 September 2018
[227], Omniglot, Cape Verdean Creole (Kriolu), http://www.omniglot.com/writing/kriol.php, 4 September 2018
[228], Omniglot, Waray-Waray, http://www.omniglot.com/writing/waray.php, 4 September 2018
[229], Omniglot, Lozi (siLozi), http://www.omniglot.com/writing/lozi.htm, 4 September 2018
[230], africanlanguages.com, Sesotho sa Leboa (Northern Sotho), http://africanlanguages.com/northern_sotho/, 4 September 2018
[231], Omniglot, Low German (Plattdüütsch / Nedderdüütsch), https://www.omniglot.com/writing/lowgerman.htm, 4 September 2018
[232], Wikipedia, Chechen language, https://en.wikipedia.org/wiki/Chechen_language, 4 September 2018
[233], Omniglot, Hungarian (magyar), http://www.omniglot.com/writing/hungarian.htm, 4 September 2018
[234], Wikipedia, Hungarian alphabet, https://en.wikipedia.org/wiki/Hungarian_alphabet, 4 September 2018
[235], Omniglot, Khoekhoe (Khoekhoegowab), http://www.omniglot.com/writing/khoekhoe.htm, 4 September 2018
[236], Omniglot, Lingala, http://www.omniglot.com/writing/lingala.htm, 4 September 2018

[237], Omniglot, Akan, https://www.omniglot.com/writing/akan.htm, 4 September 2018
[238], Wikipedia, Mossi language, https://en.wikipedia.org/wiki/Mossi_language, 4 September 2018
[239], SIL-Sudan, OCCASIONAL PAPERS in the study of SUDANESE LANGUAGES No. 9, https://www.sil.org/system/files/reapdata/10/06/46/100646256099282892829790816212446104791/OPSL_9.pdf (p. 75), 4 September 2018
[240], Omniglot, Kanuri, http://www.omniglot.com/writing/kanuri.htm, 4 September 2018
[241], Omniglot, Bugis (Basa Ugi), http://www.omniglot.com/writing/bugis.htm, 4 September 2018
[242], Omniglot, Mizo (Mizo ṭawng), http://www.omniglot.com/writing/mizo.htm, 4 September 2018
[243], Omniglot, Miskito (Mískitu), http://www.omniglot.com/writing/miskito.htm, 4 September 2018
[244], Omniglot, Zaza (Zazaki / زازاکی), http://www.omniglot.com/writing/zazaki.htm, 4 September 2018
[245], Wikipedia, Papiamento, https://en.wikipedia.org/wiki/Papiamento, 4 September 2018
[246], Omniglot, Papiamento (Papiamentu), http://www.omniglot.com/writing/papiamento.php, 4 September 2018
[247], Omniglot, Chichewa (Chicheŵa), http://www.omniglot.com/writing/chichewa.php, 4 September 2018
[248], Native Languages of the Americas website, Vocabulary in Native American Languages: Mam Words, http://www.native-languages.org/mam_words.htm, 4 September 2018
[249], Omniglot, Mam (Qyol Mam), http://www.omniglot.com/writing/mam.htm, 4 September 2018
[250], Wikipedia, Pulaar language, https://en.wikipedia.org/wiki/Pulaar_language, 4 September 2018
[251], Wikipedia, Fula language, https://en.wikipedia.org/wiki/Fula_language#Writing_systems, 4 September 2018
[252], Wikipedia, Polish alphabet, https://en.wikipedia.org/wiki/Polish_alphabet, 4 September 2018
[253], Wikipedia, French orthography, https://en.wikipedia.org/wiki/French_orthography, 4 September 2018
[254], Omniglot, Yoruba (Èdè Yorùbá), https://www.omniglot.com/writing/yoruba.htm, 4 September 2018
[255], Omniglot, Esperanto, http://www.omniglot.com/writing/esperanto.htm, 4 September 2018
[256], Omniglot, Welsh (Cymraeg), http://www.omniglot.com/writing/welsh.htm, 4 September 2018
[257], Wikipedia, List of Latin-script letters, https://en.wikipedia.org/wiki/List_of_Latin-script_letters, 4 September 2018
[258], Omniglot, Montenegrin, https://www.omniglot.com/writing/montenegrin.htm, 20 March 2019
[259], Wikipedia, Rho, https://en.wikipedia.org/wiki/Rho, 24 September 2019

9.2 Other references

[Procedure] Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013). https://www.icann.org/en/system/files/files/lgr-procedure-20mar13-en.pdf

[Requirements] Integration Panel “Requirements for LGR Proposals from Generation Panels”. https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals- 20150424.pdf

[Considerations] VIP Study Group “Considerations in the use of the Latin script in variant internationalized top-level domains” (Los Angeles, California: ICANN, October, 2011). https://archive.icann.org/en/topics/new-gtlds/latin-vip-issues-report-07oct11-en.pdf

[UCD] The Unicode Consortium, Unicode Character Database.
http://www.unicode.org/Public/UCD/latest/

[Katz & Frost 1992]]Katz, Leonard & Ram Frost. 1992. "The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis". *Haskins Laboratories Status Report on Speech Research* 111/112. 147–160.

[Wikipedia-Latin script] Latin script. Cached version retrieved 2017-02-14.
http://www.webcitation.org/6oGZwoNUu

[Wikipedia-Capital ß] Capital ß. Cached version retrieved 2018-01-17.
http://www.webcitation.org/6wXlGtfqc

[Wikipedia - Ejectives] Ejectives. Cached version retrieved 2018-01-19.
http://www.webcitation.org/6waqfVtj3

[Wikipedia - ASCII] ASCII. Cached version retrieved 2018-01-20. http://www.webcitation.org/6waqfVtj3

[Rogers] Rogers, Henry. 2005. *Writing systems: A linguistic approach*. Malden, Massachusetts: Blackwell Publishing.

[MSR] Maximal Starting Repertoire https://www.icann.org/resources/pages/msr-2015-06-21-en

[ARMENIAN] Armenian Generation Panel, "Proposal for an Armenian Script Root Zone LGR. Version 3." (Los Angeles, California: ICANN, June, 2015. https://www.icann.org/public-comments/proposal-armenian-lgr-2015-07-22-en

[CYRILLIC] Cyrillic Generation Panel, "Proposal for Cyrillic Script Root Zone Label Generation Rules. Version 1.4." (Los Angeles, California: ICANN, October, 2017. https://www.icann.org/public-comments/cyrillic-lgr-2017-10-17-en

[DANIELS] Daniels, Peter T. 1992. "The syllabic origin of writing and the segmental origin of the alphabet." *The Linguistics of Literacy* in Downing, Pamela A., Lima, Susan D., & Noonan, Micahel (Eds.), 83-110. John Benjamins, Amsterdam.

Appendix A: Updated MSR during Latin GP work

When the work of Latin Generation Panel started the Maximal Starting Repertoire (MSR) version was 2 (MSR-2). As a result of the investigation and analysis of the languages, the Panel requested an extension of MSR with the six code points in table A1 below. Three of those were accepted by the Integration Panel (IP) and could therefore be included in the repertoire. The other three were rejected and could not be included.

Table A1. Code points not found in MSR-2 and requested to be included in updated MSR.

Unicode	Glyph	Unicode name	Languages	Reference supporting inclusion	MSR-3 status

0268	ɨ	LATIN SMALL LETTER I WITH STROKE	Cubeo (3) Dagbani (4) HIxkaryána (4)	http://www.omniglot.com/writing/cubeo.htm http://www.omniglot.com/charts/dagbani.pdf http://www.omniglot.com/writing/hixkaryana.htm	INCLUDED
0272	ɲ	LATIN SMALL LETTER N WITH LEFT HOOK	Susu (4) Zarma (4)	https://www.omniglot.com/writing/susu.htm https://www.omniglot.com/writing/zarma.htm	INCLUDED
01C0	ǀ	LATIN LETTER DENTAL CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm	EXCLUDED
01C1	ǁ	LATIN LETTER LATERAL CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/ http://www.omniglot.com/writing/khoekhoe.htm	EXCLUDED
01C2	ǂ	LATIN LETTER ALVEOLAR CLICK	Khoekhoe(4)	https://www.britannica.com/topic/Khoisan-languages https://en.wikipedia.org/wiki/Khoe_languages https://www.newera.com.na/tag/khoekhoegowab/http://www.omniglot.com/writing/khoekhoe.htm	EXCLUDED
1E3D	ḽ	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW	Venda (1)	http://www.omniglot.com/writing/venda.htm	INCLUDED

MSR was upgraded to version MSR-3 on January 17, 2018, with three more Latin script code points as could be seen in table A1. A description of changes to MSR-3 can be found in https://www.icann.org/en/system/files/files/msr-3-overview-28mar18-en.pdf.

In October 2018, the Panel discovered three more code points needed for Venda language, but not included in MSR (MSR-3). The Panel then requested the inclusion of the three code points in table A2 below to the IP on 2018-10-10.

Table A2. Code points not found in MSR-3 and requested to be included in updated MSR.

Unicode	Glyph	Unicode name	Languages	Reference supporting inclusion	MSR-4 status
1E13	ḓ	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW	Venda (1)	http://www.omniglot.com/writing/venda.htm	INCLUDED
1E4B	ṋ	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW	Venda (1)	http://www.omniglot.com/writing/venda.htm	INCLUDED
1E71	ṯ	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	Venda (1)	http://www.omniglot.com/writing/venda.htm	INCLUDED

All three were included in the updated [MSR] (MSR-4).

Appendix B: Table Of Processed Languages Used to Develop Latin Script Repertoire

Table B.1. Processed Languages Used to Develop Latin Script Repertoire

	Language	ISO 639-3	EGIDS
1.	Afrikaans,	afr	1

2.	Albanian, Arbëreshë Albanian [aae] (Italy) Arvanitika Albanian [aat] (Greece) Gheg Albanian [aln] (Serbia) Tosk Albanian [als]	sqi	1
3.	Azeri, Azerbaijani	azj	1
4.	Chamorro, Chamorru Tjamoro	cha	1
5.	Croatian, Hrvatski	hrv	1
6.	Czech Bohemian Cestina	ces	1
7.	Danish, Dansk Rigsdansk	dan	1
8.	Dutch, Hollands Nederlands	nld	1
9.	English	eng	1
10.	Estonian Eesti keel	ekk	1
11.	Filipino	fil	1
12.	Finnish, Suomi	fin	1
13.	French, Français	fra	1
14.	German Deutsch Tedesco	deu	1
15.	Greenlandic Kalaallisut, Inuktitut,	kal	1
16.	Guarani Avañe'e Paraguayan	grn	1
17.	Haitian Creole, Creole, Haitian Creole Western Caribbean Creole	hat	1
18.	Hungarian Magyar	hun	1
19.	Icelandic Íslenska	isl	1
20.	Indonesian	ind	1
21.	Irish Erse Gaeilge Gaelic Irish	gle	1
22.	Italian Italiano	ita	1
23.	Kazakh, Kaisak, Kazak, Kosach, Qazaq	kaz	1
24.	Kinyarwanda, Ikinyarwanda, Orunyarwanda, Ruanda, Rwandan, Urunyaruanda	kin	1
25.	Kiribati, Gilbertese, Ikiribati, I-Kiribati, Kiribatese	gil	1
26.	Kirundi, Rundi Urundi,	run	1
27.	Latvian, "Lettisch" (pej.), "Lettish" (pej.)	lav	1

28.	Lithuanian, Lietuvi, Lietuviskai, Litauische, Litewski, Litovskiy	lit	1
29.	Malagasy, Plateau, Malagasy, Malgache, Official Malagasy, Standard Malagasy	plt	1
30.	Malay,	msa	1
31.	Maltese, Malti	mlt	1
32.	Marshallese, Ebon, Montenegrin (mne)	mah	1
33.	Ndebele, Isikhethu, IsiNdebele, Ndzundza, Nrebele, Southern Ndebele, Transvaal Ndebele	nbl	1
34.	Niuean, Niue, “Niuefekai” (pej.)	niu	1
35.	Northern Sotho, Pedi, Sepedi, Sesotho sa Leboa, Transvaal Sotho	nso	1
36.	Norwegian, Norsk	nor	1
37.	Papiamento, Papiamentu, Curaçoleño, Curassese, Papiamen, Papiamentoe	pap	1
38.	Polish, Polnisch, Polski	pol	1
39.	Portuguese,	por	1
40.	Romanian, Daco-Rumanian, Moldavian, Rumanian	ron	1
41.	Samoan,	smo	1
42.	Sango, Sangho	sag	1
43.	Serbian, srpski, српски,	srp	1
44.	Seychelles Creole, Seselwa Creole, Creole, Ilois, Kreol, Kreol Seselwa, Seselwa, Seychelles Creole French, Seychellois Creole	crs	1
45.	Slovak, Slovakian, Slovencina	slk	1
46.	Slovenian, Slovenscina, Slovene	slv	1
47.	Somali, Af-Maxaad Tiri, Af-Soomaali, Common Somali, Soomaaliga, Standard Somali	som	1
48.	Southern Sotho, Sesotho, Sisutho, Souto, Suthu, Suto	sot	1
49.	Spanish, Castellano, Castilian, Español	spa	1
50.	Swahili, Kisuaheli, Kiswahili	swh	1
51.	Swati/Swazi, Isiswazi, Ngwane, Phuthi, Siswati, Swazi, Tekela, Tekeza	ssw	1

52.	Swedish, Ruotsi, Svenska	swe	1
53.	Tahitian,	tah	1
54.	Tok Pisin, Melanesian English, Neomelanesian, New Guinea Pidgin English, Pidgin, Pisin	tpi	1
55.	Tongan, Tonga	ton	1
56.	Tsonga, Shangaan, Shangana, Shitsonga, Thonga, Tonga, Xitsonga	tso	1
57.	Tswana, Beetjuans, Chuana, Coana, Cuana, Sechuana, Setswana	tsn	1
58.	Turkish, Anatolian, Türkçe, Türkisch	tur	1
59.	Turkmen, Trukhmen, Trukhmeny, Turkmani, Turkmanian, Turkmenler, Turkomans	tuk	1
60.	Uzbek, Özbek, Usbeki, Uzbak, Uzbeki	uzb	1
61.	Venda, Chivenda, Tshivenda	ven	1
62.	Vietnamese, Annamese, Ching, Gin, Jing, Kinh, Viet	vie	1
63.	Xhosa, “Cauzuh” (pej.), Isixhosa, Koosa, Xosa	xho	1
64.	Zulu, Isizulu, Zunda	zul	1
65.	Basque, Euskara Euskera Vascuense	eus	2
66.	Catalan, Català Catalán Catalan-Valencian-Balear Catalonian Valencian	cat	2
67.	Chechen, Galancho Nokchiin Muott Nokhchiin	che	2
68.	Chuukese Chuuk Lagoon Chuukese Ruk Truk Trukese	chk	2
69.	Faroese Føroyskt	fao	2
70.	Frisian Fries Frysk	fry	2
71.	Galician Galego Gallego	glg	2
72.	Garo Garrow Mande Mandi	grt	2
73.	Hausa Abakwariga Habe Haoussa Hausawa Kado Mgbakpa	hau	2
74.	Hawaiian Olelo Hawai’i ’Olelo Hawai’i Makuahine	haw	2
75.	Igbo	ibo	2
76.	Inari Sámi Anarâškielâ Anar “Finnish Lapp” (pej.) “Inari Lappish” (pej.) “Lapp” (pej.) Saami Saame Sámi Samic	smn	2

77.	Konkani, Bankoti, Central Konkan, Concorinum, Cugani, Kathodi, Katvadi, Konkan Standard, Konkanese, Konkani Mangalorean, Kunabi, North Konkan	knn	2
78.	Kurdish,	kur	2
79.	Lingala, Ngala	lin	2
80.	Lule Sámi, “Lapp” (pej.), Lule, Saami	smj	2
81.	Mirandese, Mirandês	mwl	2
82.	Miskito, Marquito, Mískitu, Miskuto, Mísquito, Mosquito	miq	2
83.	Northern Sámi, Saami North, “Lapp” (pej.), North Sámi, “Northern Lappish” (pej.), Northern Saami, “Norwegian Lapp” (pej.), Saami, Same, Sámegiella, Samic	sme	2
84.	Palauan, Belauan, Palau	pau	2
85.	Pohnpeian, Ponapean	pon	2
86.	Skolt Sámi, “Lapp” (pej.), Southern Lapp	sma	2
87.	Tatar, Tartar	tat	2
88.	Tshiluba, Luba-Kasai, Bena-Lulua, Ciluba, Luba-Lulua, Luva, Tshiluba, Western Luba	lua	2
89.	Uyghur, Uighuir, Uighur, Uiguir, Uigur, Uygur, Weiwu’er, Wiga	uig	2
90.	Wa, Paruk, Baraog, Phalok, Praok, Standard Wa, Wa	prk	2
91.	Welsh, Cymraeg	cym	2
92.	West Frisian, Fries, Frysk	fry	2
93.	Yapese,	yap	2
94.	Yoruba, Yariba, Yooba	yor	2
95.	Akan, Twi, Ajan Twi	aka	3
96.	Bislama, Bichelamar	bis	3
97.	Bugis Basa Ugi Boegineesche Boeginezen Bugi Buginese De’ Rappang Buginese Ugi	bug	3
98.	Cebuano, Binisaya Bisayan Sebuano Sugbuanon Sugbuhanon Visayan	ceb	3
99.	Chichewa Chewa Chinyanja Nyanja Nyanja-Chewa	nya	3

100.	Cubeo Cuveo Hehenawa Hipnwa Kobeua Kobewa Kubwa Pamiwa	cub	3
101.	Duala Diwala Douala Dualla Dwala Dwela Sawa	dua	3
102.	Esperanto	epo	3
103.	Ewe Ebwe Efe Eibe Eue Eve Gbe Krepe Krepi Popo Vhe Eʋegbe	ewe	3
104.	Ewondo Ewundu Jaunde Yaounde Yaunde	ewo	3
105.	Fanagalo Fanakalo Pidgin Zulu Fanekolo Isikula Lololo or Isilololo Piki or Isipiki Silunguboi, Chilapalapa Cikabanga	fng	3
106.	Fon Dahomeen Fongbe	fon	3
107.	Fula(ni), Fulfulde Pulaar Pular' Fulaare	fuv	3
108.	Ganda Luganda	lug	3
109.	Hiligaynon Hiligainon Illogo Ilonggo	hil	3
110.	Iban Dayak	iba	3
111.	IlokoIlokano Ilocano	ilo	3
112.	Kanuri,	kau	3
113.	Kapampangan, Pampangan, Pampango, Pampangueño, Capampangan, Amanung Sisuan	pam	3
114.	Latin, Latina	let	3
115.	Manado Malay, Manadonese, Manadonese Malay, Minahasan Malay	xmm	3
116.	Masbateño, Masbatenyo, Minasbate	msb	3
117.	Mossi, Mole, Moose, More, Moshi, Mossi	mos	3
118.	Nagamese, Bodo, Kachari Bengali, Naga Creole Assamese, Naga-Assamese, Naga Pidgin	nag	3
119.	Nauruan	nau	3
120.	OshiWambo, Cuanhama, Humba, Kuanjama, Kwancama, Kwanjama, Kwanyama, Ochikwanyama, Oshikuanjama, Oshikwanyama, Ovambo, Oxikuanyama, Wambo	kua	3
121.	Pangasinan	pag	3
122.	Pijin, Neo-Solomonic, Solomons Pidgin	pis	3
123.	Quechua, Runasimi , Qhichwa simi	que	3

124.	Raga, Hano, Bwatvenua, Lamalanga, North Raga, Qatvenua, Raga, Vunmarama	lml	3
125.	Roviana, Robiana, Rubiana, Ruviana	rug	3
126.	Shona, Chishona, “Swina” (pej.), Zezuru	sna	3
127.	Sranan, Sranan Tongo, Surinaams, Suriname Creole English, Surinamese, Taki-Taki	srn	3
128.	Tagalog,	tgl	3
129.	Tausūg, Bahasa Sug, Moro Joloano, Sinug, Sulu, Suluk, Tausog, Taw Sug	tsg	3
130.	Torres-Strait Creole, Ap-Ne-Ap, Blaik, Broken, Cape York Creole, Creole, Torres Strait Broken, Torres Strait Pidgin English, West Torres, Yumplatok	tcs	3
131.	Tuvaluan, Ellice, Ellicean, Tuvalu	tvl	3
132.	Umbundu, Kimbari, Mbali, Mbari, M’bundo, Mbundu, Mbundu Benguella, Nano, Olumbali, Ovimbundu, South Mbundu, Umbundo	umb	3
133.	Waray-Waray, Binisaya, Samaran, Samareño, Samarenyo, Samar-Leyte, Waray	war	3
134.	Wolaytta, Borodda, Ometo, Ualamo, Uba, Uollamo, “Walamo” (pej.), Wallamo, Welamo, Wellamo, Wolaita, Wolaitta, Wolataita, Wolayta, Wollamo	wal	3
135.	Zhuang, Nong	zha	3
136.	Adzera, Atzera, Azera, Atsera or Acira,	adz	4
137.	Aklan, Aklan, Aklanon or AkeanonInakeanon (native)	akl	4
138.	Arrernte, Arunta, Eastern Aranda, Upper Aranda	aer	4
139.	Bambara, Bamanankan	bam	4
140.	BashkirBashkir Bashqort Basquort	bak	4
141.	Cape Verdean Creole, Creole, Kriol, “Badiu” (pej.), Caboverdiano, Criol, Crioulo, Kriol, Krioulo, Krioulu, “Sampadjudu” (pej.), Kabuverdianu	kea	4
142.	Central Sinama, “Bajaw” (pej.) Central Sinama Orang Laut Sama Dilaut Samal Siasi Sama Sinama	sml	4
143.	Chavacano, Chabacano Chabakano Zamboangueño	cbk	4

144.	CorsicanCorse Corsi Corso Corsu	cos	4
145.	DagaareDagaare Dagara Dagare Dagari Dagati Degati Dogaari Southern Dagari	dga	4
146.	DagbaniDagbamba Dagbane Dagomba	dag	4
147.	Dinka, Padang White Nile Dinka Agar Central Dinka Bor Cam Dinka Bor Eastern Dinka Rek Western Dinka	din	4
148.	DrehuDehu De'u Lifou Lifu Qene Drehu	dhv	4
149.	FijianBoumaa Fijian Eastern Fijian Fiji Standard Fijian	fij	4
150.	Friulian, Frioulan Frioulian Friulano Furlan Priulian	fur	4
151.	Ga Accra Acra Amina Gain	gaa	4
152.	HixkaryanaChawiyana Faruaru Hichkaryana Hishkariana Hishkaryana Hixkariana Hyxkaryana Kumiyana Parucutu Parukoto-Charuma Sherewyana Sokaka Wabui Xereu Xerewyana	hix	4
153.	Ifugao, Ifugaw, Mayaoyaw, Mayoyao	ifu	4
154.	Ixil	ixl	4
155.	JavaneseDjawa Jawa	jav	4
156.	Kagayanen, Cagayano, Kagay-anen, Kinagayanen	cgc	4
157.	Kaqchikel, Cakchiquel, Kaqchikel, Kaqchiquel	cak	4
158.	Khoekhoe, Bergdamara, "Hottentot" (pej.), Khoekhoegowab, Khoekhoegowap, Maqua, Nama, Namakwa, Naman, Namaqua, Tama, Tamakwa, Tamma	naq	4
159.	Ki'che', Central K'iche', Central Quiché, Chiquel, Qach'abel, Quiché	quc	4
160.	Lozi, Kololo, Kolololo, Rotse, Rozi, Rutse, Silozi, Tozvi	loz	4
161.	Luxembourgish, Frankish, Letzburgisch, Lëtzebuergesch, Luxembourgeois, Luxemburgian, Luxemburgish, Moselle Franconian	ltz	4
162.	Mam, Huehuetenango Mam	mam	4
163.	Maranao, Maranaw, Ranao	mrw	4
164.	Mbula, Kaimanga, Mangaaba, Mangaava, Mangaawa, Mangap, Mangap-Mbula	mna	4
165.	Mizo, Duhlian Twang, Dulien, Hualngo, Lukhai, Lusago, Lusai, Lusei, Lushai, Lushai-Mizo, Lushei, Sailau, Whelngo	lus	4

166.	Nuer, Naadh, Naath	nus	4
167.	Nuosu (Yi), Black Yi, Liangshan Yi, Northern Yi, Nosu Yi, Sichuan Yi	iii	4
168.	Pitjantjatjara, Pitjantjara	pjt	4
169.	Q'eqchi', Cacche’, Kekchi’, Kekchí, Ketchi’, Quecchi’	kek	4
170.	Romansh, Rhaeto-Romance, Rheto-Romance, Romanche, Romansh, Rumantsch	roh	4
171.	Scottish Gaelic, Gaelic-Scotish	gla	4
172.	Shavante, Xavante, Akuên, Akwen, A’uwe Uptabi, A’we, Chavante, Crisca, Pusciti, Shavante, Tapacua	xav	4
173.	Sorbian, Haut Sorabe, Hornjoserbski, Hornoserbski, Obersorbisch, Upper Lusatian, Wendish	hsb	4
174.	Susu, Sose, Soso, Soussou, Susoo	sus	4
175.	Tagabawà, Tagabawa Bagobo, Tagabawa Manobo	bgs	4
176.	Talysh, Talesh, Talish, Talyshi	tly	4
177.	Tumbuka, Chitumbuka, Citumbuka, Tamboka, Tambuka, Timbuka, Tombucas, Tumboka	tum	4
178.	Tuvan, Tuva, Diba, Kök Mungak, Soyod, Soyon, Soyot, Tannu-Tuva, Tofa, Tokha, Tuba, Tuvan, Tuvia, Tuvin, Tuvinian, Tyva, Uriankhai, Uriankhai-Monchak, Uryankhai	tyv	4
179.	Wolof, Ouolof, Volof, Walaf, Waro-Waro, Yallof	wol	4
180.	Zarma, Adzerma, Djerma, Dyabarma, Dyarma, Dyerma, Zabarma, Zarbarma, Zarmaci	dje	4
181.	Zazaki, Northern, Alevica, Dersimki, Dimilki, Kirmanjki, Northern Zaza, So-Bê, Zaza, Zonê Ma	kiu	4
182.	Acehnese, Achehnese AchineseAceh	ace	5
183.	Acholi, Acoli Acooli Akoli Atscholi Dok Acoli Gang Lëbacoli Log Acoli Lwo Lwoo Shuli	ach	5
184.	Afaan Oromooromo Oromiffa “Galla” (pej.) “Galligna” (pej.) “Gallinya” (pej.) Southern Oromo	orm	5
185.	Afar, Adal, ’Afar Af, Afaraf, “Danakil” (pej.), “Denkel” (pej.), Qafar	aar	5
186.	Alsatian, Elsässerdeutsche Alsacien Alemanic Alemannisch Schwyzerdütsch	gsw	5

187.	Alur, Aloro, Alua, Alulu, Dho Alur, Jo Alur, Lur, Luri	alz	5
188.	Bavarian, Bairisch Bavarian Austrian Bayerisch Ost-Oberdeutsch	bar	5
189.	Brahui, Birahui Brahuidi Brahuigi Kur Galli	brh	5
190.	Dholuo Kavirondo Luo Luo Nilotic Kavirondo	luo	5
191.	JamaicanBongo Talk Jamiekan Limon Creole English Patois Patwa Quashie Talk Western Caribbean Creole	jam	5
192.	Kabyle, Amazigh, Kabyl, Kabylia, Tamazight, Taqbaylit	kbp	5
193.	Kikuyu, Gĩkũyũ, Gekoyo, Gigikuyu,	kik	5
194.	Low Saxon, Low German, Nedderdütsch, Neddersassisch, Nedersaksisch, Niederdeutsch, Niedersaechsisch, Plattdeutsch, Plattdüütsch	nds	5
195.	Maasai, Maa, Masai	mas	5
196.	Madurese, Madura, Basa Mathura	mad	5
197.	Makhuwa, Central Makhuwa, Emakhuwa, Emakua, Macua, Makhuwa-Makhuwana, Makhuwwa of Nampula, Makoane, Makua, Maquoua	vmw	5
198.	Mandinka, Mande, Manding, Mandingo, Mandingue, Mandinque, Socé	mnk	5
199.	Minangkabau, Minang, Padang	min	5
200.	Mundari, Colh, Horo, Mandari, Mondari, Munari	unr	5
201.	Neapolitan, Napoletano, Neapolitan-Calabrese	nap	5
202.	Piedmontese, Piemontese, Piemontèis	pms	5
203.	Romany,	rom	5
204.	Sasak, Lombok	sas	5
205.	Sicilian, Calabro-Sicilian, Sicilianu, Siculu	scn	5
206.	Soga, Lusoga, Olusoga	xog	5
207.	Soninke, Aswanek, Aswanik, Azer, Ceddo, Cheddo, Gangara, Genger, Kwara, Maraka, Marka, Markaajo, Markakan, Sarakole, Sarakolle, Sarakule, Sarakulle, Sarangkole, Sarangkolle, Saraxuli, Sebbe, Serahule, Serecole, Soninkanxanne, Sooninke, Wakkore, Wankara	snk	5

208.	Tswa, Kitshwa, Sheetshwa, Shitshwa, Tshwa, Xitshwa, Xitswa	tsc	5
209.	Venetian, Talian, Venet	vec	5
210.	Zazaki, Southern, Dimili, Dimli, Southern Zaza, Zaza, Zazaca	diq	5

Appendix C: Repertoire Table Grouped by Glyph

Table C.1. Repertoire Table Grouped by Glyph

#	Unicode	Glyph	Unicode name	Languages using the code point (EGIDS)	Reference supporting inclusion (URL etc.)
1.	0061	a	LATIN SMALL LETTER A	Basic Latin	[0]
2.	0061 + 0331	a̲	LATIN SMALL LETTER A + COMBINING MACRON BELOW	Nuer (4)	[146], [129]
3.	00E0	à	LATIN SMALL LETTER A WITH GRAVE	Italian (1) Galician (2) Wolof (4)	[130], [131], [106], [132]
4.	00E1	á	LATIN SMALL LETTER A WITH ACUTE	Spanish (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Chuukese (2) Galician (2) Lule Sámi (2) Northern Sámi (2)	[100], [101], [102], [103], [104], [105], [106], [107], [108]
5.	00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	Vietnamese (1) Romanian (1) Skolt Sami (2) Kirundi (1) French (1) Galician (2) West Frisian (2) Friulian (4) Xavante (4)	[109], [110], [113], [104], [114], [106], [115], [116], [117]
6.	00E3	ã	LATIN SMALL LETTER A WITH TILDE	Umbundu (3) Guarani (1) Nauruan (3) Khoekhoe (4)	[141], [142], [143], [144], [145]

7.	00E4	ä	LATIN SMALL LETTER A WITH DIAERESIS	German (1) Finnish (1) Turkmen (1) Estonian (1) Swedish (1) Lule Sámi (2) Yapese (2) Dinka (4) Kaqchikel (4) Bashkir (4) Alsatian (5) Nuer (4)	[119], [120], [121], [122], [123], [107], [124], [125], [126], [127], [128], [129]
8.	00E5	å	LATIN SMALL LETTER A WITH RING ABOVE	Danish (1) Finnish (1) Chamorro (1) Swedish (1) Lule Sámi (2)	[139], [120], [140], [123], [107]
9.	00E6	æ	LATIN SMALL LETTER AE	Danish (1) Icelandic (1) Faroese (2)	[139], [102], [103]
10.	0101	ā	LATIN SMALL LETTER A WITH MACRON	Latvian (1) Tongan (1) Hawaiian (2) Marshallese(1)	[133], [134], [135], [136]
11.	0103	ă	LATIN SMALL LETTER A WITH BREVE	Vietnamese (1) Romanian (1)	[109], [110]
12.	0105	ą	LATIN SMALL LETTER A WITH OGONEK	Polish (1) Lithuanian (1)	[137], [138]
13.	01CE	ǎ	LATIN SMALL LETTER A WITH CARON	Kirundi (1)	[104] https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx Jean Paul Nkurunziza (personal communication)
14.	1EA1	ạ	LATIN SMALL LETTER A WITH DOT BELOW	Vietnamese (1)	[109]

15.	1EA3	ả	LATIN SMALL LETTER A WITH HOOK ABOVE	Vietnamese (1)	[109]
16.	1EA5	ấ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]
17.	1EA7	ầ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE	Vietnamese (1)	[109]
18.	1EA9	ẩ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
19.	1EAB	ẫ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]
20.	1EAD	ậ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]
21.	1EAF	ắ	LATIN SMALL LETTER A WITH BREVE AND ACUTE	Vietnamese (1)	[109]
22.	1EB1	ằ	LATIN SMALL LETTER A WITH BREVE AND GRAVE	Vietnamese (1)	[109]
23.	1EB3	ẳ	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	Vietnamese (1)	[109]
24.	1EB5	ẵ	LATIN SMALL LETTER A WITH BREVE AND TILDE	Vietnamese (1)	[109]
25.	1EB7	ặ	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	Vietnamese (1)	[109]
26.	0062	b	LATIN SMALL LETTER B	Basic Latin	[0]
27.	0253	ɓ	LATIN SMALL LETTER B WITH HOOK	Hausa (2) Dagaare - Burkina Faso (4) Pulaar, (3)	[147], [148], [250]
28.	0063	c	LATIN SMALL LETTER C	Basic Latin	[0]

29.	00E7	ç	LATIN SMALL LETTER C WITH CEDILLA	Turkish (1) Turkmen (1) Kurdish (2) French (1) Azerbaijani (1) Basque (1) Galician (2) Friulian (4) Bashkir (4)	[157], [121], [158], [114], [159], [160], [161], [106], [116], [127]
30.	0107	ć	LATIN SMALL LETTER C WITH ACUTE	Croatian (1) Serbian (1) Polish (1)	[150], [151], [152]
31.	0109	ĉ	LATIN SMALL LETTER C WITH CIRCUMFLEX	Esperanto (3)	[255]
32.	010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE	Maltese (1)	[163]
33.	010D	č	LATIN SMALL LETTER C WITH CARON	Croatian (1) Serbian (1) Latvian (1) Slovak (1) Northern Sámi (2) Lithuanian (1)	[150], [151], [133], [153], [108], [154]
34.	0064	d	LATIN SMALL LETTER D	Basic Latin	[0]
35.	00F0	ð	LATIN SMALL LETTER ETH	Faroese (2) Icelandic (1)	[103], [102]
36.	010F	ď	LATIN SMALL LETTER D WITH CARON	Czech (1) Slovak (1)	[101], [153]
37.	0111	đ	LATIN SMALL LETTER D WITH STROKE	Croatian (1) Serbian (1) Vietnamese (1) Northern Sámi Brahui (5)	[150], [151], [109], [108], [168]
38.	0256	ɖ	LATIN SMALL LETTER D WITH TAIL	Fon (3) Ewe (3)	[169], [170]
39.	0257	ɗ	LATIN SMALL LETTER D WITH HOOK	Hausa (2) Pulaar (3)	[147], [166], [250]
40.	1E13	ḓ	LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]

41.	0065	e	LATIN SMALL LETTER E	Basic Latin	[0]
42.	0065 + 0331	e̱	LATIN SMALL LETTER E + COMBINING MACRON BELOW	Nuer (4)	[146]
43.	00E8	è	LATIN SMALL LETTER E WITH GRAVE	French (1) Italian (1) Afrikaans (1) Kirundi (1) Haitian Creole (1) French (1)	[114], [130], [175], [104], [182], [183]
44.	00E9	é	LATIN SMALL LETTER E WITH ACUTE	French (1) Italian (1) Spanish (1) Czech (1) Icelandic (1) Kirundi (1) Chuukese (2) Galician (2) Wolof (4) XAVANTE (4) West Frisian (2)	[114], [130], [100], [101], [102], [104], [105], [106], [132], [117], [115]
45.	00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX	French (1) Tswana (1) Afrikaans (1) Vietnamese (1) Kurdish (2) Kirundi (1) West Frisian (2) Friulian (4)	[114], [173], [174], [175], [109], [158], [104], [115], [116]
46.	00EB	ë	LATIN SMALL LETTER E WITH DIAERESIS	Afrikaans (1) Kirundi (1) Albanian (1) French (1) Chuukese (2) Uyghur (2) Yapese (2) Wolof (4) Drehu (4) Kaqchikel (4) West Frisian (2) Nuer (4)	[175], [104], [176], [177], [114], [176], [177], [114], [178], [179], [124], [132], [180], [126], [115], [129]

47.	0113	ē	LATIN SMALL LETTER E WITH MACRON	Latvian (1) Hawaiian (2) Tongan (1) Minangkabau (5)	[133], [135], [134], [184]
48.	0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE	Lithuanian (1)	[138], [154]
49.	0119	ę	LATIN SMALL LETTER E WITH OGONEK	Polish (1) Palauan (2) Lithuanian (1)	[152], [185], [138], [154]
50.	011B	ě	LATIN SMALL LETTER E WITH CARON	Czech (1) Kirundi (1) Sorbian (4)	[101], [104], [172]
51.	01DD	ə	LATIN SMALL LETTER TURNED E	Kanuri (3)	[240]
52.	0259	ə	LATIN SMALL LETTER SCHWA	Azeri, Azerbaijani (1) Ewondo (3) Ewe (3) Bugis (3)	[159], [190], [170], [241]
53.	025B	ɛ	LATIN SMALL LETTER OPEN E	Dagaare - Burkina Faso (4) Lingala (2) Akan (3) Ewondo (3) Dagbani (Dagomba) (4) Fon (3) Mossi (3) Ga (4) Ewe (3) Duala (3) Bambara (4) Nuer (4)	[148], [236], [237], [190], [189], [169], [212], [238], [193], [170], [194], [199], [129]
54.	025B + 0308	ɛ̈	LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS	Nuer (4) Dinka (4)	[129], [146], [239], [125]
55.	025B + 0331	ɛ̱	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW	Nuer (4)	[129], [146], [239]
56.	025B + 0331 + 0308	ɛ̱̈	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS	Nuer (4)	[146], [239]

57.	1EB9	ẹ	LATIN SMALL LETTER E WITH DOT BELOW	Yoruba (2)	[181]
58.	1EB9 + 0300	ệ	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT	Yoruba (2)	[254]
59.	1EB9 + 0301	ệ	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT	Yoruba (2)	[254]
60.	1EBB	ẻ	LATIN SMALL LETTER E WITH HOOK ABOVE	Vietnamese (1)	[109]
61.	1EBD	ẽ	LATIN SMALL LETTER E WITH TILDE	Umbundu (3) Guarani (1) Cubeo (3) Xavante (4)	[141], [142], [143], [186], [187], [117]
62.	1EBF	ế	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]
63.	1EC1	ề	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	Vietnamese (1)	[109]
64.	1EC3	ể	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
65.	1EC5	ễ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]
66.	1EC7	ệ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]
67.	0066	f	LATIN SMALL LETTER F	Basic Latin	[0]
68.	0192	ƒ	LATIN SMALL LETTER F WITH HOOK	Ewe (3)	[170]
69.	0067	g	LATIN SMALL LETTER G	Basic Latin	[0]
70.	0067 + 0303	g̃	LATIN SMALL LETTER G + COMBINING TILDE	Guarani (1)	[142], [143]

71.	0067 + 0304	ḡ	LATIN SMALL LETTER G + COMBINING MACRON	Raga (Hano) (3)	[200]
72.	011D	ĝ	LATIN SMALL LETTER G WITH CIRCUMFLEX	Esperanto (3)	[255]
73.	011F	ğ	LATIN SMALL LETTER G WITH BREVE	Turkish (1) Tatar (2) Azeri (1) Bashkir (4) Zaza (5)	[157], [201], [159], [127], [202]
74.	0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE	Maltese (1)	[163]
75.	0123	ģ	LATIN SMALL LETTER G WITH CEDILLA	Latvian (1) Brahui (5)	[133], [168]
76.	01E7	ǧ	LATIN SMALL LETTER G WITH CARON	Skolt Sami (2)	[113]
77.	0263	ɣ	LATIN SMALL LETTER GAMMA	Dagbani (Dagomba) (4) Nuer (4) Dinka (4) Ewe (3) Nuer (4)	[189], [146], [125], [170], [129]
78.	0068	h	LATIN SMALL LETTER H	Basic Latin	[0]
79.	0125	ĥ	LATIN SMALL LETTER H WITH CIRCUMFLEX	Esperanto (3)	[255]
80.	0127	ħ	LATIN SMALL LETTER H WITH STROKE	Maltese (1)	[163]
81.	0069	i	LATIN SMALL LETTER I	Basic Latin	[0]
82.	0069 + 0331	i̲	LATIN SMALL LETTER I + COMBINING MACRON BELOW	Nuer (4)	[146]
83.	00EC	ì	LATIN SMALL LETTER I WITH GRAVE	Italian (1) Kirundi (1)	[130], [206], [208]
84.	00ED	í	LATIN SMALL LETTER I WITH ACUTE	Spanish (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Galician (2)	[100], [101], [102], [103], [104], [106], [127]

				Bashkir(4)	
85.	00EE	î	LATIN SMALL LETTER I WITH CIRCUMFLEX	Afrikaans (1) Romanian (1) Kurdish (2) Kirundi (1) French (1) Friulian (4)	[175], [110], [158], [104], [114], [116]
86.	00EF	ï	LATIN SMALL LETTER I WITH DIAERESIS	Afrikaans (1) French (1) Kaqchikel (4) Dinka (4) West Frisian (2)	[175], [114], [126], [125], [115]
87.	0129	ĩ	LATIN SMALL LETTER I WITH TILDE	Guarani (1) Cubeo (3) Khoekhoe (4) Kikuyu (5)	[142], [143], [186], [145], [209]
88.	012B	ī	LATIN SMALL LETTER I WITH MACRON	Latvian (1) Lithuanian (1) Hawaiian (2) Tongan (1)	[133], [138], [135], [134]
89.	012F	į	LATIN SMALL LETTER I WITH OGONEK	Lithuanian (1)	[154]
90.	0131	ı	LATIN SMALL LETTER I DOTLESS	Turkish (1) Tatar (2) Azeri (1)	[157], [203], [201], [159]
91.	0135	ĵ	LATIN SMALL LETTER J WITH CEDILLA	Esperanto (3)	[255]
92.	01D0	ǐ	LATIN SMALL LETTER I WITH CARON	Kirundi (1)	[104]
93.	0268	ɨ	LATIN SMALL LETTER I WITH STROKE	Cubeo (3) Dagbani (Dagomba) (4) HIxkaryána (4) Maasai (5)	[186], [189], [210], [211]
94.	0268 + 0303	ɨ̃	LATIN SMALL LETTER I WITH STROKE + COMBINING TILDE	Cubeo (3)	[186]
95.	1EC9	ỉ	LATIN SMALL LETTER I WITH HOOK ABOVE	Vietnamese (1)	[109]

96.	1ECB	ị	LATIN SMALL LETTER I WITH DOT BELOW	Igbo (2)	[205]
97.	006A	j	LATIN SMALL LETTER J	Basic Latin	[0]
98.	0269	ɩ	LATIN SMALL LETTER IOTA	Dagaare - Burkina Faso (4) Mossi (3)	[148], [212]
99.	006B	k	LATIN SMALL LETTER K	Basic Latin	[0]
100.	0137	ķ	LATIN SMALL LETTER K WITH CEDILLA	Latvian (1)	[133]
101.	0199	ƙ	LATIN SMALL LETTER K WITH HOOK	Hausa (2)	[147]
102.	01E9	ǩ	LATIN SMALL LETTER K WITH CARON	Skolt Sami (2)	[113]
103.	006C	l	LATIN SMALL LETTER L	Basic Latin	[0]
104.	013A	ĺ	LATIN SMALL LETTER L WITH ACUTE	Slovak (1)	[153]
105.	013C	ļ	LATIN SMALL LETTER L WITH CEDILLA	Latvian (1) Marshallese (1) Brahui (5)	[133], [213], [214], [168]
106.	013E	ľ	LATIN SMALL LETTER L WITH CARON	Slovak (1)	[153]
107.	0142	ł	LATIN SMALL LETTER L WITH STROKE	Polish (1)	[152]
108.	1E37	ḷ	LATIN SMALL LETTER L WITH DOT BELOW	Marshallese (1)	[213], [214], [215], [216]
109.	1E3D	ḽ	LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
110.	006D	m	LATIN SMALL LETTER M	Basic Latin	[0]
111.	006D + 0327	m̧	LATIN SMALL LETTER M + COMBINING CEDILLA	Marshallese (1)	[213], [136], [214]
112.	1E43	ṃ	LATIN SMALL LETTER M WITH DOT BELOW	Marshallese (1)	[213], [136], [215], [216]
113.	006E	n	LATIN SMALL LETTER N	Basic Latin	[0]

114.	006E + 0304	n̄	LATIN SMALL LETTER N + COMBINING MACRON	Raga (Hano) (3) Marshallese (1)	[200], [213], [136]
115.	006E + 0308	n̈	LATIN SMALL LETTER N + COMBINING DIAERESIS	Malagasy(1)	[230]
116.	00F1	ñ	LATIN SMALL LETTER N WITH TILDE	Spanish (1) Pulaar (3) Chamorro (1) Filipino (1) Guarani (1) Chavacano (4) Basque (1) Galician (2) Iloco (3) Quechua (3) Cape Verdean Creole (4) Waray-Waray (3) Wolof (4) Nauruan (3) Lozi (4) Bashkir (4) Marshallese (1) Mandinka (5) Igbo(2)	[221], [250] [222], [142], [143], [223], [160], [106], [224], [225], [226], [227], [228], [132], [144], [229], [127], [136], [197], [205]
117.	0144	ń	LATIN SMALL LETTER N WITH ACUTE	Polish (1) Lule Sámi (2) Sorbian (4) Brahui (5)	[152], [107], [172], [168]
118.	0146	ņ	LATIN SMALL LETTER N WITH CEDILLA	Latvian (1) Marshallese (1)	[133], [136]
119.	0148	ň	LATIN SMALL LETTER N WITH CARON	Turkmen (1) Czech (1) Slovak (1)	[121], [101], [153]
120.	014B	ŋ	LATIN SMALL LETTER ENG	Inari Sami (2) Dagaare - Burkina Faso (4) Dagbani (Dagomba) (4) Northern Sami (2) Ewondo (3) Luganda (3) Wolof (4) Adzera (4) Nuer (4)	[188], [148], [189], [108], [190], [191], [132], [192], [146], [193], [125], [194], [170], [195], [196], [197], [198], [199], [129]

				Ga (4) Dinka (4) Duala (3) Ewe (3) Soga (5) Alur (5) Mandinka (5) Acholi (5) Bambara (4) Nuer (4)	
121.	0272	ɲ	LATIN SMALL LETTER N WITH LEFT HOOK	Susu (4) Zarma (4) Bambara (4)	[218], [219], [199]
122.	1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE	Venda (1)	[164], [257]
123.	1E47	ṇ	LATIN SMALL LETTER N WITH DOT BELOW	Marshallese (1)	[136], [215], [216]
124.	1E49	n̲	LATIN SMALL LETTER N WITH LINE BELOW	Pitjantjatjara (4)	[220]
125.	1E4B	ṉ	LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
126.	006F	o	LATIN SMALL LETTER O	Basic Latin	[0]
127.	006F + 0327	o̧	LATIN SMALL LETTER O + COMBINING CEDILLA	Marshallese (1)	[136]
128.	006F + 0331	o̲	LATIN SMALL LETTER O + COMBINING MACRON BELOW	Nuer (4)	[146], [129]
129.	00F2	ò	LATIN SMALL LETTER O WITH GRAVE	Italian (1) Haitian Creole (1)	[130], [182], [183]
130.	00F3	ó	LATIN SMALL LETTER O WITH ACUTE	Spanish (1) Polish (1) Czech (1) Icelandic (1) Kirundi (1) Chuukese (2) Galician (2) Wolof (4)	[100], [152], [101], [102], [104], [105], [106], [132]

131.	00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX	Tswana (1) Afrikaans (1) Vietnamese (1) Kirundi (1) French (1) Northern Sotho (1) West Frisian (2) Galician (2) Friulian (4) Xavante(4)	[173], [174], [175], [109], [104], [114], [230], [115], [106], [116], [117]
132.	00F5	õ	LATIN SMALL LETTER O WITH TILDE	Estonian (1) Skolt Sami (2) Umbundu (3) Guarani (1) Nauruan (3) Xavante (4) Khoekhoe (4)	[122], [113], [141], [142], [143], [144], [117], [235]
133.	00F6	ö	LATIN SMALL LETTER O WITH DIAERESIS	German (1) Finnish (1) Afrikaans (1) Turkish (1) Swedish (1) Uygur (2) Yapese (2) Drehu (4) Kaqchikel (4) Dinka (4) Bashkir (4) Low German (5) Chechen (2) 1992 Version West Frisian (2) Nuer (4)	[119], [120], [175], [157], [123], [179], [124], [180], [126], [125], [127], [231], [232], [115], [129]
134.	00F8	ø	LATIN SMALL LETTER O WITH STROKE	Danish (1) Faroese (2)	[139], [103]
135.	014D	ō	LATIN SMALL LETTER O WITH MACRON	Hawaiian (2) Marshallese (1) Tongan (1)	[135], [136], [134]
136.	0151	ő	LATIN SMALL LETTER O WITH DOUBLE ACUTE	Hungarian (1)	[233], [234]
137.	0153	œ	LATIN SMALL LIGATURE OE	French (1)	[114], [253]

138.	01A1	ơ	LATIN SMALL LETTER O WITH HORN	Vietnamese (1)	[109]
139.	01D2	ǒ	LATIN SMALL LETTER O WITH CARON	Kirundi (1)	[104]
140.	0254	ɔ	LATIN SMALL LETTER OPEN O	Dagaare - Burkina Faso (4) Dagbani (Dagomba) (4) Lingala (2) Akan (3) Ewondo (3) Fon (3) Nuer (4) Ga (4) Duala (3) Ewe (3) Nuer (4)	[148], [189], [236], [237], [190], [169], [146], [193], [194], [170], [129]
141.	0254 + 0308	ɔ̈	LATIN SMALL LETTER OPEN O + COMBINING DIAERESIS	Dinka (4)	[125]
142.	0254 + 0331	ɔ̱	LATIN SMALL LETTER OPEN O + COMBINING MACRON BELOW	Nuer (4)	[129], [146]
143.	1ECD	ọ	LATIN SMALL LETTER O WITH DOT BELOW	Igbo (2) Yoruba (2) Marshallese (1)	[204], [205], [181], [136], [215], [216]
144.	1ECD + 0300	ọ̀	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT	Yoruba (2)	[254]
145.	1ECD + 0301	ọ́	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT	Yoruba (2)	[254]
146.	1ECF	ỏ	LATIN SMALL LETTER O WITH HOOK ABOVE	Vietnamese (1)	[109]
147.	1ED1	ố	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	Vietnamese (1)	[109]

148.	1ED3	ồ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	Vietnamese (1)	[109]
149.	1ED5	ổ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	Vietnamese (1)	[109]
150.	1ED7	ỗ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	Vietnamese (1)	[109]
151.	1ED9	ộ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	Vietnamese (1)	[109]
152.	1EDB	ớ	LATIN SMALL LETTER O WITH HORN AND ACUTE	Vietnamese (1)	[109]
153.	1EDD	ờ	LATIN SMALL LETTER O WITH HORN AND GRAVE	Vietnamese (1)	[109]
154.	1EDF	ở	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE	Vietnamese (1)	[109]
155.	1EE1	ỡ	LATIN SMALL LETTER O WITH HORN AND TILDE	Vietnamese (1)	[109]
156.	1EE3	ợ	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	Vietnamese (1)	[109]
157.	0070	p	LATIN SMALL LETTER P	Basic Latin	[0]
158.	00FE	þ	LATIN SMALL LETTER THORN	Icelandic (1)	[102]
159.	0071	q	LATIN SMALL LETTER Q	Basic Latin	[0]
160.	0072	r	LATIN SMALL LETTER R	Basic Latin	[0]
161.	0072 + 0303	r̃	LATIN SMALL LETTER R + COMBINING TILDE	Hausa (2)	[147]
162.	0155	ŕ	LATIN SMALL LETTER R WITH ACUTE	Slovak (1) Brahui (5)	[153], [168]
163.	0159	ř	LATIN SMALL LETTER R WITH CARON	Czech (1) Sorbian (4)	[101], [172]

164.	024D	ɼ	LATIN SMALL LETTER R WITH STROKE	Kanuri (3)	[240]
165.	0073	s	LATIN SMALL LETTER S	Basic Latin	[0]
166.	00DF	ß	LATIN SMALL LETTER SHARP S	German (1)	[119]
167.	015B	ś	LATIN SMALL LETTER S WITH ACUTE	Polish (1) Montenegrin (1)	[152], [258]
168.	015D	ŝ	LATIN SMALL LETTER S WITH CIRCUMFLEX	Esperanto (3)	[255]
169.	015F	ş	LATIN SMALL LETTER S WITH CEDILLA	Turkish (1) Turkmen (1) Kurdish (2) Tatar (2) Azeri (1) Bashkir (4) Brahui (5) Zaza (5)	[157], [121], [158], [201], [159], [127], [168], [202]
170.	0161	š	LATIN SMALL LETTER S WITH CARON	Tswana (1) Croatian (1) Serbian (1) Latvian (1) Northern Sotho (1) Northern Sami (2) Lithuanian (1)	[174], [150], [151], [133], [230], [108], [154]
171.	0219	ș	LATIN SMALL LETTER S WITH COMMA BELOW	Romanian (1)	[110]
172.	1E63	ṣ	LATIN SMALL LETTER S WITH DOT BELOW	Yoruba (2)	[181]
173.	0074	t	LATIN SMALL LETTER T	Basic Latin	[0]
174.	0165	ť	LATIN SMALL LETTER T WITH CARON	Czech (1) Slovak (1)	[101], [153]
175.	0167	ŧ	LATIN SMALL LETTER T WITH STROKE	Northern Sami (2) Brahui (5)	[108], [168]
176.	021B	ț	LATIN SMALL LETTER T WITH COMMA BELOW	Romanian (1)	[110]
177.	1E6D	ṭ	LATIN SMALL LETTER T WITH DOT BELOW	Mizo (4)	[242]

178.	1E71	ṯ	LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW	Venda (1)	[164], [257]
179.	0075	u	LATIN SMALL LETTER U	Basic Latin	[0]
180.	00F9	ù	LATIN SMALL LETTER U WITH GRAVE	Italian (1) Papiamento (1)	[130], [206], [245], [246]
181.	00FA	ú	LATIN SMALL LETTER U WITH ACUTE	Spanish (1) Czech (1) Icelandic (1) Faroese (2) Kirundi (1) Chuukese (2) West Frisian (2) Galician (2)	[100], [101], [102], [103], [104], [105], [115], [106]
182.	00FB	û	LATIN SMALL LETTER U WITH CIRCUMFLEX	Afrikaans (1) Kurdish (2) Kirundi (1) French (1) Miskito (2) West Frisian (2) Friulian (4) Zazaki (4)	[175], [158], [104], [114], [243], [115], [116], [244]
183.	00FC	ü	LATIN SMALL LETTER U WITH DIAERESIS	German (1) Spanish (1) Afrikaans (1) Turkish (1) Swedish (1) French (1) Azeri (1) Basque (1) Galician (2) Uygur (2) Kaqchikel (4) Bashkir (4)	[119], [100], [175], [157], [123], [114], [159], [161], [106], [179], [126], [127], [231]
184.	0169	ũ	LATIN SMALL LETTER U WITH TILDE	Umbundu (3) Guarani (1) Nauruan (3) Khoekhoe (4) Kikuyu (5)	[141], [142], [143], [144], [145], [209]
185.	016B	ū	LATIN SMALL LETTER U WITH MACRON	Latvian (1) Hawaiian (2) Lithuanian (1)	[133], [135], [138], [154], [136], [134]

				Marshallese (1) Tongan (1)	
186.	016D	ŭ	LATIN SMALL LETTER U WITH BREVE	Esperanto (3)	[255]
187.	016F	ů	LATIN SMALL LETTER U WITH RING ABOVE	Czech (1)	[101]
188.	0171	ű	LATIN SMALL LETTER U WITH DOUBLE ACUTE	Hungarian (1)	[233], [234]
189.	0173	ų	LATIN SMALL LETTER U WITH OGONEK	Lithuanian (1)	[154], [138]
190.	01B0	ư	LATIN SMALL LETTER U WITH HORN	Vietnamese (1)	[109]
191.	01D4	ǔ	LATIN SMALL LETTER U WITH CARON	Kirundi (1)	[104]
192.	0289	ʉ	LATIN SMALL LETTER U BAR	Cubeo (3) Maasai (5)	[186], [187], [211]
193.	0289 + 0303	ʉ̃	LATIN SMALL LETTER U BAR + COMBINING TILDE	Cubeo (3)	[186], [187]
194.	1EE5	ụ	LATIN SMALL LETTER U WITH DOT BELOW	Vietnamese (1) Igbo (2)	[109],[204], [205]
195.	1EE7	ủ	LATIN SMALL LETTER U WITH HOOK ABOVE	Vietnamese (1)	[109]
196.	1EE9	ứ	LATIN SMALL LETTER U WITH HORN AND ACUTE	Vietnamese (1)	[109]
197.	1EEB	ừ	LATIN SMALL LETTER U WITH HORN AND GRAVE	Vietnamese (1)	[109]
198.	1EED	ử	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	Vietnamese (1)	[109]
199.	1EEF	ữ	LATIN SMALL LETTER U WITH HORN AND TILDE	Vietnamese (1)	[109]
200.	1EF1	ự	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	Vietnamese (1)	[109]
201.	0076	v	LATIN SMALL LETTER V	Basic Latin	[0]

202.	028B	ʊ	LATIN SMALL LETTER V WITH HOOK	Dagaare - Burkina Faso (4) Mossi (3) Ewe (3)	[148], [212], [238], [170]
203.	0077	w	LATIN SMALL LETTER W	Basic Latin	[0]
204.	0175	ŵ	LATIN SMALL LETTER W WITH CIRCUMFLEX	Chichewa (3)	[247]
205.	0078	x	LATIN SMALL LETTER X	Basic Latin	[0]
206.	1E8D	ẍ	LATIN SMALL LETTER X WITH DIAERESIS	Mam (4)	[248], [249]
207.	0079	y	LATIN SMALL LETTER Y	Basic Latin	[0]
208.	00FD	ý	LATIN SMALL LETTER Y WITH ACUTE	Turkmen (1) Czech (1) Icelandic (1) Faroese (2) Guarani (1)	[121], [101], [102], [103], [142], [143]
209.	00FF	ÿ	LATIN SMALL LETTER Y WITH DIAERESIS	French (1)	[114], [253], [257]
210.	0177	ŷ	LATIN SMALL LETTER Y WITH CIRCUMFLEX	Welsh (2)	[256]
211.	01B4	ƴ	LATIN SMALL LETTER Y WITH HOOK	Dagaare - Burkina Faso (4)	[148], [251], [149]
212.	1EF3	ỳ	LATIN SMALL LETTER Y WITH GRAVE	Vietnamese (1)	[109]
213.	1EF5	ỵ	LATIN SMALL LETTER Y WITH DOT BELOW	Vietnamese (1)	[109]
214.	1EF7	ỷ	LATIN SMALL LETTER Y WITH HOOK ABOVE	Vietnamese (1)	[109]
215.	1EF9	ỹ	LATIN SMALL LETTER Y WITH TILDE	Vietnamese (1) Guarani (1)	[109] [142]
216.	007A	z	LATIN SMALL LETTER Z	Basic Latin	[0]
217.	017A	ź	LATIN SMALL LETTER Z WITH ACUTE	Polish (1) Brahui (5) Sorbian (4) Montenegrin(1)	[152], [252], [168], [172], [258]

218.	017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE	Polish (1) Maltese (1)	[152], [163]
219.	017E	ž	LATIN SMALL LETTER Z WITH CARON	Lithuanian (1) Croatian (1) Serbian (1) Turkmen (1) Latvian (1) Slovak (1) Northern Sami (2) Chechen (2) 1925 Version	[154], [150], [151], [121], [133], [153], [108], [232]
220.	01EF	ǯ	LATIN SMALL LETTER EZH WITH CARON	Skolt Sami (2)	[113]
221.	0292	ʒ	LATIN SMALL LETTER EZH	Skolt Sami (2) Dagbani (Dagomba) (4)	[113], [189]

Appendix D: Variants Analysis

Below all shortlisted variant candidates are presented. Effectively these tables are a superset of all variant candidates summarized above in section 6.5. Below these are given in different categories based on the main criteria used for comparison following the principles for variant analysis established above in section 6.1. These categories however served only as initial motivation for consideration as variant candidates, and in several cases further variant candidates evolved out of the original set of candidates or the rationale for analysis was changed based on the data gathered (the final rationale for inclusion in the variant sets is given above for each pair in section 6.5).
As an aid to the reader, the lines have been color coded, where by yellow indicates that a potential variant pair was identified, and green indicates that a potential variant pair was confirmed.

D.1 Shaping of Base Characters

D.1.1 Latin Small Letter F vs. Latin Small Letter F with Hook

Code Points Considered:

Code Points	Glyph	Name
0066	f	Latin Small Letter F
0192	ƒ	Latin Small Letter F With Hook

Example from Swedish Newspaper:


DAGENS NYHETER. Nyheter Ekonomi Kultur Sthlm Sport ♥ Mitt DN

● Så var de första larmsamtalen till 112.

● TV Minidokumentär: Här är minuterna som förändrade landet.

Analys

Stefan Lisinski: Allt talar för livstids fängelse

Stockholm terror suspect formally charged

Findings:
Swedish uses a shape of “LATIN SMALL LETTER F” (0066) that is identical to “LATIN SMALL LETTER F WITH HOOK” (0192) in italic style. Example from a large, daily newspaper, in which all instances of “ƒ” are just variants of “f”.

Conclusions:
These two Code Points should be treated as variants

D.1.2 Latin Small Letter A vs. Latin Small Letter Alpha

Code Points Considered:

Code Points	Glyph	Name
0061	a	Latin Small Letter A
0251	ɑ	Latin Small Letter Alpha

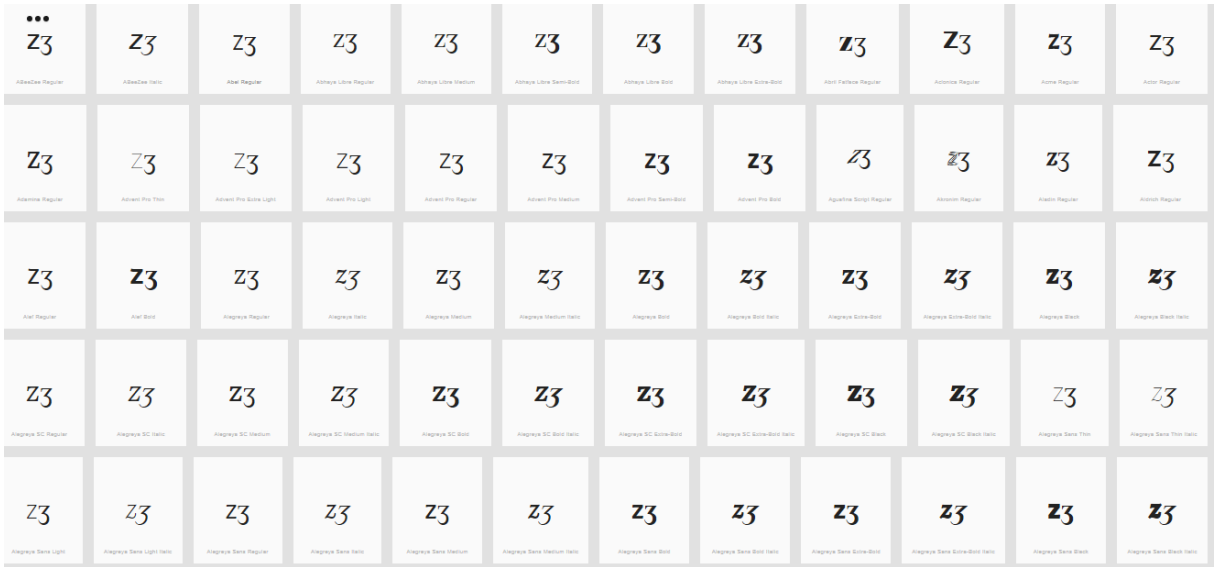
Findings:
Latin Small Letter Alpha is not in the Repertoire

D.1.3 Letter Z vs. Letter Ezh

Code Points Considered:

Code Points	Glyph	Name
007A	z	Letter Z
0292	ʒ	Letter Ezh

Sequence zʒ (007A 0292) compared using Google Fonts in https://wordmark.it/:



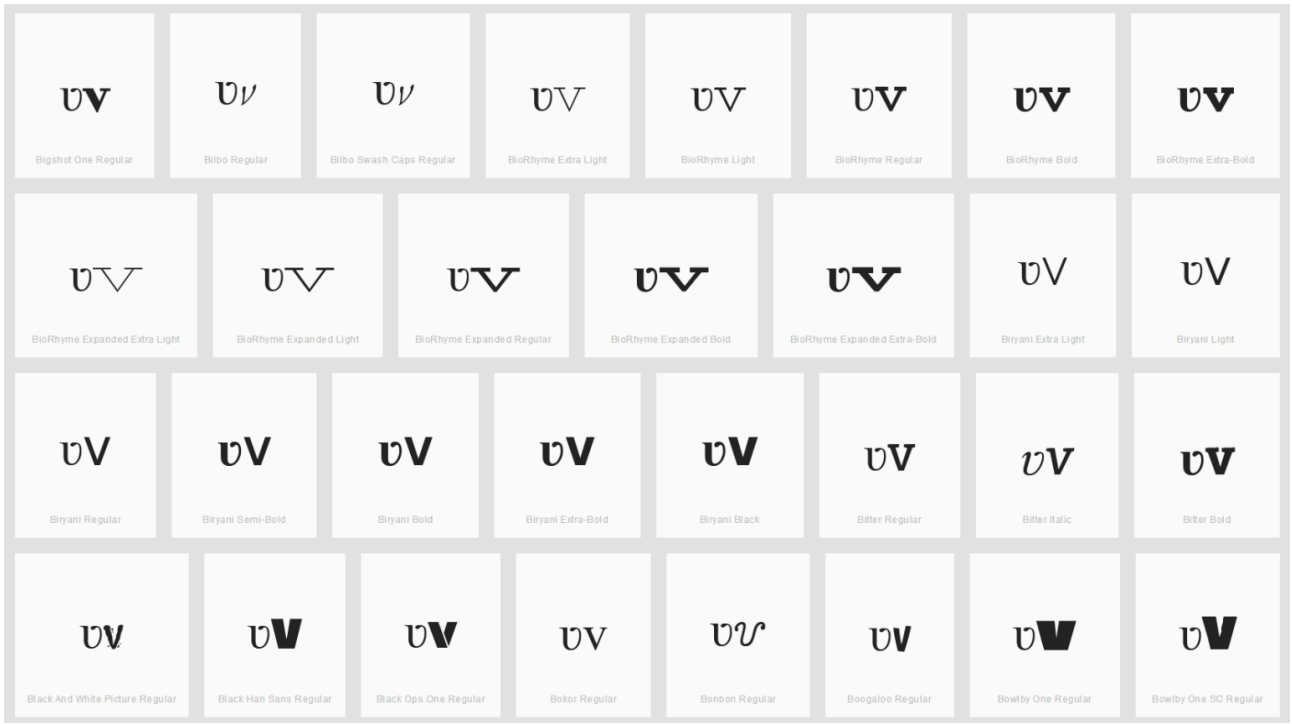
Findings:
Glyphs are distinguishable. In a large number of fonts, the two letters are consistently different.

D.1.4 Latin Small Letter V With Hook vs. Latin Small Letter V

Code Points Considered:

Code Points	Glyph	Name
028B	ʋ	Latin Small Letter V With Hook
0076	v	Latin Small Letter V

Sequence ʋv (028B 0067) compared using Google Fonts in https://wordmark.it/ :



Findings:
All cases I viewed on wordmark.it looked more or less similar to the above screenshot. In particular the ʋ looks more like a u than a v at the bottom in the sense that it never has a sharp angle, but always a

curve (whereas v has almost always a sharp angle). Furthermore, the top right corner of the ʊ is always turned visibly to the left. Even in cases where the v has some serif this is distinguishable from the ʊ hook as the serif is always in both directions (left an right).

D.3.5 I vs. Dotless I vs. Iota

Code Points Considered:

Code Points	Glyph	Name
0069	i	Latin Small Letter I
0131	ı	Latin Small Letter Dotless I
0269	ɩ	Latin Small Letter Iota

Sequence ıi (0131 0069) compared using Google Fonts in https://wordmark.it/ :

Findings:

Both glyphs are distinguishable when written in lower case. I could not find a font, where the dot on the i was missing or almost invisible. However, some fonts displayed the lower case characters in upper case instead. In those examples, the letters were exactly the same (see red marked examples).

Sequence ıɩ Dotless (0131 0269) compared using Google Fonts in https://wordmark.it/:

Findings:
In the italic versions of any of the serif fonts (e.g. Times New Roman or Consolas) these are identical.

D.1.5 Letter E vs. Open E

Code Points Considered:

Code Points	Glyph	Name
0065	e	Letter E
025B	ɛ	Letter Open E

Glyph Representations per https://wordmark.it/:

Findings:
Glyphs are distinguishable. In a large number of fonts, the two letters are consistently different.

D.1.6 Letter K vs. Letter K With Hook

Code Points Considered:

Code Points	Glyph	Name
006B	k	Letter K
0199	ƙ	Letter K with Hook

Sequence K (006B) and K with hook: ƙ (0199) compared using Google Fonts in https://wordmark.it/:

Findings:
Variant – indistinguishable in some fonts

D.1.7 Latin Small Letter Y vs. Latin Small Letter Y With Hook

Code Points Considered:

Code Points	Glyph	Name
01B4	y	Latin Small Letter Y With Hook
0079	y	Latin Small Letter Y

Sequence yyyy (0079 + 0079 + 01B4 + 0079) compared using Google Fonts in https://wordmark.it/:

Findings:

As expected, there is a large degree in variation in the rendering of the glyphs of 0079.
Two essential differences between 01B4 and 0079 are recognized. 01B4 tends to be tilted or italicized and the key difference is the extended diagonal line turning into a right hand side hook.
As demonstrated by the examples, a number of fonts show a similar tilting, not only in italic fonts, as well as an extension of lines.
However, no example was found where the right hand-side line is extended right-wards (but only left-wards - generally also in cursive handwriting the letter doesn't connect right-wards at the top to following letter), and only one font (highlighted in blue) was shown where the two renderings are visually (nearly) identical.

Conclusions:
Since the two code-points are different in a large number of fonts (albeit inconsistently) no variant pair is warranted in this case.

D.1.8 Letter D With Caron vs. Letter D With Hook

Code Points Considered:

Code Points	Glyph	Name
010F	ď	Letter D with Caron
0257	ɗ	Letter D with Hook

Sequence D with Caron vs D with hook compared using Google Fonts in https://wordmark.it/:

Findings:
Variant – indistinguishable, depending on font design.

D.1.9 Latin Small Letter T vs. Latin Small Letter L With Stroke

Code Points Considered:

Code Points	Glyph	Name

0074	t	Latin Small Letter T
0142	ł	Latin Small Letter L With Stroke

Sequence (t ł) (0074 0142) compared using Google Fonts in https://wordmark.it/:

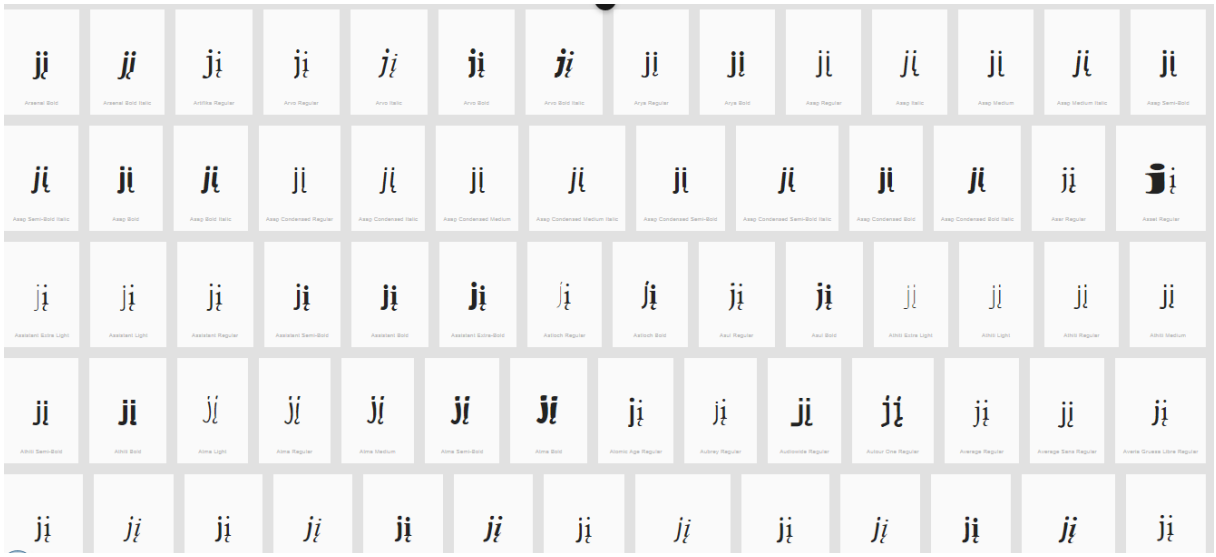
Findings:
Glyphs are distinguishable

D.1.10 Letter J vs. Letter I With Ogonek

Code Points Considered:

Code Points	Glyph	Name
006A	j	LETTER J
012F	į	LETTER I WITH OGONEK

Sequence jį (006A 012F) compared using Google Fonts in https://wordmark.it/:



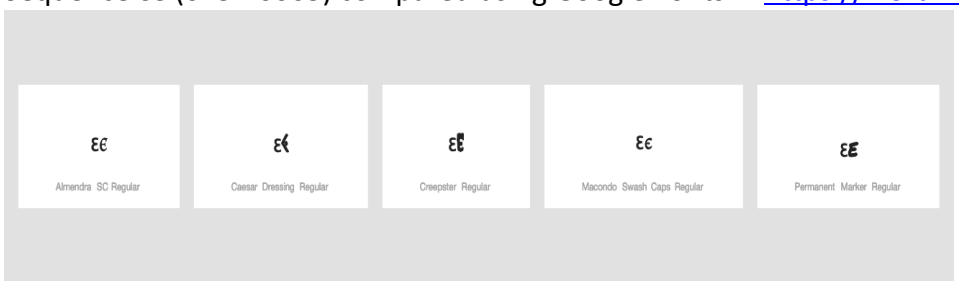
Findings:
Glyphs are distinguishable
In a large number of fonts, the two letters are consistently different.

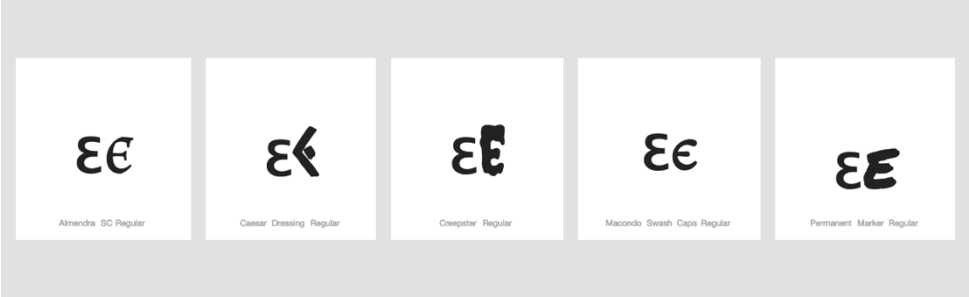
D.1.11 Latin Small Letter Open E vs. Latin Small Letter E

Code Points Considered:

Code Points	Glyph	Name
025B	ɛ	LATIN SMALL LETTER OPEN E
0065	e	LATIN SMALL LETTER E

Sequence ɛe (025B 0065) compared using Google Fonts in https://wordmark.it/





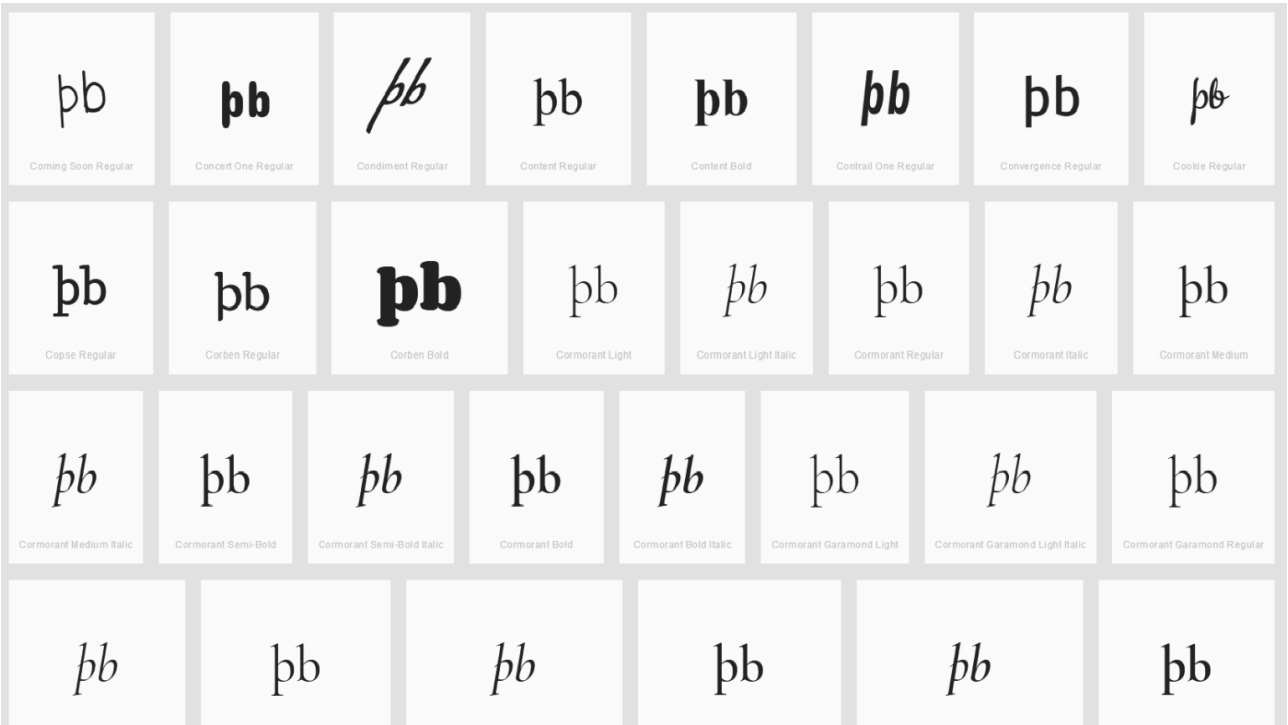
Findings:
Glyphs are distinguishable

D.1.12 Latin Small Letter B vs. Latin Small Letter Thorn vs. Latin Small Letter P

Code Points Considered:

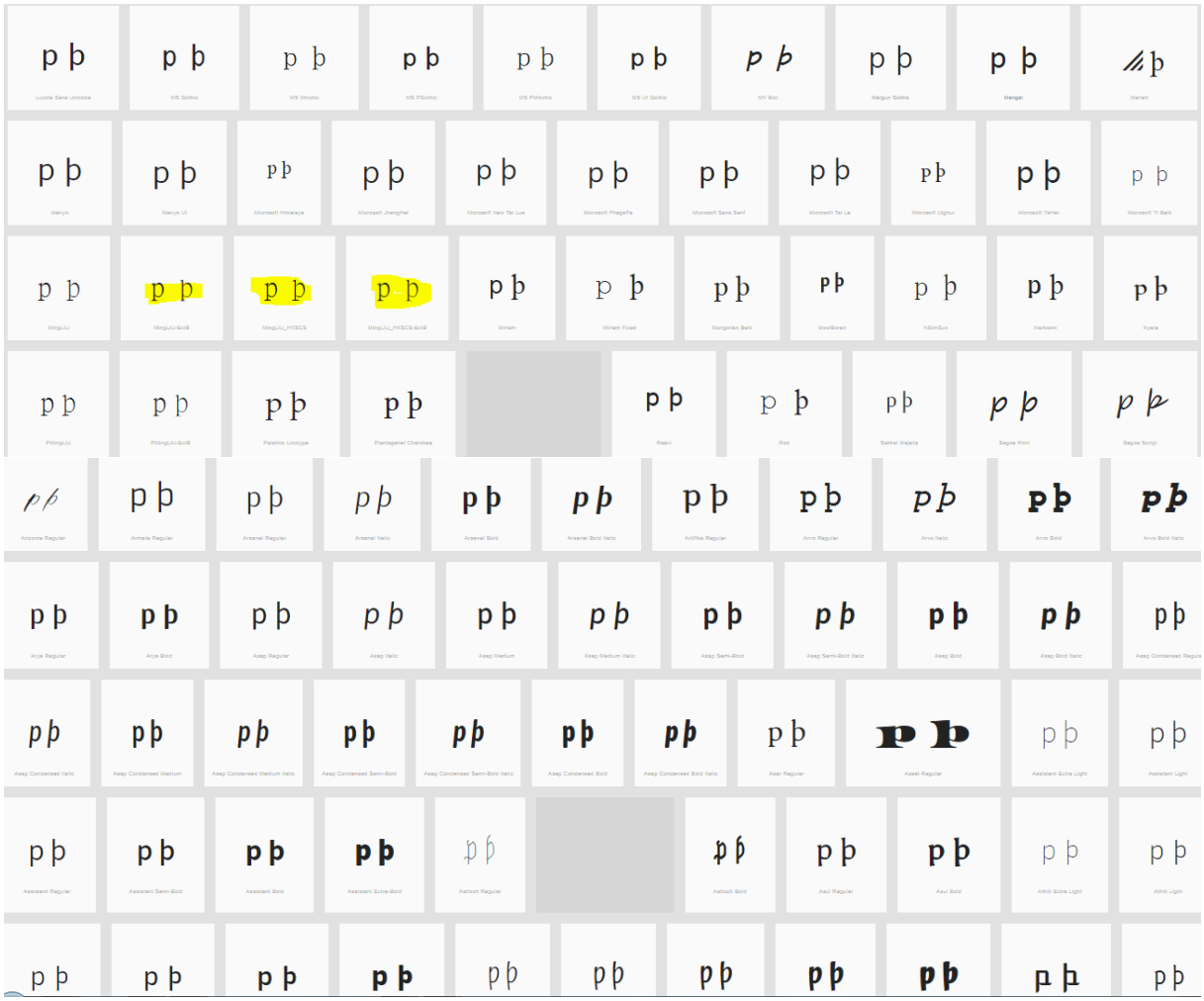
Code Points	Glyph	Name
00FE	þ	LATIN SMALL LETTER THORN
0062	b	LATIN SMALL LETTER B
0070	p	LATIN SMALL LETTER P

Sequence þb (00FE 0062) compared using Google Fonts in https://wordmark.it/ :



Findings:
All cases I viewed on wordmark.it looked similar to the above screenshot. The þ and b always appear quite distinguishable as the þ always has a stroke below the base line and the b never crosses the base line.

Sequence (p þ) (0070 00FE) compared using Google Fonts in https://wordmark.it/:



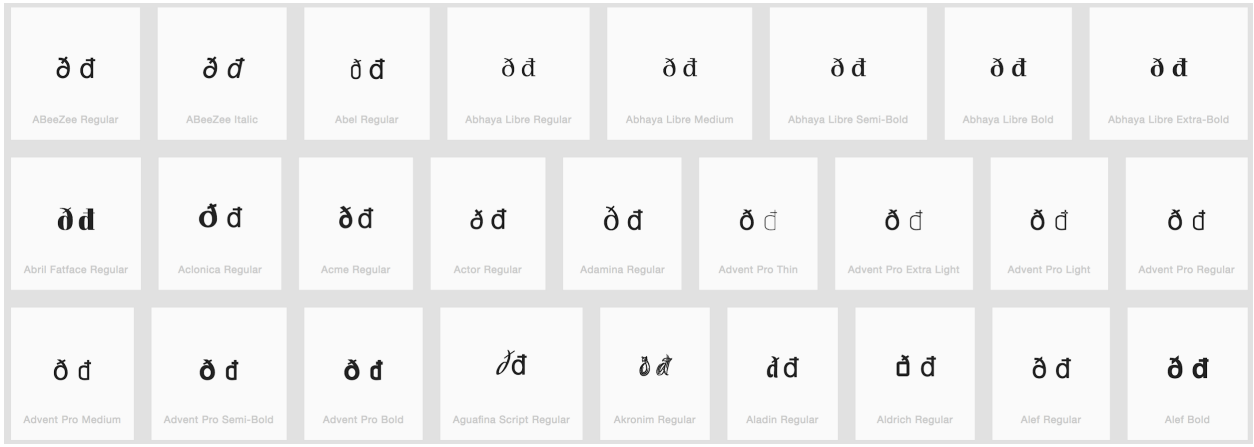
Findings:

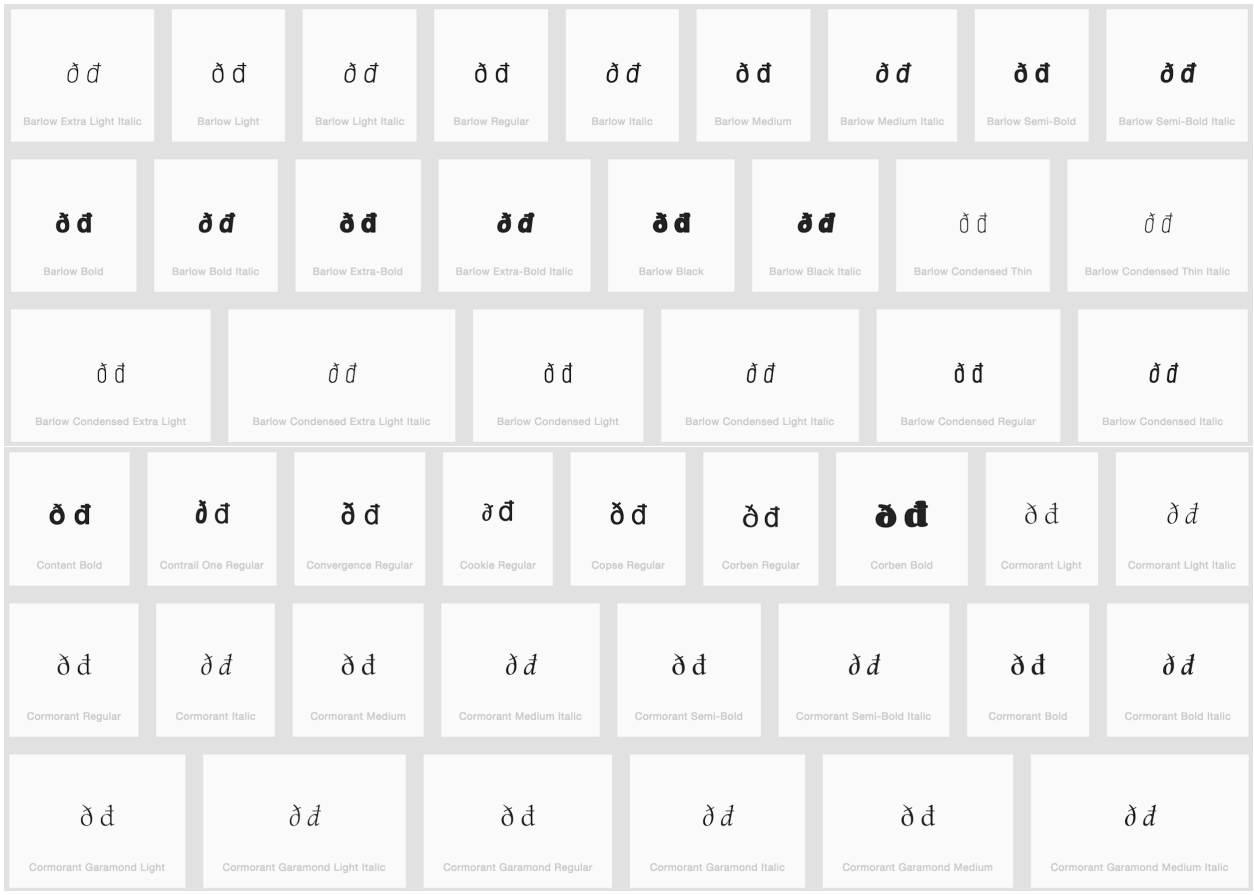
The upper leg of the Thorn is visible in most fonts (except those highlighted) can be somewhat unclear.

D.1.13 Letter Eth Versus Letter D With Stroke

Code Points Considered:

Code Points	Glyph	Name
00F0	ð	LETTER ETH
0111	đ	LETTER D WITH STROKE





Findings:

The two letters are consistently rendered with their distinguishable features.

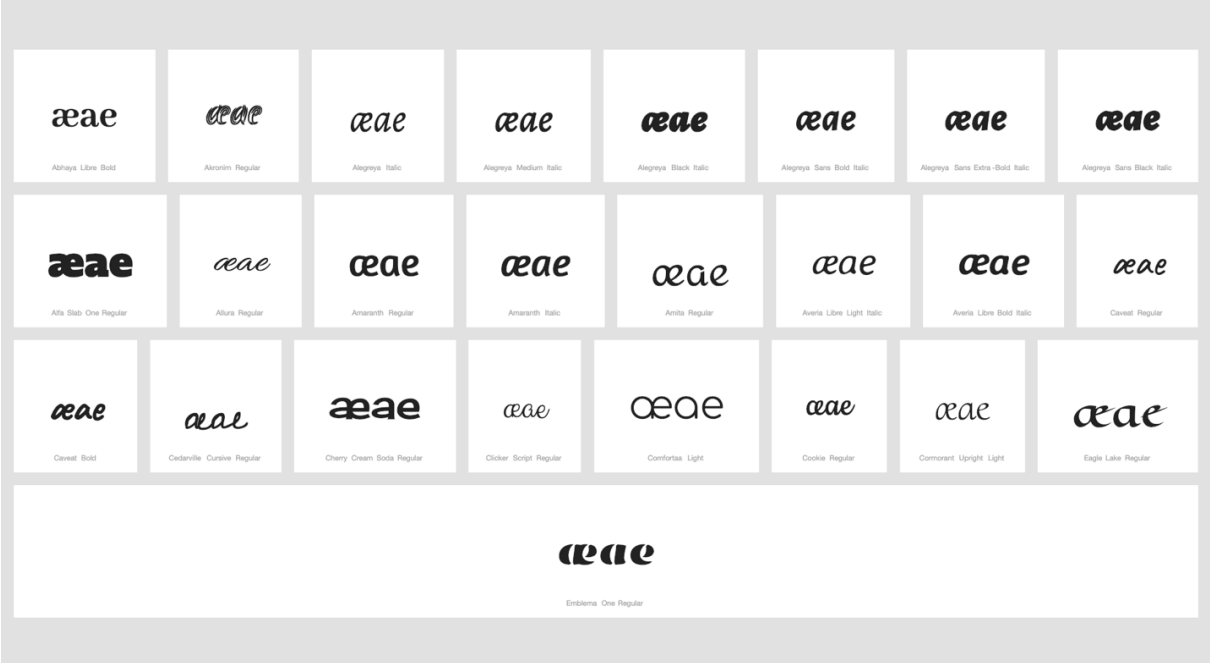
D.2 Spacing of Base Characters

D.2.1 AE Ligature vs. Sequence AE

Code Points Considered:

Code Points	Glyph	Name
00E6	æ	Latin Small Letter Ae
0061	a	Latin Small Letter A
0065	e	Latin Small Letter E
0153	œ	Latin Small Ligature Oe
0251	ɑ	Latin Small Letter Alpha

Sequence æae (00E6 + 0061 + 0065) compared using Google Fonts in https://wordmark.it/:



Findings:
In some fonts, in which the a-glyph takes a shape similar to that of 0251 ɑ LATIN SMALL LETTER ALPHA, the ligature and the sequence bare some similarity but are distinguishable.
In a large number of fonts, the ligature and the sequence are consistently different.

Additional Findings:
In fonts, in which the a-glyph takes a shape similar to that of 0251 ɑ LATIN SMALL LETTER ALPHA, the ligature 00E6 becomes nearly visually identical with the o-e ligature (0153 œ LATIN SMALL LIGATURE OE) as demonstrated below.

Sequence æaeœoe (00E6+0061+0065+0153+006F+0065) compared using Google Fonts in https://wordmark.it/:

Conclusion:
Suggestion to consider 00E6 LATIN SMALL LETTER AE and 0153 LATIN SMALL LIGATURE OE as variant pair or add to the string similarity list on the grounds
of them being visually nearly identical

AND being similar on non-visual grounds because of conceptional identity of 0251 ɑ LATIN SMALL LETTER ALPHA and 0061 a 0061 LATIN SMALL LETTER A
in a significant number of fonts.

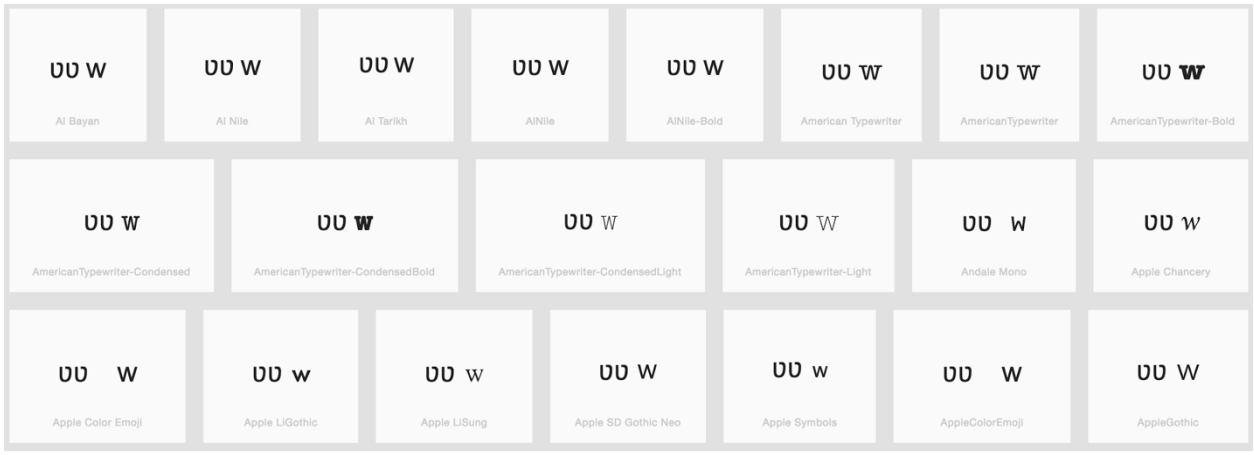
D.2.2 OE Ligature vs. Sequence OE

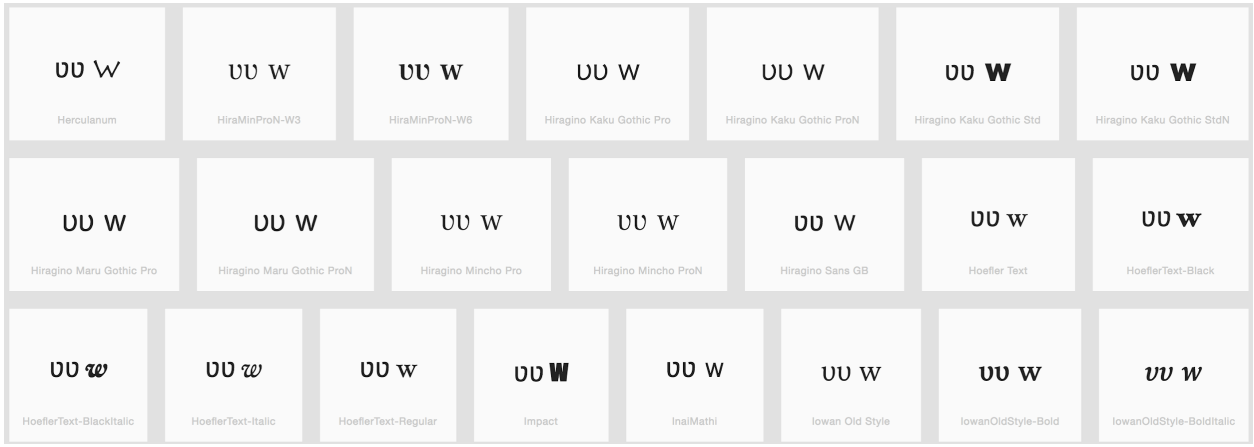
D.2.3 Sequence of Two Letter V With Hook vs. Letter W

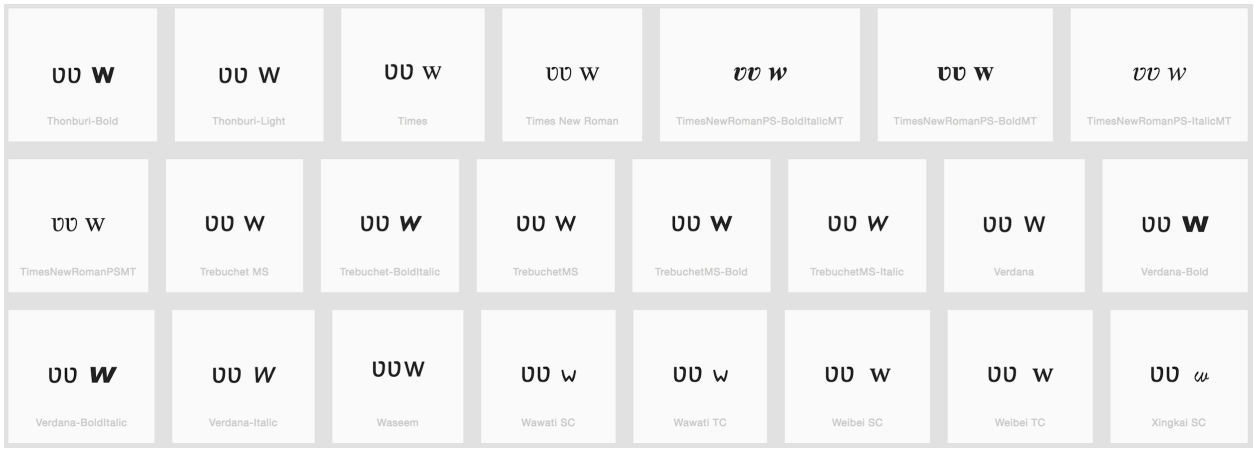
Code Points Considered:

Code Points	Glyph	Name
028B 028B	ʋʋ	Letter V with Hook (x2)
0077	w	Letter W

Sequence ʋʋ w (028B028B 0077) compared using Google Fonts in https://wordmark.it/:







Findings:

Sequence of two Letters V with Hook is different than Letter W

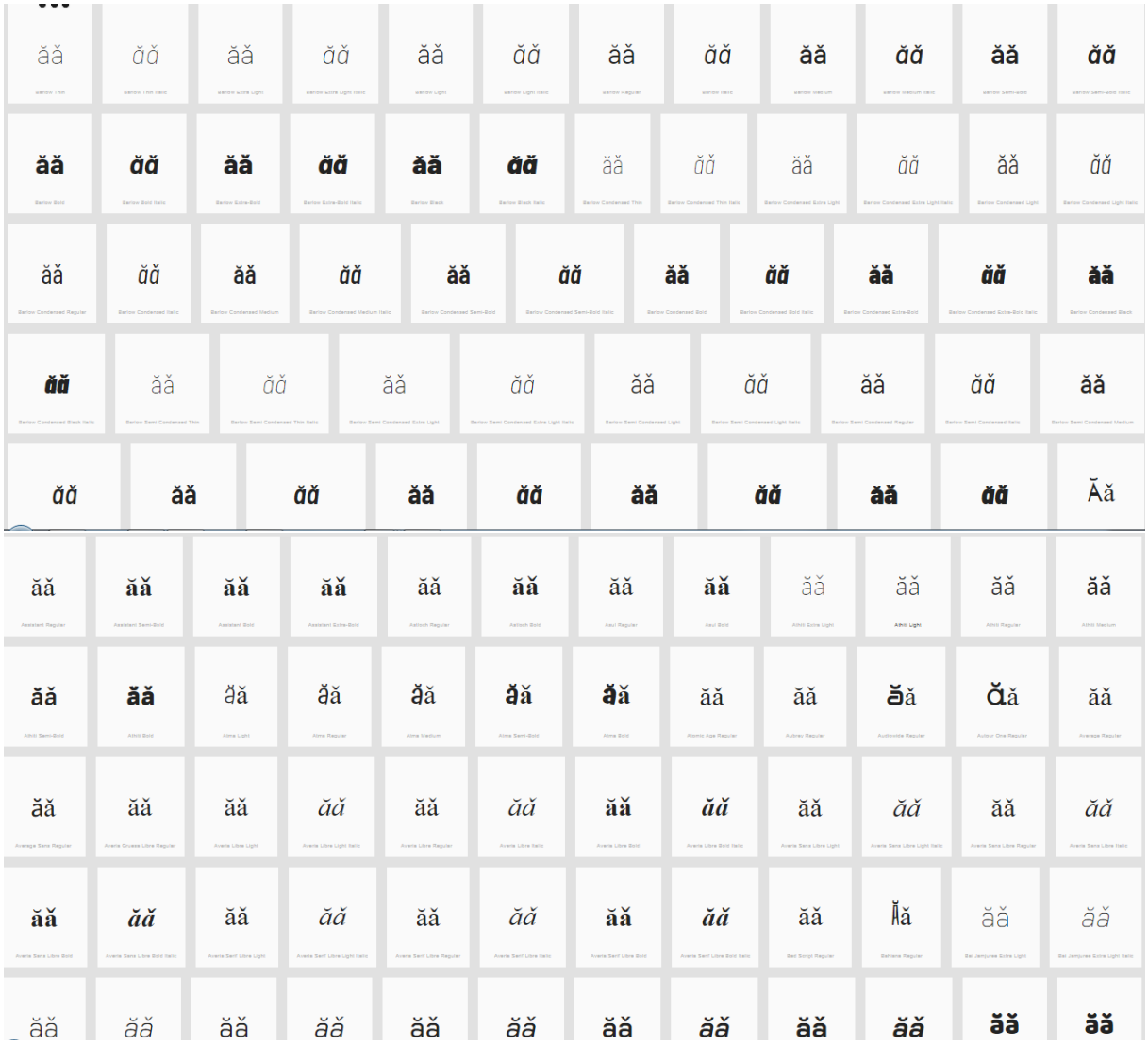
D.3 Shaping of Diacritics

D.3.1 Caron (Above) vs. Breve

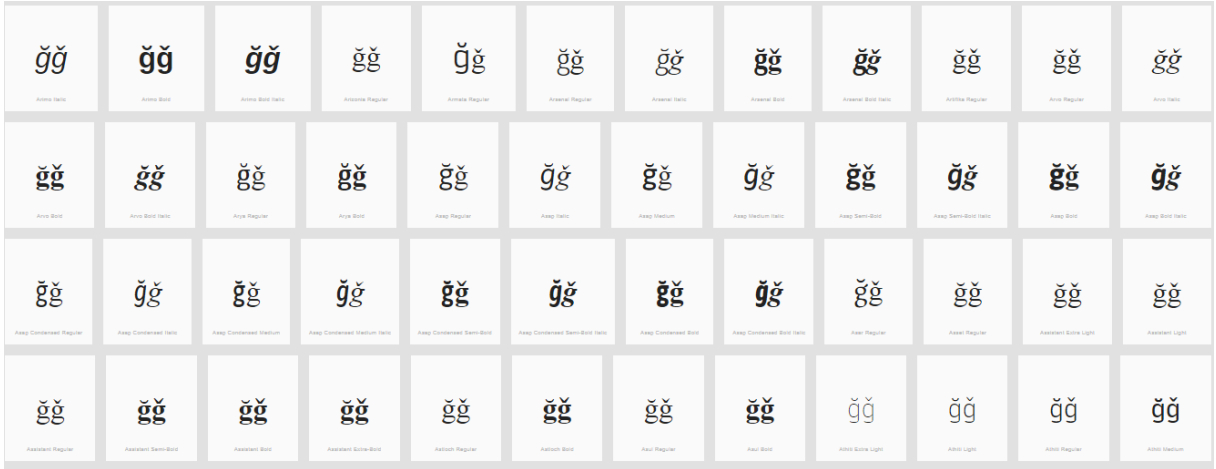
Code Points Considered:

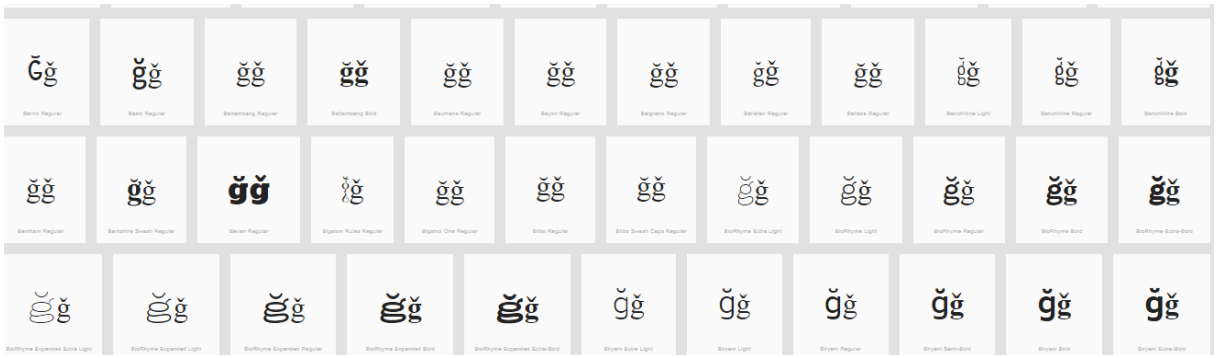
Code Points	Glyph	Name
0103	ă	LETTER A WITH BREVE
011F	ğ	LETTER G WITH BREVE
016D	ŭ	LETTER U WITH BREVE
010D	č	LETTER C WITH CARON
011B	ě	LETTER E WITH CARON
0148	ň	LETTER N WITH CARON
0159	ř	LETTER R WITH CARON
0161	š	LETTER S WITH CARON
017E	ž	LETTER Z WITH CARON
01CE	ǎ	LETTER A WITH CARON
01D0	ǐ	LETTER I WITH CARON
01D2	ǒ	LETTER O WITH CARON
01D4	ǔ	LETTER U WITH CARON
01E7	ǧ	LETTER G WITH CARON
01E9	ǩ	LETTER K WITH CARON
01EF	ǯ	LETTER EZH WITH CARON

Sequence ăǎ (0103 01CE) compared using Google Fonts in https://wordmark.it/:



Sequence ğğ (011F 01E7) compared using Google Fonts in https://wordmark.it/:





Sequence ŭ ǔ (016D 01D4) compared using Google Fonts in https://wordmark.it/:

Findings:
The representations of the Breve and the Caron in Letters A, G and U are distinguishable and undistinguishable in a number of fonts (see pictures above); depending on the font and size.

D.3.2 Tilde vs. Macron (Above)

Code Points Considered:

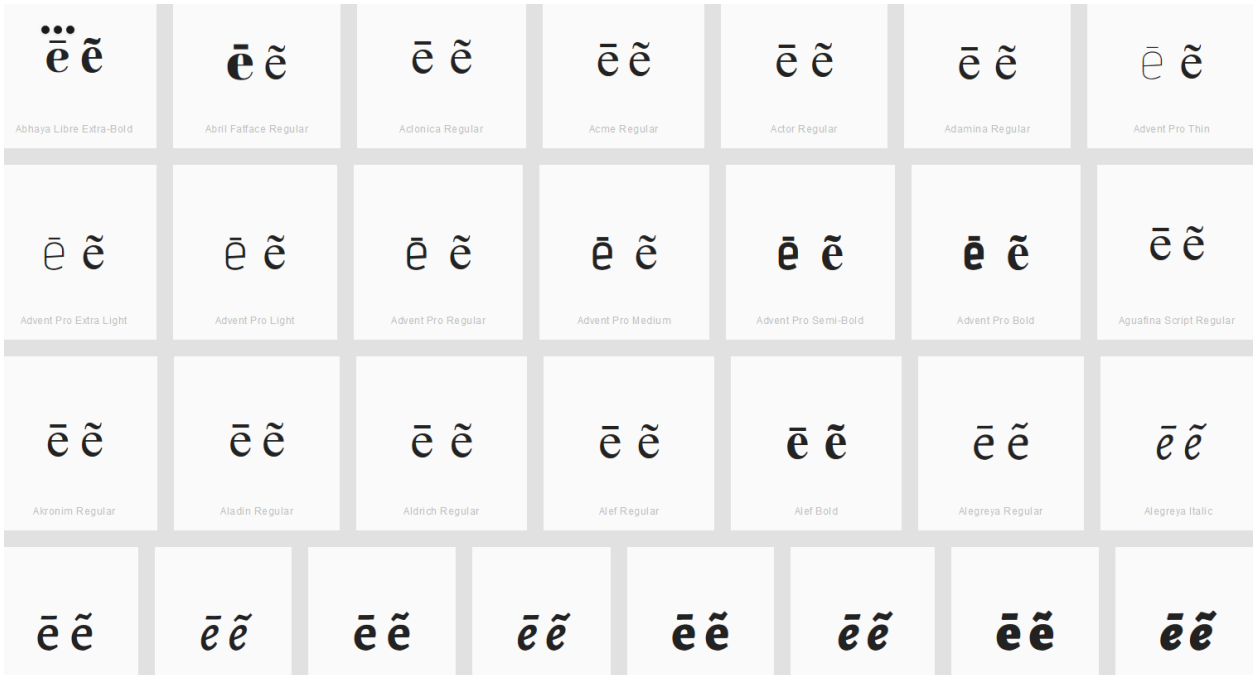
Code Points	Glyph	Name
0067 + 0303	g̃	LATIN SMALL LETTER G + COMBINING TILDE
006E + 0304	n̄	LATIN SMALL LETTER N + COMBINING MACRON
0072 + 0303	r̃	LATIN SMALL LETTER R WITH TILDE
00E3	ã	LATIN SMALL LETTER A WITH TILDE
00F1	ñ	LATIN SMALL LETTER N WITH TILDE

00F5	õ	LATIN SMALL LETTER O WITH TILDE
0101	ā	LATIN SMALL LETTER A WITH MACRON
0113	ē	LATIN SMALL LETTER E WITH MACRON
0129	ĩ	LATIN SMALL LETTER I WITH TILDE
012B	ī	LATIN SMALL LETTER I WITH MACRON
014D	ō	LATIN SMALL LETTER O WITH MACRON
0169	ũ	LATIN SMALL LETTER U WITH TILDE
016B	ū	LATIN SMALL LETTER U WITH MACRON
1E21	ḡ	LATIN SMALL LETTER G + MACRON
1EBD	ẽ	LATIN SMALL LETTER E WITH TILDE
1EF9	ỹ	LATIN SMALL LETTER Y WITH TILDE

Sequence ãā (00E3 0101) compared using Google Fonts in https://wordmark.it/:

Findings:
Macron and Tilde are distinguishable for the viewed fonts.

Sequence ēẽ (0113 1EBD) compared using Google Fonts in https://wordmark.it/:



Findings:
Macron and Tilde are distinguishable for the viewed fonts.

Sequence g̃ḡ (0067+0303 1E21) compared using Google Fonts in https://wordmark.it/:

Findings:
Macron and Tilde are distinguishable for the viewed fonts.

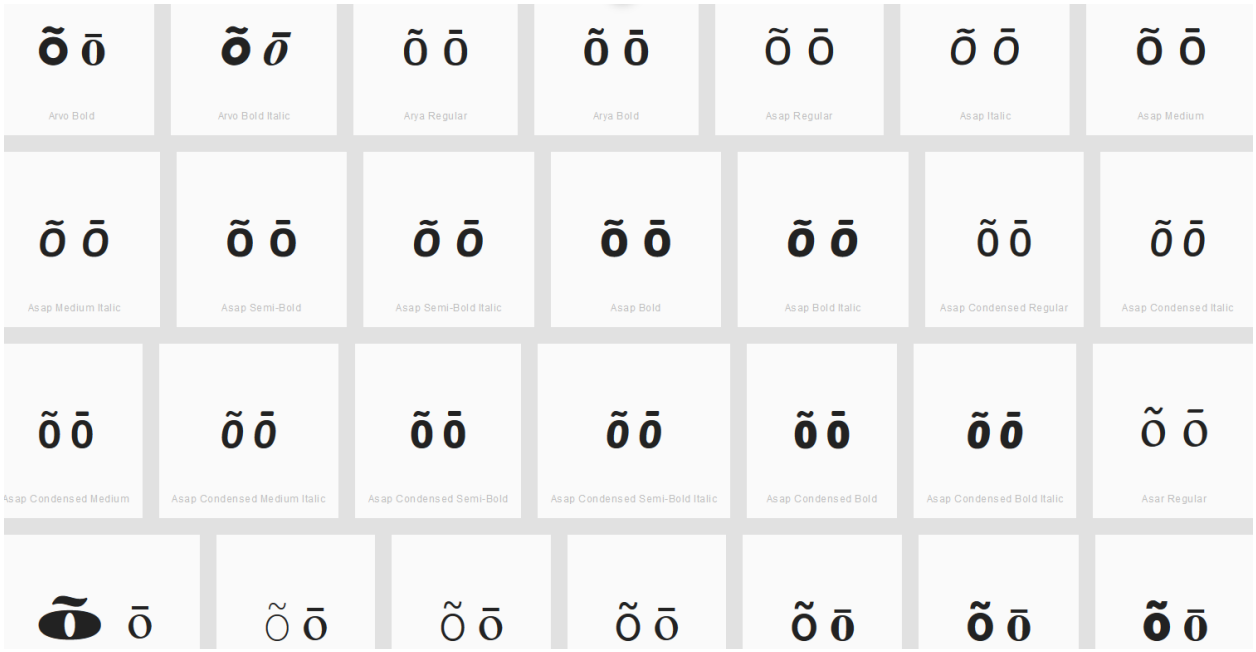
Sequence ĩī (0129 012B) compared using Google Fonts in https://wordmark.it/:

Findings:
Macron and Tilde are distinguishable for the viewed fonts.

Sequence ñ̄ñ (006E+0304 00F1) compared using Google Fonts in https://wordmark.it/:

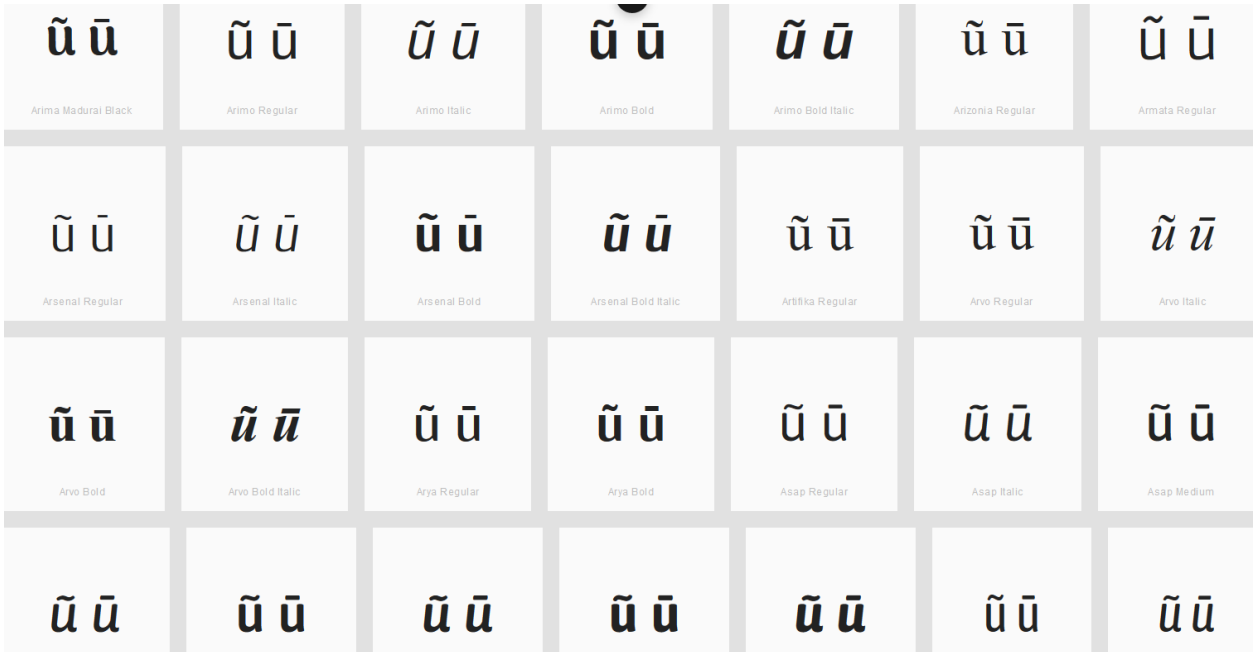
Findings:
Macron and Tilde are distinguishable for almost all viewed fonts. I found very few examples where they are not. In the example below the second pair (marked red) is distinguishable, but only because the macron above is moved to the right.

Sequence õō (00F5 014D) compared using Google Fonts in https://wordmark.it/:



Findings:
Macron and Tilde are distinguishable for the viewed fonts.

Sequence ũū (0169 016B) compared using Google Fonts in https://wordmark.it/:



Findings:
Macron and Tilde are distinguishable for the viewed fonts.

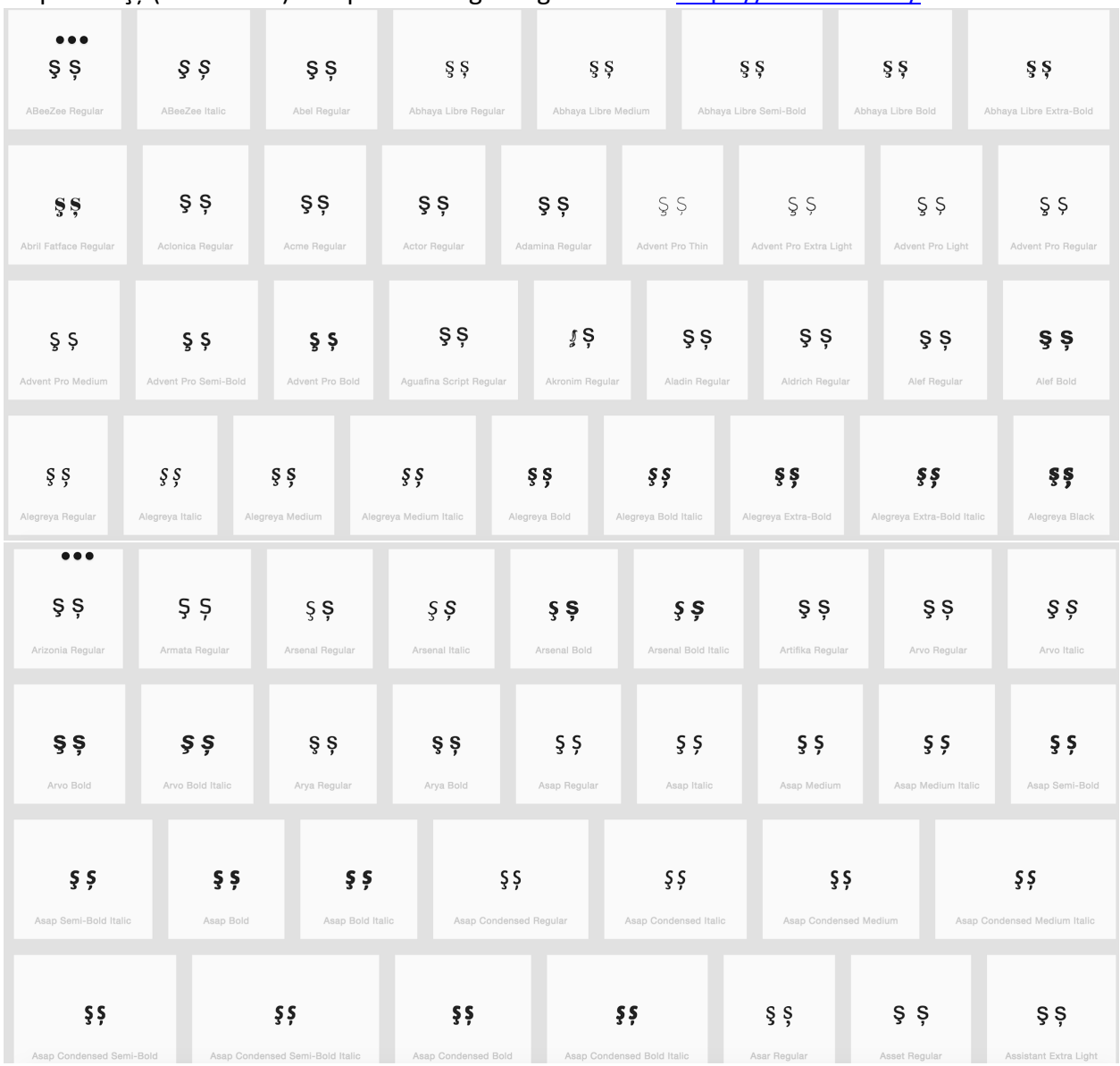
D.3.3 Combining Cedilla (Below), Ogonek And Comma Below

Code Points Considered:

Code Points	Glyph	Name
006D 0327	m̧	LETTER M WITH COMBINING CEDILLA
006F 0327	o̧	LETTER O WITH COMBINING CEDILLA

00E7	Ç	LETTER C WITH CEDILLA
0105	Ą	LETTER A WITH OGONEK
0119	Ę	LETTER E WITH OGONEK
012F	Į	LETTER I WITH OGONEK
0137	Ķ	LETTER K WITH CEDILLA
013C	ļ	LETTER L WITH CEDILLA
0146	ņ	LETTER N WITH CEDILLA
015F	ş	LETTER S WITH CEDILLA
0173	ų	LETTER U WITH OGONEK
0219	ș	LETTER S WITH COMMA BELOW
021B	ț	LETTER T WITH COMMA BELOW

Sequence şș (015F 0219) compared using Google Fonts in https://wordmark.it/:



Findings:
The representations of the Cedilla and the Comma Below in Letter S are distinguishable in a number of fonts (see pictures below); in a large number of fonts, the two diacritics are consistently different. No

other point base character (except for Letter S) uses two different diacritics (i.e., Letter M only exists with a Combining Cedilla, but not with Ogonek or Comma Below).

D.3.4 Circle above vs. Ring

Code Points Considered:

Code Points	Glyph	Name
00E5	å	LATIN SMALL LETTER A WITH RING ABOVE
016F	ů	LATIN SMALL LETTER U WITH RING ABOVE
017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE
010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE
0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE
0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE
1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE

Findings:

No eligible candidates.

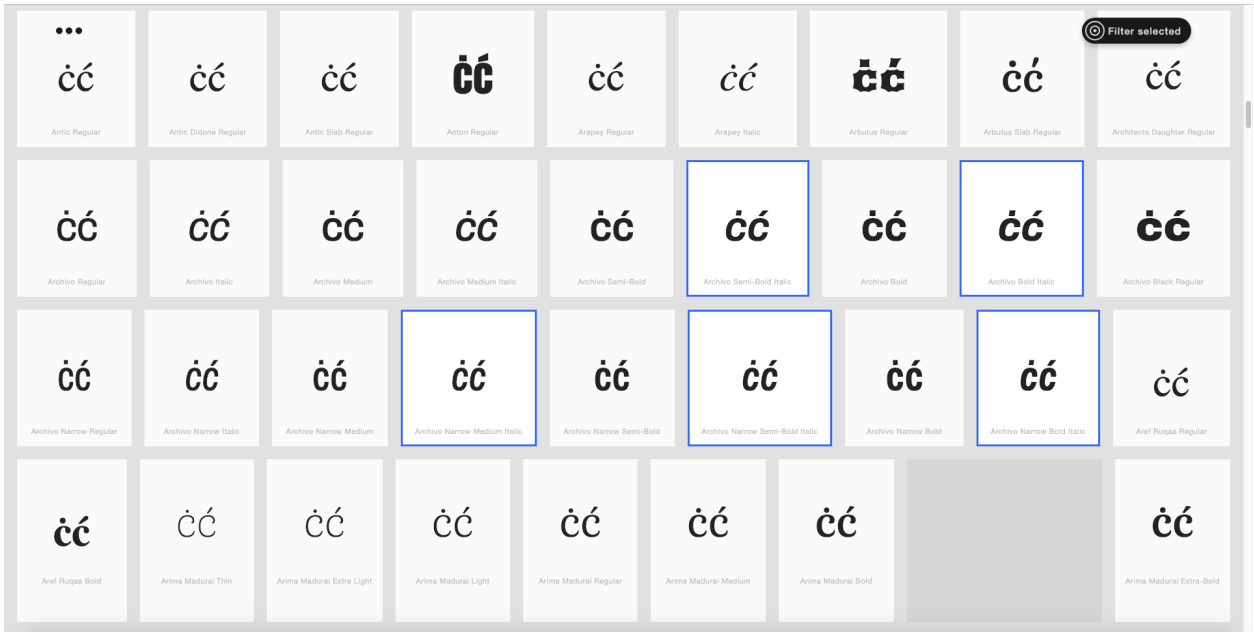
D.3.5 Acute Above vs. Dot Above

Code Points Considered:

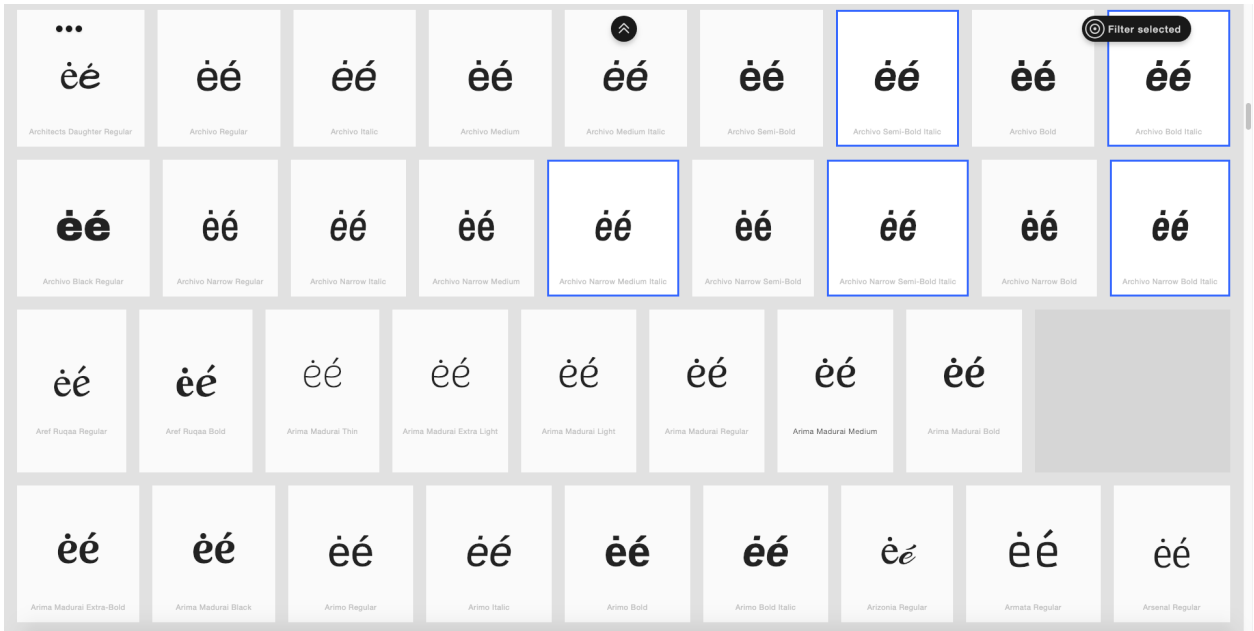
Code Points	Glyph	Name
0069	i	LATIN SMALL LETTER I
00E1	á	LATIN SMALL LETTER A WITH ACUTE
00E9	é	LATIN SMALL LETTER E WITH ACUTE
00ED	í	LATIN SMALL LETTER I WITH ACUTE
00F3	ó	LATIN SMALL LETTER O WITH ACUTE
00FA	ú	LATIN SMALL LETTER U WITH ACUTE
00FD	ý	LATIN SMALL LETTER Y WITH ACUTE
0107	ć	LATIN SMALL LETTER C WITH ACUTE
013A	Í	LATIN SMALL LETTER L WITH ACUTE
0144	ń	LATIN SMALL LETTER N WITH ACUTE
0155	ŕ	LATIN SMALL LETTER R WITH ACUTE
015B	ś	LATIN SMALL LETTER S WITH ACUTE
017A	ź	LATIN SMALL LETTER Z WITH ACUTE

010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE
0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE
0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE
017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE
1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE

Sequence ċć (010B+ 0107) compared using Google Fonts in https://wordmark.it/:

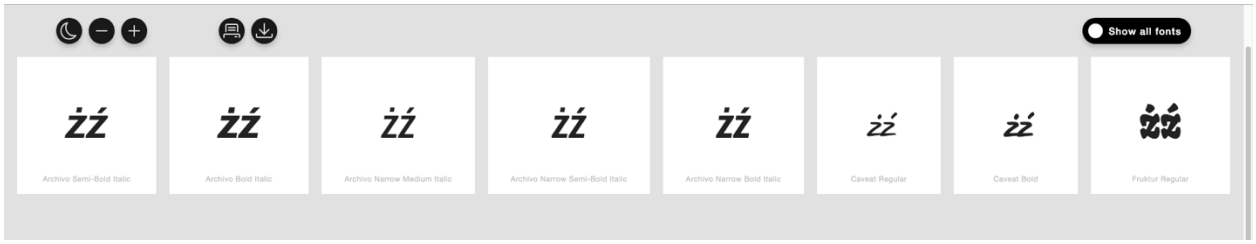


Sequence ėé (0117 + 00E9) compared using Google Fonts in https://wordmark.it/:



Sequence ṅń (1E45+ 0144) compared using Google Fonts in https://wordmark.it/:

Sequence żź (017C+ 017A) compared using Google Fonts in https://wordmark.it/:



Findings:
ċć,ėé, ṅń, and żź were considered as potential variant pairs
The representations of the acute and the dot above in these pairs are distinguishable in a number of fonts.
In a large number of fonts, the two diacritics are consistently different.

Conclusion:
No variant pairs are warranted.

D.3.6 Grave vs. Dot above

Code Points Considered:

Code Points	Glyph	Name
00E8	è	LATIN SMALL LETTER E WITH GRAVE
00EC	ì	LATIN SMALL LETTER I WITH GRAVE
00F2	ò	LATIN SMALL LETTER O WITH GRAVE
00F9	ù	LATIN SMALL LETTER U WITH GRAVE
1EF3	ỳ	LATIN SMALL LETTER Y WITH GRAVE

010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE
0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE
0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE
017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE
1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE

Sequence èė (00E8 + 0117) compared using Google Fonts in https://wordmark.it/ :

Findings:
Despite variation in the shaping of e, as well as occasional clippings, the representations of the grave and the dot remain distinguishable.
In a large number of fonts, the two diacritics are consistently different.

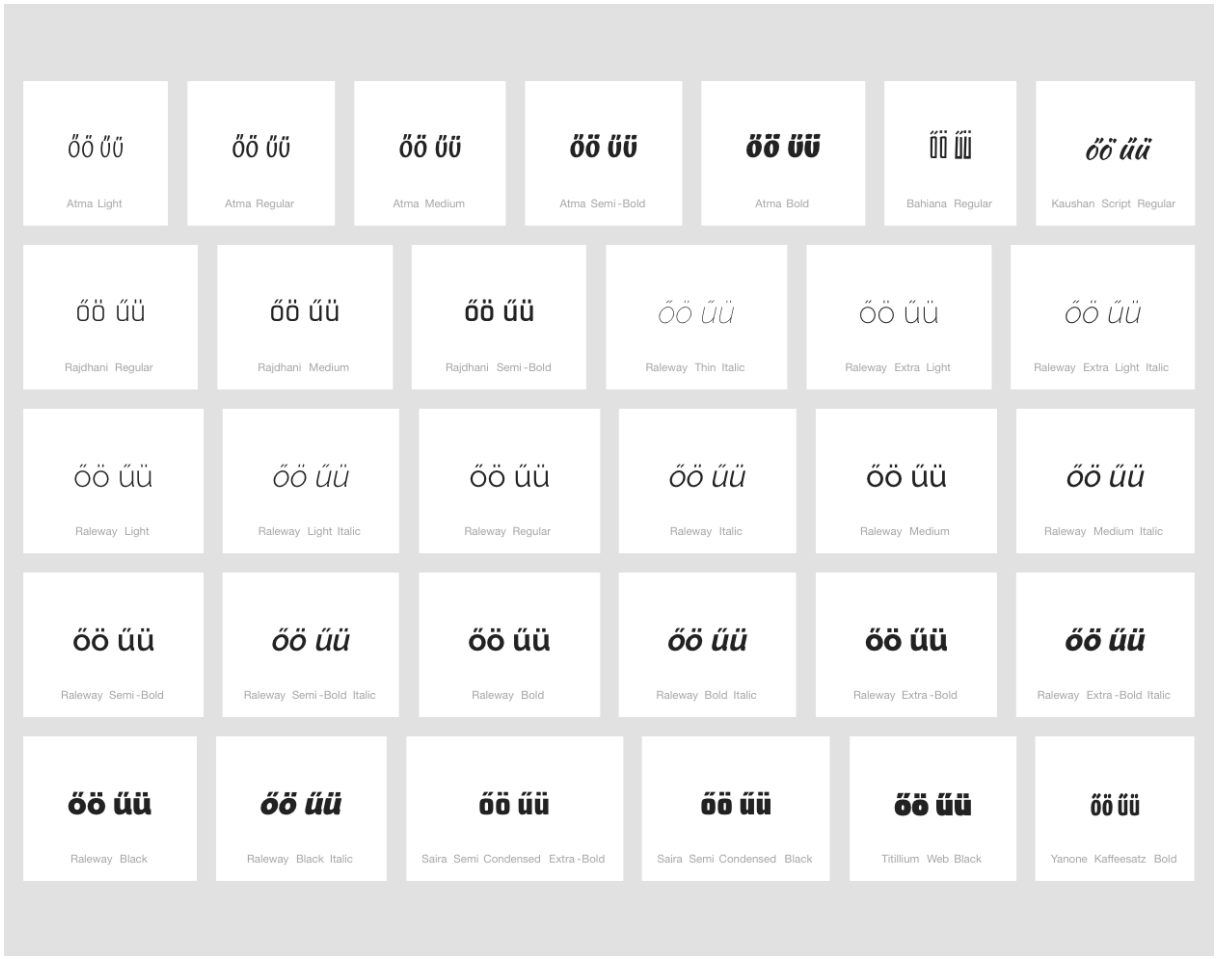
Conclusion:
No variant pairs are warranted.

D.3.7 Double Acute vs. Diaresis

Code Points Considered:

Code Points	Glyph	Name
006E + 0308	n̈	LATIN SMALL LETTER N + COMBINING DIAERESIS
00E4	ä	LATIN SMALL LETTER A WITH DIAERESIS
00EB	ë	LATIN SMALL LETTER E WITH DIAERESIS
00EF	ï	LATIN SMALL LETTER I WITH DIAERESIS
00F6	ö	LATIN SMALL LETTER O WITH DIAERESIS
00FC	ü	LATIN SMALL LETTER U WITH DIAERESIS
00FF	ÿ	LATIN SMALL LETTER Y WITH DIAERESIS
0151	ő	LATIN SMALL LETTER O WITH DOUBLE ACUTE
0171	ű	LATIN SMALL LETTER U WITH DOUBLE ACUTE
0254 + 0308	ɔ̈	LATIN SMALL LETTER OPEN O + COMBINING DIAERESIS
025B + 0308	ɛ̈	LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS
025B + 0331 + 0308	ɛ̱̈	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS
1E8D	ẍ	LATIN SMALL LETTER X WITH DIAERESIS

Sequence őö and üű (00F6 0151 and 00FC 0171) compared using Google Fonts in https://wordmark.it/:



Findings:
The representations of the Double Acute vs Diaresis in these pairs are distinguishable in a number of fonts. In some fonts, the two diacritics look similar.

Conclusion:
Code points őö and üű should be investigated for visual similarity

D.3.8 Dot Below vs. Comma Below

Code Points Considered:

Code Points	Glyph	Name
1E37	ḷ	LETTER L WITH DOT BELOW
1E43	ṃ	LETTER M WITH DOT BELOW
1E47	ṇ	LETTER N WITH DOT BELOW
1E63	ṣ	LETTER S WITH DOT BELOW
1E6D	ṭ	LETTER T WITH DOT BELOW
1EA1	ạ	LETTER A WITH DOT BELOW
1EB9	ẹ	LETTER E WITH DOT BELOW
1ECB	ị	LETTER I WITH DOT BELOW
1ECD	ọ	LETTER O WITH DOT BELOW

1EE5	ụ	LETTER U WITH DOT BELOW
1EF5	ỵ	LETTER Y WITH DOT BELOW
0219	ș	LETTER S WITH COMMA BELOW
021B	ț	LETTER T WITH COMMA BELOW

Sequence ṣș (1E63 0219) compared using Google Fonts in https://wordmark.it/:

Sequence ṭț (1E6D 021B) compared using Google Fonts in https://wordmark.it/:

Findings:
The representations of the Dot below and Comma below in Letters S and T are distinguishable in a number of fonts (see pictures above); in a large number of fonts, the two diacritics are consistently different.

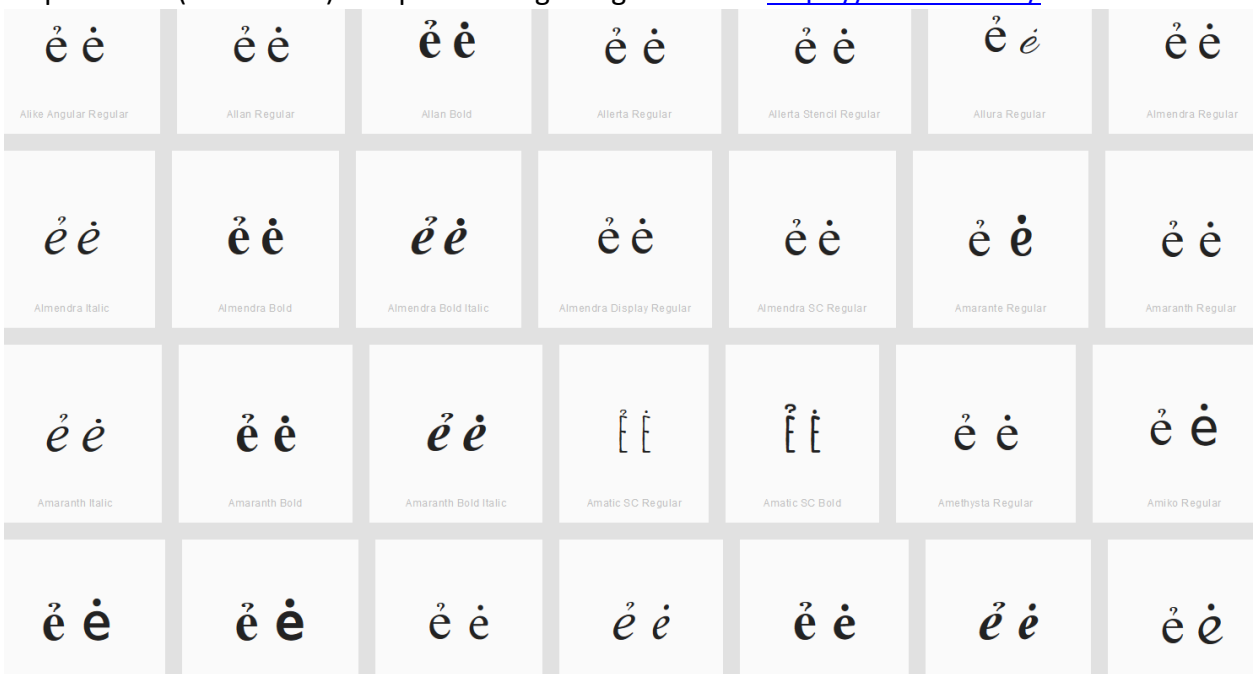
D.3.9 Hook vs. Dot (Above)

Code Points Considered:

Code Points	Glyph	Name
0069	i	LATIN SMALL LETTER I
010B	ċ	LATIN SMALL LETTER C WITH DOT ABOVE
0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE
0121	ġ	LATIN SMALL LETTER G WITH DOT ABOVE
017C	ż	LATIN SMALL LETTER Z WITH DOT ABOVE
0199	ƙ	LATIN SMALL LETTER K WITH HOOK
01B4	y	LATIN SMALL LETTER Y WITH HOOK
0253	ɓ	LATIN SMALL LETTER B WITH HOOK
0257	ɗ	LATIN SMALL LETTER D WITH HOOK
1E45	ṅ	LATIN SMALL LETTER N WITH DOT ABOVE
1EA3	ả	LATIN SMALL LETTER A WITH HOOK ABOVE
1EBB	ẻ	LATIN SMALL LETTER E WITH HOOK ABOVE

1EC9	ỉ	LATIN SMALL LETTER I WITH HOOK ABOVE
1ECF	ỏ	LATIN SMALL LETTER O WITH HOOK ABOVE
1EE7	ủ	LATIN SMALL LETTER U WITH HOOK ABOVE
1EF7	ỷ	LATIN SMALL LETTER Y WITH HOOK ABOVE

Sequence ẻė (1EBB 0117) compared using Google Fonts in https://wordmark.it/:



Findings:
Dot and Hook are distinguishable for the viewed fonts.

Sequence iỉ (0069 1EC9) compared using Google Fonts in https://wordmark.it/:

Findings:
Dot and Hook are distinguishable for the viewed fonts.

Sequence yỷ (01B4 1EF7) compared using Google Fonts in https://wordmark.it/:

Findings:

Hook and Hook Above are distinguishable for the viewed fonts.

D.3.10 Caron vs. Hook

Code Points Considered:

Code Points	Glyph	Name
010F	ď	LETTER D WITH CARON
0257	ɗ	LETTER D WITH HOOK

Sequence D with Caron vs D with hook compared using Google Fonts in https://wordmark.it/:

Findings:
Variant – indistinguishable, depending on font design.

D.3.11 Caron vs. Horn

Code Points Considered:

Code Points	Glyph	Name
01CE	ǎ	LATIN SMALL LETTER A WITH CARON
010D	č	LATIN SMALL LETTER C WITH CARON
010F	ď	LATIN SMALL LETTER D WITH CARON
011B	ě	LATIN SMALL LETTER E WITH CARON
01E7	ǧ	LATIN SMALL LETTER G WITH CARON
01D0	ǐ	LATIN SMALL LETTER I WITH CARON
01E9	ǩ	LATIN SMALL LETTER K WITH CARON
013E	ľ	LATIN SMALL LETTER L WITH CARON
0148	ň	LATIN SMALL LETTER N WITH CARON
01D2	ǒ	LATIN SMALL LETTER O WITH CARON
01A1	ơ	LATIN SMALL LETTER O WITH HORN
1EDB	ớ	LATIN SMALL LETTER O WITH HORN AND ACUTE

1EDD	ờ	LATIN SMALL LETTER O WITH HORN AND GRAVE
1EE1	ỡ	LATIN SMALL LETTER O WITH HORN AND TILDE
1EDF	ở	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
1EE3	ợ	LATIN SMALL LETTER O WITH HORN AND DOT BELOW
0159	ř	LATIN SMALL LETTER R WITH CARON
0161	š	LATIN SMALL LETTER S WITH CARON
0165	ť	LATIN SMALL LETTER T WITH CARON
01D4	ǔ	LATIN SMALL LETTER U WITH CARON
01B0	ư	LATIN SMALL LETTER U WITH HORN
1EE9	ứ	LATIN SMALL LETTER U WITH HORN AND ACUTE
1EEB	ừ	LATIN SMALL LETTER U WITH HORN AND GRAVE
1EEF	ữ	LATIN SMALL LETTER U WITH HORN AND TILDE
1EED	ử	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
1EF1	ự	LATIN SMALL LETTER U WITH HORN AND DOT BELOW
017E	ž	LATIN SMALL LETTER Z WITH CARON
01EF	ǯ	LATIN SMALL LETTER EZH WITH CARON

D.4 Stacking of Diacritics

D.4.1 Circumflex And Tilde

Code Points Considered:

Code Points	Glyph	Name
00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX
00E3	ã	LATIN SMALL LETTER A WITH TILDE
00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX
00EE	î	LATIN SMALL LETTER I WITH CIRCUMFLEX
1EAB	ẫ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
00F1	ñ	LATIN SMALL LETTER N WITH TILDE
00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX

00F5	õ	LATIN SMALL LETTER O WITH TILDE
00FB	û	LATIN SMALL LETTER U WITH CIRCUMFLEX
1EC5	ễ	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
006F	o	LATIN SMALL LETTER O
1ED7	ỗ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
1EF9	ỹ	LATIN SMALL LETTER Y WITH TILDE
011D	ĝ	LATIN SMALL LETTER G WITH CIRCUMFLEX
015D	ŝ	LATIN SMALL LETTER S WITH CIRCUMFLEX
0061	a	LATIN SMALL LETTER A
0065	e	LATIN SMALL LETTER E
0109	ĉ	LATIN SMALL LETTER C WITH CIRCUMFLEX
0125	ĥ	LATIN SMALL LETTER H WITH CIRCUMFLEX
0129	ĩ	LATIN SMALL LETTER I WITH TILDE
0135	ĵ	LATIN SMALL LETTER J WITH CIRCUMFLEX
0169	ũ	LATIN SMALL LETTER U WITH TILDE
0175	ŵ	LATIN SMALL LETTER W WITH CIRCUMFLEX
0177	ŷ	LATIN SMALL LETTER Y WITH CIRCUMFLEX
0067 + 0303	g̃	LATIN SMALL LETTER G + COMBINING TILDE
0072 + 0303	r̃	LATIN SMALL LETTER R WITH COMBINING TILDE
0268 + 0303	ɨ̃	LATIN SMALL LETTER I WITH STROKE + COMBINING TILDE
0289 + 0303	ʉ̃	LATIN SMALL LETTER U BAR + COMBINING TILDE

Sequence aa̋a (0061 1EAB 0061) compared using Google Fonts in https://wordmark.it/:

Findings:
Stacking diacritics are always in place

Sequence eə̃e (0065 1EC5 0065) compared using Google Fonts in https://wordmark.it/:

Findings:
Stacking diacritics are always in place

Sequence oo͂o (006F 1EC5 006F) compared using Google Fonts in https://wordmark.it/:

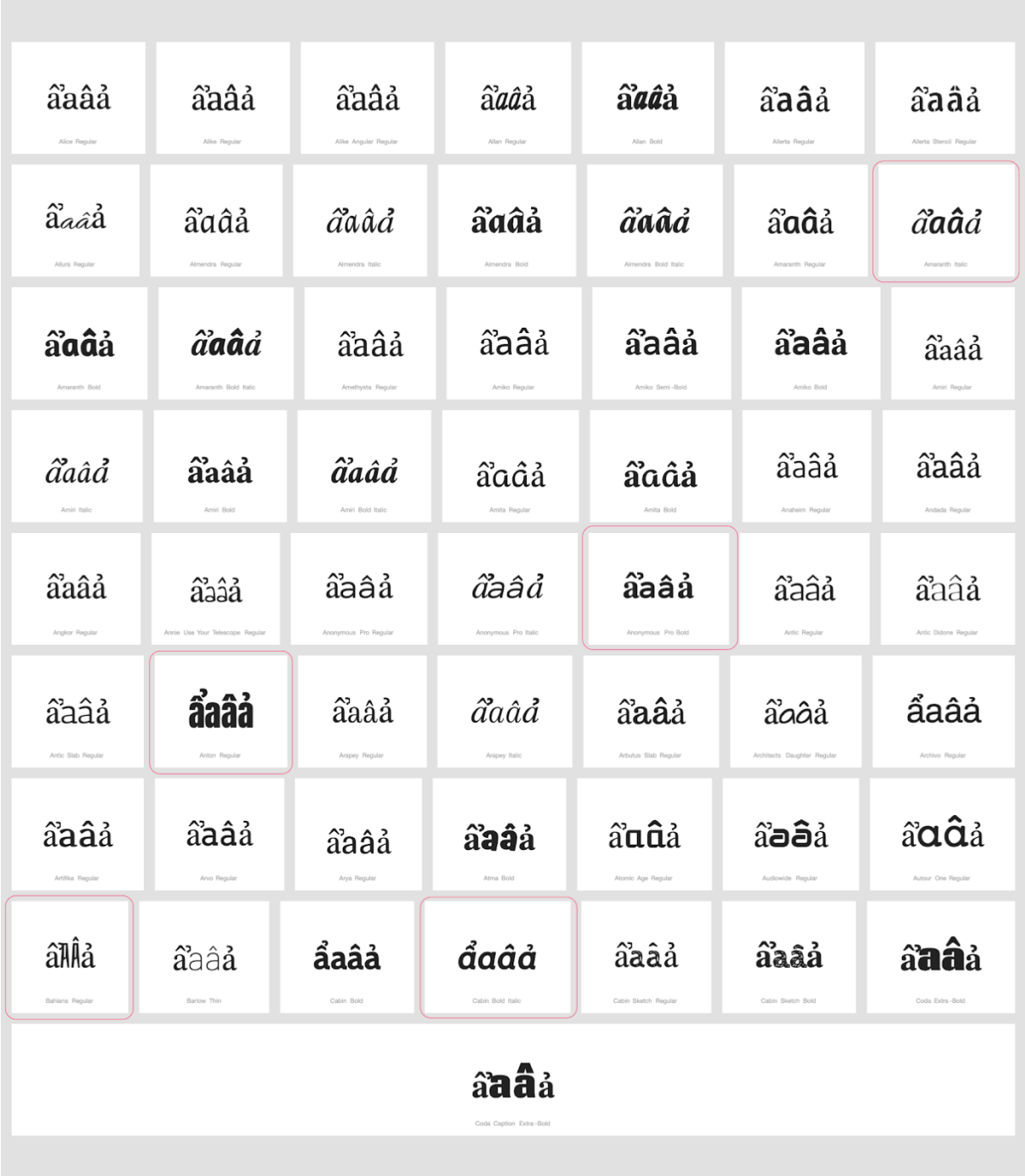
Findings:
Stacking diacritics are always in place

D.4.2 Circumflex and Hook Above

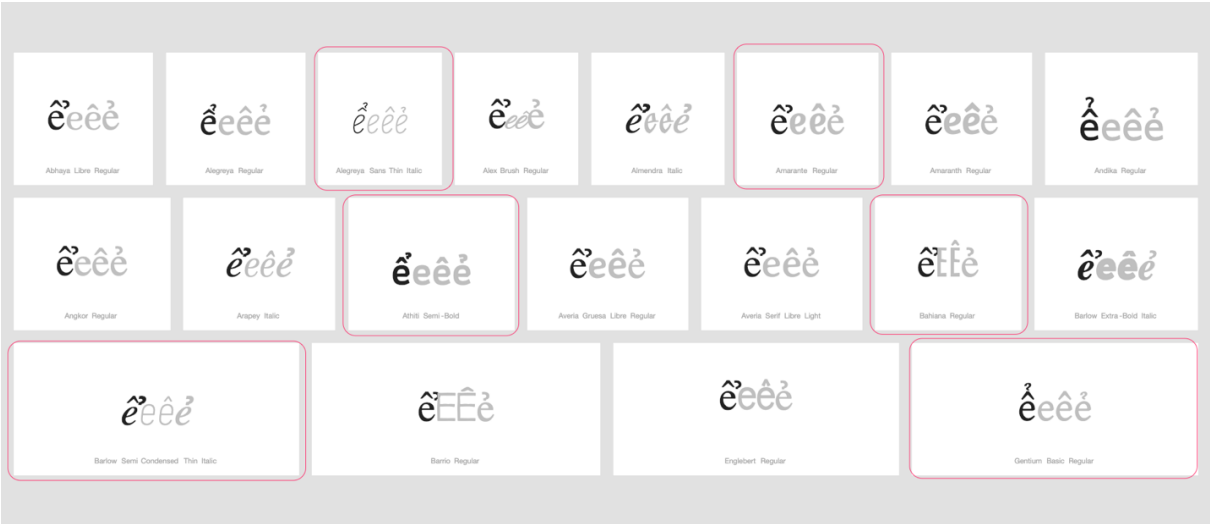
Code Points Considered:

Code Points	Glyph	Name
1EA9	ẩ	Latin Small Letter A With Circumflex And Hook Above
00E2	â	Latin Small Letter A With Circumflex
1EA3	ả	Latin Small Letter A With Hook Above
1EC3	ể	Latin Small Letter E With Circumflex And Hook Above
00EA	ê	Latin Small Letter E With Circumflex
1EBB	ẻ	Latin Small Letter E With Hook Above
1ED5	ổ	Latin Small Letter O With Circumflex And Hook Above
00F4	ô	Latin Small Letter O With Circumflex
1ECF	ỏ	Latin Small Letter O With Hook Above

Sequence ẩaâả (1EA9 + 0061 + 00E2 + 1EA3) compared using Google Fonts in https://wordmark.it/:



Sequence ểeêẻ (1EC3 + 0065 + 00EA + 1EBB) compared using Google Fonts in https://wordmark.it/:



Sequence ổoôỏ (1ED5 + 006F + 00F4 + 1ECF) compared using Google Fonts in https://wordmark.it/:

Findings:
In a large number of fonts, the two letters are consistently different. However, in a significant number of fonts, renderings are very diverse. In some case the hook as secondary modifier is placed vertically above, in others it is set horizontally next to the circumflex as primary modifier, in some fonts it is spaced so far horizontally to the right that it becomes unclear if it is a modifier belonging to the first or the second code point, and yet in other cases it even overlaps with the glyph of the following code point.

Conclusion:
Suggestion to add to shortlist for the string similarity list or create three variant pairs on the ground of them being visually similar to the level of being nearly identical or confusable.
ẩ 1EA9 and âả 00E2 + 1EA3
ể 1EC3 and êẻ 00EA + 1EBB
ổ 1ED5 and ôỏ 00F4 + 1ECF

D.4.3 Breve + Grave above
Code Points Considered:

1EB1	ắ	LATIN SMALL LETTER A WITH BREVE AND GRAVE

Sequence aắa (0061 1EB1 0061) compared using Google Fonts in https://wordmark.it/ :


aăà
Noto Sans HK Thin


aăà
Noto Sans HK Light


aăà
Noto Sans HK Regular


aăà
Noto Sans HK Medium


aăà
Noto Sans HK Bold


aăà
Noto Sans HK Black


aǎa
Quicksand Regular


aǎa
Roboto Thin Italic

Findings:
Stacking diacritics are in place in most cases
One font namely Noto Sans HK has an error in design, or there are some errors in wordmark.it software: on the screen diacritics are not positioned properly, in .png downloaded from wordmark.it diacritics are positioned properly, in .pdf presentation of the same web page diacritics are not positioned properly

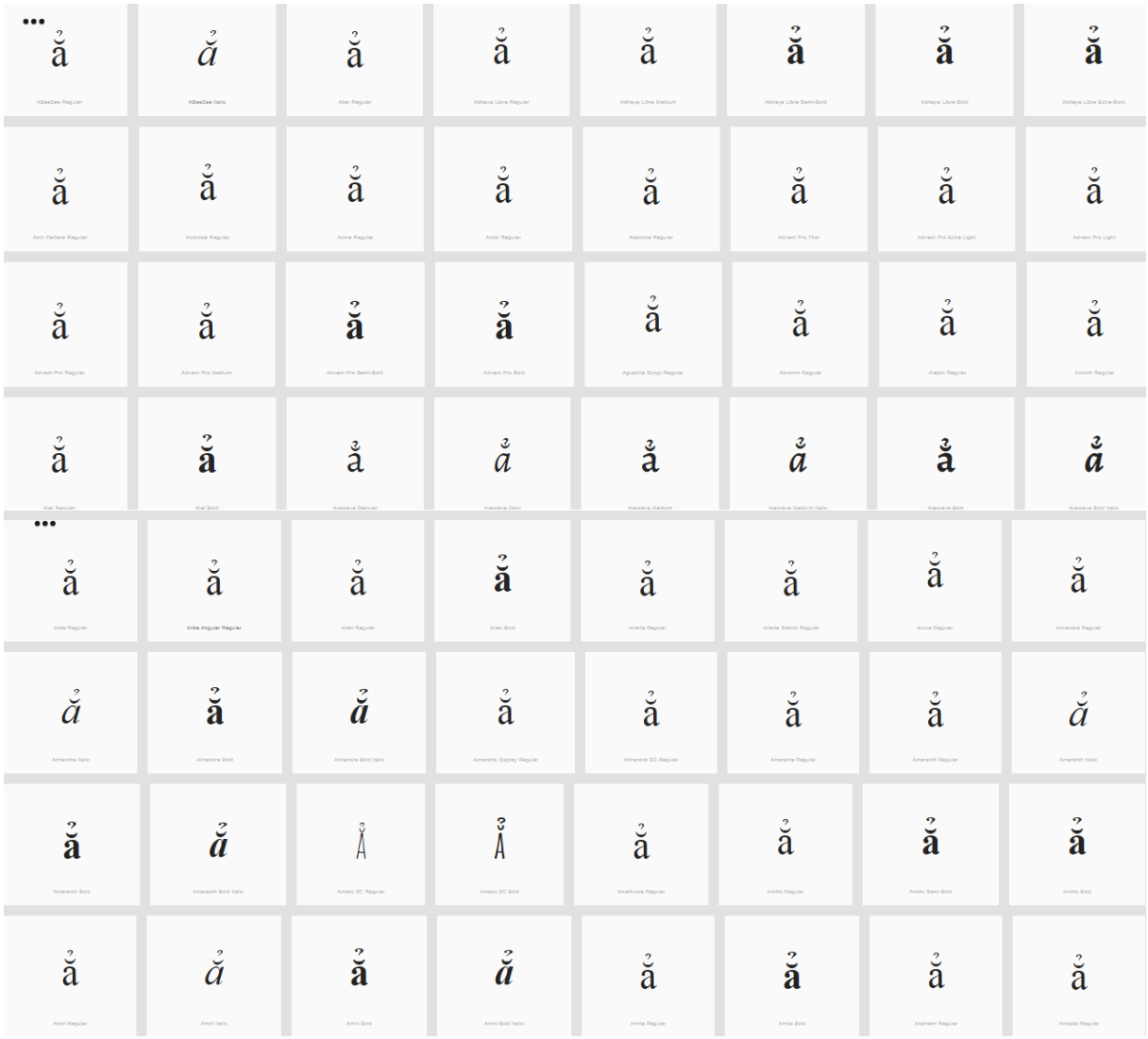
Conclusion:
Stacking diacritics are almost always in place

D.4.4 Breve and Hook Above

Code Points Considered:

Code Points	Glyph	Name
1EB5	ẵ	LATIN SMALL LETTER A WITH BREVE AND TILDE

Sequence (ẵ) (1EB3) compared using Google Fonts in https://wordmark.it/:



Findings:
Stacking diacritics are always in place

D.4.5 Breve and Tilde

Code Points Considered:

Code Points	Glyph	Name
1EB5	ắ	Latin Small Letter A With Breve And Tilde

Sequence ắ (1EB5) compared using Google Fonts in https://wordmark.it/:

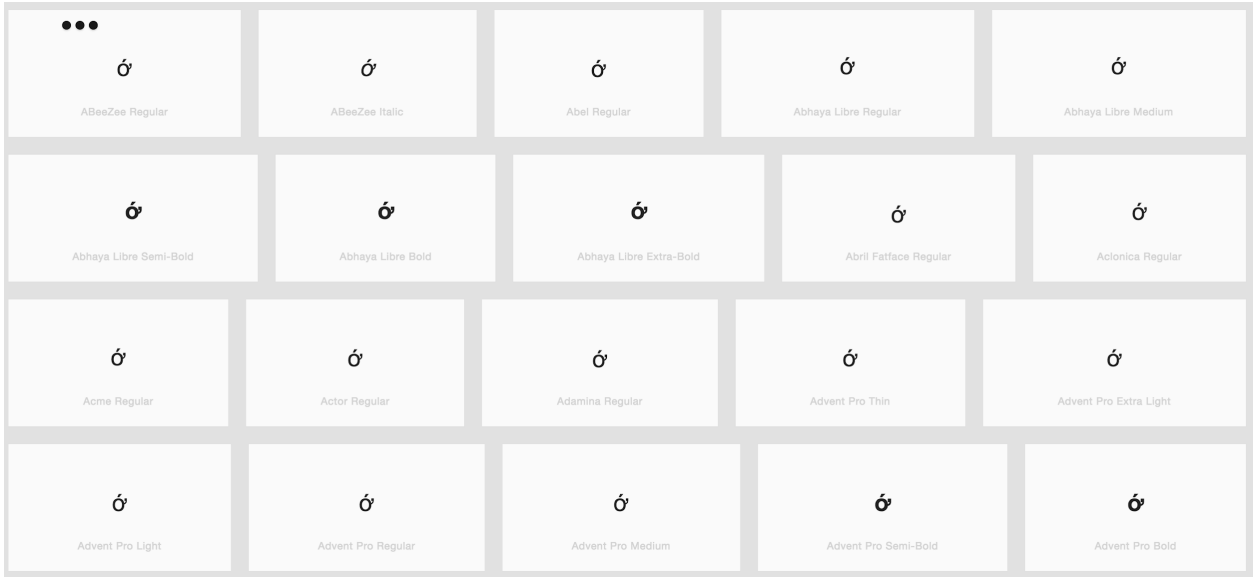
Findings:
The double diacritics stay at the base character and thus will not be confused with characters next to it having just one of the diacritics.

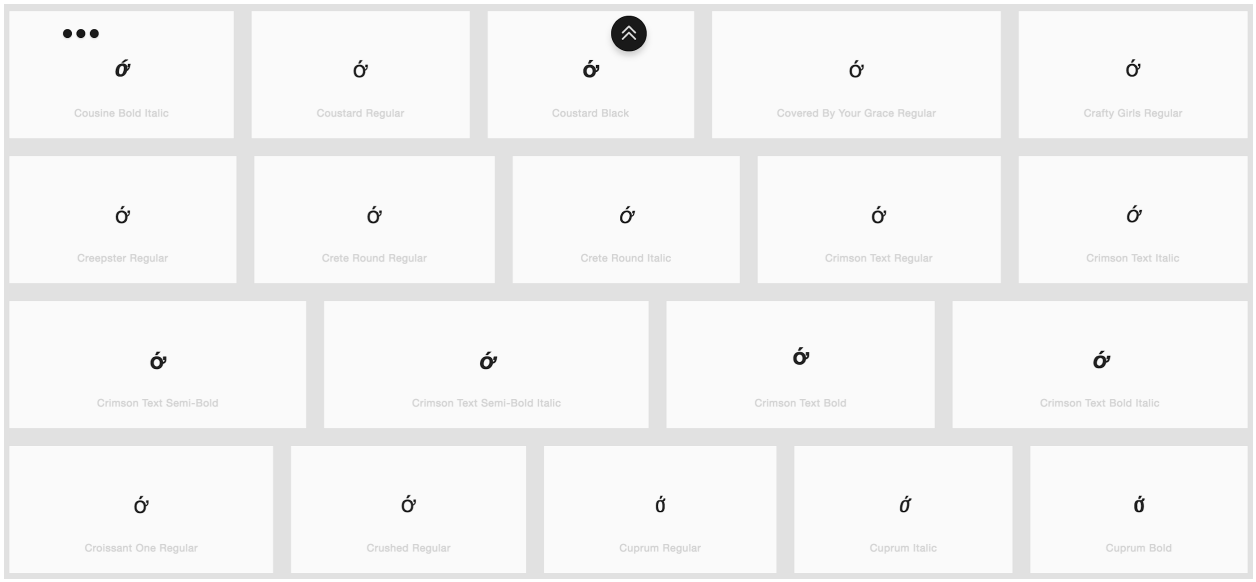
D.4.6 Horn and Acute

Code Points Considered:

1EDB	ớ	LATIN SMALL LETTER O WITH HORN AND ACUTE
1EE9	ứ	LATIN SMALL LETTER U WITH HORN AND ACUTE

Sequence ớ (1EDB) compared using Google Fonts in https://wordmark.it/:





Sequence ứ (1EE9) compared using Google Fonts in https://wordmark.it/:

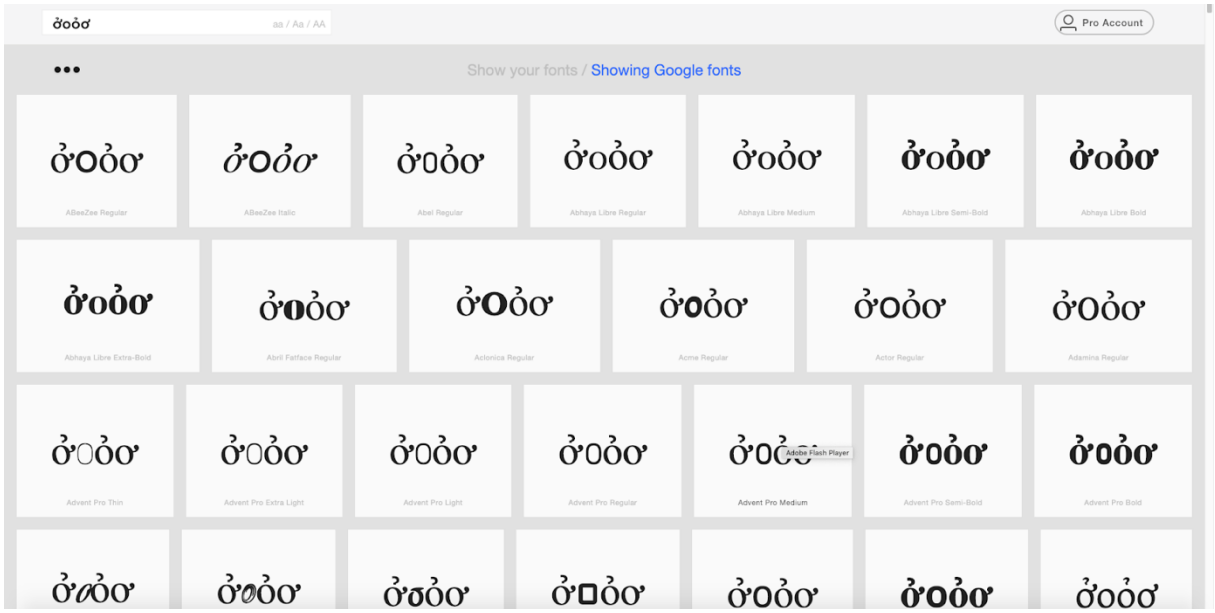
Finding: Diacritics are rendered in a consistent manner

D.4.7 Horn and Hook Above

Code Points Considered:

Code Points	Glyph	Name
1EDF	ở	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
1ECF	ỏ	LATIN SMALL LETTER O WITH HOOK ABOVE
01A1	ơ	LATIN SMALL LETTER O WITH HORN
1EED	ử	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
1EE7	ủ	LATIN SMALL LETTER U WITH HOOK ABOVE
01B0	ư	LATIN SMALL LETTER U WITH HORN

Sequence ởoỏơ (1EDF + 006F + 1ECF + 01A1) compared using Google Fonts in https://wordmark.it/:



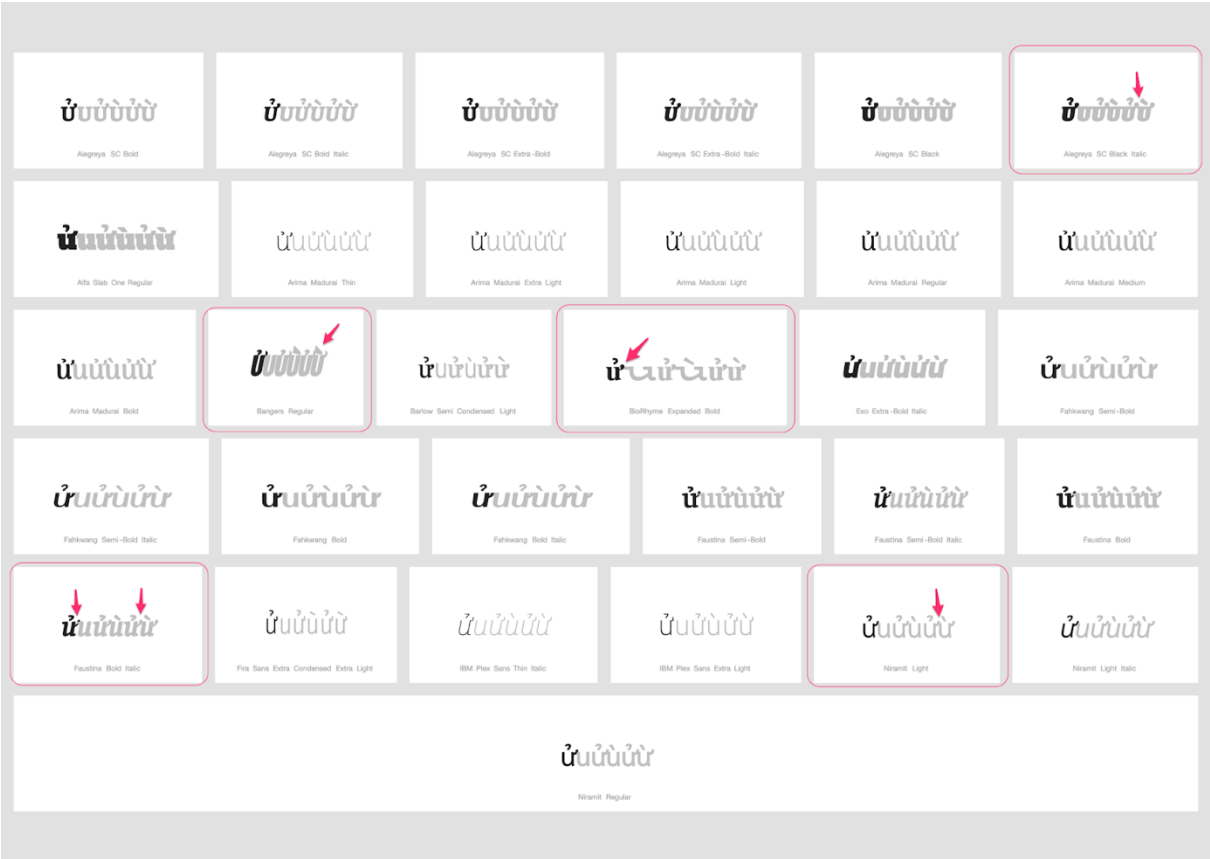
Sequence ửuủư (1EED + 0075+ 1EE7+ 01B0) compared using Google Fonts in https://wordmark.it/:

Findings:
In the case of 1EDF, renderings are considerably homogenous and clearly discernible from adjacent glyphs.

In the case of 1EED however, renderings are rather heterogeneous and there is a significant number of fonts in which it is not clear whether the modifying hook is a modifier of 1EED, a ligature between 1EED and the following code point, or a left hand-side modifier of a subsequently following code point to the right.

Therefore, additional analysis is warranted of a sequence of 1EED followed by u-shape based Code Points featuring a left-hand side modifier, i.e. 00F9 (ù LATIN SMALL LETTER U WITH GRAVE) and 1EEB (ừ LATIN SMALL LETTER U WITH HORN AND GRAVE), which was conducted as demonstrated below:

Sequence ửuửùửừ (1EED + 0075 + 1EED + 00F9 + 1EED + 1EEB) compared using Google Fonts in https://wordmark.it/:



Additional *Findings:*
In some fonts, it remains unclear whether the right-hand side hook of 1EED belongs to that glyph or the code point following to the right. Given however two facts, namely that no code point exists in with a left-hand modifier similar enough, and that these Code Points are used only in a minority of language communities, the readers of which should be attuned to such differences, this would not seem to cross the threshold to constitute a variant. It may however be advisable to pay attention to these inconsistencies in a string-similarity review before admission to the root zone.

Conclusion:
Highlight the inconsistencies of the rendering of 1EED in the string-similarity shortlist.
If a u-based shape with a left-hand side modifier is suggested for a future revision of the LGR, particular attention needs to be paid to that code point in sequence with 1EED.

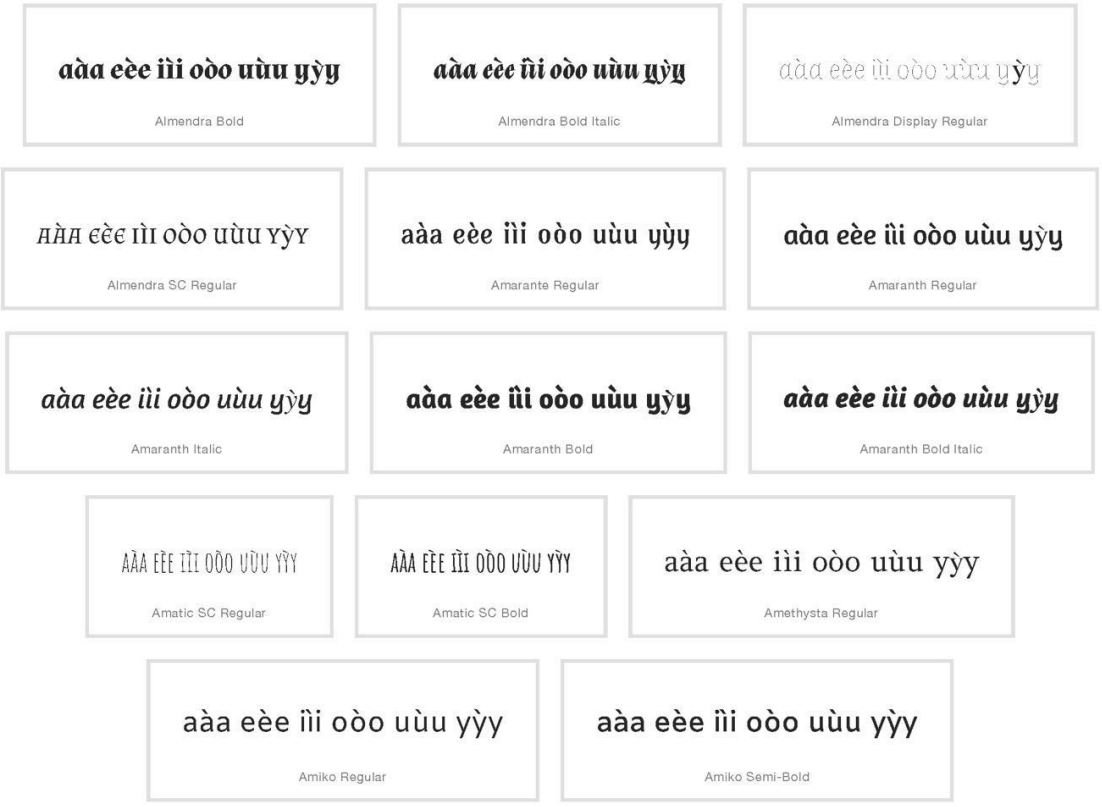
D.4.8 Diacritic Grave

Code Points Considered:

Code Points	Glyph	Name
0061	a	LATIN SMALL LETTER A
0065	e	LATIN SMALL LETTER E
0069	i	LATIN SMALL LETTER I
006F	o	LATIN SMALL LETTER E
0075	u	LATIN SMALL LETTER U
0079	y	LATIN SMALL LETTER Y
00E0	à	LATIN SMALL LETTER A WITH GRAVE
00E8	è	LATIN SMALL LETTER E WITH GRAVE

00EC	ì	LATIN SMALL LETTER I WITH GRAVE
00F2	ò	LATIN SMALL LETTER O WITH GRAVE
00F9	ù	LATIN SMALL LETTER U WITH GRAVE
1EF3	ỳ	LATIN SMALL LETTER Y WITH GRAVE

Sequence aàa, eèe, iìi, oòo, uùu, and yỳy (0061 00E0 0061, 0065 00E8 0065, 0069 00EC 0069, 006F 00F2 006F, 0075 00F9 0075, and 0079 1EF3 0079) compared using Google Fonts in https://wordmark.it/:



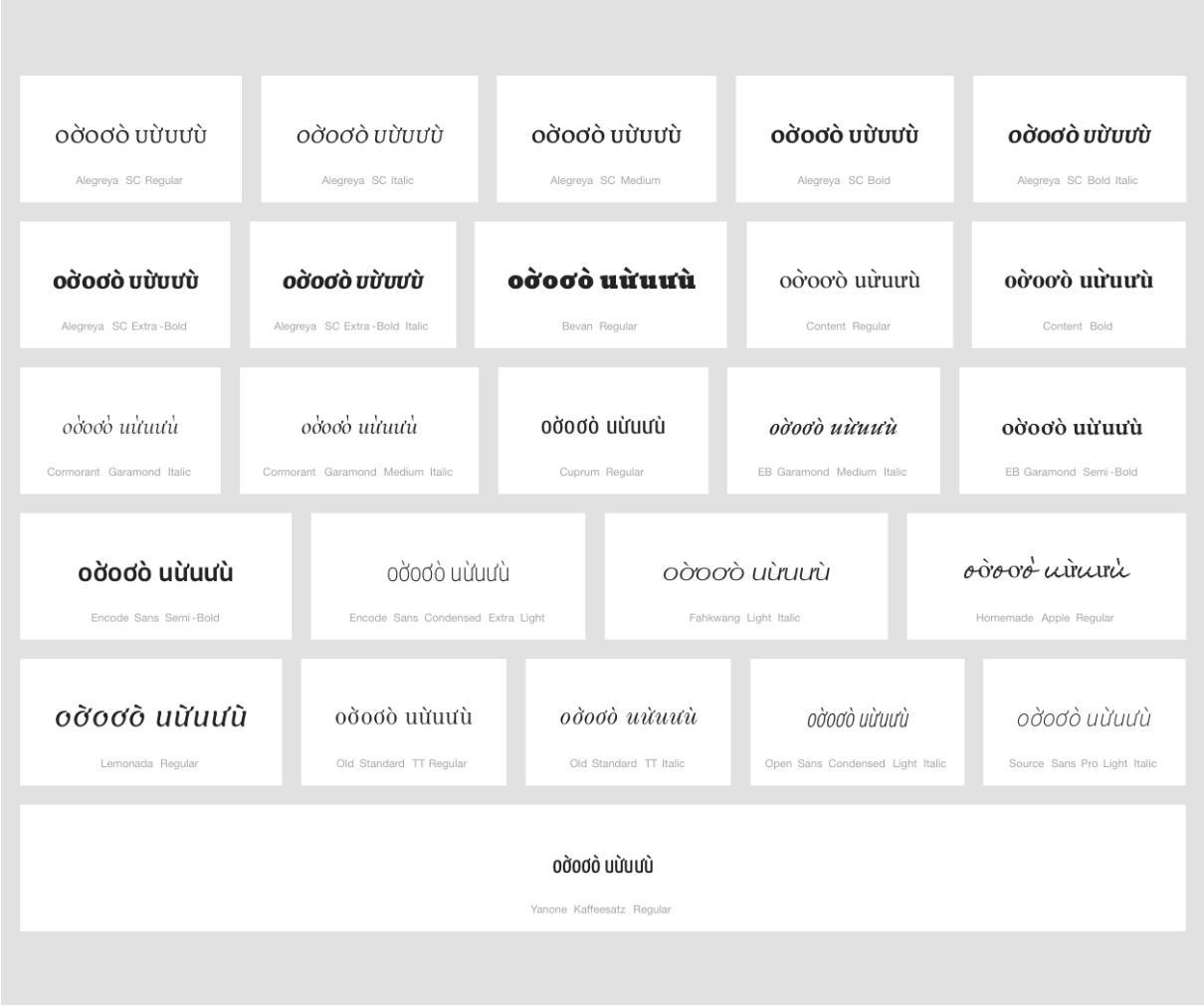
Findings:
Diacritics are always in place

D.4.16 Diacritics Horn And Grave

Code Points Considered:

Code Points	Glyph	Name
006F	o	LATIN SMALL LETTER E
0075	u	LATIN SMALL LETTER U
00F2	ò	LATIN SMALL LETTER O WITH GRAVE
00F9	ù	LATIN SMALL LETTER U WITH GRAVE
01A1	ơ	LATIN SMALL LETTER O WITH HORN
01B0	ư	LATIN SMALL LETTER U WITH HORN
1EDD	ờ	LATIN SMALL LETTER O WITH HORN AND GRAVE
1EEB	ừ	LATIN SMALL LETTER U WITH HORN AND GRAVE

Sequence ờoơò and ừuưù (1EDD 006F 01A1 00F2 and 1EEB 0075 01B0 00F9) compared using Google Fonts in https://wordmark.it/:



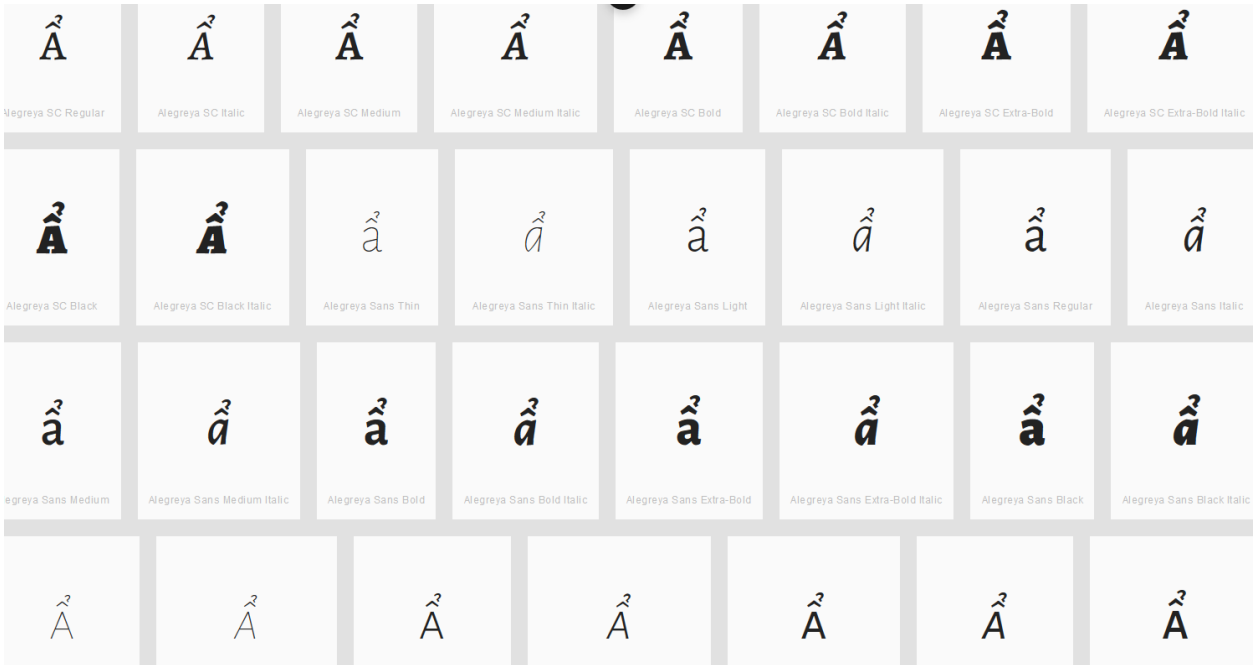
Findings:
Diacritics are always in place
Additional *Findings:*
In some fonts, especially in letter "u" case, it seems that horn belongs to the next character. There is no character with horn to the left in Repertoire.

D.4.17 Circumflex And Hook Above

Code Points Considered:

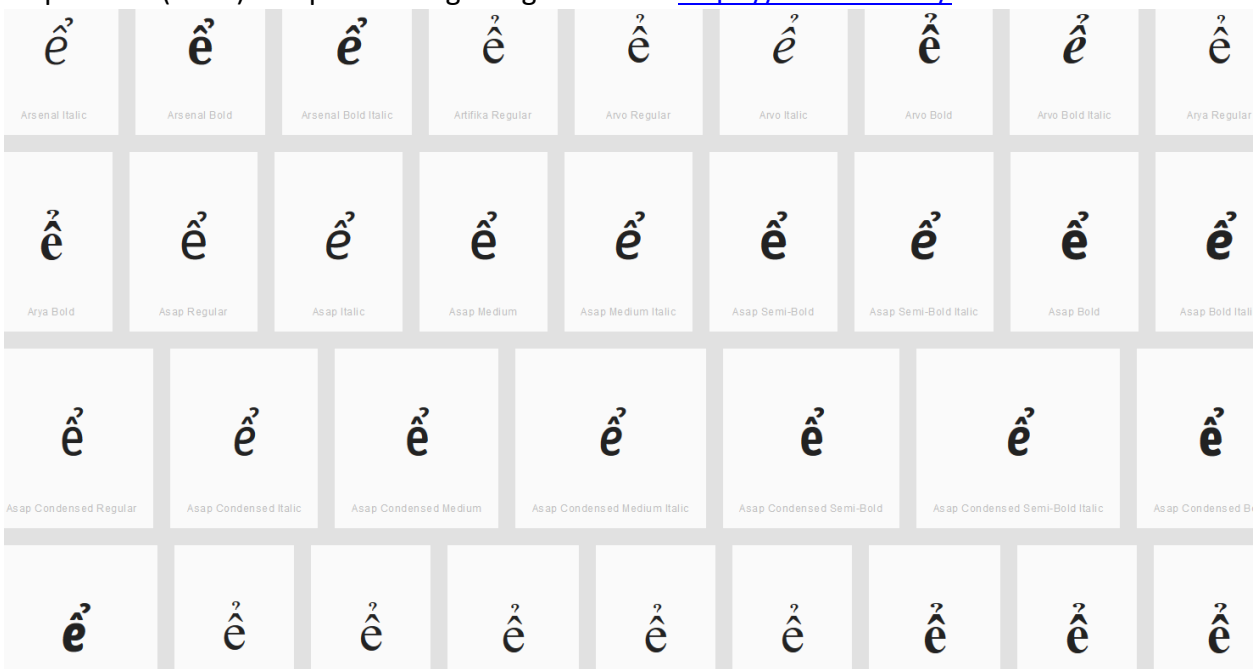
Code Points	Glyph	Name
1EA9	ẩ	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE
1EC3	ể	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
1ED5	ổ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE

Sequence ẩ (1EA9) compared using Google Fonts in https://wordmark.it/:



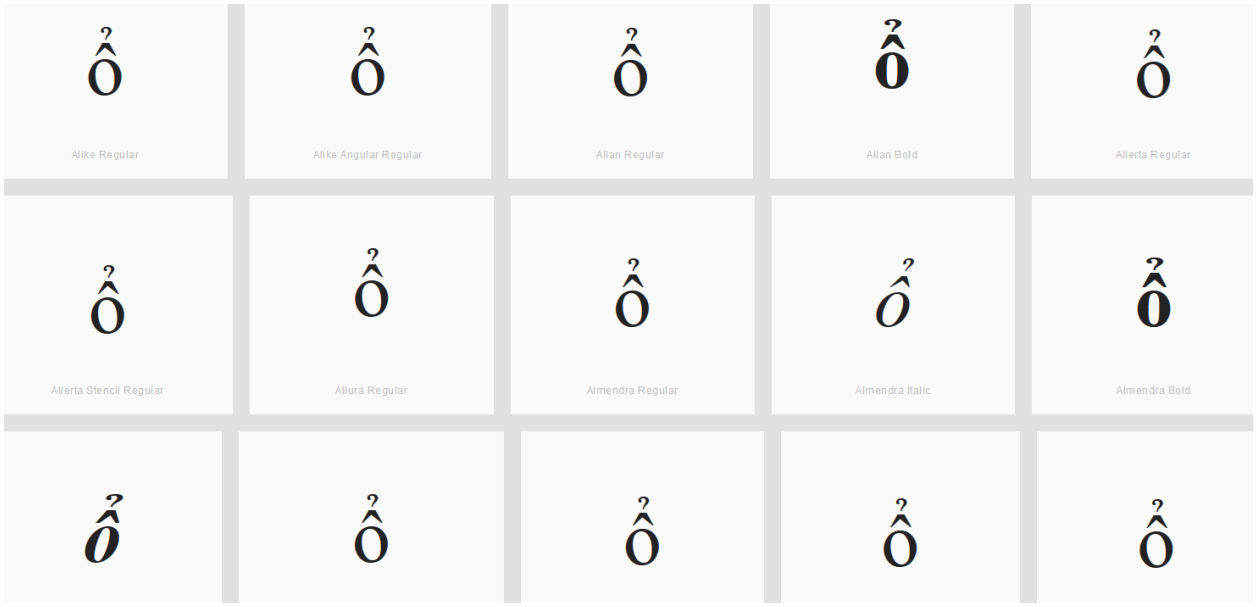
Findings:
The double diacritics stay at the base character and thus will not be confused with characters next to it having just one of the diacritics.

Sequence ể (1EC3) compared using Google Fonts in https://wordmark.it/:



Findings:
The double diacritics stay at the base character and thus will not be confused with characters next to it having just one of the diacritics.

Sequence ổ (1ED5) compared using Google Fonts in https://wordmark.it/:



Findings:
The double diacritics stay at the base character and thus will not be confused with characters next to it having just one of the diacritics.

D.4.9 Circumflex + Dot Below

D.4.10 Breve + Dot Below

D.4.11 Acute + Dot Below

D.4.12 Grave (vs. Non-Grave)

D.4.13 Acute (vs. Non-Acute)

Code Points Considered:

Code Points	Glyph	Name
00E1	á	LATIN SMALL LETTER A WITH ACUTE
00E9	é	LATIN SMALL LETTER E WITH ACUTE
00ED	í	LATIN SMALL LETTER I WITH ACUTE
00F3	ó	LATIN SMALL LETTER O WITH ACUTE
00FA	ú	LATIN SMALL LETTER U WITH ACUTE
00FD	ý	LATIN SMALL LETTER Y WITH ACUTE
0107	ć	LATIN SMALL LETTER C WITH ACUTE
013A	ĺ	LATIN SMALL LETTER L WITH ACUTE
0144	ń	LATIN SMALL LETTER N WITH ACUTE
0155	ŕ	LATIN SMALL LETTER R WITH ACUTE
015B	ś	LATIN SMALL LETTER S WITH ACUTE
017A	ź	LATIN SMALL LETTER Z WITH ACUTE
0061	a	LATIN SMALL LETTER A
0065	e	LATIN SMALL LETTER E

0069	i	LATIN SMALL LETTER I
006F	o	LATIN SMALL LETTER O
0075	u	LATIN SMALL LETTER U
0079	y	LATIN SMALL LETTER Y
0063	c	LATIN SMALL LETTER C
006C	l	LATIN SMALL LETTER L
006E	n	LATIN SMALL LETTER N
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
007A	z	LATIN SMALL LETTER Z

D.4.14 Stacking in Courier New (And Perhaps Other Fonts)

We have seen that, with precomposed Code Points, there is no stacking problem. However, when we have not had a precomposed Code Points available, we have necessarily used combining diacritics. Then, the situation changes. In particular, when using the Courier New font (which is one of our three standard fonts for analysis), there is sometimes a problem. Sometimes, the combining mark simply gets its own space, with the following letter shifter right to make room – which is irritating, but not confusing. However in other cases the combining mark appears to be associated with the following letter.

Code Points Considered:

1EB9 + 0301	ẹ́	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT
1EB9 + 0300	ẹ̀	LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT
0067 + 0303	g̃	LATIN SMALL LETTER G + COMBINING TILDE
0268 + 0303	ɨ̃	LATIN SMALL LETTER I WITH STROKE + COMBINING TILDE
1ECD + 0300	ọ̀	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT
1ECD + 0301	ọ́	LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT
025B + 0331 + 0308	ɛ̱̈	LATIN SMALL LETTER OPEN E + COMBINING DIARESIS + COMBINING MACRON BELOW
025B + 0331	ɛ̱	LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW
0254 + 0331	ɔ̱	LATIN SMALL LETTER OPEN O + COMBINING MACRON BELOW
0072 + 0303	r̃	LATIN SMALL LETTER R + COMBINING TILDE
0289 + 0303	ʉ̃	LATIN SMALL LETTER U WITH BAR + COMBINING TILDE

In each case below, the letter is followed by another letter (or two, in the case of two combining marks. (In each case shown, the letters were simply copied, then the font changed.)

Ariel	Courier New	
a̲a	a_a	Latin Small Letter A + Combining Macron Below
e̲e	e_e	Latin Small Letter E + Combining Macron Below
g̃g	g̃g <=	Latin Small Letter G + Combining Tilde
i̲i	i_i	Latin Small Letter I + Combining Macron Below
m̧m	m̧m	Latin Small Letter M + Combining Cedilla
n̈n	n̈n	Latin Small Letter N + Combining Dieresis
o̧o	o̧o	Latin Small Letter O + Combining Cedilla
o̲o	o_o	Latin Small Letter O + Combining Macron Below
r̃r	r̃r <=	Latin Small Letter R + Combining Tilde
ɔ̈o	ɔ̈o	Latin Small Letter Open O + Combining Dieresis
ɔ̲o	ɔ_o <=	Latin Small Letter Open O + Combining Macron Below
ɛ̈e	ɛ̈e	Latin Small Letter Open E + Combining Dieresis
ɛ̲e	ɛ_e <=	Latin Small Letter Open E + Combining Macron Below
ɛ̲̈ee	ɛ_̈ee <=	Latin Small Letter Open E + Combining Dieresis + Combining Macron Below **
ɨ̃i	ɨ̃i <=	Latin Small Letter I with Stroke + Combining Tilde
ʉ̃u	ʉ̃u <=	Latin Small Letter U with Bar + Combining Tilde
ẹ̀e	ẹ̀e <=	Latin Small Letter E with Dot Below + Combining Grave Accent
ẹ́e	ẹ́e <=	Latin Small Letter E with Dot Below + Combining Acute Accent
ọ̀o	ọ̀o <=	Latin Small Letter O with Dot Below + Combining Grave Accent
ọ́o	ọ́o <=	Latin Small Letter O with Dot Below + Combining Acute Accent
ộo	ộo	Latin Small Letter O with Circumflex + Combining Dot Below
ộo	ộo	Latin Small Letter O with Dot Below + Combining Circumflex
ợo	ợo	Latin Small Letter O with Horn + Combining Dot Below
ợo	ợo	Latin Small Letter O with Dot Below + Combining Horn **

Findings:
With each of these cases, error is a certainty. The ideal solution, of course, would be for the Unicode Consortium to create new pre-composed Code Points for these problem cases. But I suspect there is little chance of them doing so before our report is due. So we will have to figure out an alternate approach to recommend.

D.5 IDNA 2003 Compatibility

D.5.1 LATIN SMALL LETTER SHARP S (ß) 00DF

IDNA2003 Versus IDNA2008
One of the differences between IDNA2008 and IDNA2003 is the treatment of four characters, one of which is relevant to the Latin Script LGR, the Latin Small Letter Sharp S or 00DF. Despite the fact

IDNA2008 superseded IDNA2003, some applications continued to apply the character mapping from IDNA2003 resulting in DNS lookup queries that look like the following:

Table D.1. DNS resolution comparison for Sharp S (00DF)

Char	Example	IDNA2003 Result	IDNA2008 Result
ß 00DF	href="**http://faß.de"**	**http://faß.de → http://fass.de**	**http://faß.de → http://xn--fa-hia.de**

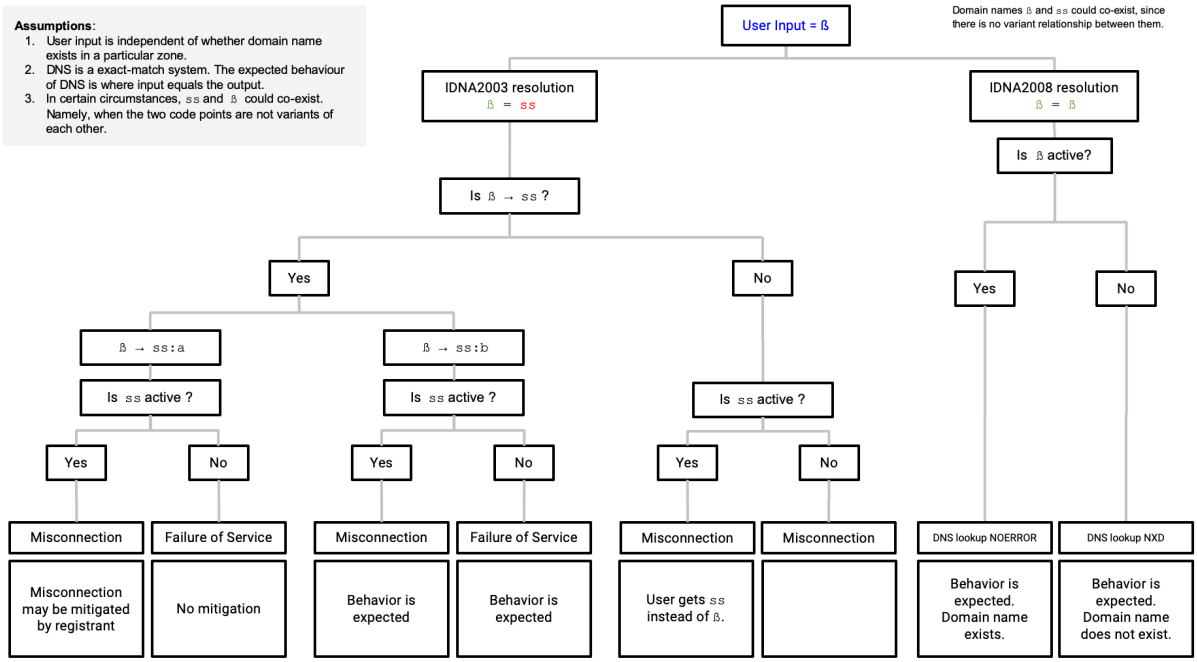
Source: https://unicode.org/reports/tr46/#Transition_Considerations

The difference in application behavior relative to DNS labels containing the code point 00DF causes two types of problems:

1. **Failure of service**. The user intends to navigate to “example.faß” but the application sends the user to “example.fass” which doesn’t exist, because the domain name is not registered or is blocked or withheld.
2. **Misconnection**. The user intends to navigate to “example.faß” but the browser returns “example.fass” which is controlled by a different registrant.

The situation is summarized in Diagram D.1 below:

Diagram D.1: Resolution of LATIN SMALL LETTER SHARP S (ß) 00DF in Different Enviroments



Internet Browser Support

As of the writing of this proposal, certain Internet browsers process 00DF using the IDNA2003 mapping mechanism instead of doing the IDNA2008 conversion. A test with the four major Internet browsers shows that Google Chrome and Microsoft Edge have not fully implemented IDNA2008; they still are in

what is called “transitional mode”. For more information about IDNA2008 transitional mode, see Unicode Technical Standard #46 at https://unicode.org/reports/tr46/.

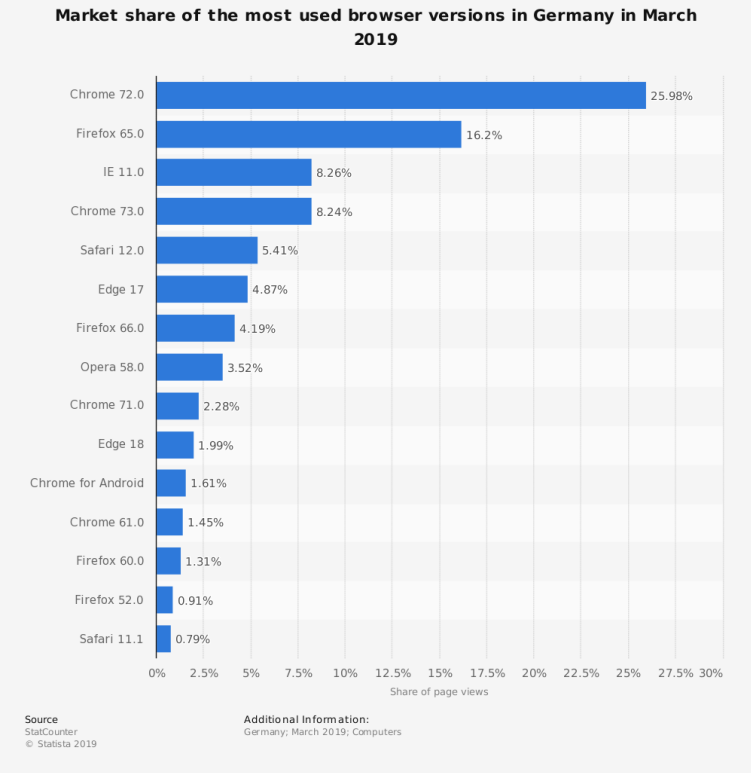
Table D.2. Resolution of **http://faß.de** by Different Internet Browsers

Internet Browser	http://faß.de resolves to
Microsoft Edge/Explorer	**http://fass.de**
Apple Safari	**http://xn--fa-hia.de**
Firefox	**http://xn--fa-hia.de**
Google Chrome	**http://fass.de**

The trend of browser implementation seems to be towards full IDNA2008 compliance (given that Apple Safari and Firefox did migrate from IDNA2003 to IDNA2008). However, it is not clear how soon or late Google Chrome or Microsoft Edge will fully transition to IDNA2008. See for example, https://bugs.chromium.org/p/chromium/issues/detail?id=941691

As of March 2019, Chrome has the largest browser market share in Germany, which suggests an important part of the end-user population is exposed to the problem with DNS lookups when utilizing the non-IDNA2008-conforming browsers when the label contains code point 00DF.

Diagram D.2: Market Share of the Most Used Browser Versions in Germany in March 2019



Registry Implementation at the Second Level

Latin GP sought the input of TLD registries serving the German-speaking communities, namely DENIC (www.denic.de), NIC.AT (www.nic.at), and SWITCH (www.nic.ch) to inform Latin GP's solution regarding the IDNA2003 compatibility issue.

At the second level, the .DE registry (DENIC) offers 00DF as a separate, stand-alone code point[10]; in consequence these hypothetical domain names "straße.de" and "strasse.de" would be offered for registration as two separate domains[11]. The .CH registry (SWITCH) and the .AT registry (nic.at) do not offer 00DF in their repertoires for the second level per their published policies[12] [13].

Input from the German User Community

The GP has sought input from experts of the three major German-speaking ccTLDs (namely Denic, nic.at, and switch, for Germany, Austria, and Switzerland, respectively) on the topic of whether ß and ss should be considered variants. After some discussions, these experts found the following consensus solution, which they suggested to the GP for use at LGR level:

Table D.3 Solution Suggested by the German User Community

Group	ß vs ss			
Target	Source			Rationale

[10] DENIC Domain Name Guidelines: https://www.denic.de/fileadmin/public/documents/DENIC_Domainrichtlinien_EN.pdf
[11] https://www.denic.de/en/know-how/idn-domains/
[12] SWITCH IDN Policy: https://www.nic.ch/faqs/idn/
[13] NIC.AT Repertoire: https://www.nic.at/media/files/pdf/IDN-Zeichentabelle.pdf

Code Point	Glyph	Name	Code Point	Glyph	Name	Variant Candidate [Yes/No]	Disposition [Allocatable/ Blocked]	
00DF	ß	LATIN SMALL LETTER SHARP S	0073 + 0073	ss	LATIN SMALL LETTER S + LATIN SMALL LETTER S	YES	Blocked	See Section 6.7.2
0073 + 0073	ss	LATIN SMALL LETTER S + LATIN SMALL LETTER S	00DF	ß	LATIN SMALL LETTER SHARP S	YES	Allocatable	See Section 6.7.2

The experts from the German-speaking ccTLD of German users suggested two main reasons for creating this variant relation:

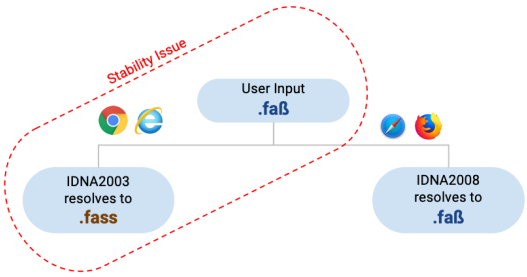
1. There are still browsers (e.g. Chrome) that apply IDNA2003 at the time of writing. Users of such browsers have each ß automatically replaced by a sequence of two s.
2. Swiss users do not use ß and consider it as equivalent to ss, even where they are able to recognize and point out the differences, when pressed to do so. By consequence, a Swiss user would e.g. very likely rewrite an IDN as .strasse even where it had been presented to the same user .straße before. Therefore, a variant relationship is warranted on non-visual grounds.

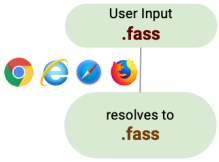
For the variant disposition, the same experts were of the opinion that ß needs to be allocatable towards ss, since the same transformation is done by IDNA2003 and since the same is a long-standing and widely-applied orthographic solution by the German-language community also outside of IDNs, considered valid by all users, especially in the context of domain names. For the other direction, however, the experts were of the opinion ,that the disposition should be blocked since there are many non-German words having a double ss (e.g., cross, process, discussion) for which the same label with ß makes no sense (e.g., croß, proceß, discußion), which would lead to the generation of too many invalid variants otherwise.

Possible Solutions to Address the IDNA2003 Compatibility Issue for LATIN SMALL LETTER SHARP S (ß) 00DF: Pros and Cons

Based on the evidence presented, the GP tried to weigh different solutions to address the IDNA 2003 Compatibility issues, which are summarized in Diagram D.3:

Diagram D.3: General Factors to Resolving the IDNA2003 Compatibility Issue in the Case of LATIN SMALL LETTER SHARP S (ß) 00DF





	Option 1	Option 2	Option 3	Option 4
	Exclude ß	**Include ß with allocatable variant to "ss" (ß → ss: a)**	**Include ß with blocked variant to "ss" (ß → ss: b)**	**Include ß without variant to "ss"**
Eliminates stability issue	🟢	🟠	🟠	🔴
Addresses failure of service	🟢	🟠	🔴	🔴
Addresses misconnection	🟢	🟢	🟢	🔴

The pros and cons for each solution are presented in more detail in the following tables:

Table D.3. Solution to Exclude 00DF from the Latin script repertoire

Pros	Cons
• Most conservative option; removes the option to have DNS labels with code point 00DF. The possibility of landing at the "wrong" website is greatly diminished because there would be only one version of the website (i.e. the one using 'ss' (0073 0073)).	• Misconnection or failure of service is still possible when using Chrome or Edge (albeit only one domain name would actually exists) because user input is independent of whether a domain name exists or not. • Code point 00DF is used in the orthography of German as written in Germany and Austria (but not in Switzerland). German is an EGIDS level 1 language. • It would restrict the freedom of expression for the German-speaking part of the user community, due to the lack of 00DF in the LGR

Table D.3. Solution to Include 00DF with variant relationship with 'ss' (ß → ss)

Pros	Cons

• The possibility of landing at the "wrong" website is diminished provided the two versions of domain names are controlled by the same entity. • Enables freedom of expression for the German-speaking part of the user community; code point 00DF is used in the orthography of German as written in Germany and Austria (but not in Switzerland). German is an EGIDS level 1 language.	• Limits registration choices. • Due to transitivity there will be a variant relationship ß (Latin Sharp S, 00DF) → 'ss' → β (Greek Beta, 03B2), therefore imposing a cross-script variant on the Greek script LGR. • Failure of service or misconnection may occur depending on application's implementation (IDNA2003 or IDNA2008 + TR46).

Table D.4. Solution for Disposition: Allocatable versus Blocked ß → ss

2.1 ß → ss: Allocatable	**2.2 ß → ss:** Blocked
• It would be possible for a registry operator to apply for the variant label. Per the latest IDN variant TLD Management Framework recommendation, each TLD variant should be evaluated and processed as a stand alone TLD (i.e. separate application fee, evaluation process, etc.) • If registry operator does not apply for the variant label, the label will remain reserved for said registry operator. • Misconnection cannot occur but failure of service can.	• With a "blocked" disposition, the variant label would remain withheld from registration by any registry operator. • Misconnection cannot occur but failure of service can.

Table D.5. Solution for Disposition: Allocatable versus Blocked ss → ß

2.3 ss → ß: Allocatable	**2.4 ss → ß:** Blocked
• Simpler solution for TLD applicant; the TLD applicant does not need to be concerned about asymmetrical relationship. Can	• Alignment with LGR procedure (i.e. minimize allocatable variants)

apply for the ‘ss’ version first and apply for the 00DF version at a later point in time. ● German-language users do not expect a label which is spelled with double ‘ss’ be represented with a label with letter Sharp S (00DF), the user does expect a label with Sharp S (00DF) to sometimes be represented with a label with double ‘ss’.	● No linguistic expectations on the side of the users. Most conservative option according to the LGR Procedure ● Denies the opportunity to apply for the 00DF version, if ‘ss’ is registered first.

Table D.6. Solution to Include 00DF without variant relationship with ‘ss’

Pros	Cons
● Option is consistent with implementation by DENIC (German registry); German users have been conditioned to this behavior.	● Failure of service or misconnection may occur depending on the application's implementation (IDNA2003 or IDNA2008 + TR46) with respect to ß. ● Confusing for Swiss people as they generally use ‘ss’ in all cases for Sharp S (00DF).

Conclusion: Inclusion of 00DF with Variant Mechanism

The Latin GP proposes a solution that balances the needs of certain parts of the Latin script community while minimizing security and stability issues introduced by applications outside the DNS. The solution is to include Latin Small Letter Sharp S (00DF) with a variant relationship with the sequence of letters ‘ss’ (0073 0073), as follows:

Table D.7. Final Variant Solution for Latin Small Letter Sharp S (00DF)

Source Code Point	Variant Relationship	Target Code Point	Disposition
00DF Latin Small Letter Sharp S	→	0073 0073 Latin Small Letter S + Latin Small Letter S	Allocatable
0073 0073 Latin Small Letter S + Latin Small Letter S	→	00DF Latin Small Letter Sharp S	Blocked

This LGR solution along with the appropriate policies (i.e. TLD variant labels managed by the same entity, and second level variant labels managed by the same registrant) would not solve the failure of service problems but would mitigate the issues of misconnection.

D.5.2. LATIN SMALL LETTER DOTLESS I (ı) 0131

There are four Latin code points that have special case (upper case/lower case) relationship:

- U+0069 LATIN SMALL LETTER I ("i")
- U+0049 LATIN CAPITAL LETTER I ("I")
- U+0131 LATIN SMALL LETTER DOTLESS I ("ı")
- U+0130 LATIN CAPITAL LETTER I WITH DOT ABOVE ("İ")

In most locales SMALL LETTER I is lower case of CAPITAL LETTER I, and reverse CAPITAL LETTER I (U+0069) is upper case of SMALL LETTER I (U+0069). In those locales, CAPITAL LETTER I (U+0049) is also upper case of SMALL LETTER DOTLESS I. It could be described as in the following chart:

Table D.8. Case Relationships for 0069, 0049, , 0130, and 0131

Character	Process	Resulting Character	Process	Resulting Character
SMALL LETTER I U+0069	up case →	CAPITAL LETTER I U+0049	down case →	SMALL LETTER I U+0069
SMALL LETTER DOTLESS I U+0131	up case →	CAPITAL LETTER I U+0049	down case →	SMALL LETTER I U+0069
CAPITAL LETTER I WITH DOT ABOVE U+0130	down case →	SMALL LETTER I U+0069	up case →	CAPITAL LETTER I U+0049

In two locales, Turkish and Azeri, respectively, the case relationship is different. In those two, SMALL LETTER I and CAPITAL LETTER I WITH DOT ABOVE are in mutual upcase/downcase relationship to each other, as well as SMALL LETTER DOTLESS I and LATIN CAPITAL LETTER I, which could be described as in the following chart:

Table D.9. Case Relationships in Turkish and Azeri Locales

Character	Process	Resulting Character	Process	Resulting Character
SMALL LETTER I	up case →	CAPITAL LETTER I WITH DOT ABOVE	down case →	SMALL LETTER I

SMALL LETTER DOTLESS I	up case →	CAPITAL LETTER I	down case →	SMALL LETTER DOTLESS I

If we look at the repertoire of Latin code points for the root zone, as proposed by the Latin Generation Panel, SMALL LETTER I and SMALL LETTER DOTLESS I are included, whereas the capital letters are excluded. Capital letters are not even valid in IDNA2008, so the question is, is the case relationship described here a problem or even relevant?

Before IDNA2008, there was IDNA2003. Even though IDNA2003 has been replaced by IDNA2008 it is still implemented. For example, the web browser Google Chrome to date remains IDNA2003 compliant but not fully IDNA 2008 compliant. In IDNA2003 there is a pre-process, normalization, of domain names before conversion to punycode. That normalization includes down casing of Latin characters. For ASCII labels there is already an equivalence between upper case and lower case letters. And this is what users, based on decades of experience, expect to happen.

In an IDNA2003-compliant web browser it is expected that "EXÄMPEL" and "EXAMPLE" are equivalent to "exämpel" and "example", respectively. In an IDNA2008 browser "EXAMPLE" must be accepted, but "EXÄMPLE" could be rejected since "Ä" is not valid, but that is not how e.g. Mozilla Firefox and Apple Safari have been designed to handle the problem. They too do down casing before the formal IDNA2008 process.

Even though down casing is not part of the formal IDNA2008 process, one of the IDNA2008 documents, RFC 5894, states that the user interface of an application, before IDNA2008 processing, can do normalization. The down casing in IDNA2008 browsers should probably seen in that light.

It is quite simple that "TÄT" will probably be down cased to "tät" in the browser, but what should the browser do with "TIT"? Depending on the locale that the browser is running in, it may be down cased to either "tit" or "tıt".

The casing, in an application, is expected to go in one direction, from upper case to lower case. When domain names are presented in text, however, it is common that domain names are presented in upper or mixed case. So "ice" might become "Ice" or "İce".

It is quite obvious from the text above that case shift of dotted or dotless I could create erroneous lookup, but the question is how large threat it would be to the users. Since the applications are expected to go from upper case to lower case, when they handle domain names, we should consider a situation where down casing could result in different lower case letters, i.e. when CAPITAL LETTER I is down cased.

With a non-Turkish and non-Azeri locale, a CAPITAL LETTER I in a domain name is either down cased to LATIN SMALL LETTER I (IDN label) or equivalent to LATIN SMALL LETTER I (ASCII label).

With a Turkish or Azeri locale, a CAPITAL LETTER I is expected to be down cased to SMALL LETTER DOTLESS I, but in an ASCII label in a domain name, it is still expected to be equivalent with LATIN SMALL LETTER I, because that is what the DNS standards says.

There is an obvious risk that, in a Turkish or Azeri locale that the two letters are confused or mistreated due to the case folding, and this confusion could be misused. To be on the safe side LATIN SMALL LETTER I and SMALL LETTER DOTLESS I should be variants. Accordingly, the following variant set could be the optimal solution:

Table D.10. Possible Variant Relationships for 0069 and 0131

Group	Dotless i vs. i							
Target			Source			Variant Candidate [Yes/No]	Disposition [Allocatable/ Blocked]	Rationale
Code Point	Glyph	Name	Code Point	Glyph	Name			
0069	i	LATIN SMALL LETTER I	0131	ı	LATIN SMALL LETTER DOTLESS I	YES	Blocked	Risk of confusion due to inconsistent case folding
0131	ı	LATIN SMALL LETTER DOTLESS I	0069	i	LATIN SMALL LETTER I	YES	Blocked	Risk of confusion due to inconsistent case folding

D.6 Underlining Evaluation Process

Because it is common for domain names to be presented as underlined by applications making use or representing IDNs, we evaluated those code points which included diacritics below the line and those which extend below the line. Code points were again displayed in the same three common fonts used for cross-script variants analysis, i.e. Arial, Courier New, and Times New Roman. Each pair was then evaluated by two members of the GP; if they agreed that the pair were variants, in any of the fonts, that finding was adopted. When there was disagreement, the pairs were evaluated by each of the members of the GP, and the median finding was adopted.

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter A with Circumflex	00E2	â **â** â	â **â** â	1EAD	Latin Small Letter A with Circumflex and Dot Below	Variant due to underlining
Latin Small Letter A	0061	a **a** a	a **a** a	0061 + 0331	Latin Small Letter A + Combining Macron Below	Variant due to underlining
Latin Small Letter A	0061	a **a** a	a **a** a	0105	Latin Small Letter A with Ogonek	Variant due to underlining
Latin Small Letter A	0061	a **a** a	a **a** a	1EA1	Latin Small Letter A with Dot Below	Variant due to underlining
Latin Small Letter A with Breve	0103	ă **ă** ă	ă **ă** ă	1EA7	Latin Small Letter A with Breve and Dot Below	Variant due to underlining
Latin Small Letter B	0062	b **b** b	b **b** b	00FE	Latin Small Letter Thorn	Variant due to underlining
Latin Small Letter C	0063	c **c** c	c **c** c	00E7	Latin Small Letter C with Cedilla	Distinguishable

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter D	0064	d	d	0256	Latin Small Letter D with Tail	Distinguishable
		d	d			
		d	d			
Latin Small Letter D	0064	d	d	1E13	Latin Small Letter D with Circumflex Below	Variant due to underlining
		d	d			
		d	d			
Latin Small Letter E	0065	e	e	0065 + 0331	Latin Small Letter E + Combining Macron Below	Variant due to underlining
		e	e			
		e	e			
Latin Small Letter E	0065	e	e	0119	Latin Small Letter E with Ogonek	Variant due to underlining
		e	e			
		e	e			
Latin Small Letter Open E	025B	ɛ	ɛ	025B + 0331	Latin Small Letter Open E + Combining Macron Below	Variant due to underlining
		ɛ	ɛ			
		ɛ	ɛ			
Latin Small Letter Open E + Combining Diaeresis	025B + 0308	ɛ̈	ɛ̈	025B + 0331 + 0308	Latin Small Letter Open E + Combining Macron Below + Combining Diaeresis	Variant due to underlining
		ɛ̈	ɛ̈			
		ɛ̈	ɛ̈			
Latin Small Letter E	0065	e	e	1EB9	Latin Small Letter E with Dot Below	Variant due to underlining
		e	e			
		e	e			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter E with Grave	00E8	è	è	1EB9 + 0300	Latin Small Letter E with Dot Below + Combining Grave Accent	Variant due to underlining
		è	ẹ̀			
		è	ẹ̀			
Latin Small Letter E with Acute	00E9	é	é	1EB9 + 0301	Latin Small Letter E with Dot Below + Combining Acute Accent	Variant due to underlining
		é	ẹ́			
		é	ẹ́			
Latin Small Letter E with Circumflex	00EA	ê	ê	1EC7	Latin Small Letter E with Circumflex and Dot Below	Variant due to underlining
		ê	ệ			
		ê	ệ			
Latin Small Letter G	00EC	g	q	0071	Latin Small Letter Q	Distinguishable/Out of Scope (ASCII)
		g	q			
		g	q			
Latin Small Letter Gamma	0263	ɣ	y	0079	Latin Small Letter Y	Variant due to underlining
		ɣ	y			
		ɣ	y			
Latin Small Letter I	0069	i	i	0069 + 0331	Latin Small Letter I + Combining Macron Below	Variant due to underlining
		i	i			
		i	i			
Latin Small Letter I	0069	i	i	1ECB	Latin Small Letter I with Dot Below	Variant due to underlining
		i	i			
		i	i			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter I	0069	i	į	012F	Latin Small Letter I with Ogonek	Distinguishable
		i	į			
		i	į			
Latin Small Letter J	007A	j	į	012F	Latin Small Letter I with Ogonek	Variant due to underlining
		j	į			
		j	į			
Latin Small Letter K	006B	k	k	0137	Latin Small Letter K with Cedilla	Variant due to underlining
		k	k			
		k	k			
Latin Small Letter L	006C	l	l	013C	Latin Small Letter L with Cedilla	Variant due to underlining
		l	l			
		l	l			
Latin Small Letter L	006C	l	l	1E37	Latin Small Letter L with Dot Below	Variant due to underlining
		l	l			
		l	l			
Latin Small Letter L	006C	l	l	1E3D	Latin Small Letter L with Circumflex Below	Variant due to underlining
		l	l			
		l	l			
Latin Small Letter L with Circumflex Below	1E3D	l	l	013C	Latin Small Letter L with Cedilla	Variant due to underlining
		l	l			
		l	l			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter M	006D	m	m	006D + 0327	Latin Small Letter M + Combining Cedilla	Distinguishable
		m	m̧			
		m	m̧			
Latin Small Letter M	006D	m	m	1E43	Latin Small Letter M with Dot Below	Variant due to underlining
		m	m			
		m	m			
Latin Small Letter N	006E	n	n	0146	Latin Small Letter N with Cedilla	Variant due to underlining
		n	n			
		n	n			
Latin Small Letter N	006E	n	n	1E47	Latin Small Letter N with Dot Below	Variant due to underlining
		n	n			
		n	n			
Latin Small Letter N	006E	n	n	1E49	Latin Small Letter N with Line Below	Variant due to underlining
		n	n			
		n	n			
Latin Small Letter N	006E	n	n	1E4B	Latin Small Letter N with Circumflex Below	Distinguishable
		n	n			
		n	n			
Latin Small Letter N with Cedilla	0146	n	n	1E4B	Latin Small Letter N with Circumflex Below	Variant due to underlining
		n	n			
		n	n			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter N	006E	n̲	ŋ̲	014B	Latin Small Letter Eng	Variant due to underlining
		n̲	ŋ̲			
		n̲	ŋ̲			
Latin Small Letter O	006F	o̲	o̦	006F + 0327	Latin Small Letter O + Combining Cedilla	Distinguishable
		o̲	o̦			
		o̲	o̦			
Latin Small Letter O	006F	o̲	o̲	006F + 0331	Latin Small Letter O + Combining Macron Below	Variant due to underlining
		o̲	o̲			
		o̲	o̲			
Latin Small Letter O	006F	o̲	ọ̲	1ECD	Latin Small Letter O with Dot Below	Variant due to underlining
		o̲	ọ̲			
		o̲	ọ̲			
Latin Small Letter O with Grave	00F2	ò̲	ọ̀	1ECD + 0300	Latin Small Letter O with Dot Below + Combining Grave Accent	Variant due to underlining
		ò̲	ọ̀			
		ò̲	ọ̀			
Latin Small Letter O with Acute	00F3	ó̲	ọ́	1ECD + 0301	Latin Small Letter O with Dot Below + Combining Acute Accent	Variant due to underlining
		ó̲	ọ́			
		ó̲	ọ́			
Latin Small Letter Open O	0254	ɔ̲	ɔ̲	0254 + 0331	Latin Small Letter Open O + Combining Macron Below	Variant due to underlining
		ɔ̲	ɔ̲			
		ɔ̲	ɔ̲			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter O with Circumflex	00F4	ô	ô	1ED9	Latin Small Letter O with Circumflex and Dot Below	Variant due to underlining
		ô	ô			
		ô	ô			
Latin Small Letter O with Horn	01A1	ơ	ơ	1EE3	Latin Small Letter O with Horn and Dot Below	Variant due to underlining
		ơ	ơ			
		ơ	ơ			
Latin Small Letter S	0073	s	s	015F	Latin Small Letter S with Cedilla	Variant due to underlining
		s	s			
		s	s			
Latin Small Letter S	0073	s	s	0219	Latin Small Letter S with Comma Below	Distinguishable
		s	s			
		s	s			
Latin Small Letter S with Cedilla	015F	s	s	0219	Latin Small Letter S with Comma Below	Variant due to underlining
		s	s			
		s	s			
Latin Small Letter S	0073	s	s	1E63	Latin Small Letter S with Dot Below	Variant due to underlining
		s	s			
		s	s			
Latin Small Letter T	0074	t	t	021B	Latin Small Letter T with Comma Below	Variant due to underlining
		t	t			
		t	t			

Unicode Name	Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Panel Decision
Latin Small Letter T	0074	t̲	t̲	1E6D	Latin Small Letter T with Dot Below	Variant due to underlining
		ṯ	ṯ			
		t̲	t̲			
Latin Small Letter T with Comma Below	021B	t̲	t̲	1E71	Latin Small Letter T with Circumflex Below	Variant due to underlining
		ț	ț			
		t̲	t̲			
Latin Small Letter T	0074	t̲	t̲	1E71	Latin Small Letter T with Circumflex Below	Variant due to underlining
		ṯ	ṯ			
		t̲	t̲			
Latin Small Letter U	0075	u̲	u̲	0173	Latin Small Letter U with Ogonek	Variant due to underlining
		u̲	u̲			
		u̲	u̲			
Latin Small Letter U	0075	u̲	u̲	1EE5	Latin Small Letter U with Dot Below	Variant due to underlining
		u̲	u̲			
		u̲	u̲			
Latin Small Letter U with Horn	01B0	u̲	u̲	1EF1	Latin Small Letter U with Horn and Dot Below	Variant due to underlining
		u̲	u̲			
		u̲	u̲			
Latin Small Letter Y	0079	y̲	y̲	1EF5	Latin Small Letter Y with Dot Below	Variant due to underlining
		y̲	y̲			
		y̲	y̲			

D.7 Generic Glyphs

Latin GP has tentatively identified the following variant sets for future analysis based on generic glyph shapes. Combining mark code points are indicated in the tables below by a dotted circle to the left of the glyph.

Table D.12. Generic Glyphs - Straight vertical line, full length

Glyph	Unicode	Name
L	006C	Latin Small Letter L
I	04CF	Cyrillic Small Letter Palochka
ا	0627	Arabic Letter Alef

Table D.13. Generic Glyphs - Straight vertical line, half length

Glyph	Unicode	Name
ı	0131	Latin Small Letter Dotless I
ו	05D5	Hebrew Letter Vav
◌ၤ	1062	Myanmar Vowel Sign Sgaw Karen Eu

Table D.14. Generic Glyphs - Circle

Glyph	Unicode	Unicode Name
O	006F	Latin Small Letter O
ο	03BF	Greek Small Letter Omicron
о	043F	Cyrillic Small Letter O
օ	0585	Armenian Small Letter Oh
ס	05E1	Hebrew Letter Samekh
ଠ	0B20	Oriya Letter Ttha
ഠ	0D20	Malayalam Letter Tta
ဝ	101D	Myanmar Letter Wa
ዐ	12D0	Ethiopic Syllable Pharyngeal A

Note that the Latin script only includes crescents with openings to the left and right, not to the top and bottom. So only those are included here.

Table D.15. Generic Glyphs - Crescent - Open to right

Glyph	Unicode	Name
C	0053	Latin Small Letter C
с	0441	Cyrillic Small Letter ES

ɕ	0EC0	Lao Vowel Sign E
ငꩻ	1004	Myanmar Letter Nga

Table D.16. Generic Glyphs - Crescent - Open to left

Glyph	Unicode	Name
ɔ	0254	Latin Small Letter Open O
ɔ	0EA7	Lao Letter Wo
◌ɔ	102C	Myanmar Vowel Sign Aa

Appendix E: Confusables

The Latin GP is clear that identification of Confusable is not part of our mandate. However, in the course of evaluating potential Variants we identified a number of cases which were not quite close enough to be designated as variants, but still close enough to cause confusion. (We have taken a relatively broad view of Confusables. Basically, if one of our members found them to be confusable, the pair has been included.)
These are provided in this Appendix. Note however that this list is neither comprehensive nor definitive.

Table E.1. Latin – Armenian Confusables

Unicode name	Unicode	Glyph	Glyph	Unicode	Unicode Name
Latin Small Letter A with Breve	0103	ă	ձ	0571	Armenian Small Letter Ja
Latin Small Letter B with Hook	0253	ɓ	ճ	0573	Armenian Small Letter Cheh
Latin Small Letter D	0064	d	ժ	056A	Armenian Small Letter Zhe

Latin Small Letter D with Hook	0257	ɗ	ժ	056A	Armenian Small Letter Zhe
Latin Small Letter D with Stroke	0111	đ	ժ	056A	Armenian Small Letter Zhe
Latin Small Letter Eng	014B	ŋ	ղ	0564	Armenian Small Letter Da
Latin Small Letter Eng	014B	ŋ	ղ	0572	Armenian Small Letter Ghad
Latin Small Letter Eth	00F0	ð	ծ	056E	Armenian Small Letter Ca
Latin Small Letter H	0068	h	հ	056B	Armenian Small Letter Ini
Latin Small Letter H + Latin Small Letter U	0068 0075	hu	խ	056D	Armenian Small Letter Xeh
Latin Small Letter H + Latin Small Letter U with Grave	0068 00F9	hù	խ	056D	Armenian Small Letter Xeh
Latin Small Letter H + Latin Small Letter U with Ogonek	0068 0173	hų	խ	056D	Armenian Small Letter Xeh
Latin Small Letter H + Latin Small Letter V with Hook	0068 028B	hʋ	խ	056D	Armenian Small Letter Xeh

Latin Small Letter I + Combining Macron Below	0069 0331	i̱	l	056C	Armenian Small Letter Liwn
Latin Small Letter Iota + Latin Small Letter H	0269 0068	ɩh	վհ	0583	Armenian Small Letter Piwr
Latin Small Letter J	006A	j	ȷ	0575	Armenian Small Letter Yi
Latin Small Letter L	006C	l	l	056C	Armenian Small Letter Liwn
Latin Small Letter N with Left Hook	0272	ɲ	ը	0568	Armenian Small Letter Et
Latin Small Letter N with Left Hook	0272	ɲ	ր	0580	Armenian Small Letter Reh
Latin Small Letter O with Dot Below with Combining Grave Accent	1ECD 0300	ọ̀	ծ	056E	Armenian Small Letter Ca
Latin Small Letter O with Dot Below with Combining Grave Accent	1ECD 0300	ọ̀	ձ	0571	Armenian Small Letter Ja
Latin Small Letter P	0070	p	բ	0562	Armenian Small Letter Ben

Latin Small Letter P	0070	p	թ	0569	Armenian Small Letter To
Latin Small Letter T	0074	t	է	0567	Armenian Small Letter Eh
Latin Small Letter T + Latin Small Letter Dotless I	0074 0131	tı	ե	0565	Armenian Small Letter Ech
Latin Small Letter T + Latin Small Letter Iota	0074 0269	tɩ	ե	0565	Armenian Small Letter Ech
Latin Small Letter Thorn	00FE	þ	խ	056B	Armenian Small Letter Ini
Latin Small Letter Thorn + Latin Small Letter U	00FE 0075	þu	խu	056D	Armenian Small Letter Xeh
Latin Small Letter Thorn + Latin Small Letter U with Grave	00FE 00F9	þù	խu	056D	Armenian Small Letter Xeh
Latin Small Letter U + Latin Small Letter N	0075 006E	un	տ	057F	Armenian Small Letter Tiwn
Latin Small Letter U with Horn	01B0	ư	մ	0574	Armenian Small Letter Men
Latin Small Letter U with Ogonek	0173	ų	կ	056F	Armenian Small Letter Ken

In addition, we have this pair:

Latin Small Letter Q	0071	q	ɋ	0563	Armenian Small Letter Gim

There is substantial opinion within the Latin GP that these two *should* be considered variants. However, we have already identified the Armenian small letter Za (0566) as a variant of the Latin small letter Q. If we were to designate this pair as variants, transitivity would impose an in-script variant on Armenian, one which was not identified by the Armenian GP. Since the Armenian GP is no longer available to negotiate the issue, we restrict ourselves to including this pair among the Confusables.

Table E.2 Latin – Cyrillic Confusables

Latin Small Letter B	0062	b	ь	044C	Cyrillic Small Letter Soft Sign
Latin Small Letter B + Latin Small Letter L	0062 006C	bl	ы	044B	Cyrillic Small Letter Yeru
Latin Small Letter B with Stroke	0253	ɓ	ҕ	0495	Cyrillic Small Letter Ghe with Middle Hook
Latin Small Letter E	0065	e	ҽ	04BD	Cyrillic Small Letter Abkhasian Che
Latin Small Letter E with Dot Below	1EB9	ẹ	ҿ	04BF	Cyrillic Small Letter Abkhasian Che with Descender
Latin Small Letter E with Dot Below + Combining Grave Accent	1EB9 + 0300	ẹ̀	ҿ	04BF	Cyrillic Small Letter Abkhasian Che with Descender
Latin Small Letter H with Stroke	0127	ħ	ђ	0452	Cyrillic Small Letter Dje
Latin Small Letter Iota	0269	ɩ	I	04CF	Cyrillic Small Letter Palochka

Latin Small Letter N	006E	n	п̣	0525	Cyrillic Small Letter Pe with Descender
Latin Small Letter Open E	025B	ɛ	є	0454	Cyrillic Small Letter Ukrainian Ie
Latin Small Letter U with Ogonek	0173	ų	ч	0447	Cyrillic Small Letter Che
Latin Small Letter X	0078	x	ҳ̣	04B3	Cyrillic Small Letter Ha with Descender
Latin Small Letter Y with Tilde	1EF9	ỹ	ӯ	04EF	Cyrillic Small Letter U with Macron
Latin Small Letter Y with Tilde	1EF9	ỹ	ӱ	04F1	Cyrillic Small Letter U with Diaeresis
Latin Small Letter Y with Tilde	1EF9	ỹ	ӳ	04F3	Cyrillic Small Letter U with Double Acute

In addition, we have these pairs where the Cyrillic lower case looks like the Latin upper case.

Table E.3. Latin - Cyrillic Lower Case

Latin Small Letter B	0062	b	в	0432	Cyrillic Small Letter Ve
Latin Small Letter H	0068	h	н	043D	Cyrillic Small Letter En
Latin Small Letter K	006B	k	к	043A	Cyrillic Small Letter Ka
Latin Small Letter M	006D	m	м	043C	Cyrillic Small Letter Em
Latin Small Letter T	0074	t	т	0442	Cyrillic Small Letter Te

While domain name labels are, by definition, strictly lower case, general Internet users (with the exception or the technical community) have decades of experience that teaches them that Latin upper and lower case are interchangeable.

The potential for substantial confusion is obvious. For example, a user encountering a Cyrillic TLD of .сом for the first time would naturally assume that what he was seeing was a .com TLD, merely rendered in upper case as .COM. Accordingly it seems appropriate to treat these as Confusables.

Table E.4. Latin – Greek Confusables

Latin Small Letter C with Cedilla	00E7	ç	ς	03C2	Greek Small Letter Final Sigma
Latin Small Letter Eng	014B	ŋ	η	03B7	Greek Small Letter Eta
Latin Small Letter Eth	00F0	ð	δ	03B4	Greek Small Letter Delta
Latin Small Letter I with Diaeresis	00EF	ï	ΐ	0390	Greek Small Letter Iota with Dialytika and Tonos
Latin Small Letter L	006C	l	ι	03B9	Greek Small Letter Iota
Latin Small Letter L with Acute	013A	Ĺ	ί	03AF	Greek Small Letter Iota with Tonos
Latin Small Letter N with Acute	0144	ń	ή	03AE	Greek Small Letter Iota with Tonos
Latin Small Letter Open E	025B	ɛ	έ	03AD	Greek Small Letter Epsilon with Tonos
Latin Small Letter T	0074	t	τ	03C4	Greek Small Letter Tau
Latin Small Letter T + Latin Small Letter T	0074 0074	tt	π	03C0	Greek Small Letter Pi
Latin Small Letter U	0075	u	μ	03BC	Greek Small Letter Mu

Latin Small Letter U with Acute	00FA	ú	ύ	03CD	Greek Small Letter Upsilon with Tonos
Latin Small Letter U with Horn	01B0	ư	υ	03C5	Greek Small Letter Upsilon
Latin Small Letter U with Diaeresis	00FC	ü	ϋ	03CB	Greek Small Letter Upsilon with Dialytika
Latin Small Letter U with Diaeresis	00FC	ü	ΰ	03B0	Greek Small Letter Upsilon with Dialytika and Tonos
Latin Small Letter V with Hook + Latin Small Letter V with Hook	028B 028B	υυ	ω	03C9	Greek Small Letter Omega
Latin Small Letter W	0077	w	ω	03C9	Greek Small Letter Omega
Latin Small Letter X	0078	x	χ	03C7	Greek Small Letter Chi
Latin Small Letter Y with Hook	01B4	ƴ	γ	03B3	Greek Small Letter Gamma

As with Cyrillic, we have cases where the Greek lower case looks like a Latin upper case:
Table E.5. Latin - Greek Lower Case

Latin Small Letter K	006B	k	κ	03BA	Greek Small Letter Kappa
Latin Small Letter K with Hook	0199	ƙ	κ	03BA	Greek Small Letter Kappa

E.1 Latin In-Script Confusables

Key

V Variants

C Confusables

Distinguishable

A

[This is intended to illustrate the FORMAT for displaying the information. Actual content for the cells necessarily awaits final decisions on which pairs are variants.]

		à	á	â	ã	ä	å	ā	ă
		00E0	00E1	00E2	00E3	00E4	00E5	0101	0103
a	0061	C	C						
à	00E0		V						
á	00E1								
â	00E2								
ã	00E3					V		V	C
ä	00E4								
å	00E5								
ā	0101								C
ă	0103								

B

C

D

E

F

G

		ĝ	ğ	ġ	ģ	ǧ	ḡ	g̃	q
		011D	011F	0121	0123	01E7	1E21	0067 + 0303	0071
g	0067			**C**					**C**
ĝ	011D								
ğ	011F					V	**C**		
ġ	0121				**C**				**C**
ģ	0123								
ǧ	01E7						**C**		
ḡ	1E21							**C**	
g̃	0067 + 0303								
q	0071								

The Latin Small Letter G can have two very different forms, depending on the font used. In some fonts, it appears as g, in others it appears as g. When the latter form occurs, and we have underlining (as generally happens with domain names), the underlining obscures the difference. Consider, for example, .qov vs .gov. By rule, two ASCII letters cannot be variants. But the potential for massive confusion is obvious.

H

I

J

K

L

M

N

O

P

Q

		g
		0067
q	0071	**C**
g	0067	

The Latin Small Letter G can have two very different forms, depending on the font used. In some fonts, it appears as g, in others it appears as g. When the latter form occurs, and we have underlining (as generally happens with domain names), the underlining obscures the difference. Consider, for example, .qov vs .gov. By rule, two ASCII letters cannot be variants. But the potential for massive confusion is obvious.

R

S

T

U

V

W

X

Y

Z

Other