**Proposal for a Latin Script Root Zone LGR**

*LGR Version 4.0*

*Date: 2019-01-17*

*Document version:* 4.0

*Authors:* Latin Generation Panel

Table of Contents

[1. General Information 3](#_Toc535527059)

[2. Script for which the LGR is proposed 3](#_Toc535527060)

[3. Background on Script and Principal Languages Using It 5](#_Toc535527061)

[3.1 Principal languages using Latin script 5](#_Toc535527062)

[3.2 Geographic territories or countries with significant user communities 5](#_Toc535527063)

[3.3 Related scripts 6](#_Toc535527064)

[4. Overall Development Process and Methodology 6](#_Toc535527065)

[5. Repertoire 6](#_Toc535527066)

[5.1 Definitions 7](#_Toc535527067)

[5.2 Principles for developing Repertoire 7](#_Toc535527068)

[5.2.1 Inclusion Principles 7](#_Toc535527069)

[5.2.2 Exclusion Principles 8](#_Toc535527070)

[5.3 Code points included 9](#_Toc535527071)

[5.3.1 Combining marks 32](#_Toc535527072)

[5.4 Code points excluded 33](#_Toc535527073)

[5.4.1 Punctuation marks or punctuation mark look-alikes 33](#_Toc535527074)

[5.4.2 Letters combined with punctuation marks or punctuation mark look-alikes 34](#_Toc535527075)

[5.4.3 Other Excluded Letters 37](#_Toc535527076)

[6. Variants 37](#_Toc535527077)

[6.1 Principles for developing variants 38](#_Toc535527078)

[6.1.1 Distinguishing Visual from Non-Visual Variants 38](#_Toc535527079)

[6.1.2 Visual Variants 39](#_Toc535527080)

[6.1.3 Non-Visual Variant 39](#_Toc535527081)

[6.2 Methodology for developing cross-script variants 42](#_Toc535527082)

[6.3 Cross-script variants 42](#_Toc535527083)

[6.3.1 Armenian 42](#_Toc535527084)

[6.3.2 Cyrillic 44](#_Toc535527085)

[6.3.3 Greek 46](#_Toc535527086)

[6.4 In-Script Latin variants imposed by cross-script variant relationships 52](#_Toc535527087)

[6.5 Methodology for developing in-script variants 53](#_Toc535527088)

[6.6 In-Script Latin Variants 54](#_Toc535527089)

[6.7 Other Considerations for Variant Analysis (Placeholder) 54](#_Toc535527090)

[6.7.1 HTML Underlining 54](#_Toc535527091)

[6.7.2 IDNA2003 Compatibility 54](#_Toc535527092)

[7 Whole Label Evaluation Rules (WLE) and contextual rules 54](#_Toc535527093)

[8. Contributors 55](#_Toc535527094)

[9 References 55](#_Toc535527095)

[9.1 References used in developing Repertoire 55](#_Toc535527096)

[9.2 Other references 62](#_Toc535527097)

[Appendix A: Updated MSR during Latin GP work 63](#_Toc535527098)

[Appendix B: Table of processed languages used to develop Latin Script Repertoire 65](#_Toc535527099)

[Appendix C: Repertoire table grouped by Glyph 75](#_Toc535527100)

# 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*

# 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.

|  |  |
| --- | --- |
| **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 only lowercase | 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 1. Unicode ranges included in [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 |

Table 2. Unicode ranges excluded from [MSR].

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].

# Background on Script and Principal Languages Using It

The Latin script[[1]](#footnote-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]](#footnote-2) [Wikipedia-Latin script].

## 3.1 Principal languages using Latin script

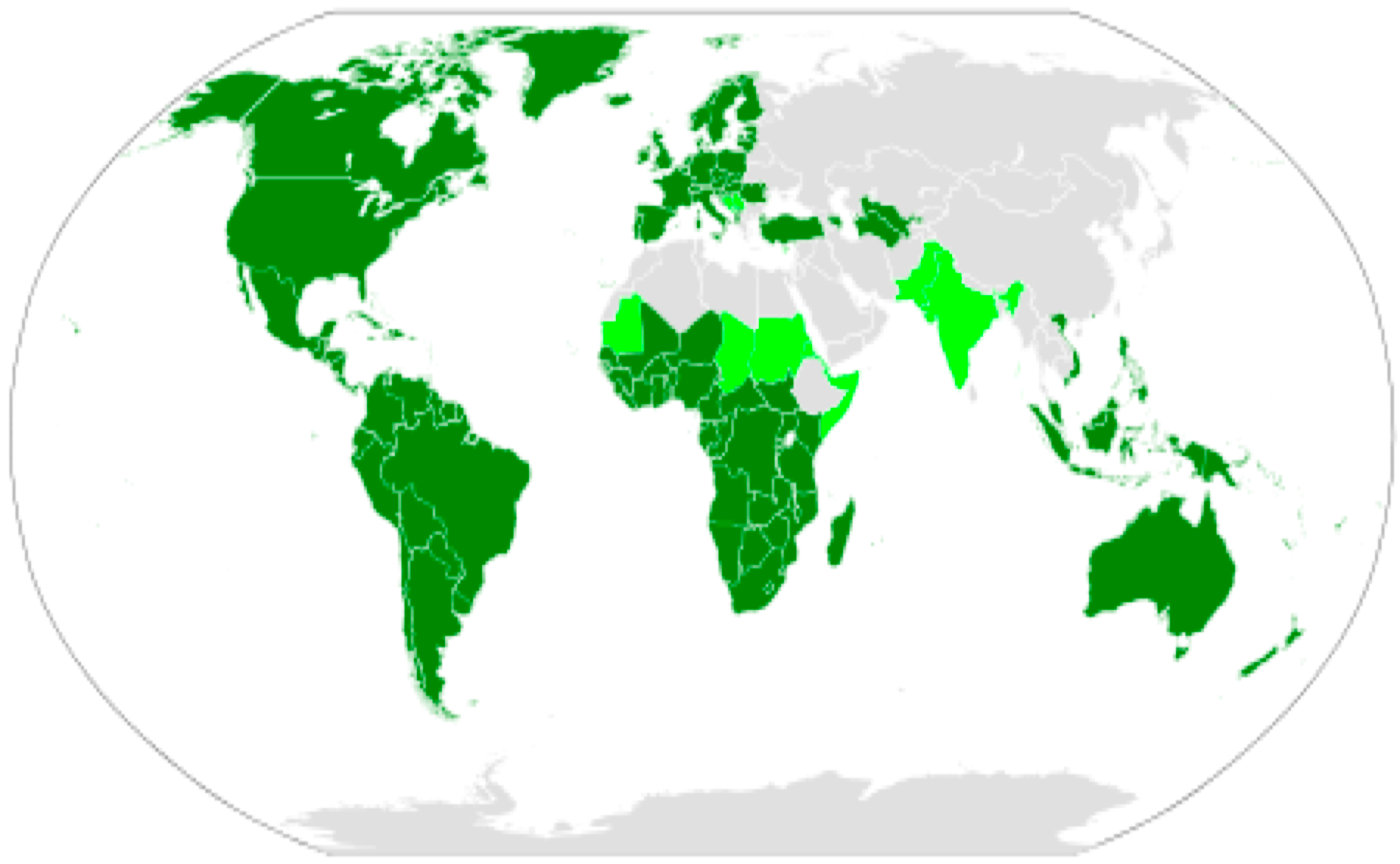
The list of languages taken into consideration contains relevant data for 455 languages using Latin script. The table with languages using Latin script was derived using data from

<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](https://en.wikipedia.org/wiki/Latin_script) the distribution of the Latin script on the world map is:



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 related to Latin script:

* Cyrillic
* Greek
* Armenian

# 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.

# 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]](#footnote-3) today. Languages are restricted tonatural 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).

## 

## 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)](https://www.ethnologue.com/enterprise-faq/what-egids-how-it-used) 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**:** 
   1. Formatting character or mark
   2. Numerical digit
   3. Punctuation mark
   4. Honorific mark or symbol
   5. 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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Unicode | Glyph | Unicode name | Languages using the code point (EGIDS) | Reference supporting inclusion (URL etc.) |
|  | **0061** | **a** | LATIN SMALL LETTER A | Basic Latin | [0] |
|  | **0061 + 0331** | **a̱** | LATIN SMALL LETTER A + COMBINING MACRON BELOW | Nuer (4) | [146], [129] |
|  | **0062** | **b** | LATIN SMALL LETTER B | Basic Latin | [0] |
|  | **0063** | **c** | LATIN SMALL LETTER C | Basic Latin | [0] |
|  | **0064** | **d** | LATIN SMALL LETTER D | Basic Latin | [0] |
|  | **0065** | **e** | LATIN SMALL LETTER E | Basic Latin | [0] |
|  | **0065 + 0331** | **e̱** | LATIN SMALL LETTER E + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  | **0066** | **f** | LATIN SMALL LETTER F | Basic Latin | [0] |
|  | **0067** | **g** | LATIN SMALL LETTER G | Basic Latin | [0] |
|  | **0067 + 0303** | **g̃** | LATIN SMALL LETTER G + COMBINING TILDE | Guarani (1) | [142], [143] |
|  | **0068** | **h** | LATIN SMALL LETTER H | Basic Latin | [0] |
|  | **0069** | **i** | LATIN SMALL LETTER I | Basic Latin | [0] |
|  | **0069 + 0331** | **i̲** | LATIN SMALL LETTER I + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  | **006A** | **j** | LATIN SMALL LETTER J | Basic Latin | [0] |
|  | **006B** | **k** | LATIN SMALL LETTER K | Basic Latin | [0] |
|  | **006C** | **l** | LATIN SMALL LETTER L | Basic Latin | [0] |
|  | **006D** | **m** | LATIN SMALL LETTER M | Basic Latin | [0] |
|  | **006D + 0327** | **m̦** | LATIN SMALL LETTER M + COMBINING CEDILLA | Marshallese (1) | [213], [136], [214] |
|  | **006E** | **n** | LATIN SMALL LETTER N | Basic Latin | [0] |
|  | **006E + 0304** | **n̄** | LATIN SMALL LETTER N + COMBINING MACRON | Raga (Hano) (3)  Marshallese (1) | [200], [213], [136] |
|  | **006E + 0308** | **n̈** | LATIN SMALL LETTER N + COMBINING DIAERESIS | Malagasy(1) | [230] |
|  | **006F** | **o** | LATIN SMALL LETTER O | Basic Latin | [0] |
|  | **006F + 0327** | **o̧** | LATIN SMALL LETTER O + COMBINING CEDILLA | Marshallese (1) | [136] |
|  | **006F + 0331** | **o̱** | LATIN SMALL LETTER O + COMBINING MACRON BELOW | Nuer (4) | [146], [129] |
|  | **0070** | **p** | LATIN SMALL LETTER P | Basic Latin | [0] |
|  | **0071** | **q** | LATIN SMALL LETTER Q | Basic Latin | [0] |
|  | **0072** | **r** | LATIN SMALL LETTER R | Basic Latin | [0] |
|  | **0072 + 0303** | **r̃** | LATIN SMALL LETTER R WITH TILDE | Hausa (2) | [147] |
|  | **0073** | **s** | LATIN SMALL LETTER S | Basic Latin | [0] |
|  | **0074** | **t** | LATIN SMALL LETTER T | Basic Latin | [0] |
|  | **0075** | **u** | LATIN SMALL LETTER U | Basic Latin | [0] |
|  | **0076** | **v** | LATIN SMALL LETTER V | Basic Latin | [0] |
|  | **0077** | **w** | LATIN SMALL LETTER W | Basic Latin | [0] |
|  | **0078** | **x** | LATIN SMALL LETTER X | Basic Latin | [0] |
|  | **0079** | **y** | LATIN SMALL LETTER Y | Basic Latin | [0] |
|  | **007A** | **z** | LATIN SMALL LETTER Z | Basic Latin | [0] |
|  | **00DF** | **ß** | LATIN SMALL LETTER SHARP S | German (1) | [119] |
|  | **00E0** | **à** | LATIN SMALL LETTER A WITH GRAVE | Italian (1)  Galician (2)  Wolof (4) | [130], [131], [106], [132] |
|  | **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] |
|  | **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] |
|  | **00E3** | **ã** | LATIN SMALL LETTER A WITH TILDE | Umbundu (3)  Guarani (1)  Nauruan (3)  Khoekhoe (4) | [141], [142], [143], [144], [145] |
|  | **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] |
|  | **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] |
|  | **00E6** | **æ** | LATIN SMALL LETTER AE | Danish (1)  Icelandic (1)  Faroese (2) | [139], [102], [103] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **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], [178], [179], [124], [132], [180], [126], [115], [129] |
|  | **00EC** | **ì** | LATIN SMALL LETTER I WITH GRAVE | Italian (1)  Kirundi (1) | [130], [206], [208] |
|  | **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] |
|  | **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] |
|  | **00EF** | **ï** | LATIN SMALL LETTER I WITH DIAERESIS | Afrikaans (1)  French (1)  Kaqchikel (4)  Dinka (4)  West Frisian (2) | [175], [114], [126], [125], [115] |
|  | **00F0** | **ð** | LATIN SMALL LETTER ETH | Faroese (2)  Icelandic (1) | [103], [102] |
|  | **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] |
|  | **00F2** | **ò** | LATIN SMALL LETTER O WITH GRAVE | Italian (1)  Haitian Creole (1) | [130], [182], [183] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **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)  Chechen (2) 1992 Version  West Frisian (2)  Nuer (4) | [119], [120], [175], [157], [123], [179], [124], [180], [126], [125], [127], [231], [232], [115], [129] |
|  | **00F8** | **ø** | LATIN SMALL LETTER O WITH STROKE | Danish (1)  Faroese (2) | [139], [103] |
|  | **00F9** | **ù** | LATIN SMALL LETTER U WITH GRAVE | Italian (1)  Papiamento (1) | [130], [206], [245], [246] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **00FD** | **ý** | LATIN SMALL LETTER Y WITH ACUTE | Turkmen (1)  Czech (1)  Icelandic (1)  Faroese (2)  Guarani (1) | [121], [101], [102], [103], [142], [143] |
|  | **00FE** | **þ** | LATIN SMALL LETTER THORN | Icelandic (1) | [102] |
|  | **00FF** | **ÿ** | LATIN SMALL LETTER Y WITH DIAERESIS | French (1) | [114], [253], [257] |
|  | **0101** | **ā** | LATIN SMALL LETTER A WITH MACRON | Latvian (1)  Tongan (1)  Hawaiian (2)  Marshallese (1) | [133], [134], [135], [136] |
|  | **0103** | **ă** | LATIN SMALL LETTER A WITH BREVE | Vietnamese (1)  Romanian (1) | [109], [110] |
|  | **0105** | **ą** | LATIN SMALL LETTER A WITH OGONEK | Polish (1)  Lithuanian (1) | [137], [138] |
|  | **0107** | **ć** | LATIN SMALL LETTER C WITH ACUTE | Croatian (1)  Serbian (1)  Polish (1) | [150], [151], [152] |
|  | **0109** | **ĉ** | LATIN SMALL LETTER C WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **010B** | **ċ** | LATIN SMALL LETTER C WITH DOT ABOVE | Maltese(1) | [163] |
|  | **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] |
|  | **010F** | **ď** | LATIN SMALL LETTER D WITH CARON | Czech (1)  Slovak (1) | [101], [153] |
|  | **0111** | **đ** | LATIN SMALL LETTER D WITH STROKE | Croatian (1)  Serbian (1)  Vietnamese (1)  Northern Sámi (2) | [150], [151], [109], [108], [168] |
|  | **0113** | **ē** | LATIN SMALL LETTER E WITH MACRON | Latvian (1)  Hawaiian (2)  Tongan (1)  Minangkabau (5) | [133], [135], [134], [184] |
|  | **0117** | **ė** | LATIN SMALL LETTER E WITH DOT ABOVE | Lithuanian (1) | [138], [154] |
|  | **0119** | **ę** | LATIN SMALL LETTER E WITH OGONEK | Polish (1)  Palauan (2)  Lithuanian (1) | [152], [185], [138], [154] |
|  | **011B** | **ě** | LATIN SMALL LETTER E WITH CARON | Czech (1)  Kirundi (1)  Sorbian (4) | [101], [104], [172] |
|  | **011D** | **ĝ** | LATIN SMALL LETTER G WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **011F** | **ğ** | LATIN SMALL LETTER G WITH BREVE | Turkish (1)  Tatar (2)  Azer i(1)  Bashkir (4)  Zaza (5) | [157], [201], [159], [127], [202] |
|  | **0121** | **ġ** | LATIN SMALL LETTER G WITH DOT ABOVE | Maltese (1) | [163] |
|  | **0123** | **ģ** | LATIN SMALL LETTER G WITH CEDILLA | Latvian (1)  Brahui (5) | [133], [168] |
|  | **0125** | **ĥ** | LATIN SMALL LETTER H WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **0127** | **ħ** | LATIN SMALL LETTER H WITH STROKE | Maltese (1) | [163] |
|  | **0129** | **ĩ** | LATIN SMALL LETTER I WITH TILDE | Guarani (1)  Cubeo (3)  Khoekhoe (4)  Kikuyu ( 5) | [142], [143], [186], [145], [209] |
|  | **012B** | **ī** | LATIN SMALL LETTER I WITH MACRON | Latvian (1)  Lithuanian (1)  Hawaiian (2)  Tongan (1) | [133], [138], [135], [134] |
|  | **012F** | **į** | LATIN SMALL LETTER I WITH OGONEK | Lithuanian (1) | [154] |
|  | **0131** | **ı** | LATIN SMALL LETTER DOTLESS I | Turkish (1)  Tatar (2)  Azeri(1) | [157], [203], [201], [159] |
|  | **0135** | **ĵ** | LATIN SMALL LETTER J WITH CEDILLA | Esperanto (3) | [255] |
|  | **0137** | **ķ** | LATIN SMALL LETTER K WITH CEDILLA | Latvian (1) | [133] |
|  | **013A** | **ĺ** | LATIN SMALL LETTER L WITH ACUTE | Slovak (1) | [153] |
|  | **013C** | **ļ** | LATIN SMALL LETTER L WITH CEDILLA | Latvian (1)  Marshallese (1)  Brahui (5) | [133], [213], [214], [168] |
|  | **013E** | **ľ** | LATIN SMALL LETTER L WITH CARON | Slovak (1) | [153] |
|  | **0142** | **ł** | LATIN SMALL LETTER L WITH STROKE | Polish (1) | [152] |
|  | **0144** | **ń** | LATIN SMALL LETTER N WITH ACUTE | Polish (1)  Lule Sámi (2)  Sorbian (4)  Brahui (5) | [152], [107], [172], [168] |
|  | **0146** | **ņ** | LATIN SMALL LETTER N WITH CEDILLA | Latvian (1)  Marshallese (1) | [133], [136] |
|  | **0148** | **ň** | LATIN SMALL LETTER N WITH CARON | Turkmen (1)  Czech (1)  Slovak (1) | [121], [101], [153] |
|  | **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] |
|  | **014D** | **ō** | LATIN SMALL LETTER O WITH MACRON | Hawaiian (2)  Marshallese (1)  Tongan (1) | [135], [136], [134] |
|  | **0151** | **ő** | LATIN SMALL LETTER O WITH DOUBLE ACUTE | Hungarian (1) | [233], [234] |
|  | **0153** | **œ** | LATIN SMALL LIGATURE OE | French (1) | [114], [253] |
|  | **0155** | **ŕ** | LATIN SMALL LETTER R WITH ACUTE | Slovak (1)  Brahui (5) | [153], [168] |
|  | **0159** | **ř** | LATIN SMALL LETTER R WITH CARON | Czech (1)  Sorbian (4) | [101], [172] |
|  | **015B** | **ś** | LATIN SMALL LETTER S WITH ACUTE | Polish (1) | [152] |
|  | **015D** | **ŝ** | LATIN SMALL LETTER S WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **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] |
|  | **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] |
|  | **0165** | **ť** | LATIN SMALL LETTER T WITH CARON | Czech (1)  Slovak (1) | [101], [153] |
|  | **0167** | **ŧ** | LATIN SMALL LETTER T WITH STROKE | Northern Sami (2)  Brahui (5) | [108], [168] |
|  | **0169** | **ũ** | LATIN SMALL LETTER U WITH TILDE | Umbundu (3)  Guarani (1)  Nauruan (3)  Khoekhoe (4) Kikuyu (5) | [141], [142], [143], [144], [145], [209] |
|  | **016B** | **ū** | LATIN SMALL LETTER U WITH MACRON | Latvian (1)  Hawaiian (2)  Lithuanian (1)  Marshallese (1)  Tongan (1) | [133], [135], [138], [154], [136], [134] |
|  | **016D** | **ŭ** | LATIN SMALL LETTER U WITH BREVE | Esperanto (3) | [255] |
|  | **016F** | **ů** | LATIN SMALL LETTER U WITH RING ABOVE | Czech (1) | [101] |
|  | **0171** | **ű** | LATIN SMALL LETTER U WITH DOUBLE ACUTE | Hungarian (1) | [233], [234] |
|  | **0173** | **ų** | LATIN SMALL LETTER U WITH OGONEK | Lithuanian (1) | [154], [138] |
|  | **0175** | **ŵ** | LATIN SMALL LETTER W WITH CIRCUMFLEX | Chichewa (3)  Welsh (2) | [247], [256] |
|  | **0177** | **ŷ** | LATIN SMALL LETTER Y WITH CIRCUMFLEX | Welsh (2) | [256] |
|  | **017A** | **ź** | LATIN SMALL LETTER Z WITH ACUTE | Polish (1)  Brahui (5)  Sorbian (4) | [152], [252], [168], [172] |
|  | **017C** | **ż** | LATIN SMALL LETTER Z WITH DOT ABOVE | Polish (1)  Maltese (1) | [152], [163] |
|  | **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] |
|  | **0192** | **ƒ** | LATIN SMALL LETTER F WITH HOOK | Ewe (3) | [170] |
|  | **0199** | **ƙ** | LATIN SMALL LETTER K WITH HOOK | Hausa (2) | [147] |
|  | **01A1** | **ơ** | LATIN SMALL LETTER O WITH HORN | Vietnamese (1) | [109] |
|  | **01B0** | **ư** | LATIN SMALL LETTER U WITH HORN | Vietnamese (1) | [109] |
|  | **01B4** | **ƴ** | LATIN SMALL LETTER Y WITH HOOK | Dagaare - Burkina Faso (4) | [148], [251], [149] |
|  | **01CE** | **ǎ** | LATIN SMALL LETTER A WITH CARON | Kirundi (1) | [104] [**https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx**](https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx)  **Jean Paul Nkurunziza (personal communication)** |
|  | **01D0** | **ǐ** | LATIN SMALL LETTER I WITH CARON | Kirundi (1) | [104] |
|  | **01D2** | **ǒ** | LATIN SMALL LETTER O WITH CARON | Kirundi (1) | [104] |
|  | **01D4** | **ǔ** | LATIN SMALL LETTER U WITH CARON | Kirundi (1) | [104] |
|  | **01DD** | **ǝ** | LATIN SMALL LETTER TURNED E | Kanuri (3) | [240] |
|  | **01E7** | **ǧ** | LATIN SMALL LETTER G WITH CARON | Skolt Sami (2) | [113] |
|  | **01E9** | **ǩ** | LATIN SMALL LETTER K WITH CARON | Skolt Sami (2) | [113] |
|  | **01EF** | **ǯ** | LATIN SMALL LETTER EZH WITH CARON | Skolt Sami (2) | [113] |
|  | **0219** | **ș** | LATIN SMALL LETTER S WITH COMMA BELOW | Romanian (1) | [110] |
|  | **021B** | **ț** | LATIN SMALL LETTER T WITH COMMA BELOW | Romanian (1) | [110] |
|  | **024D** | **ɍ** | LATIN SMALL LETTER R WITH STROKE | Kanuri (3) | [240] |
|  | **0253** | **ɓ** | LATIN SMALL LETTER B WITH HOOK | Hausa (2)  Dagaare - Burkina Faso (4)  Pulaar, (3) | [147], [148], [250] |
|  | **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] |
|  | **0254 + 0308** | **ɔ̈** | LATIN SMALL LETTER OPEN O + COMBINING DIAERESIS | Dinka (4) | [125] |
|  | **0254 + 0331** | **ɔ̱** | LATIN SMALL LETTER OPEN O + COMBINING MACRON BELOW | Nuer (4) | [129], [146] |
|  | **0256** | **ɖ** | LATIN SMALL LETTER D WITH TAIL | Fon (3)  Ewe (3) | [169], [170] |
|  | **0257** | **ɗ** | LATIN SMALL LETTER D WITH HOOK | Hausa (2)  Pulaar (3) | [147], [166], [250] |
|  | **0259** | **ə** | LATIN SMALL LETTER SCHWA | Azeri, Azerbaijani (1)  Ewondo (3)  Ewe (3)  Bugis (3) | [159], [190], [170], [241] |
|  | **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] |
|  | **025B + 0308** | **ɛ̈** | LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS | Nuer (4)  Dinka (4) | [129], [146], [239], [125] |
|  | **025B + 0331** | **ɛ̱** | LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW | Nuer (4) | [129], [146], [239] |
|  | **025B + 0331 + 0308** | **ɛ̱̈** | LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS | Nuer (4) | [146], [239] |
|  | **0263** | **ɣ** | LATIN SMALL LETTER GAMMA | Dagbani (Dagomba) (4)  Nuer (4)  Dinka (4)  Ewe (3)  Nuer (4) | [189], [146], [125], [170], [129] |
|  | **0268** | **ɨ** | LATIN SMALL LETTER I WITH STROKE | Cubeo (3)  Dagbani (Dagomba) (4)  HIxkaryána (4)  Maasai (5) | [186], [189], [210], [211] |
|  | **0268 + 0303** | **ɨ̃** | LATIN SMALL LETTER I WITH STROKE + COMBINING TILDE | Cubeo (3) | [186] |
|  | **0269** | **ɩ** | LATIN SMALL LETTER IOTA | Dagaare - Burkina Faso (4)  Mossi (3) | [148], [212] |
|  | **0272** | **ɲ** | LATIN SMALL LETTER N WITH LEFT HOOK | Susu (4)  Zarma (4)  Bambara (4) | [218], [219], [199] |
|  | **0289** | **ʉ** | LATIN SMALL LETTER U BAR | Cubeo (3)  Maasai (5) | [186], [187], [211] |
|  | **0289 + 0303** | **ʉ̃** | LATIN SMALL LETTER U BAR + COMBINING TILDE | Cubeo (3) | [186], [187] |
|  | **028B** | **ʋ** | LATIN SMALL LETTER V WITH HOOK | Dagaare - Burkina Faso (4)  Mossi (3)  Ewe (3) | [148], [212], [238], [170] |
|  | **0292** | **ʒ** | LATIN SMALL LETTER EZH | Skolt Sami (2)  Dagbani (Dagomba) (4) | [113], [189] |
|  | **1E13** | **ḓ** | LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **1E21** | **ḡ** | LATIN SMALL LETTER G + MACRON | Raga (Hano) (3) | [200] |
|  | **1E37** | **ḷ** | LATIN SMALL LETTER L WITH DOT BELOW | Marshallese (1) | [213], [214], [215], [216] |
|  | **1E3D** | **ḽ** | LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **1E43** | **ṃ** | LATIN SMALL LETTER M WITH DOT BELOW | Marshallese (1) | [213], [136], [215], [216] |
|  | **1E45** | **ṅ** | LATIN SMALL LETTER N WITH DOT ABOVE | Venda (1) | [164], [257] |
|  | **1E47** | **ṇ** | LATIN SMALL LETTER N WITH DOT BELOW | Marshallese (1) | [136], [215], [216] |
|  | **1E49** | **ṉ** | LATIN SMALL LETTER N WITH LINE BELOW | Pitjantjatjara (4) | [220] |
|  | **1E4B** | **ṋ** | LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **1E63** | **ṣ** | LATIN SMALL LETTER S WITH DOT BELOW | Yoruba (2) | [181] |
|  | **1E6D** | **ṭ** | LATIN SMALL LETTER T WITH DOT BELOW | Mizo (4) | [242] |
|  | **1E71** | **ṱ** | LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **1E8D** | **ẍ** | LATIN SMALL LETTER X WITH DIAERESIS | Mam (4) | [248], [249] |
|  | **1EA1** | **ạ** | LATIN SMALL LETTER A WITH DOT BELOW | Vietnamese (1) | [109] |
|  | **1EA3** | **ả** | LATIN SMALL LETTER A WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EA5** | **ấ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1EA7** | **ầ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1EA9** | **ẩ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EAB** | **ẫ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1EAD** | **ậ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EAF** | **ắ** | LATIN SMALL LETTER A WITH BREVE AND ACUTE | Vietnamese (1) | [109] |
|  | **1EB1** | **ằ** | LATIN SMALL LETTER A WITH BREVE AND GRAVE | Vietnamese (1) | [109] |
|  | **1EB3** | **ẳ** | LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EB5** | **ẵ** | LATIN SMALL LETTER A WITH BREVE AND TILDE | Vietnamese (1) | [109] |
|  | **1EB7** | **ặ** | LATIN SMALL LETTER A WITH BREVE AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EB9** | **ẹ** | LATIN SMALL LETTER E WITH DOT BELOW | Yoruba(2) | [181] |
|  | **1EB9 + 0300** | **ẹ̀** | LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT | Yoruba (2) | [254] |
|  | **1EB9 + 0301** | **ẹ́** | LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT | Yoruba (2) | [254] |
|  | **1EBB** | **ẻ** | LATIN SMALL LETTER E WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EBD** | **ẽ** | LATIN SMALL LETTER E WITH TILDE | Umbundu (3)  Guarani (1)  Cubeo (3)  Xavante (4) | [141], [142], [143], [186], [187], [117] |
|  | **1EBF** | **ế** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1EC1** | **ề** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1EC3** | **ể** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EC5** | **ễ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1EC7** | **ệ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EC9** | **ỉ** | LATIN SMALL LETTER I WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ECB** | **ị** | LATIN SMALL LETTER I WITH DOT BELOW | Igbo (2) | [205] |
|  | **1ECD** | **ọ** | LATIN SMALL LETTER O WITH DOT BELOW | Igbo (2)  Yoruba (2)  Marshallese (1) | [204], [205], [181], [136], [215], [216] |
|  | **1ECD + 0300** | **ọ̀** | LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT | Yoruba (2) | [254] |
|  | **1ECD + 0301** | **ọ́** | LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT | Yoruba (2) | [254] |
|  | **1ECF** | **ỏ** | LATIN SMALL LETTER O WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ED1** | **ố** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1ED3** | **ồ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1ED5** | **ổ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ED7** | **ỗ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1ED9** | **ộ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EDB** | **ớ** | LATIN SMALL LETTER O WITH HORN AND ACUTE | Vietnamese (1) | [109] |
|  | **1EDD** | **ờ** | LATIN SMALL LETTER O WITH HORN AND GRAVE | Vietnamese (1) | [109] |
|  | **1EDF** | **ở** | LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EE1** | **ỡ** | LATIN SMALL LETTER O WITH HORN AND TILDE | Vietnamese (1) | [109] |
|  | **1EE3** | **ợ** | LATIN SMALL LETTER O WITH HORN AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EE5** | **ụ** | LATIN SMALL LETTER U WITH DOT BELOW | Vietnamese (1]  Igbo (2) | [109],[204], [205] |
|  | **1EE7** | **ủ** | LATIN SMALL LETTER U WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EE9** | **ứ** | LATIN SMALL LETTER U WITH HORN AND ACUTE | Vietnamese (1) | [109] |
|  | **1EEB** | **ừ** | LATIN SMALL LETTER U WITH HORN AND GRAVE | Vietnamese (1) | [109] |
|  | **1EED** | **ử** | LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EEF** | **ữ** | LATIN SMALL LETTER U WITH HORN AND TILDE | Vietnamese (1) | [109] |
|  | **1EF1** | **ự** | LATIN SMALL LETTER U WITH HORN AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EF3** | **ỳ** | LATIN SMALL LETTER Y WITH GRAVE | Vietnamese (1) | [109] |
|  | **1EF5** | **ỵ** | LATIN SMALL LETTER Y WITH DOT BELOW | Vietnamese (1) | [109] |
|  | **1EF7** | **ỷ** | LATIN SMALL LETTER Y WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **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. They are not listed individually in the repertoire, since they can not 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.)

The code points are:

|  |  |  |
| --- | --- | --- |
| **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, codepoints 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.

### 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** | **’** | 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** | **ꞌ** | LATIN SMALL LETTER SALTILLO | Central Sinama (4)  Guarani (1)  Kaqchikel (4)  Oromo (Afaan) (5)  Pangasinan (3) | [https://www.omniglot.com/writing/centralsinama.htm http://sinama.org/bahasa-sinama/sama-alphabet/](http://www.omniglot.com/writing/guarani.htm)  [http://www.omniglot.com/writing/guarani.htm https://en.wikipedia.org/wiki/Guarani\_alphabet](http://www.omniglot.com/writing/guarani.htm)  [https://www.omniglot.com/writing/kaqchikel.htm](https://www.omniglot.com/writing/centralsinama.htm)  [https://www.omniglot.com/writing/oromo.htm](http://www.omniglot.com/writing/uzbek.htm)  [https://www.omniglot.com/writing/pangasinan.htm](http://www.omniglot.com/charts/dagbani.pdf) |
| **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> |

### Letters combined with punctuation marks or punctuation mark look-alikes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Unicode** | **Glyph** | **Unicode Name** | **Language** | **Reference** |
| **0063 + 0068 + A78C** | **chꞌ** | 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](https://en.wikipedia.org/wiki/Uzbek_alphabet" \l "Distinct_characters) |
| **02BC + 0068** | **ʼh** | LATIN MODIFIER LETTER APOSTROPHE WITH LATIN SMALL LETTER H | Dagaare - Burkina Faso (4) | <http://www.omniglot.com/writing/dagaare.htm> |
| **006B + A78C** | **kꞌ** | LATIN SMALL LETTER K + LATIN SMALL LETTER SALTILLO | Quechua (3) | <https://www.omniglot.com/writing/quechua.htm> |
| **02BC + 006C** | **ʼl** | 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](http://www.webcitation.org/6sl20cbZO) |
| **006D + 02BC** | **m’** | LATIN SMALL LETTER M + MODIFIER LETTER APOSTROPHE | Garo (2) | [http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm](http://www.webcitation.org/6sl20cbZO) |
| **006E + 02BC** | **n’** | LATIN SMALL LETTER N + MODIFIER LETTER APOSTROPHE | Garo (2) | [http://www.webcitation.org/6sl20cbZO https://www.omniglot.com/writing/garo.htm](http://www.webcitation.org/6sl20cbZO) |
| **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](http://www.webcitation.org/6sl20cbZO) |
| **014B + 02BC** | **ŋʼ** | 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](https://en.wikipedia.org/wiki/Uzbek_alphabet" \l "Distinct_characters) |
| **0070 + A78C** | **pꞌ** | LATIN SMALL LETTER O + LATIN SMALL LETTER SALTILLO | Quechua (3) | <https://www.omniglot.com/writing/quechua.htm> |
| **0071 + A78C** | **qꞌ** | LATIN SMALL LETTER Q + LATIN SMALL LETTER SALTILLO | Quechua (3) | <https://www.omniglot.com/writing/quechua.htm> |
| **0074 + A78C** | **tꞌ** | LATIN SMALL LETTER T + LATIN SMALL LETTER SALTILLO | Quechua (3) | <https://www.omniglot.com/writing/quechua.htm> |
| **02BC + 0077** | **ʼw** | LATIN MODIFIER LETTER APOSTROPHE WITH LATIN SMALL LETTER W | Dagaare - Burkina Faso (4) | <http://www.omniglot.com/writing/dagaare.htm> |

### Other Excluded Letters

The Integration Panel has declined to include three letters, proposed by Latin GP for inclusion in [MSR], because of 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).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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](https://www.britannica.com/topic/Khoisan-languages) |
| **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](https://www.britannica.com/topic/Khoisan-languages) |
| **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](https://www.britannica.com/topic/Khoisan-languages) |

# Variants

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

In accordance to 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 that do not clearly fall underneath visual or non-visual categories, such as IDNA2003 compatibility and HTML underlining.

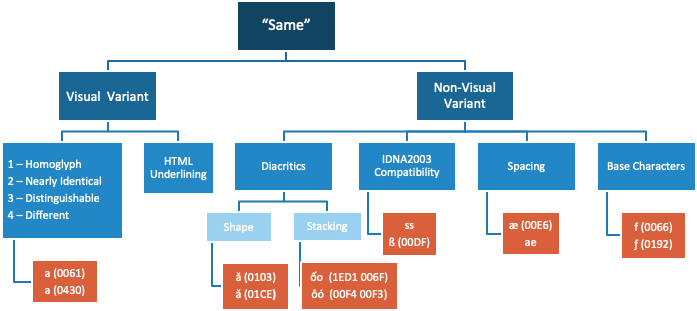
### 6.1.1 Distinguishing Visual from Non-Visual Variants

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, situational/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 the several aspects, as summarized in the following illustration:



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

### 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).**

Latin GP used a four-point score system to determine whether a given pair of candidate characters is in the “essentially identical appearance by design” group or in the “merely similar appearance”, as follows:

|  |  |
| --- | --- |
| 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. |

The types of non-visual similarity (on the right-hand branch of the illustration) are discussed in below in 6.2.2.

### 6.1.3 Non-Visual Variant

#### 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 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. ı). 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 of 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 hand-writing 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 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 kerning), ligatures may become indistinguishable from a sequence of letters of which the same ligature is composed.

#### 6.1.3.3 IDNA Compatibility Issues

In Section 5.5 of [Maximal Starting Repertoire — MSR-4 Overview and Rationale](https://www.icann.org/en/system/files/files/msr-4-overview-09nov18-en.pdf), Integration Panel highlighted risks due to IDNA compatibility issues, and two of those code points mentioned belong to the Latin-script repertoire, namely 00DF LATIN SMALL LETTER SHARP S and 0131 LATIN SMALL LETTER DOTLESS I. In the case of 00DF, Unicode has retained the German orthographic practice to replace ß by a sequence of ss in uppercase strings through case folding, with 00DF therefore becoming inter-changeable in upper case strings with 0073+0073. Accordingly, Integration Panel asked the GP to ”consider the inclusion of this code point in the LGR, to investigate the case for or against making it a blocked variant of ss.”

In the case of 0131, case folding generally applies to the upper-case forms of both 0131 and 0069 LATIN SMALL LETTER I. As a consequence, 0131 is not distinguishable from 0069 in upper case forms and would generally be converted back to 0069 instead of 0131 when changing the case back from upper to lower. Accordingly, ”Integration Panel expects the Latin Generation Panel to investigate the need to address any compatibility issues related to this code point, and if found, suggest means to mitigate them.“

#### 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 in fact 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 only in 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 which such modifiers:

First, certain diacritics may be considered to be 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 being often 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 also the diacritics become visually identical or confusable with others, such as a diaresis 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 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 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 print and in Unicode may, in hand-writing, be considered as being interchangeable or even the same, or may become visually identical or confusable with others by 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.

In 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 by readers with sequences of glyphs featuring the same diacritics on two separate code points or may even become effectively invisible in context by lapping 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.

A Latin code point will be deemed a visual cross-script variant with a code point from another script when the two code points or sequence of code points are

* homoglyphs (i.e. visual score = 1) or
* nearly identical (i.e. visual score = 2) when the difference can be attributed to different font styles.

To test this, Latin GP compared glyphs across three different wide-spread fonts, i.e. Arial, Courier New, and Times New Roman.

## 6.3 Cross-script variants

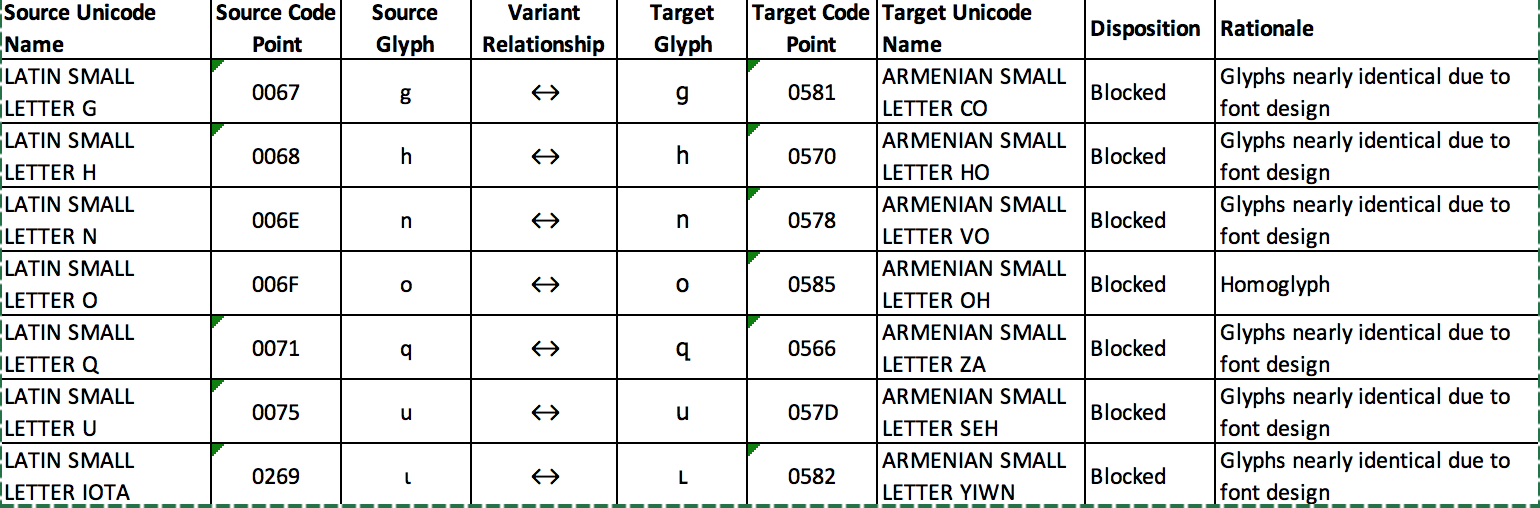
### 6.3.1 Armenian

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 snapshot image taken from Microsoft Excel to demonstrate the glyphs used during the cross-script variant analysis[[4]](#footnote-4).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | ↔ | ց | 0581 | ARMENIAN SMALL LETTER CO | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER H | 0068 | h | ↔ | հ | 0570 | ARMENIAN SMALL LETTER HO | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER N | 006E | n | ↔ | ո | 0578 | ARMENIAN SMALL LETTER VO | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER O | 006F | o | ↔ | օ | 0585 | ARMENIAN SMALL LETTER OH | Blocked | Homoglyph |
| LATIN SMALL LETTER Q | 0071 | q | ↔ | զ | 0566 | ARMENIAN SMALL LETTER ZA | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER U | 0075 | u | ↔ | ս | 057D | ARMENIAN SMALL LETTER SEH | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER IOTA | 0269 | ɩ | ↔ | ւ | 0582 | ARMENIAN SMALL LETTER YIWN | Blocked | Glyphs nearly identical due to font design |

Snapshot taken from Microsoft Excel

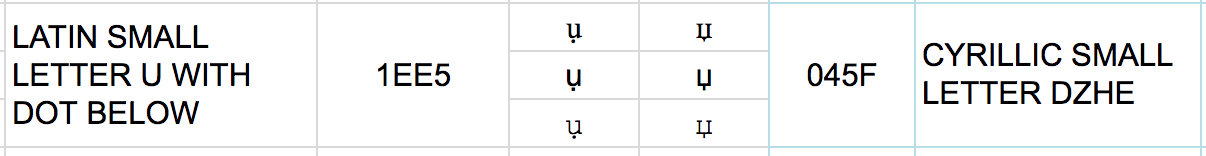


### Cyrillic

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | ↔ | г | 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 |
| LATIN SMALL LETTER C WITH CEDILLA | 00E7 | ç | ↔ | ҫ | 04AB | CYRILLIC SMALL LETTER ES WITH DESCENDER | Blocked | Glyphs nearly identical due to font design |
| LATIN SMALL LETTER Y WITH DIAERESIS | 00FF | ÿ | ↔ | Ӱ | 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 **[C1]** 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 BYELORUSSIAN-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 | ↔ | о | 043E | CYRILLIC SMALL LETTER O | Blocked | Homoglyph |
| LATIN SMALL LETTER P | 0070 | p | ↔ | р | 0440 | CYRILLIC SMALL LETTER ER | Blocked | Homoglyph |
| LATIN SMALL LETTER S | 0073 | s | ↔ | ѕ | 0455 | CYRILLIC SMALL LETTER DZE | Blocked | Homoglyph |
| LATIN SMALL LETTER X | 0078 | x | ↔ | х | 0445 | CYRILLIC SMALL LETTER HA | Blocked | Homoglyph |
| LATIN SMALL LETTER Y | 0079 | y | ↔ | у | 0443 | CYRILLIC SMALL LETTER U | Blocked | Homoglyph |
| LATIN SMALL LETTER A WITH DIAERESIS | 00E4 | ä | ↔ | ӓ | 04D3 | CYRILLIC SMALL LETTER A WITH DIAERESIS | Blocked | Homoglyph |
| LATIN SMALL LETTER AE | 00E6 | æ | ↔ | ӕ | 04D5 | CYRILLIC SMALL LIGATURE A IE | Blocked | Homoglyph |
| LATIN SMALL LETTER E WITH DIAERESIS | 00EB | ë | ↔ | ё | 0451 | CYRILLIC SMALL LETTER IO | Blocked | Homoglyph |
| LATIN SMALL LETTER I WITH DIAERESIS | 00EF | ï | ↔ | ї | 0457 | CYRILLIC SMALL LETTER YI | Blocked | Homoglyph |
| LATIN SMALL LETTER O WITH DIAERESIS | 00F6 | ö | ↔ | ӧ | 04E7 | CYRILLIC SMALL LETTER O WITH DIAERESIS | Blocked | Homoglyph |
| LATIN SMALL LETTER A WITH BREVE | 0103 | ă | ↔ | ӑ | 04D1 | CYRILLIC SMALL LETTER A WITH BREVE | Blocked | Homoglyph |
| LATIN SMALL LETTER H WITH STROKE | 0127 | ħ | ↔ | ћ | 045B | CYRILLIC SMALL LETTER TSHE | Blocked | Homoglyph |
| LATIN SMALL LETTER TURNED E | 01DD | ǝ | ↔ | ә | 04D9 | CYRILLIC SMALL LETTER SCHWA | Blocked | Homoglyph |
| LATIN SMALL LETTER SCHWA | 0259 | ə | ↔ | ә | 04D9 | CYRILLIC SMALL LETTER SCHWA | Blocked | Homoglyph |
| LATIN SMALL LETTER EZH | 0292 | ʒ | ↔ | ӡ | 04E1 | CYRILLIC SMALL LETTER ABKHASIAN DZE | Blocked | Homoglyph |

**[C1]** In certain fonts (e.g. Arial and Courier New, second and third rows respectively) the two glyphs look nearly identical. See screenshot below.



### Greek

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | ↔ | ϊ | 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; based on security |
| 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. |

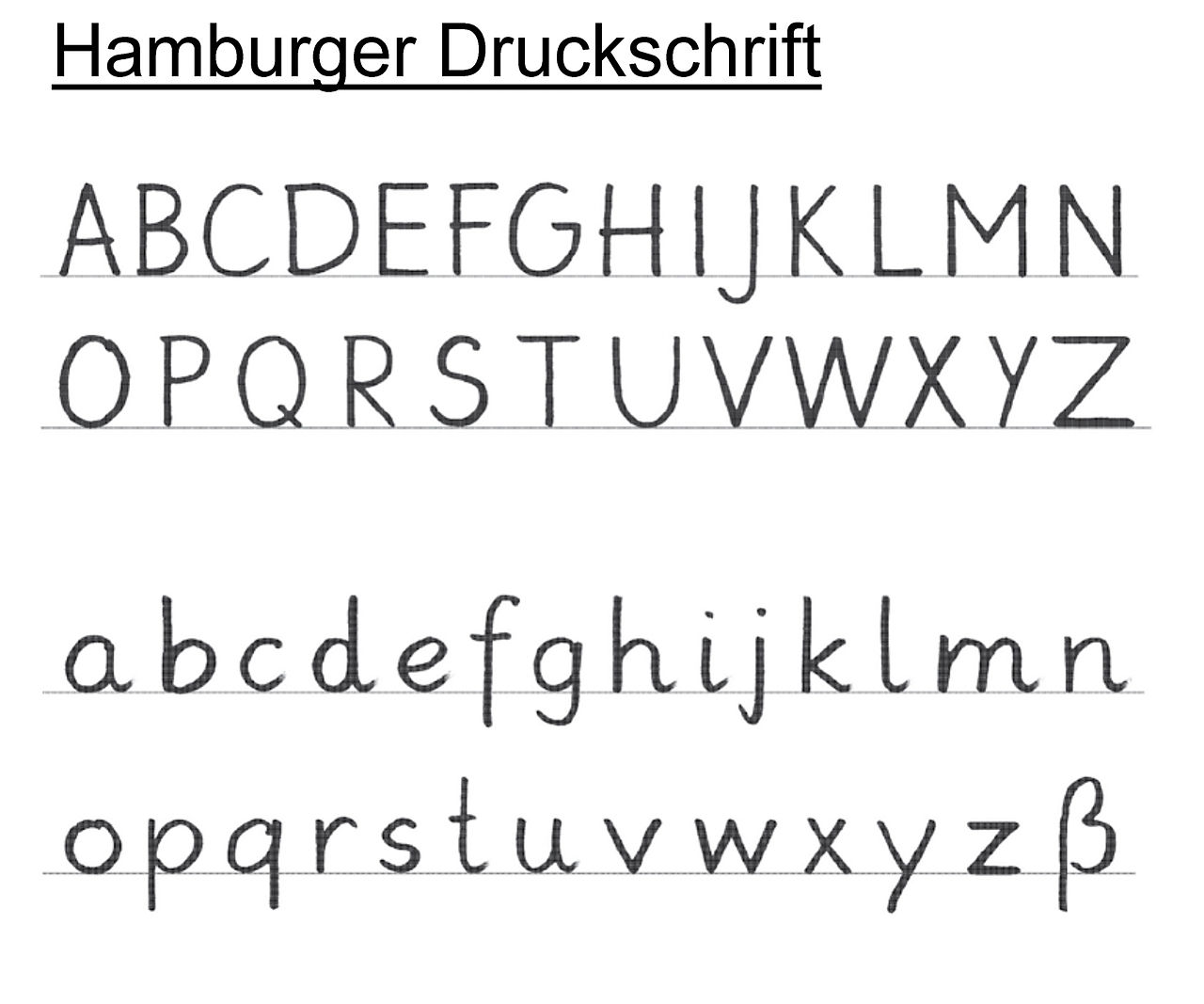
**[G1]** LATIN SMALL LETTER A 0061 and GREEK SMALL LETTER ALPHA 03B1 are visually different. However 03B1 - which is considered as equivalent to 0061 LATIN SMALL LETTER A by the Latin script-using community - is visually nearly identical to LATIN SMALL LETTER ALPHA 0251, in a significant number of font renderings, despite not being included in the current repertoire because of a lack of supporting evidence.

While Greek 03B1 is often set in cursive and has a continuing stroke on the top right-hand side, such Graphic cues are not retained across most bold, italic, and light fonts, as demonstrated by the following screenshot of renderings of the string 0061 + 0251 + 03B1 across a number of different fonts example.

Figure G01: Screenshot from Wordmark.it of a rendering of 0061 + 0251 + 03B1 using Google Fonts  
  
  
Accordingly, the visual differences between 0251 and 03B1 disappear in a significant number of fonts, which makes the codepoints visually nearly identical.

Furthermore 0061 is regularly represented using a glyph (nearly) identical with 0251 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 0251 large parts of the script-using community, and a shape similar to 0251 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>) demonstrated in Figure G02, which is the current standard repertoire for the German state of Hamburg (and which is currently being endorsed for use across all German states - similar glyphs are used by other German-speaking countries such as the Swiss Basisschrift - <https://www.basisschrift.ch/aufbau-und-didaktik>).

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>

By consequence, a significant number of fonts also represents 0061 with a glyph similar to 0251 as can be observed e.g. in the lower left-most rendering of Figure G01 above, and which would produce an in-script variant for Latin script, if 0251 was part of the suggested repertoire (which it is not because of a lack of supporting evidence - this produces however another in-script variant by proxy in between 00E6 LATIN SMALL LETTER AE and 0153 LATIN SMALL LIGATURE OE; (Section will be added).

Meanwhile, 0061 is considered the block- or print-letter shape to the hand-written shape of 0251 across significant if not most parts of the script-using community. For example, a shape similar to 0251 is used in standard primers and repertoire of handwriting as taught to school children, such as the Grundschrift (<https://en.wikipedia.org/wiki/Grundschrift>), which is the is a current standard repertoire for the German state of Hamburg (and which is currently being endorsed for use across all German states) and similar glyphs are used by other German-speaking countries such as the Swiss Basisschrift (<https://www.basisschrift.ch/aufbau-und-didaktik>).

This variation between glyphs similar to both 0061 and 0251 is encountered not only in a significant number of fonts in rare font designs, but is also constantly reiterated to the minds of readers by being featured prominently in also graphic design of logos of international brand names, such as:

* 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>)

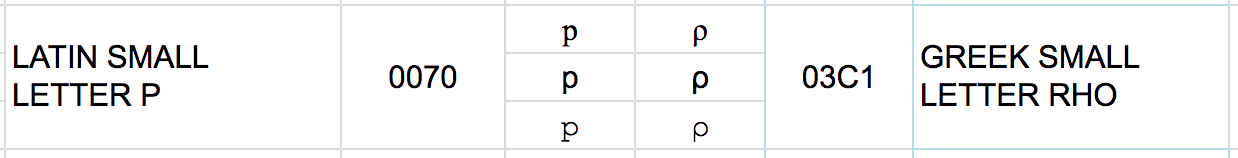
The variation between occurs also within the same logos

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

Also there are some historically established, as in typography employed in movies (cf. the initial “a” Paramount movie openers ([http://logos.wikia.comwiki/Paramount\_Cartoon\_Studios](http://logos.wikia.com/wiki/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 warrant a variant relationship.

In summary, accordingly, users of Latin script may not be able to differentiate 03B1 from 0251, and since they consider 0251 to be a non-visual variant of 0061, by transitivity 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 wide-spread fonts (such as Times New Roman and Courier New, first and third row respectively).  
  
   
  
In such cases, the two code-points are visually only distinguishable in context because of their respective positioning towards the baseline, whereas 0070 crosses below the baseline but 03C1 does not. Given that there are several variant candidates among the cross-script variants, plausible labels such as .ρορ, .ραγ , in which most Latin-script users would be hard-pressed to distinguish them in context.  
  
Furthermore, designers from the Latin-script using community have exploited the historical relation and visual similarity 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>).

This use in logo designs is not limited to the Pepsi-brand logo or the English-using community - cf.

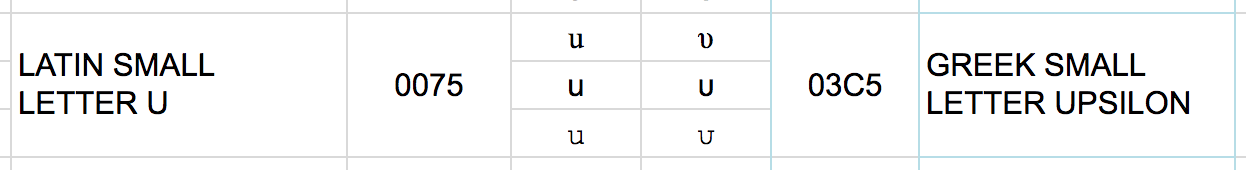
* <http://logos.wikia.com/wiki/Logopedia:Theme/Logos_with_the_letter_P?file=Publix_logo.png>,
* <http://logos.wikia.com/wiki/File:150px-Android_P_logo.png>
* <http://logos.wikia.com/wiki/File:Vpf.png>,
* <http://logos.wikia.com/wiki/Category:Red_PAT> -

and it is historically established – cf

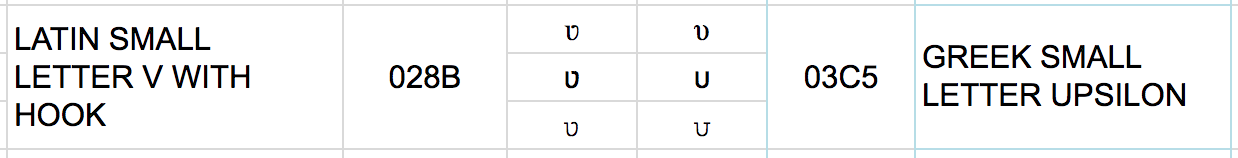
* <http://logos.wikia.com/wiki/File:Pba_83_on_city_2_Vintage_Sports.jpg>.

In 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 non-visual variant.

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



028B-03C5: The two glyphs look “nearly identical” in Times New Roman font (as shown in the first row in the image below). By the same principle, 00FA and 03CD are considered nearly identical



These two variant relationships impose an in-script variant relationship (through transitivity property) between 0075 and 028B (both Latin code points).

Mossi Language:

The Mossi – a language of Burkina Faso - uses both Latin U (0075) and Latin V with Hook (028B) in it’s alphabet (however French is the official language of the country, cf. [https](file:///Users/meikal/Downloads/https)[://en.wikipedia.org/wiki/Burkina\_Faso](https://en.wikipedia.org/wiki/Burkina_Faso)). Given that the variant relationship between U and V with Hook will have a disposition of “blocked” the Latin GP foresees no issues and accepts the imposed variant relationship between the aforementioned letters.

**[G4]** The Greek script codepoint 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 codepoints 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 codepoint 00DF ß across the German-speaking part of the script-using community, as demonstrated by Figure G03 below.

Figure G03: A handwritten form of the German Lexeme Grüße ‘Greetings’, taken from

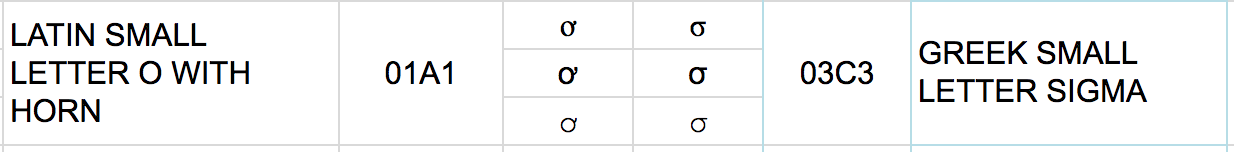
[https](file:///Users/meikal/Downloads/https)[://de.wikipedia.org/wiki/%C3%9F#/media/File:Gruesse-Schneidler-Legende.png](https://de.wikipedia.org/wiki/%C3%9F#/media/File:Gruesse-Schneidler-Legende.png) (Cf. e.g. <https://de.wikipedia.org/wiki/%C3%9F#/media/File:Gruesse-Schneidler-Legende.png>,

where the German lexem Grüße is spelled with such a variant).

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 codepoints which are variants of Latin codepoints used in German orthography, there are again plausible labels which could be constructed, such as Greek .νοβ, which may be identified with the common 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 is also affected by the in-script variant IDNA cases – c.f. Figure G03. Accordingly, there is specific risk 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 number of fonts.

**[G5]** In the Courier New font (third row in the picture) the glyphs are deemed nearly identical, difference due to font design. See Figure G04 below:

Figure G04: 01A1 vs. 03C3



## 

## 6.4 In-Script Latin variants imposed by cross-script variant relationships

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | ↔ | ʋ | 028B | LATIN SMALL LETTER V WITH HOOK | Blocked | In-script variant due to transitivity relationship of 03C5 Greek Small Letter Upsilon |

## 6.5 Methodology for developing in-script variants

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

* 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 studied both handwriting samples as well as font design. Handwriting samples were looked at to help shortlist certain candidates for a more careful analysis, but handwriting was never taken as a sole decisionmaker. 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.

The GP used the website<https://wordmark.it/> to compare strings across 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](https://fonts.google.com/), a font library employed by many APIs, instead of system fonts as rendered by that website.

Where shaping of base characters or diacritics were assumed to give way to variant candidates, strings containing the two code points, such as fƒ or vice versa, i.e. 0066 + 0192, or strings containing code points featuring the two diacritics, such as ăǎ 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 codepoints 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. ốoôó, i.e. 1ED1 + 006F + 00F4 + 00F3.

INSERT DISCUSSION ON METHODOLOGY FOR IDNA COMPATIBILITY CASES ABOUT HERE

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.

## 6.6 In-Script Latin Variants

Latin GP preliminarily identified the following variant candidates:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | Breve vs. Caron | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 0103 | ă | LATIN SMALL LETTER A WITH BREVE | 01CE | ǎ | LATIN SMALL LETTER A WITH CARON | YES | Blocked | Glyphs nearly identical due to font design |
| 011F | ğ | LATIN SMALL LETTER G WITH BREVE | 01E7 | ǧ | LATIN SMALL LETTER G WITH CARON | YES | Blocked | Glyphs nearly identical due to font design |
| 016D | ŭ | LATIN SMALL LETTER U WITH BREVE | 01D4 | ǔ | LATIN SMALL LETTER U WITH CARON | YES | Blocked | Glyphs nearly identical due to font design |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | Caron vs. Hook | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 010F | ď | LATIN SMALL LETTER D WITH CARON | 0257 | ɗ | LATIN SMALL LETTER D WITH HOOK | YES | Blocked | Glyphs nearly identical due to font design |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | Macron vs. Tilde | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 006E + 0304 | n̄ | LATIN SMALL LETTER N + COMBINING MACRON | 00F1 | ñ | LATIN SMALL LETTER N WITH TILDE | YES | Blocked | Glyphs nearly identical due to font design |

Comment: The tilde of 00F1 ñ LATIN SMALL LETTER N WITH TILDE is rendered as a macron in some fonts.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | Glyphs nearly identical due to font design |

Comment: Some fonts display the lower case characters in upper case instead, causing those two codepoints to become the visually the same.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | Double Acute vs. Diaresis | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 0151 | ő | LATIN SMALL LETTER O WITH DOUBLE ACUTE | 00F6 | ö | LATIN SMALL LETTER O WITH DIAERESIS | YES | Blocked | Glyphs nearly identical due to font design |
| 0171 | ű | LATIN SMALL LETTER U WITH DOUBLE ACUTE | 00FC | ü | LATIN SMALL LETTER U WITH DIAERESIS | YES | Blocked | Glyphs nearly identical due to font design |

Comment: Double acute is rendered nearly identical in a significant number of fonts, and readers of languages such as German consider them equivalent due to handwriting traditions.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | F. vs. F with hook | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 0066 | f | LATIN SMALL LETTER F | 0192 | ƒ | LATIN SMALL LETTER F WITH HOOK | YES | Blocked | Used interchangeably in languages such as Swedish |

Comment: Swedish uses a shape of “LATIN SMALL LETTER F” (U+0066) that is identical to “LATIN SMALL

LETTER F WITH HOOK” (U+0192) in italic style. Example from a large, daily newspaper, in

which all instances of “ƒ” are just variants of “f”.



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | ligature ae vs. ligature oe vs. sequence oe | | | | | | | |
| **Target** | | | **Source** | | | **Variant Candidate [Yes/No]** | **Disposition [Allocatable/Blocked]** | **Rationale** |
| **Code Point** | **Glyph** | **Name** | **Code Point** | **Glyph** | **Name** |
| 00E6 | æ | LATIN SMALL LETTER AE | 0153 | œ | LATIN SMALL LIGATURE OE | YES | Blocked | Glyphs nearly identical due to font design |
| 0153 | œ | LATIN SMALL LIGATURE OE | 006F+0065 | oe | LATIN SMALL LETTER O + LATIN SMALL LETTER E | YES |  | Glyphs nearly identical due to font design; Used interchangeably in English |

Comment: 0251 ɑ LATIN SMALL LETTER ALPHA and 0061 a 0061 LATIN SMALL LETTER A are considered equivalent by readers, however 0251 is not in repertoire because of a lack of evidence. Since the a-glyph in 00E6 æ LATIN SMALL LETTER AE takes the shape of 0251 in a significant number of fonts, it becomes visually confusable with 0153 œ LATIN SMALL LIGATURE OE.

Furthermore, 0153 œ LATIN SMALL LIGATURE OE i used interchangeably with a sequence 006F+0065 oe LATIN SMALL LETTER O + LATIN SMALL LETTER E in languages such as English. Furthermore, they are visually nearly identical in a the rendering by a significant number of fonts if the distance is disregarded.

### 6.7 Other Considerations for Variant Analysis (Placeholder)

### HTML Underlining

### IDNA2003 Compatibility

# 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. The only code points that need contextual restrictions are the non-space marks (se section 5.3.1). The restriction of those is that they are only allowed, in the Latin LRG, 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

# 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 / 㗂越), 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 (Sääˊmǩiõll / 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 (ʻŌlelo Hawaiʻi), http://www.omniglot.com/writing/hawaiian.htm, 4 September 2018

[136], Omniglot, Marshallese (kajin m̧ajeļ), 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 هَوْسَ), 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 (آذربايجانجا ديلي / 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], Cmpiled 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 (Deʼu), 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 مينڠكاباو), 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/10/58/27/10582785843693992331766506069073895620/40337\_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

## **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](https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-)

[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>

# 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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |

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

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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 | **ṱ** | 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

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Language** | **ISO 639-3** | **EGIDS** |
|  | [Afrikaans,](http://www.omniglot.com/writing/afrikaans.htm) | [afr](http://www-01.sil.org/iso639-3/documentation.asp?id=afr) | 1 |
|  | Albanian, Arbëreshë Albanian [aae] (Italy) Arvanitika Albanian [aat] (Greece) Gheg Albanian [aln] (Serbia) Tosk Albanian [als] | [sqi](http://www-01.sil.org/iso639-3/documentation.asp?id=sqi) | 1 |
|  | [Azeri, Azerbaijani](http://www.omniglot.com/writing/azeri.htm) | [azj](http://www-01.sil.org/iso639-3/documentation.asp?id=azj) | 1 |
|  | [Chamorro, Chamorru Tjamoro](http://www.omniglot.com/writing/chamorro.htm) | [cha](http://www-01.sil.org/iso639-3/documentation.asp?id=cha) | 1 |
|  | [Croatian, Hrvatski](http://www.omniglot.com/writing/croatian.htm) | [hrv](http://www-01.sil.org/iso639-3/documentation.asp?id=hrv) | 1 |
|  | [Czech Bohemian Cestina](http://www.omniglot.com/writing/czech.htm) | [ces](http://www-01.sil.org/iso639-3/documentation.asp?id=ces) | 1 |
|  | [Danish, Dansk Rigsdansk](http://www.omniglot.com/writing/danish.htm) | [dan](http://www-01.sil.org/iso639-3/documentation.asp?id=dan) | 1 |
|  | [Dutch, Hollands Nederlands](http://www.omniglot.com/writing/dutch.htm) | [nld](http://www-01.sil.org/iso639-3/documentation.asp?id=nld) | 1 |
|  | [English](http://www.omniglot.com/writing/english.htm) | [eng](http://www-01.sil.org/iso639-3/documentation.asp?id=eng) | 1 |
|  | [Estonian Eesti keel](http://www.omniglot.com/writing/estonian.htm) | [ekk](http://www-01.sil.org/iso639-3/documentation.asp?id=ekk) | 1 |
|  | [Filipino](http://www.omniglot.com/writing/filipino.htm) | [fil](http://www-01.sil.org/iso639-3/documentation.asp?id=fil) | 1 |
|  | [Finnish, Suomi](http://www.omniglot.com/writing/finnish.htm) | [fin](http://www-01.sil.org/iso639-3/documentation.asp?id=fin) | 1 |
|  | [French, Français](http://www.omniglot.com/writing/french.htm) | [fra](http://www-01.sil.org/iso639-3/documentation.asp?id=fra) | 1 |
|  | [German Deutsch Tedesco](http://www.omniglot.com/writing/german.htm) | [deu](http://www-01.sil.org/iso639-3/documentation.asp?id=deu) | 1 |
|  | [Greenlandic Kalaallisut, Inuktitut,](http://www.omniglot.com/writing/greenlandic.htm) | [kal](http://www-01.sil.org/iso639-3/documentation.asp?id=kal) | 1 |
|  | [Guarani Avañe’e Paraguayan](http://www.omniglot.com/writing/guarani.htm) | grn | 1 |
|  | [Haitian Creole, Creole, Haitian Creole Western Caribbean Creole](http://www.omniglot.com/writing/haitiancreole.htm) | [hat](http://www-01.sil.org/iso639-3/documentation.asp?id=hat) | 1 |
|  | [Hungarian Magyar](http://www.omniglot.com/writing/hungarian.htm) | [hun](http://www-01.sil.org/iso639-3/documentation.asp?id=hun) | 1 |
|  | [Icelandic Íslenska](http://www.omniglot.com/writing/icelandic.htm) | [isl](http://www-01.sil.org/iso639-3/documentation.asp?id=isl) | 1 |
|  | [Indonesian](http://www.omniglot.com/writing/indonesian.htm) | [ind](http://www-01.sil.org/iso639-3/documentation.asp?id=ind) | 1 |
|  | [Irish Erse Gaeilge Gaelic Irish](http://www.omniglot.com/writing/irish.htm) | [gle](http://www-01.sil.org/iso639-3/documentation.asp?id=gle) | 1 |
|  | [Italian Italiano](http://www.omniglot.com/writing/italian.htm) | [ita](http://www-01.sil.org/iso639-3/documentation.asp?id=ita) | 1 |
|  | [Kazakh, Kaisak, Kazak, Kosach, Qazaq](http://www.omniglot.com/writing/kazakh.htm) | [kaz](http://www-01.sil.org/iso639-3/documentation.asp?id=kaz) | 1 |
|  | [Kinyarwanda, Ikinyarwanda, Orunyarwanda, Ruanda, Rwandan, Urunyaruanda](http://www.omniglot.com/writing/kinyarwanda.htm) | [kin](http://www-01.sil.org/iso639-3/documentation.asp?id=kin) | 1 |
|  | [Kiribati, Gilbertese, Ikiribati, I-Kiribati, Kiribatese](http://www.omniglot.com/writing/kiribati.htm) | [gil](http://www-01.sil.org/iso639-3/documentation.asp?id=gil) | 1 |
|  | [Kirundi, Rundi Urundi,](http://www.omniglot.com/writing/kirundi.php) | [run](http://www-01.sil.org/iso639-3/documentation.asp?id=run) | 1 |
|  | [Latvian, “Lettisch” (pej.), “Lettish” (pej.)](http://www.omniglot.com/writing/latvian.htm) | lav | 1 |
|  | [Lithuanian, Lietuvi, Lietuviskai, Litauische, Litewski, Litovskiy](http://www.omniglot.com/writing/lithuanian.htm) | [lit](http://www-01.sil.org/iso639-3/documentation.asp?id=lit) | 1 |
|  | [Malagasy, Plateau, Malagasy, Malgache, Official Malagasy, Standard Malagasy](http://www.omniglot.com/writing/malagasy.htm) | [plt](http://www-01.sil.org/iso639-3/documentation.asp?id=plt) | 1 |
|  | [Malay,](http://www.omniglot.com/writing/malay.htm) | msa | 1 |
|  | [Maltese, Malti](http://www.omniglot.com/writing/maltese.htm) | [mlt](http://www-01.sil.org/iso639-3/documentation.asp?id=mlt) | 1 |
|  | [Marshallese, Ebon](http://www.omniglot.com/writing/marshallese.php) | [mah](http://www-01.sil.org/iso639-3/documentation.asp?id=mah) | 1 |
|  | [Ndebele, Isikhethu, IsiNdebele, Ndzundza, Nrebele, Southern Ndebele, Transvaal Ndebele](http://www.omniglot.com/writing/ndebele.php) | [nbl](http://www-01.sil.org/iso639-3/documentation.asp?id=nbl) | 1 |
|  | [Niuean, Niue, “Niuefekai” (pej.)](http://www.omniglot.com/writing/niuean.php) | [niu](http://www-01.sil.org/iso639-3/documentation.asp?id=niu) | 1 |
|  | [Northern Sotho, Pedi, Sepedi, Sesotho sa Leboa, Transvaal Sotho](http://www.omniglot.com/writing/nsotho.htm) | [nso](http://www-01.sil.org/iso639-3/documentation.asp?id=nso) | 1 |
|  | [Norwegian, Norsk](http://www.omniglot.com/writing/norwegian.htm) | [nor](http://www-01.sil.org/iso639-3/documentation.asp?id=nor) | 1 |
|  | [Papiamento, Papiamentu, Curaçoleño, Curassese, Papiamen, Papiamentoe](http://www.omniglot.com/writing/papiamento.php) | [pap](http://www-01.sil.org/iso639-3/documentation.asp?id=pap) | 1 |
|  | [Polish, Polnisch, Polski](http://www.omniglot.com/writing/polish.htm) | [pol](http://www-01.sil.org/iso639-3/documentation.asp?id=pol) | 1 |
|  | [Portuguese,](http://www.omniglot.com/writing/portuguese.htm) | [por](http://www-01.sil.org/iso639-3/documentation.asp?id=por) | 1 |
|  | [Romanian, Daco-Rumanian, Moldavian, Rumanian](http://www.omniglot.com/writing/romanian.htm) | [ron](http://www-01.sil.org/iso639-3/documentation.asp?id=ron) | 1 |
|  | [Samoan,](http://www.omniglot.com/writing/samoan.htm) | [smo](http://www-01.sil.org/iso639-3/documentation.asp?id=smo) | 1 |
|  | [Sango, Sangho](http://www.omniglot.com/writing/sango.php) | [sag](http://www-01.sil.org/iso639-3/documentation.asp?id=sag) | 1 |
|  | [Serbian, srpski, српски,](http://www.omniglot.com/writing/serbian.htm) | [srp](http://www-01.sil.org/iso639-3/documentation.asp?id=srp) | 1 |
|  | [Seychelles Creole, Seselwa Creole, Creole, Ilois, Kreol, Kreol Seselwa, Seselwa, Seychelles Creole French, Seychellois Creole](http://www.omniglot.com/writing/seselwa.htm) | [crs](http://www-01.sil.org/iso639-3/documentation.asp?id=crs) | 1 |
|  | [Slovak, Slovakian, Slovencina](http://www.omniglot.com/writing/slovak.htm) | [slk](http://www-01.sil.org/iso639-3/documentation.asp?id=slk) | 1 |
|  | [Slovenian, Slovenscina, Slovene](http://www.omniglot.com/writing/slovene.htm) | [slv](http://www-01.sil.org/iso639-3/documentation.asp?id=slv) | 1 |
|  | [Somali, Af-Maxaad Tiri, Af-Soomaali, Common Somali, Soomaaliga, Standard Somali](http://www.omniglot.com/writing/somali.htm) | [som](http://www-01.sil.org/iso639-3/documentation.asp?id=som) | 1 |
|  | [Southern Sotho, Sesotho, Sisutho, Souto, Suthu, Suto](http://www.omniglot.com/writing/sesotho.htm) | [sot](http://www-01.sil.org/iso639-3/documentation.asp?id=sot) | 1 |
|  | [Spanish, Castellano, Castilian, Español](http://www.omniglot.com/writing/spanish.htm) | [spa](http://www-01.sil.org/iso639-3/documentation.asp?id=spa) | 1 |
|  | [Swahili, Kisuaheli, Kiswahili](http://www.omniglot.com/writing/swahili.htm) | [swh](http://www-01.sil.org/iso639-3/documentation.asp?id=swh) | 1 |
|  | [Swati/Swazi, Isiswazi, Ngwane, Phuthi, Siswati, Swazi, Tekela, Tekeza](http://www.omniglot.com/writing/swati.php) | [ssw](http://www-01.sil.org/iso639-3/documentation.asp?id=ssw) | 1 |
|  | [Swedish, Ruotsi, Svenska](http://www.omniglot.com/writing/swedish.htm) | [swe](http://www-01.sil.org/iso639-3/documentation.asp?id=swe) | 1 |
|  | [Tahitian,](http://www.omniglot.com/writing/tahitian.htm) | [tah](http://www-01.sil.org/iso639-3/documentation.asp?id=tah) | 1 |
|  | [Tok Pisin, Melanesian English, Neomelanesian, New Guinea Pidgin English, Pidgin, Pisin](http://www.omniglot.com/writing/tokpisin.htm) | [tpi](http://www-01.sil.org/iso639-3/documentation.asp?id=tpi) | 1 |
|  | [Tongan, Tonga](http://www.omniglot.com/writing/tongan.htm) | [ton](http://www-01.sil.org/iso639-3/documentation.asp?id=ton) | 1 |
|  | [Tsonga, Shangaan, Shangana, Shitsonga, Thonga, Tonga, Xitsonga](http://www.omniglot.com/writing/tsonga.php) | [tso](http://www-01.sil.org/iso639-3/documentation.asp?id=tso) | 1 |
|  | [Tswana, Beetjuans, Chuana, Coana, Cuana, Sechuana, Setswana](http://www.omniglot.com/writing/tswana.php) | [tsn](http://www-01.sil.org/iso639-3/documentation.asp?id=tsn) | 1 |
|  | [Turkish, Anatolian, Türkçe, Türkisch](http://www.omniglot.com/writing/turkish.htm) | [tur](http://www-01.sil.org/iso639-3/documentation.asp?id=tur) | 1 |
|  | [Turkmen, Trukhmen, Trukhmeny, Turkmani, Turkmanian, Turkmenler, Turkomans](http://www.omniglot.com/writing/turkmen.htm) | [tuk](http://www-01.sil.org/iso639-3/documentation.asp?id=tuk) | 1 |
|  | [Uzbek, Özbek, Usbeki, Uzbak, Uzbeki](http://www.omniglot.com/writing/uzbek.htm) | [uzb](http://www-01.sil.org/iso639-3/documentation.asp?id=uzb) | 1 |
|  | [Venda, Chivenda, Tshivenda](http://www.omniglot.com/writing/venda.htm) | [ven](http://www-01.sil.org/iso639-3/documentation.asp?id=ven) | 1 |
|  | [Vietnamese, Annamese, Ching, Gin, Jing, Kinh, Viet](http://www.omniglot.com/writing/vietnamese.htm) | [vie](http://www-01.sil.org/iso639-3/documentation.asp?id=vie) | 1 |
|  | [Xhosa, “Cauzuh” (pej.), Isixhosa, Koosa, Xosa](http://www.omniglot.com/writing/xhosa.htm) | [xho](http://www-01.sil.org/iso639-3/documentation.asp?id=xho) | 1 |
|  | [Zulu, Isizulu, Zunda](http://www.omniglot.com/writing/zulu.htm) | [zul](http://www-01.sil.org/iso639-3/documentation.asp?id=zul) | 1 |
|  | [Basque, Euskara Euskera Vascuense](http://www.omniglot.com/writing/basque.htm) | [eus](http://www-01.sil.org/iso639-3/documentation.asp?id=eus) | 2 |
|  | [Catalan, Català Catalán Catalan-Valencian-Balear Catalonian Valencian](http://www.omniglot.com/writing/catalan.htm) | [cat](http://www-01.sil.org/iso639-3/documentation.asp?id=cat) | 2 |
|  | [Chechen, Galancho Nokchiin Muott Nokhchiin](http://www.omniglot.com/writing/chechen.htm) | [che](http://www-01.sil.org/iso639-3/documentation.asp?id=che) | 2 |
|  | [Chuukese Chuuk Lagoon Chuukese Ruk Truk Trukese](http://www.omniglot.com/writing/chuukese.htm) | [chk](http://www-01.sil.org/iso639-3/documentation.asp?id=chk) | 2 |
|  | [Faroese Føroyskt](http://www.omniglot.com/writing/faroese.htm) | [fao](http://www-01.sil.org/iso639-3/documentation.asp?id=fao) | 2 |
|  | [Frisian Fries Frysk](http://www.omniglot.com/writing/westfrisian.htm) | [fry](http://www-01.sil.org/iso639-3/documentation.asp?id=fry) | 2 |
|  | [Galician Galego Gallego](http://www.omniglot.com/writing/galician.htm) | [glg](http://www-01.sil.org/iso639-3/documentation.asp?id=glg) | 2 |
|  | [Garo Garrow Mande Mandi](http://www.omniglot.com/writing/garo.htm) | [grt](http://www-01.sil.org/iso639-3/documentation.asp?id=grt) | 2 |
|  | [Hausa Abakwariga Habe Haoussa Hausawa Kado Mgbakpa](http://www.omniglot.com/writing/hausa.htm) | [hau](http://www-01.sil.org/iso639-3/documentation.asp?id=hau) | 2 |
|  | [Hawaiian Olelo Hawai’i ’Olelo Hawai’i Makuahine](http://www.omniglot.com/writing/hawaiian.htm) | [haw](http://www-01.sil.org/iso639-3/documentation.asp?id=haw) | 2 |
|  | [Igbo](http://www.omniglot.com/writing/igbo.htm) | [ibo](http://www-01.sil.org/iso639-3/documentation.asp?id=ibo) | 2 |
|  | [Inari Sámi Anarâškielâ Anar “Finnish Lapp” (pej.) “Inari Lappish” (pej.) “Lapp” (pej.) Saami Saame Sámi Samic](http://www.omniglot.com/writing/inarisami.htm) | [smn](http://www-01.sil.org/iso639-3/documentation.asp?id=smn) | 2 |
|  | [Konkani, Bankoti, Central Konkan, Concorinum, Cugani, Kathodi, Katvadi, Konkan Standard, Konkanese, Konkani Mangalorean, Kunabi, North Konkan](http://www.omniglot.com/writing/konkani.htm) | [knn](http://www-01.sil.org/iso639-3/documentation.asp?id=knn) | 2 |
|  | [Kurdish,](http://www.omniglot.com/writing/kurdish.htm) | [kur](http://www-01.sil.org/iso639-3/documentation.asp?id=kur) | 2 |
|  | [Lingala, Ngala](http://www.omniglot.com/writing/lingala.htm) | [lin](http://www-01.sil.org/iso639-3/documentation.asp?id=lin) | 2 |
|  | [Lule Sámi, “Lapp” (pej.), Lule, Saami](http://www.omniglot.com/writing/lulesami.htm) | [smj](http://www-01.sil.org/iso639-3/documentation.asp?id=smj) | 2 |
|  | [Mirandese, Mirandês](http://www.omniglot.com/writing/mirandese.htm) | [mwl](http://www-01.sil.org/iso639-3/documentation.asp?id=mwl) | 2 |
|  | [Miskito, Marquito, Mískitu, Miskuto, Mísquito, Mosquito](http://www.omniglot.com/writing/miskito.htm) | [miq](http://www-01.sil.org/iso639-3/documentation.asp?id=miq) | 2 |
|  | [Northern Sámi, Saami North, “Lapp” (pej.), North Sámi, “Northern Lappish” (pej.), Northern Saami, “Norwegian Lapp” (pej.), Saami, Same, Sámegiella, Samic](http://www.omniglot.com/writing/northernsami.htm) | [sme](http://www-01.sil.org/iso639-3/documentation.asp?id=sme) | 2 |
|  | [Palauan, Belauan, Palau](http://www.omniglot.com/writing/palauan.htm) | [pau](http://www-01.sil.org/iso639-3/documentation.asp?id=pau) | 2 |
|  | [Pohnpeian, Ponapean](http://www.omniglot.com/writing/pohnpeian.htm) | [pon](http://www-01.sil.org/iso639-3/documentation.asp?id=pon) | 2 |
|  | [Skolt Sámi, “Lapp” (pej.), Southern Lapp](http://www.omniglot.com/writing/skoltsami.htm) | [sma](http://www-01.sil.org/iso639-3/documentation.asp?id=sma) | 2 |
|  | [Tatar, Tartar](http://www.omniglot.com/writing/tatar.htm) | [tat](http://www-01.sil.org/iso639-3/documentation.asp?id=tat) | 2 |
|  | [Tshiluba, Luba-Kasai, Bena-Lulua, Ciluba, Luba-Lulua, Luva, Tshiluba, Western Luba](http://www.omniglot.com/writing/tshiluba.php) | [lua](http://www-01.sil.org/iso639-3/documentation.asp?id=lua) | 2 |
|  | [Uyghur, Uighuir, Uighur, Uiguir, Uigur, Uygur, Weiwu’er, Wiga](http://www.omniglot.com/writing/uyghur.htm) | [uig](http://www-01.sil.org/iso639-3/documentation.asp?id=uig) | 2 |
|  | [Wa, Paruk, Baraog, Phalok, Praok, Standard Wa, Wa](http://www.omniglot.com/writing/wa.htm) | [prk](http://www-01.sil.org/iso639-3/documentation.asp?id=prk) | 2 |
|  | [Welsh, Cymraeg](http://www.omniglot.com/writing/welsh.htm) | [cym](http://www-01.sil.org/iso639-3/documentation.asp?id=cym) | 2 |
|  | [West Frisian, Fries, Frysk](http://www.omniglot.com/writing/westfrisian.htm) | [fry](http://www-01.sil.org/iso639-3/documentation.asp?id=fry) | 2 |
|  | [Yapese,](http://www.omniglot.com/writing/yapese.htm) | [yap](http://www-01.sil.org/iso639-3/documentation.asp?id=yap) | 2 |
|  | [Yoruba, Yariba, Yooba](http://www.omniglot.com/writing/yoruba.htm) | [yor](http://www-01.sil.org/iso639-3/documentation.asp?id=yor) | 2 |
|  | Akan, Twi, Ajan Twi | [aka](http://www-01.sil.org/iso639-3/documentation.asp?id=aka) | 3 |
|  | Bislama, Bichelamar | [bis](http://www-01.sil.org/iso639-3/documentation.asp?id=bis) | 3 |
|  | [Bugis Basa Ugi Boegineesche Boeginezen Bugi Buginese De’ Rappang Buginese Ugi](http://www.omniglot.com/writing/bugis.htm) | [bug](http://www-01.sil.org/iso639-3/documentation.asp?id=bug) | 3 |
|  | Cebuano, Binisaya Bisayan Sebuano Sugbuanon Sugbuhanon Visayan | [ceb](http://www-01.sil.org/iso639-3/documentation.asp?id=ceb) | 3 |
|  | [Chichewa Chewa Chinyanja Nyanja Nyanja-Chewa](http://www.omniglot.com/writing/chichewa.php) | [nya](http://www-01.sil.org/iso639-3/documentation.asp?id=nya) | 3 |
|  | Cubeo Cuveo Hehenawa Hipnwa Kobeua Kobewa Kubwa Pamiwa | [cub](http://www-01.sil.org/iso639-3/documentation.asp?id=cub) | 3 |
|  | [Duala Diwala Douala Dualla Dwala Dwela Sawa](http://www.omniglot.com/writing/duala.php) | [dua](http://www-01.sil.org/iso639-3/documentation.asp?id=dua) | 3 |
|  | [Esperanto](http://www.omniglot.com/writing/esperanto.htm) | [epo](https://iso639-3.sil.org/code/epo) | 3 |
|  | [Ewe Ebwe Efe Eibe Eue Eve Gbe Krepe Krepi Popo Vhe Eʋegbe](http://www.omniglot.com/writing/ewe.htm) | [ewe](http://www-01.sil.org/iso639-3/documentation.asp?id=ewe) | 3 |
|  | [Ewondo Ewundu Jaunde Yaounde Yaunde](http://www.omniglot.com/writing/ewondo.php) | [ewo](http://www-01.sil.org/iso639-3/documentation.asp?id=ewo) | 3 |
|  | [Fanagalo Fanakalo Pidgin Zulu Fanekolo Isikula Lololo or Isilololo Piki or Isipiki Silunguboi, Chilapalapa Cikabanga](http://www.omniglot.com/writing/fanagalo.htm) | [fng](http://www-01.sil.org/iso639-3/documentation.asp?id=fng) | 3 |
|  | [Fon Dahomeen Fongbe](http://www.omniglot.com/writing/fon.htm) | [fon](http://www-01.sil.org/iso639-3/documentation.asp?id=fon) | 3 |
|  | [Fula(ni), Fulfulde Pulaar Pular' Fulaare](http://www.omniglot.com/writing/fula.htm) | [fuv](http://www-01.sil.org/iso639-3/documentation.asp?id=fuv) | 3 |
|  | [Ganda Luganda](http://www.omniglot.com/writing/ganda.php) | [lug](http://www-01.sil.org/iso639-3/documentation.asp?id=lug) | 3 |
|  | [Hiligaynon Hiligainon Illogo Ilonggo](http://www.omniglot.com/writing/hiligaynon.htm) | [hil](http://www-01.sil.org/iso639-3/documentation.asp?id=hil) | 3 |
|  | [Iban Dayak](http://www.omniglot.com/writing/iban.htm) | [iba](http://www-01.sil.org/iso639-3/documentation.asp?id=iba) | 3 |
|  | [IlokoIlokano Ilocano](http://www.omniglot.com/writing/ilocano.htm) | [ilo](http://www-01.sil.org/iso639-3/documentation.asp?id=ilo) | 3 |
|  | [Kanuri,](http://www.omniglot.com/writing/kanuri.htm) | [kau](http://www-01.sil.org/iso639-3/documentation.asp?id=kau) | 3 |
|  | [Kapampangan, Pampangan, Pampango, Pampangueño, Capampangan, Amanung Sisuan](http://www.omniglot.com/writing/kapampangan.php) | [pam](http://www-01.sil.org/iso639-3/documentation.asp?id=pam) | 3 |
|  | [Latin, Latina](http://www.omniglot.com/writing/latin2.htm) | [let](http://www-01.sil.org/iso639-3/documentation.asp?id=lat) | 3 |
|  | [Manado Malay, Manadonese, Manadonese Malay, Minahasan Malay](http://www.omniglot.com/writing/manadomalay.htm) | [xmm](http://www-01.sil.org/iso639-3/documentation.asp?id=xmm) | 3 |
|  | [Masbateño, Masbatenyo, Minasbate](http://www.omniglot.com/writing/masbateno.htm) | [msb](http://www-01.sil.org/iso639-3/documentation.asp?id=msb) | 3 |
|  | [Mossi, Mole, Moose, More, Moshi, Mossi](http://www.omniglot.com/writing/mossi.htm) | [mos](http://www-01.sil.org/iso639-3/documentation.asp?id=mos) | 3 |
|  | [Nagamese, Bodo, Kachari Bengali, Naga Creole Assamese, Naga-Assamese, Naga Pidgin](http://www.omniglot.com/writing/nagamese.php) | [nag](http://www-01.sil.org/iso639-3/documentation.asp?id=nag) | 3 |
|  | Nauruan | [nau](http://www-01.sil.org/iso639-3/documentation.asp?id=nau) | 3 |
|  | [OshiWambo, Cuanhama, Humba, Kuanjama, Kwancama, Kwanjama, Kwanyama, Ochikwanyama, Oshikuanjama, Oshikwanyama, Ovambo, Oxikuanyama, Wambo](http://www.omniglot.com/writing/oshiwambo.php) | [kua](http://www-01.sil.org/iso639-3/documentation.asp?id=kua) | 3 |
|  | Pangasinan | [pag](http://www-01.sil.org/iso639-3/documentation.asp?id=pag) | 3 |
|  | [Pijin, Neo-Solomonic, Solomons Pidgin](http://www.omniglot.com/writing/pijin.htm) | [pis](http://www-01.sil.org/iso639-3/documentation.asp?id=pis) | 3 |
|  | [Quechua, Runasimi , Qhichwa simi](http://www.omniglot.com/writing/quechua.htm) | [que](http://www-01.sil.org/iso639-3/documentation.asp?id=que) | 3 |
|  | [Raga, Hano, Bwatvenua, Lamalanga, North Raga, Qatvenua, Raga, Vunmarama](http://www.omniglot.com/writing/raga.htm) | [lml](http://www-01.sil.org/iso639-3/documentation.asp?id=lml) | 3 |
|  | Roviana, Robiana, Rubiana, Ruviana | [rug](http://www-01.sil.org/iso639-3/documentation.asp?id=rug) | 3 |
|  | [Shona, Chishona, “Swina” (pej.), Zezuru](http://www.omniglot.com/writing/shona.php) | [sna](http://www-01.sil.org/iso639-3/documentation.asp?id=sna) | 3 |
|  | [Sranan, Sranan Tongo, Surinaams, Suriname Creole English, Surinamese, Taki-Taki](http://www.omniglot.com/writing/sranan.htm) | [srn](http://www-01.sil.org/iso639-3/documentation.asp?id=srn) | 3 |
|  | [Tagalog,](http://www.omniglot.com/writing/tagalog.htm) | [tgl](http://www-01.sil.org/iso639-3/documentation.asp?id=tgl) | 3 |
|  | [Tausūg, Bahasa Sug, Moro Joloano, Sinug, Sulu, Suluk, Tausog, Taw Sug](http://www.omniglot.com/writing/tausug.htm) | [tsg](http://www-01.sil.org/iso639-3/documentation.asp?id=tsg) | 3 |
|  | [Torres-Strait Creole, Ap-Ne-Ap, Blaik, Broken, Cape York Creole, Creole, Torres Strait Broken, Torres Strait Pidgin English, West Torres, Yumplatok](http://www.omniglot.com/writing/tsc.htm) | [tcs](http://www-01.sil.org/iso639-3/documentation.asp?id=tcs) | 3 |
|  | [Tuvaluan, Ellice, Ellicean, Tuvalu](http://www.omniglot.com/writing/tuvaluan.htm) | [tvl](http://www-01.sil.org/iso639-3/documentation.asp?id=tvl) | 3 |
|  | [Umbundu, Kimbari, Mbali, Mbari, M’bundo, Mbundu, Mbundu Benguella, Nano, Olumbali, Ovimbundu, South Mbundu, Umbundo](http://www.omniglot.com/writing/umbundu.htm) | [umb](http://www-01.sil.org/iso639-3/documentation.asp?id=umb) | 3 |
|  | [Waray-Waray, Binisaya, Samaran, Samareño, Samarenyo, Samar-Leyte, Waray](http://www.omniglot.com/writing/waray.php) | [war](http://www-01.sil.org/iso639-3/documentation.asp?id=war) | 3 |
|  | [Wolaytta, Borodda, Ometo, Ualamo, Uba, Uollamo, “Walamo” (pej.), Wallamo, Welamo, Wellamo, Wolaita, Wolaitta, Wolataita, Wolayta, Wollamo](http://www.omniglot.com/writing/wolaytta.htm) | [wal](http://www-01.sil.org/iso639-3/documentation.asp?id=wal) | 3 |
|  | [Zhuang, Nong](http://www.omniglot.com/writing/zhuang.htm) | [zha](http://www-01.sil.org/iso639-3/documentation.asp?id=zha) | 3 |
|  | [Adzera, Atzera, Azera, Atsera or Acira,](http://www.omniglot.com/writing/adzera.htm) | [adz](http://www-01.sil.org/iso639-3/documentation.asp?id=adz) | 4 |
|  | [Aklan, Aklan, Aklanon or AkeanonInakeanon (native)](http://www.omniglot.com/writing/aklan.htm) | [akl](http://www-01.sil.org/iso639-3/documentation.asp?id=akl) | 4 |
|  | [Arrernte, Arunta, Eastern Aranda, Upper Aranda](http://www.omniglot.com/writing/arrernte.htm) | [aer](http://www-01.sil.org/iso639-3/documentation.asp?id=aer) | 4 |
|  | [Bambara, Bamanankan](http://www.omniglot.com/writing/bambara.htm) | [bam](http://www-01.sil.org/iso639-3/documentation.asp?id=bam) | 4 |
|  | [BashkirBashkir Bashqort Basquort](http://www.omniglot.com/writing/bashkir.htm) | [bak](http://www-01.sil.org/iso639-3/documentation.asp?id=bak) | 4 |
|  | [Cape Verdean Creole, Creole, Kriol, “Badiu” (pej.), Caboverdiano, Criol, Crioulo, Kriol, Krioulo, Krioulu, “Sampadjudu” (pej.), Kabuverdianu](http://www.omniglot.com/writing/kriol.php) | [kea](http://www-01.sil.org/iso639-3/documentation.asp?id=kea) | 4 |
|  | [Central Sinama, “Bajaw” (pej.) Central Sinama Orang Laut Sama Dilaut Samal Siasi Sama Sinama](http://www.omniglot.com/writing/centralsinama.htm) | [sml](http://www-01.sil.org/iso639-3/documentation.asp?id=sml) | 4 |
|  | [Chavacano, Chabacano Chabakano Zamboangueño](http://www.omniglot.com/writing/chavacano.php) | [cbk](http://www-01.sil.org/iso639-3/documentation.asp?id=cbk) | 4 |
|  | [CorsicanCorse Corsi Corso Corsu](http://www.omniglot.com/writing/corsican.htm) | [cos](http://www-01.sil.org/iso639-3/documentation.asp?id=cos) | 4 |
|  | [DagaareDagaare Dagara Dagare Dagari Dagati Degati Dogaari Southern Dagari](http://www.omniglot.com/writing/dagaare.htm) | [dga](http://www-01.sil.org/iso639-3/documentation.asp?id=dga) | 4 |
|  | [DagbaniDagbamba Dagbane Dagomba](http://www.omniglot.com/writing/dagbani.htm) | [dag](http://www-01.sil.org/iso639-3/documentation.asp?id=dag) | 4 |
|  | Dinka, Padang White Nile Dinka Agar Central Dinka Bor Cam Dinka Bor Eastern Dinka Rek Western Dinka | [din](http://www-01.sil.org/iso639-3/documentation.asp?id=din) | 4 |
|  | [DrehuDehu De’u Lifou Lifu Qene Drehu](http://www.omniglot.com/writing/drehu.php) | [dhv](http://www-01.sil.org/iso639-3/documentation.asp?id=dhv) | 4 |
|  | [FijianBoumaa Fijian Eastern Fijian Fiji Standard Fijian](http://www.omniglot.com/writing/fijian.htm) | [fij](http://www-01.sil.org/iso639-3/documentation.asp?id=fij) | 4 |
|  | [Friulian, Frioulan Frioulian Friulano Furlan Priulian](http://www.omniglot.com/writing/friulian.htm) | [fur](http://www-01.sil.org/iso639-3/documentation.asp?id=fur) | 4 |
|  | [Ga Accra Acra Amina Gain](http://www.omniglot.com/writing/ga.htm) | [gaa](http://www-01.sil.org/iso639-3/documentation.asp?id=gaa) | 4 |
|  | [HixkaryanaChawiyana Faruaru Hichkaryana Hishkariana Hishkaryana Hixkariana Hyxkaryana Kumiyana Parucutu Parukoto-Charuma Sherewyana Sokaka Wabui Xereu Xerewyana](http://www.omniglot.com/writing/hixkaryana.htm) | [hix](http://www-01.sil.org/iso639-3/documentation.asp?id=hix) | 4 |
|  | [Ifugao, Ifugaw, Mayaoyaw, Mayoyao](http://www.omniglot.com/writing/ifugao.htm) | [ifu](http://www-01.sil.org/iso639-3/documentation.asp?id=ifu) | 4 |
|  | [Ixil](http://www.omniglot.com/writing/ixil.htm) | [ixl](http://www-01.sil.org/iso639-3/documentation.asp?id=ixl) | 4 |
|  | [JavaneseDjawa Jawa](http://www.omniglot.com/writing/javanese.htm) | [jav](http://www-01.sil.org/iso639-3/documentation.asp?id=jav) | 4 |
|  | [Kagayanen, Cagayano, Kagay-anen, Kinagayanen](http://www.omniglot.com/writing/kagayanen.htm) | [cgc](http://www-01.sil.org/iso639-3/documentation.asp?id=cgc) | 4 |
|  | [Kaqchikel, Cakchiquel, Kaqchikel, Kaqchiquel](http://www.omniglot.com/writing/kaqchikel.htm) | [cak](http://www-01.sil.org/iso639-3/documentation.asp?id=cak) | 4 |
|  | [Khoekhoe, Bergdamara, “Hottentot” (pej.), Khoekhoegowab, Khoekhoegowap, Maqua, Nama, Namakwa, Naman, Namaqua, Tama, Tamakwa, Tamma](http://www.omniglot.com/writing/khoekhoe.htm) | [naq](http://www-01.sil.org/iso639-3/documentation.asp?id=naq) | 4 |
|  | [Ki'che', Central K’iche’, Central Quiché, Chiquel, Qach’abel, Quiché](http://www.omniglot.com/writing/kiche.htm) | [quc](http://www-01.sil.org/iso639-3/documentation.asp?id=quc) | 4 |
|  | [Lozi, Kololo, Kolololo, Rotse, Rozi, Rutse, Silozi, Tozvi](http://www.omniglot.com/writing/lozi.htm) | [loz](http://www-01.sil.org/iso639-3/documentation.asp?id=loz) | 4 |
|  | [Luxembourgish, Frankish, Letzburgisch, Lëtzebuergesch, Luxembourgeois, Luxemburgian, Luxemburgish, Moselle Franconian](http://www.omniglot.com/writing/luxembourgish.htm) | [ltz](http://www-01.sil.org/iso639-3/documentation.asp?id=ltz) | 4 |
|  | [Mam, Huehuetenango Mam](http://www.omniglot.com/writing/mam.htm) | [mam](http://www-01.sil.org/iso639-3/documentation.asp?id=mam) | 4 |
|  | [Maranao, Maranaw, Ranao](http://www.omniglot.com/writing/maranao.htm) | [mrw](http://www-01.sil.org/iso639-3/documentation.asp?id=mrw) | 4 |
|  | [Mbula, Kaimanga, Mangaaba, Mangaava, Mangaawa, Mangap, Mangap-Mbula](http://www.omniglot.com/writing/mbula.htm) | [mna](http://www-01.sil.org/iso639-3/documentation.asp?id=mna) | 4 |
|  | [Mizo, Duhlian Twang, Dulien, Hualngo, Lukhai, Lusago, Lusai, Lusei, Lushai, Lushai-Mizo, Lushei, Sailau, Whelngo](http://www.omniglot.com/writing/mizo.htm) | [lus](http://www-01.sil.org/iso639-3/documentation.asp?id=lus) | 4 |
|  | [Nuer, Naadh, Naath](http://www.omniglot.com/writing/nuer.htm) | [nus](http://www-01.sil.org/iso639-3/documentation.asp?id=nus) | 4 |
|  | [Nuosu (Yi), Black Yi, Liangshan Yi, Northern Yi, Nosu Yi, Sichuan Yi](http://www.omniglot.com/writing/yi.htm) | [iii](http://www-01.sil.org/iso639-3/documentation.asp?id=iii) | 4 |
|  | [Pitjantjatjara, Pitjantjara](http://www.omniglot.com/writing/pitjantjatjara.htm) | [pjt](http://www-01.sil.org/iso639-3/documentation.asp?id=pjt) | 4 |
|  | [Q'eqchi', Cacche’, Kekchi’, Kekchí, Ketchi’, Quecchi’](http://www.omniglot.com/writing/qeqchi.htm) | [kek](http://www-01.sil.org/iso639-3/documentation.asp?id=kek) | 4 |
|  | [Romansh, Rhaeto-Romance, Rheto-Romance, Romanche, Romansh, Rumantsch](http://www.omniglot.com/writing/romansh.htm) | [roh](http://www-01.sil.org/iso639-3/documentation.asp?id=roh) | 4 |
|  | [Scottish Gaelic, Gaelic-Scotish](http://www.omniglot.com/writing/gaelic.htm) | [gla](http://www-01.sil.org/iso639-3/documentation.asp?id=gla) | 4 |
|  | [Shavante, Xavante, Akuên, Akwen, A’uwe Uptabi, A’we, Chavante, Crisca, Pusciti, Shavante, Tapacua](http://www.omniglot.com/writing/shavante.php) | [xav](http://www-01.sil.org/iso639-3/documentation.asp?id=xav) | 4 |
|  | [Sorbian, Haut Sorabe, Hornjoserbski, Hornoserbski, Obersorbisch, Upper Lusatian, Wendish](http://www.omniglot.com/writing/sorbian.htm) | [hsb](http://www-01.sil.org/iso639-3/documentation.asp?id=hsb) | 4 |
|  | [Susu, Sose, Soso, Soussou, Susoo](http://www.omniglot.com/writing/susu.htm) | [sus](http://www-01.sil.org/iso639-3/documentation.asp?id=sus) | 4 |
|  | [Tagabawà, Tagabawa Bagobo, Tagabawa Manobo](http://www.omniglot.com/writing/tagabawa.htm) | [bgs](http://www-01.sil.org/iso639-3/documentation.asp?id=bgs) | 4 |
|  | [Talysh, Talesh, Talish, Talyshi](http://www.omniglot.com/writing/talysh.htm) | [tly](http://www-01.sil.org/iso639-3/documentation.asp?id=tly) | 4 |
|  | [Tumbuka, Chitumbuka, Citumbuka, Tamboka, Tambuka, Timbuka, Tombucas, Tumboka](http://www.omniglot.com/writing/tumbuka.htm) | [tum](http://www-01.sil.org/iso639-3/documentation.asp?id=tum) | 4 |
|  | [Tuvan, Tuva, Diba, Kök Mungak, Soyod, Soyon, Soyot, Tannu-Tuva, Tofa, Tokha, Tuba, Tuvan, Tuvia, Tuvin, Tuvinian, Tyva, Uriankhai, Uriankhai-Monchak, Uryankhai](http://www.omniglot.com/writing/tuvan.php) | [tyv](http://www-01.sil.org/iso639-3/documentation.asp?id=tyv) | 4 |
|  | [Wolof, Ouolof, Volof, Walaf, Waro-Waro, Yallof](http://www.omniglot.com/writing/wolof.htm) | [wol](http://www-01.sil.org/iso639-3/documentation.asp?id=wol) | 4 |
|  | [Zarma, Adzerma, Djerma, Dyabarma, Dyarma, Dyerma, Zabarma, Zarbarma, Zarmaci](http://www.omniglot.com/writing/zarma.htm) | [dje](http://www-01.sil.org/iso639-3/documentation.asp?id=dje) | 4 |
|  | [Zazaki, Northern, Alevica, Dersimki, Dimilki, Kirmanjki, Northern Zaza, So-Bê, Zaza, Zonê Ma](http://www.omniglot.com/writing/zazaki.htm) | [kiu](http://www-01.sil.org/iso639-3/documentation.asp?id=kiu) | 4 |
|  | Acehnese, Achehnese AchineseAceh | [ace](http://www-01.sil.org/iso639-3/documentation.asp?id=ace) | 5 |
|  | Acholi, Acoli Acooli Akoli Atscholi Dok Acoli Gang Lëbacoli Log Acoli Lwo Lwoo Shuli | [ach](http://www-01.sil.org/iso639-3/documentation.asp?id=ach) | 5 |
|  | [Afaan Oromooromo Oromiffa “Galla” (pej.) “Galligna” (pej.) “Gallinya” (pej.) Southern Oromo](http://www.omniglot.com/writing/oromo.htm) | [orm](http://www-01.sil.org/iso639-3/documentation.asp?id=orm) | 5 |
|  | [Afar, Adal, ’Afar Af, Afaraf, “Danakil” (pej.), “Denkel” (pej.), Qafar](http://www.omniglot.com/writing/afar.htm) | [aar](http://www-01.sil.org/iso639-3/documentation.asp?id=aar) | 5 |
|  | Alsatian, Elsässerdeutsche Alsacien Alemanic Alemannisch Schwyzerdütsch | [gsw](http://www-01.sil.org/iso639-3/documentation.asp?id=gsw) | 5 |
|  | [Alur, Aloro, Alua, Alulu, Dho Alur, Jo Alur, Lur, Luri](http://www.omniglot.com/writing/alur.htm) | [alz](http://www-01.sil.org/iso639-3/documentation.asp?id=alz) | 5 |
|  | Bavarian, Bairisch Bavarian Austrian Bayerisch Ost-Oberdeutsch | [bar](http://www-01.sil.org/iso639-3/documentation.asp?id=bar) | 5 |
|  | Brahui, Birahui Brahuidi Brahuigi Kur Galli | [brh](http://www-01.sil.org/iso639-3/documentation.asp?id=brh) | 5 |
|  | [Dholuo Kavirondo Luo Luo Nilotic Kavirondo](http://www.omniglot.com/writing/dholuo.php) | [luo](http://www-01.sil.org/iso639-3/documentation.asp?id=luo) | 5 |
|  | [JamaicanBongo Talk Jamiekan Limon Creole English Patois Patwa Quashie Talk Western Caribbean Creole](http://www.omniglot.com/writing/jamaican.php) | [jam](http://www-01.sil.org/iso639-3/documentation.asp?id=jam) | 5 |
|  | [Kabyle, Amazigh, Kabyl, Kabylia, Tamazight, Taqbaylit](http://www.omniglot.com/writing/kabyle.php) | [kbp](http://www-01.sil.org/iso639-3/documentation.asp?id=kbp) | 5 |
|  | [Kikuyu, Gĩkũyũ, Gekoyo, Gigikuyu,](http://www.omniglot.com/writing/kikuyu.htm) | [kik](http://www-01.sil.org/iso639-3/documentation.asp?id=kik) | 5 |
|  | [Low Saxon, Low German, Nedderdütsch, Neddersassisch, Nedersaksisch, Niederdeutsch, Niedersaechsisch, Plattdeutsch, Plattdüütsch](http://www.omniglot.com/writing/lowgerman.htm) | [nds](http://www-01.sil.org/iso639-3/documentation.asp?id=nds) | 5 |
|  | [Maasai, Maa, Masai](http://www.omniglot.com/writing/maasai.htm) | [mas](http://www-01.sil.org/iso639-3/documentation.asp?id=mas) | 5 |
|  | [Madurese, Madura, Basa Mathura](http://www.omniglot.com/writing/madurese.htm) | [mad](http://www-01.sil.org/iso639-3/documentation.asp?id=mad) | 5 |
|  | [Makhuwa, Central Makhuwa, Emakhuwa, Emakua, Macua, Makhuwa-Makhuwana, Makhuwwa of Nampula, Makoane, Makua, Maquoua](http://www.omniglot.com/writing/makhuwa.php) | [vmw](http://www-01.sil.org/iso639-3/documentation.asp?id=vmw) | 5 |
|  | [Mandinka, Mande, Manding, Mandingo, Mandingue, Mandinque, Socé](http://www.omniglot.com/writing/mandinka.htm) | [mnk](http://www-01.sil.org/iso639-3/documentation.asp?id=mnk) | 5 |
|  | [Minangkabau, Minang, Padang](http://www.omniglot.com/writing/minangkabau.htm) | [min](http://www-01.sil.org/iso639-3/documentation.asp?id=min) | 5 |
|  | [Mundari, Colh, Horo, Mandari, Mondari, Munari](http://www.omniglot.com/writing/mundari.htm) | [unr](http://www-01.sil.org/iso639-3/documentation.asp?id=unr) | 5 |
|  | [Neapolitan, Napoletano, Neapolitan-Calabrese](http://www.omniglot.com/writing/neapolitan.php) | [nap](http://www-01.sil.org/iso639-3/documentation.asp?id=nap) | 5 |
|  | [Piedmontese, Piemontese, Piemontèis](http://www.omniglot.com/writing/piedmontese.htm) | [pms](http://www-01.sil.org/iso639-3/documentation.asp?id=pms) | 5 |
|  | [Romany,](http://www.omniglot.com/writing/romany.htm) | [rom](http://www-01.sil.org/iso639-3/documentation.asp?id=rom) | 5 |
|  | [Sasak, Lombok](http://www.omniglot.com/writing/sasak.htm) | [sas](http://www-01.sil.org/iso639-3/documentation.asp?id=sas) | 5 |
|  | [Sicilian, Calabro-Sicilian, Sicilianu, Siculu](http://www.omniglot.com/writing/sicilian.htm) | [scn](http://www-01.sil.org/iso639-3/documentation.asp?id=scn) | 5 |
|  | [Soga, Lusoga, Olusoga](http://www.omniglot.com/writing/soga.htm) | [xog](http://www-01.sil.org/iso639-3/documentation.asp?id=xog) | 5 |
|  | [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](http://www.omniglot.com/writing/soninke.htm) | [snk](http://www-01.sil.org/iso639-3/documentation.asp?id=snk) | 5 |
|  | [Tswa, Kitshwa, Sheetshwa, Shitshwa, Tshwa, Xitshwa, Xitswa](http://www.omniglot.com/writing/tswa.htm) | [tsc](http://www-01.sil.org/iso639-3/documentation.asp?id=tsc) | 5 |
|  | [Venetian, Talian, Venet](http://www.omniglot.com/writing/venetian.htm) | [vec](http://www-01.sil.org/iso639-3/documentation.asp?id=vec) | 5 |
|  | [Zazaki, Southern, Dimili, Dimli, Southern Zaza, Zaza, Zazaca](https://www.omniglot.com/writing/zazaki.htm) | [diq](http://www-01.sil.org/iso639-3/documentation.asp?id=diq) | 5 |

# Appendix C: Repertoire table grouped by Glyph

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Unicode** | **Glyph** | **Unicode name** | **Languages using the code point (EGIDS)** | **Reference supporting inclusion (URL etc.)** |
|  | **0061** | **a** | LATIN SMALL LETTER A | Basic Latin | [0] |
|  | **0061 + 0331** | **a̱** | LATIN SMALL LETTER A + COMBINING MACRON BELOW | Nuer (4) | [146], [129] |
|  | **00E0** | **à** | LATIN SMALL LETTER A WITH GRAVE | Italian (1)  Galician (2)  Wolof (4) | [130], [131], [106], [132] |
|  | **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] |
|  | **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] |
|  | **00E3** | **ã** | LATIN SMALL LETTER A WITH TILDE | Umbundu (3)  Guarani (1)  Nauruan (3)  Khoekhoe (4) | [141], [142], [143], [144], [145] |
|  | **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] |
|  | **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] |
|  | **00E6** | **æ** | LATIN SMALL LETTER AE | Danish (1)  Icelandic (1)  Faroese (2) | [139], [102], [103] |
|  | **0101** | **ā** | LATIN SMALL LETTER A WITH MACRON | Latvian (1)  Tongan (1)  Hawaiian (2)  Marshallese(1) | [133], [134], [135], [136] |
|  | **0103** | **ă** | LATIN SMALL LETTER A WITH BREVE | Vietnamese (1)  Romanian (1) | [109], [110] |
|  | **0105** | **ą** | LATIN SMALL LETTER A WITH OGONEK | Polish (1)  Lithuanian (1) | [137], [138] |
|  | **01CE** | **ǎ** | LATIN SMALL LETTER A WITH CARON | Kirundi (1) | [104] [**https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx**](https://www.dropbox.com/s/ptfclojxkmbceyf/Kirundi%20and%20its%20tonal%20diacritics.docx)  **Jean Paul Nkurunziza (personal communication)** |
|  | **1EA1** | **ạ** | LATIN SMALL LETTER A WITH DOT BELOW | Vietnamese (1) | [109] |
|  | **1EA3** | **ả** | LATIN SMALL LETTER A WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EA5** | **ấ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1EA7** | **ầ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1EA9** | **ẩ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EAB** | **ẫ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1EAD** | **ậ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EAF** | **ắ** | LATIN SMALL LETTER A WITH BREVE AND ACUTE | Vietnamese (1) | [109] |
|  | **1EB1** | **ằ** | LATIN SMALL LETTER A WITH BREVE AND GRAVE | Vietnamese (1) | [109] |
|  | **1EB3** | **ẳ** | LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EB5** | **ẵ** | LATIN SMALL LETTER A WITH BREVE AND TILDE | Vietnamese (1) | [109] |
|  | **1EB7** | **ặ** | LATIN SMALL LETTER A WITH BREVE AND DOT BELOW | Vietnamese (1) | [109] |
|  | **0062** | **b** | LATIN SMALL LETTER B | Basic Latin | [0] |
|  | **0253** | **ɓ** | LATIN SMALL LETTER B WITH HOOK | Hausa (2)  Dagaare - Burkina Faso (4)  Pulaar, (3) | [147], [148], [250] |
|  | **0063** | **c** | LATIN SMALL LETTER C | Basic Latin | [0] |
|  | **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] |
|  | **0107** | **ć** | LATIN SMALL LETTER C WITH ACUTE | Croatian (1)  Serbian (1)  Polish (1) | [150], [151], [152] |
|  | **0109** | **ĉ** | LATIN SMALL LETTER C WITH CIRCUMFLEX | Esperanto (3) | [[255]](http://www.omniglot.com/writing/esperanto.htm) |
|  | **010B** | **ċ** | LATIN SMALL LETTER C WITH DOT ABOVE | Maltese (1) | [163] |
|  | **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] |
|  | **0064** | **d** | LATIN SMALL LETTER D | Basic Latin | [0] |
|  | **00F0** | **ð** | LATIN SMALL LETTER ETH | Faroese (2)  Icelandic (1) | [103], [102] |
|  | **010F** | **ď** | LATIN SMALL LETTER D WITH CARON | Czech (1)  Slovak (1) | [101], [153] |
|  | **0111** | **đ** | LATIN SMALL LETTER D WITH STROKE | Croatian (1)  Serbian (1)  Vietnamese (1)  Northern Sámi  Brahui (5) | [150], [151], [109], [108], [168] |
|  | **0256** | **ɖ** | LATIN SMALL LETTER D WITH TAIL | Fon (3)  Ewe (3) | [169], [170] |
|  | **0257** | **ɗ** | LATIN SMALL LETTER D WITH HOOK | Hausa (2)  Pulaar (3) | [147], [166], [250] |
|  | **1E13** | **ḓ** | LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **0065** | **e** | LATIN SMALL LETTER E | Basic Latin | [0] |
|  | **0065 + 0331** | **e̱** | LATIN SMALL LETTER E + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **0113** | **ē** | LATIN SMALL LETTER E WITH MACRON | Latvian (1)  Hawaiian (2)  Tongan (1)  Minangkabau (5) | [133], [135], [134], [184] |
|  | **0117** | **ė** | LATIN SMALL LETTER E WITH DOT ABOVE | Lithuanian (1) | [138], [154] |
|  | **0119** | **ę** | LATIN SMALL LETTER E WITH OGONEK | Polish (1)  Palauan (2)  Lithuanian (1) | [152], [185], [138], [154] |
|  | **011B** | **ě** | LATIN SMALL LETTER E WITH CARON | Czech (1)  Kirundi (1)  Sorbian (4) | [101], [104], [172] |
|  | **01DD** | **ǝ** | LATIN SMALL LETTER TURNED E | Kanuri (3) | [240] |
|  | **0259** | **ə** | LATIN SMALL LETTER SCHWA | Azeri, Azerbaijani (1)  Ewondo (3)  Ewe (3)  Bugis (3) | [159], [190], [170], [241] |
|  | **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] |
|  | **025B + 0308** | **ɛ̈** | LATIN SMALL LETTER OPEN E + COMBINING DIAERESIS | Nuer (4)  Dinka (4) | [129], [146], [239], [125] |
|  | **025B + 0331** | **ɛ̱** | LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW | Nuer (4) | [129], [146], [239] |
|  | **025B + 0331 + 0308** | **ɛ̱̈** | LATIN SMALL LETTER OPEN E + COMBINING MACRON BELOW + COMBINING DIAERESIS | Nuer (4) | [146], [239] |
|  | **1EB9** | **ẹ** | LATIN SMALL LETTER E WITH DOT BELOW | Yoruba (2) | [181] |
|  | **1EB9 + 0300** | **ẹ̀** | LATIN SMALL LETTER E WITH DOT BELOW + COMBINING GRAVE ACCENT | Yoruba (2) | [254] |
|  | **1EB9 + 0301** | **ẹ́** | LATIN SMALL LETTER E WITH DOT BELOW + COMBINING ACUTE ACCENT | Yoruba (2) | [254] |
|  | **1EBB** | **ẻ** | LATIN SMALL LETTER E WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EBD** | **ẽ** | LATIN SMALL LETTER E WITH TILDE | Umbundu (3)  Guarani (1)  Cubeo (3)  Xavante (4) | [141], [142], [143], [186], [187], [117] |
|  | **1EBF** | **ế** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1EC1** | **ề** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1EC3** | **ể** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EC5** | **ễ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1EC7** | **ệ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **0066** | **f** | LATIN SMALL LETTER F | Basic Latin | [0] |
|  | **0192** | **ƒ** | LATIN SMALL LETTER F WITH HOOK | Ewe (3) | [170] |
|  | **0067** | **g** | LATIN SMALL LETTER G | Basic Latin | [0] |
|  | **0067 + 0303** | **g̃** | LATIN SMALL LETTER G + COMBINING TILDE | Guarani (1) | [142], [143] |
|  | **0067 + 0304** | **ḡ** | LATIN SMALL LETTER G + COMBINING MACRON | Raga (Hano) (3) | [200] |
|  | **011D** | **ĝ** | LATIN SMALL LETTER G WITH CIRCUMFLEX | Esperanto (3) | [[255]](http://www.omniglot.com/writing/esperanto.htm) |
|  | **011F** | **ğ** | LATIN SMALL LETTER G WITH BREVE | Turkish (1)  Tatar (2)  Azeri (1)  Bashkir (4)  Zaza (5) | [157], [201], [159], [127], [202] |
|  | **0121** | **ġ** | LATIN SMALL LETTER G WITH DOT ABOVE | Maltese (1) | [163] |
|  | **0123** | **ģ** | LATIN SMALL LETTER G WITH CEDILLA | Latvian (1)  Brahui (5) | [133], [168] |
|  | **01E7** | **ǧ** | LATIN SMALL LETTER G WITH CARON | Skolt Sami (2) | [113] |
|  | **0263** | **ɣ** | LATIN SMALL LETTER GAMMA | Dagbani (Dagomba) (4)  Nuer (4)  Dinka (4)  Ewe (3)  Nuer (4) | [189], [146], [125], [170], [129] |
|  | **0068** | **h** | LATIN SMALL LETTER H | Basic Latin | [0] |
|  | **0125** | **ĥ** | LATIN SMALL LETTER H WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **0127** | **ħ** | LATIN SMALL LETTER H WITH STROKE | Maltese (1) | [163] |
|  | **0069** | **i** | LATIN SMALL LETTER I | Basic Latin | [0] |
|  | **0069 + 0331** | **i̲** | LATIN SMALL LETTER I + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  | **00EC** | **ì** | LATIN SMALL LETTER I WITH GRAVE | Italian (1)  Kirundi (1) | [130], [206], [208] |
|  | **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] |
|  | **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] |
|  | **00EF** | **ï** | LATIN SMALL LETTER I WITH DIAERESIS | Afrikaans (1)  French (1)  Kaqchikel (4)  Dinka (4)  West Frisian (2) | [175], [114], [126], [125], [115] |
|  | **0129** | **ĩ** | LATIN SMALL LETTER I WITH TILDE | Guarani (1)  Cubeo (3)  Khoekhoe (4)  Kikuyu ( 5) | [142], [143], [186], [145], [209] |
|  | **012B** | **ī** | LATIN SMALL LETTER I WITH MACRON | Latvian (1)  Lithuanian (1)  Hawaiian (2)  Tongan (1) | [133], [138], [135], [134] |
|  | **012F** | **į** | LATIN SMALL LETTER I WITH OGONEK | Lithuanian (1) | [154] |
|  | **0131** | **ı** | LATIN SMALL LETTER I DOTLESS | Turkish (1)  Tatar (2)  Azeri (1) | [157], [203], [201], [159] |
|  | **0135** | **ĵ** | LATIN SMALL LETTER J WITH CEDILLA | Esperanto (3) | [255] |
|  | **01D0** | **ǐ** | LATIN SMALL LETTER I WITH CARON | Kirundi (1) | [104] |
|  | **0268** | **ɨ** | LATIN SMALL LETTER I WITH STROKE | Cubeo (3)  Dagbani (Dagomba) (4)  HIxkaryána (4)  Maasai (5) | [186], [189], [210], [211] |
|  | **0268 + 0303** | **ɨ̃** | LATIN SMALL LETTER I WITH STROKE + COMBINING TILDE | Cubeo (3) | [186] |
|  | **1EC9** | **ỉ** | LATIN SMALL LETTER I WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ECB** | **ị** | LATIN SMALL LETTER I WITH DOT BELOW | Igbo (2) | [205] |
|  | **006A** | **j** | LATIN SMALL LETTER J | Basic Latin | [0] |
|  | **0269** | **ɩ** | LATIN SMALL LETTER IOTA | Dagaare - Burkina Faso (4)  Mossi (3) | [148], [212] |
|  | **006B** | **k** | LATIN SMALL LETTER K | Basic Latin | [0] |
|  | **0137** | **ķ** | LATIN SMALL LETTER K WITH CEDILLA | Latvian (1) | [133] |
|  | **0199** | **ƙ** | LATIN SMALL LETTER K WITH HOOK | Hausa (2) | [147] |
|  | **01E9** | **ǩ** | LATIN SMALL LETTER K WITH CARON | Skolt Sami (2) | [113] |
|  | **006C** | **l** | LATIN SMALL LETTER L | Basic Latin | [0] |
|  | **013A** | **ĺ** | LATIN SMALL LETTER L WITH ACUTE | Slovak (1) | [153] |
|  | **013C** | **ļ** | LATIN SMALL LETTER L WITH CEDILLA | Latvian (1)  Marshallese (1)  Brahui (5) | [133], [213], [214], [168] |
|  | **013E** | **ľ** | LATIN SMALL LETTER L WITH CARON | Slovak (1) | [153] |
|  | **0142** | **ł** | LATIN SMALL LETTER L WITH STROKE | Polish (1) | [152] |
|  | **1E37** | **ḷ** | LATIN SMALL LETTER L WITH DOT BELOW | Marshallese (1) | [213], [214], [215], [216] |
|  | **1E3D** | **ḽ** | LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **006D** | **m** | LATIN SMALL LETTER M | Basic Latin | [0] |
|  | **006D + 0327** | **m̦** | LATIN SMALL LETTER M + COMBINING CEDILLA | Marshallese (1) | [213], [136], [214] |
|  | **1E43** | **ṃ** | LATIN SMALL LETTER M WITH DOT BELOW | Marshallese (1) | [213], [136], [215], [216] |
|  | **006E** | **n** | LATIN SMALL LETTER N | Basic Latin | [0] |
|  | **006E + 0304** | **n̄** | LATIN SMALL LETTER N + COMBINING MACRON | Raga (Hano) (3)  Marshallese (1) | [200], [213], [136] |
|  | **006E + 0308** | **n̈** | LATIN SMALL LETTER N + COMBINING DIAERESIS | Malagasy(1) | [230] |
|  | **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] |
|  | **0144** | **ń** | LATIN SMALL LETTER N WITH ACUTE | Polish (1)  Lule Sámi (2)  Sorbian (4)  Brahui (5) | [152], [107], [172], [168] |
|  | **0146** | **ņ** | LATIN SMALL LETTER N WITH CEDILLA | Latvian (1)  Marshallese (1) | [133], [136] |
|  | **0148** | **ň** | LATIN SMALL LETTER N WITH CARON | Turkmen (1)  Czech (1)  Slovak (1) | [121], [101], [153] |
|  | **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] |
|  | **0272** | **ɲ** | LATIN SMALL LETTER N WITH LEFT HOOK | Susu (4)  Zarma (4)  Bambara (4) | [218], [219], [199] |
|  | **1E45** | **ṅ** | LATIN SMALL LETTER N WITH DOT ABOVE | Venda (1) | [164], [257] |
|  | **1E47** | **ṇ** | LATIN SMALL LETTER N WITH DOT BELOW | Marshallese (1) | [136], [215], [216] |
|  | **1E49** | **ṉ** | LATIN SMALL LETTER N WITH LINE BELOW | Pitjantjatjara (4) | [220] |
|  | **1E4B** | **ṋ** | LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW | Venda (1) | [164], [257] |
|  | **006F** | **o** | LATIN SMALL LETTER O | Basic Latin | [0] |
|  | **006F + 0327** | **o̧** | LATIN SMALL LETTER O + COMBINING CEDILLA | Marshallese (1) | [136] |
|  | **006F + 0331** | **o̱** | LATIN SMALL LETTER O + COMBINING MACRON BELOW | Nuer (4) | [146], [129] |
|  | **00F2** | **ò** | LATIN SMALL LETTER O WITH GRAVE | Italian (1)  Haitian Creole (1) | [130], [182], [183] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **00F8** | **ø** | LATIN SMALL LETTER O WITH STROKE | Danish (1)  Faroese (2) | [139], [103] |
|  | **014D** | **ō** | LATIN SMALL LETTER O WITH MACRON | Hawaiian (2)  Marshallese (1)  Tongan (1) | [135], [136], [134] |
|  | **0151** | **ő** | LATIN SMALL LETTER O WITH DOUBLE ACUTE | Hungarian (1) | [233], [234] |
|  | **0153** | **œ** | LATIN SMALL LIGATURE OE | French (1) | [114], [253] |
|  | **01A1** | **ơ** | LATIN SMALL LETTER O WITH HORN | Vietnamese (1) | [109] |
|  | **01D2** | **ǒ** | LATIN SMALL LETTER O WITH CARON | Kirundi (1) | [104] |
|  | **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] |
|  | **0254 + 0308** | **ɔ̈** | LATIN SMALL LETTER OPEN O + COMBINING DIAERESIS | Dinka (4) | [125] |
|  | **0254 + 0331** | **ɔ̱** | LATIN SMALL LETTER OPEN O + COMBINING MACRON BELOW | Nuer (4) | [129], [146] |
|  | **1ECD** | **ọ** | LATIN SMALL LETTER O WITH DOT BELOW | Igbo (2)  Yoruba (2)  Marshallese (1) | [204], [205], [181], [136], [215], [216] |
|  | **1ECD + 0300** | **ọ̀** | LATIN SMALL LETTER O WITH DOT BELOW + COMBINING GRAVE ACCENT | Yoruba (2) | [254] |
|  | **1ECD + 0301** | **ọ́** | LATIN SMALL LETTER O WITH DOT BELOW + COMBINING ACUTE ACCENT | Yoruba (2) | [254] |
|  | **1ECF** | **ỏ** | LATIN SMALL LETTER O WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ED1** | **ố** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [109] |
|  | **1ED3** | **ồ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [109] |
|  | **1ED5** | **ổ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1ED7** | **ỗ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [109] |
|  | **1ED9** | **ộ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [109] |
|  | **1EDB** | **ớ** | LATIN SMALL LETTER O WITH HORN AND ACUTE | Vietnamese (1) | [109] |
|  | **1EDD** | **ờ** | LATIN SMALL LETTER O WITH HORN AND GRAVE | Vietnamese (1) | [109] |
|  | **1EDF** | **ở** | LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EE1** | **ỡ** | LATIN SMALL LETTER O WITH HORN AND TILDE | Vietnamese (1) | [109] |
|  | **1EE3** | **ợ** | LATIN SMALL LETTER O WITH HORN AND DOT BELOW | Vietnamese (1) | [109] |
|  | **0070** | **p** | LATIN SMALL LETTER P | Basic Latin | [0] |
|  | **00FE** | **þ** | LATIN SMALL LETTER THORN | Icelandic (1) | [102] |
|  | **0071** | **q** | LATIN SMALL LETTER Q | Basic Latin | [0] |
|  | **0072** | **r** | LATIN SMALL LETTER R | Basic Latin | [0] |
|  | **0072 + 0303** | **r̃** | LATIN SMALL LETTER R + COMBINING TILDE | Hausa (2) | [147] |
|  | **0155** | **ŕ** | LATIN SMALL LETTER R WITH ACUTE | Slovak (1)  Brahui (5) | [153], [168] |
|  | **0159** | **ř** | LATIN SMALL LETTER R WITH CARON | Czech (1)  Sorbian (4) | [101], [172] |
|  | **024D** | **ɍ** | LATIN SMALL LETTER R WITH STROKE | Kanuri (3) | [240] |
|  | **0073** | **s** | LATIN SMALL LETTER S | Basic Latin | [0] |
|  | **00DF** | **ß** | LATIN SMALL LETTER SHARP S | German (1) | [119] |
|  | **015B** | **ś** | LATIN SMALL LETTER S WITH ACUTE | Polish (1) | [152] |
|  | **015D** | **ŝ** | LATIN SMALL LETTER S WITH CIRCUMFLEX | Esperanto (3) | [255] |
|  | **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] |
|  | **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] |
|  | **0219** | **ș** | LATIN SMALL LETTER S WITH COMMA BELOW | Romanian (1) | [110] |
|  | **1E63** | **ṣ** | LATIN SMALL LETTER S WITH DOT BELOW | Yoruba (2) | [181] |
|  | **0074** | **t** | LATIN SMALL LETTER T | Basic Latin | [0] |
|  | **0165** | **ť** | LATIN SMALL LETTER T WITH CARON | Czech (1)  Slovak (1) | [101], [153] |
|  | **0167** | **ŧ** | LATIN SMALL LETTER T WITH STROKE | Northern Sami (2)  Brahui (5) | [108], [168] |
|  | **021B** | **ț** | LATIN SMALL LETTER T WITH COMMA BELOW | Romanian (1) | [110] |
|  | **1E6D** | **ṭ** | LATIN SMALL LETTER T WITH DOT BELOW | Mizo (4) | [242] |
|  | **1E71** | **ṱ** | LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW | Venda (1) | [164]**, [257]** |
|  | **0075** | **u** | LATIN SMALL LETTER U | Basic Latin | [0] |
|  | **00F9** | **ù** | LATIN SMALL LETTER U WITH GRAVE | Italian (1)  Papiamento (1) | [130], [206], [245], [246] |
|  | **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] |
|  | **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] |
|  | **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] |
|  | **0169** | **ũ** | LATIN SMALL LETTER U WITH TILDE | Umbundu (3)  Guarani (1)  Nauruan (3)  Khoekhoe (4) Kikuyu (5) | [141], [142], [143], [144], [145], [209] |
|  | **016B** | **ū** | LATIN SMALL LETTER U WITH MACRON | Latvian (1)  Hawaiian (2)  Lithuanian (1)  Marshallese (1)  Tongan (1) | [133], [135], [138], [154], [136], [134] |
|  | **016D** | **ŭ** | LATIN SMALL LETTER U WITH BREVE | Esperanto (3) | [255] |
|  | **016F** | **ů** | LATIN SMALL LETTER U WITH RING ABOVE | Czech (1) | [101] |
|  | **0171** | **ű** | LATIN SMALL LETTER U WITH DOUBLE ACUTE | Hungarian (1) | [233], [234] |
|  | **0173** | **ų** | LATIN SMALL LETTER U WITH OGONEK | Lithuanian (1) | [154], [138] |
|  | **01B0** | **ư** | LATIN SMALL LETTER U WITH HORN | Vietnamese (1) | [109] |
|  | **01D4** | **ǔ** | LATIN SMALL LETTER U WITH CARON | Kirundi (1) | [104] |
|  | **0289** | **ʉ** | LATIN SMALL LETTER U BAR | Cubeo (3)  Maasai (5) | [186], [187], [211] |
|  | **0289 + 0303** | **ʉ̃** | LATIN SMALL LETTER U BAR + COMBINING TILDE | Cubeo (3) | [186], [187] |
|  | **1EE5** | **ụ** | LATIN SMALL LETTER U WITH DOT BELOW | Vietnamese (1)  Igbo (2) | [109],[204], [205] |
|  | **1EE7** | **ủ** | LATIN SMALL LETTER U WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EE9** | **ứ** | LATIN SMALL LETTER U WITH HORN AND ACUTE | Vietnamese (1) | [109] |
|  | **1EEB** | **ừ** | LATIN SMALL LETTER U WITH HORN AND GRAVE | Vietnamese (1) | [109] |
|  | **1EED** | **ử** | LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EEF** | **ữ** | LATIN SMALL LETTER U WITH HORN AND TILDE | Vietnamese (1) | [109] |
|  | **1EF1** | **ự** | LATIN SMALL LETTER U WITH HORN AND DOT BELOW | Vietnamese (1) | [109] |
|  | **0076** | **v** | LATIN SMALL LETTER V | Basic Latin | [0] |
|  | **028B** | **ʋ** | LATIN SMALL LETTER V WITH HOOK | Dagaare - Burkina Faso (4)  Mossi (3)  Ewe (3) | [148], [212], [238], [170] |
|  | **0077** | **w** | LATIN SMALL LETTER W | Basic Latin | [0] |
|  | **0175** | **ŵ** | LATIN SMALL LETTER W WITH CIRCUMFLEX | Chichewa (3) | [247] |
|  | **0078** | **x** | LATIN SMALL LETTER X | Basic Latin | [0] |
|  | **1E8D** | **ẍ** | LATIN SMALL LETTER X WITH DIAERESIS | Mam (4) | [248], [249] |
|  | **0079** | **y** | LATIN SMALL LETTER Y | Basic Latin | [0] |
|  | **00FD** | **ý** | LATIN SMALL LETTER Y WITH ACUTE | Turkmen (1)  Czech (1)  Icelandic (1)  Faroese (2)  Guarani (1) | [121], [101], [102], [103], [142], [143] |
|  | **00FF** | **ÿ** | LATIN SMALL LETTER Y WITH DIAERESIS | French (1) | [114], [253], [257] |
|  | **0177** | **ŷ** | LATIN SMALL LETTER Y WITH CIRCUMFLEX | Welsh (2) | [256] |
|  | **01B4** | **ƴ** | LATIN SMALL LETTER Y WITH HOOK | Dagaare - Burkina Faso (4) | [148], [251], [149] |
|  | **1EF3** | **ỳ** | LATIN SMALL LETTER Y WITH GRAVE | Vietnamese (1) | [109] |
|  | **1EF5** | **ỵ** | LATIN SMALL LETTER Y WITH DOT BELOW | Vietnamese (1) | [109] |
|  | **1EF7** | **ỷ** | LATIN SMALL LETTER Y WITH HOOK ABOVE | Vietnamese (1) | [109] |
|  | **1EF9** | **ỹ** | LATIN SMALL LETTER Y WITH TILDE | Vietnamese (1)  Guarani (1) | [109] [142] |
|  | **007A** | **z** | LATIN SMALL LETTER Z | Basic Latin | [0] |
|  | **017A** | **ź** | LATIN SMALL LETTER Z WITH ACUTE | Polish (1)  Brahui (5)  Sorbian (4) | [152], [252], [168], [172] |
|  | **017C** | **ż** | LATIN SMALL LETTER Z WITH DOT ABOVE | Polish (1)  Maltese (1) | [152], [163] |
|  | **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] |
|  | **01EF** | **ǯ** | LATIN SMALL LETTER EZH WITH CARON | Skolt Sami (2) | [113] |
|  | **0292** | **ʒ** | LATIN SMALL LETTER EZH | Skolt Sami (2)  Dagbani (Dagomba) (4) | [113], [189] |

1. *Script* is used here to indicate the whole writing system including basic letters, ligatures and diacritics. See also RFC 6365 and ISO 15924. [↑](#footnote-ref-1)
2. However, several orthographies on the basis of different scripts are frequently used simultaneously, both historically and contemporarily. [↑](#footnote-ref-2)
3. Latin script is also known as Roman script in academic literature. [↑](#footnote-ref-3)
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. [↑](#footnote-ref-4)