Proposal for a Latin Script Root Zone LGR

*LGR Version 3.0*

*Date: 2018-05-25*

*Document version:* 1.0 Repertoire Only

*Authors:* Latin Generation Panel

Table of Contents

[1 General Information 3](#_Toc514961742)

[2 Script for which the LGR is proposed 3](#_Toc514961743)

[3 Background on Script and Principal Languages Using It 4](#_Toc514961744)

[3.1 Principal languages using Latin script 5](#_Toc514961745)

[3.2 Geographic territories or countries with significant user communities 5](#_Toc514961746)

[3.3 Related scripts 5](#_Toc514961747)

[4 Overall Development Process and Methodology 5](#_Toc514961748)

[5 Repertoire 6](#_Toc514961749)

[5.1 Definitions 6](#_Toc514961750)

[5.2 Principles for developing Repertoire 7](#_Toc514961751)

[5.2.1 Inclusion Principles 7](#_Toc514961752)

[5.2.2 Exclusion Principles 8](#_Toc514961753)

[5.3 Code points included 8](#_Toc514961754)

[5.3.1 Code points proposed for inclusion in LGR 8](#_Toc514961755)

[5.4 Code points proposed to be added to MSR-3 27](#_Toc514961756)

[5.5 Code points excluded 28](#_Toc514961757)

[5.5.1 Punctuation marks 28](#_Toc514961758)

[5.5.2 Letters combined with punctuation marks 30](#_Toc514961759)

[5.5.3 Other Excluded Letters 32](#_Toc514961760)

[6 Variants 33](#_Toc514961761)

[6.1 Principles for developing in-script Variants 33](#_Toc514961762)

[6.2 In-script Latin variants 33](#_Toc514961763)

[6.3 Principles for developing Cross script variants 33](#_Toc514961764)

[6.4 Cross-script variants 33](#_Toc514961765)

[6.4.1 Armenian variants 33](#_Toc514961766)

[6.4.2 Cyrillic variants 34](#_Toc514961767)

[6.4.3 Greek variants 34](#_Toc514961768)

[6.4.4 Other Cross-script variants 34](#_Toc514961769)

[7 Whole Label Evaluation Rules (WLE) 34](#_Toc514961770)

[8 Contributors 34](#_Toc514961771)

[9 References 34](#_Toc514961772)

[9.1 References for code points inclusion 34](#_Toc514961773)

[9.2 Other references 38](#_Toc514961774)

[Appendix A: Table of processed languages used to develop Latin Script Repertoire 39](#_Toc514961775)

[Appendix B: Statistics about processed languages using Latin script 48](#_Toc514961776)

[B.1 Number of languages processed per EGIDS 48](#_Toc514961777)

[B.2 Number of languages processed per continent 48](#_Toc514961778)

[B.3 Number of processed languages per language family 49](#_Toc514961779)

[B.4 Number of states represented per continent 50](#_Toc514961780)

[Appendix C: Repertoire table grouped by Glyph 51](#_Toc514961781)

# 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-latn-2018MMDD.xml

The test labels of the LGR can be found in the accompanying file:

Proposed-LGR-latn-2018MMDD.xml

# 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-3

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 Maximal Starting Repertoire, Second Version (Referred to simply as [MSR-2] in the following) were considered.

When the work of Latin Generation Panel started Maximal Starting Repertoire (MSR) version was MSR-2, but MSR was upgraded to version MSR-3 on January 17, 2018. Latin script part in MSR-3 contains three more letters than the Latin script part in MSR-2. Three new code points are added in Latin script part in MSR-3 as the result of the work of Latin Generation Panel.

The set of code points in the Latin script, as specified by [MSR-3], contains 343 selected code points, i.e. 323 letters and 20 Combining Diacritical Marks. Code points are from the following Unicode ranges:

|  |  |
| --- | --- |
| **Latin Script** | **Range of Unicode code points** |
| Controls and Basic Latin | U+0061 – U+007A |
| Controls and Latin-1 Supplement | U+00DF - U+00F6U+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 |

MSR- 3 excluded the following Latin script UNICODE ranges:

|  |  |
| --- | --- |
| **Latin Script** | **Range of Unicode code points** |
| Latin Extended-D; technical use (phonetic)/obsolete/punctuation | U+A720 – U+A7FF |
| Latin Ligatures; compatibility characters not PVALID in IDNA 2008 | U+FB00 – U+FB0F |
| Full-width Latin Letters; compatibility characters not PVALID in IDNA 2008 | U+FF00 – U+FF5E |

 When there is a single code point which gives the combined letter and diacritical mark, that will be used in preference to use the combining diacritic mark code point. Furthermore, only lowercase letters are considered in creating the repertoire, as uppercase 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].

## 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.

## 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.

## 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 work plan given in “Proposal for the Generation Panel (GP) for the Latin Script Label Generation Ruleset (LGR) for the Root Zone”.

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 to public unofficial comment. Comments from Integration panel were encompassed in final version of Principles.

During Repertoire definition phase Panel reviewed and processed 180 languages with EGIDS level 1 through 4, and 29 languages with EGIDS Level 5, which have more than 1,000,000 speakers.

The Latin Generation Panel used MSR-2 as the starting point and after processing 209 languages Latin GP found:

* 195 MSR-2 Unicode code points verified
* 22 code point sequences detected
* 6 new code points detected which were not included in MSR-2. Latin GP recommended these code points to be added to MSR-3.
* 3 code points of these 6 were accepted by IP and added to MSR-3

Second phase of Latin GP work is mainly devoted to definition in-script and cross-script Variants. Tables presented in Appendix B should give an insight in the comprehensiveness of the job done.

Data about population, used in creating tables in Appendix B, are collected in the year 2015. Total world population in 2015 was 7.3 billion according to [Population Reference Bureau](http://www.prb.org/pdf15/2015-world-population-data-sheet_eng.pdf). The work of Latin GP encompassed languages spoken by almost one third of the world population.

# Repertoire

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

## 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. (See Appendix A)

**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. (See Appendix A)

**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).

## Principles for developing Repertoire

### 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.
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 Criteria 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.

### 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

## Code points included

The table below lists the code points recommended 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.

### Code points proposed for inclusion in LGR

This table is sorted by Unicode column. The table with the same data, but sorted by glyph, can be found in Appendix C."

Description of References supporting inclusion of code point are in section 9.1

| # | Comment on inclusion | Unicode | Glyph | Unicode name | Languages using the code point (EGIDS) | Reference supporting inclusion (URL etc.) |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **0061** | **a** | LATIN SMALL LETTER A | Many languages | **ASCII a-z** |
|  |  | **0061 + 0331** | **a̱** | LATIN SMALL LETTER A + COMBINING MACRON BELOW | Nuer (4) | [146], [129] |
|  |  | **0062** | **b** | LATIN SMALL LETTER B | Many languages | **ASCII a-z** |
|  |  | **0063** | **c** | LATIN SMALL LETTER C | Many languages | **ASCII a-z** |
|  |  | **0064** | **d** | LATIN SMALL LETTER D | Many languages | **ASCII a-z** |
|  |  | **0065** | **e** | LATIN SMALL LETTER E | Many languages | **ASCII a-z** |
|  |  | **0065 + 0331** | **e̱** | LATIN SMALL LETTER E + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  |  | **0066** | **f** | LATIN SMALL LETTER F | Many languages | **ASCII a-z** |
|  |  | **0067** | **g** | LATIN SMALL LETTER G | Many languages | **ASCII a-z** |
|  |  | **0067 + 0303** | **g̃** | LATIN SMALL LETTER G + COMBING TILDE | Guarani (1) | [142], [143] |
|  |  | **0067 + 0304** | **ḡ** | LATIN SMALL LETTER G + COMBINING MACRON | Raga (Hano) (3) | [200] |
|  |  | **0068** | **h** | LATIN SMALL LETTER H | Many languages | **ASCII a-z** |
|  |  | **0069** | **i** | LATIN SMALL LETTER I | Many languages | **ASCII a-z** |
|  |  | **0069 + 0331** | **i̲** | LATIN SMALL LETTER I + COMBINING MACRON BELOW | Nuer (4) | [146] |
|  |  | **006A** | **j** | LATIN SMALL LETTER J | Many languages | **ASCII a-z** |
|  |  | **006B** | **k** | LATIN SMALL LETTER K | Many languages | **ASCII a-z** |
|  |  | **006C** | **l** | LATIN SMALL LETTER L | Many languages | **ASCII a-z** |
|  |  | **006D** | **m** | LATIN SMALL LETTER M | Many languages | **ASCII a-z** |
|  |  | **006D + 0327** | **m̦** | LATIN SMALL LETTER M + COMBINING CEDILLA | Marshallese (1) | [213], [136], [214] |
|  |  | **006E** | **n** | LATIN SMALL LETTER N | Many languages | **ASCII a-z** |
|  |  | **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 | Many languages | **ASCII a-z** |
|  |  | **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  | Many languages | **ASCII a-z** |
|  |  | **0071** | **q** | LATIN SMALL LETTER Q | Many languages | **ASCII a-z** |
|  |  | **0072** | **r** | LATIN SMALL LETTER R | Many languages | **ASCII a-z** |
|  |  | **0072 + 0303** | **r̃** | LATIN SMALL LETTER R + COMBINING TILDE | Hausa (2) | [147] |
|  |  | **0073** | **s** | LATIN SMALL LETTER S | Many languages | **ASCII a-z** |
|  |  | **0074** | **t** | LATIN SMALL LETTER T | Many languages | **ASCII a-z** |
|  |  | **0075** | **u** | LATIN SMALL LETTER U | Many languages | **ASCII a-z** |
|  |  | **0076** | **v** | LATIN SMALL LETTER V | Many languages | **ASCII a-z** |
|  |  | **0077** | **w** | LATIN SMALL LETTER W | Many languages | **ASCII a-z** |
|  |  | **0078** | **x** | LATIN SMALL LETTER X | Many languages | **ASCII a-z** |
|  |  | **0079** | **y** | LATIN SMALL LETTER Y | Many languages | **ASCII a-z** |
|  |  | **007A** | **z** | LATIN SMALL LETTER Z | Many languages | **ASCII a-z** |
|  |  | **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 (1)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], [162], [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], [162], [132], [117], [171] |
|  |  | **00EA** | **ê** | LATIN SMALL LETTER E WITH CIRCUMFLEX | French (1)Tswana (1)Afrikaans (1)Vietnamese (1)Kurdish (2)Kirundi (1)West Frisian (1)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], [171], [129] |
|  |  | **00EC** | **ì** | LATIN SMALL LETTER I WITH GRAVE | Italian (1)Kirundi (1) | [130], [206], [207], [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], [162], [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], [171] |
|  |  | **00F0** | **ð** | LATIN SMALL LETTER ETH | Faroese (2)Icelandic (1) | [103], [102] |
|  |  | **00F1** | **ñ** | LATIN SMALL LETTER N WITH TILDE | Spanish (1)Pulaar (3)Chomorro (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], [222], [142], [143], [223], [160], [162], [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], [162], [132] |
|  |  | **00F4** | **ô** | LATIN SMALL LETTER O WITH CIRCUMFLEX | Tswana (1)Afrikaans (1)Vietnamese (1)Kirundi (1)French (1)Northern Sotho(1)West Frisian (1)Galician (2)Friulian (4)Xavante(4) | [173], [174], [175], [109], [104], [114], [230], [115], [162], [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 VersionWest Frisian (2)Nuer (4) | [119], [120], [175], [157], [123], [179], [124], [180], [126], [125], [127], [231], [232], [171], [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 (1)Galician (2) | [100], [101], [102], [103], [104], [105], [115], [162] |
|  |  | **00FB** | **û** | LATIN SMALL LETTER U WITH CIRCUMFLEX | Afrikaans (1) Kurdish (2)Kirundi (1)French (1)Miskito (2)West Frisian (1)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)Low German (5) | [119], [100], [175], [157], [123], [114], [159], [161], [162], [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] |
|  |  | **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)Bavarian ( 5) | [109], [110], [111],, [112] |
|  |  | **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] |
|  |  | **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)Kabyle (5) | [150], [151], [152], [133], [153], [108], [154], [155], [156] |
|  |  | **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 SámiBrahui (5) | [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] |
|  |  | **0115** | **ĕ** | LATIN SMALL LETTER E WITH BREVE | Bavarian(5) | [111], [112] |
|  |  | **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] |
|  |  | **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] |
|  |  | **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 I DOTLESS | Turkish (1)Tatar (2) Azeri(1) | [157], [203], [201], [159] |
|  |  | **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 | Tswa (5) | [217] |
|  |  | **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)Nothert 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] |
|  |  | **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) | [247] |
|  |  | **017A** | **ź** | LATIN SMALL LETTER Z WITH ACUTE | Polish (1)Brahui (5)(Lower) 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) | [118] |
|  |  | **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] |
|  | **Note complexity of "schwa" and "turned E" and that those are homeglyphs in lower case. 2017-12-31/MD** | **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], [149] |
|  |  | **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], [167] |
|  | **Note complexity of "schwa" and "turned E", and that those are homeglyphs in lower case. 2017-12-31/MD** | **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] |
|  | **Included in IP's proposal for MSR3.** | **0268** | **ɨ** | LATIN SMALL LETTER I WITH STROKE | Cubeo (3)Dagbani (Dagomba) (4)HIxkaryána (4)Maasai (5) | [186], [189], [210], [211] |
|  | **U+0268 included in IP's proposal for MSR3.** | **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] |
|  | **Included in IP's proposal for MSR3.** | **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] |
|  |  | **1E0D** | **ḍ** | LATIN SMALL LETTER D WITH DOT BELOW | Mundari (5)Kabyle (5) | [165], [155], [156]  |
|  |  | **1E13** | **ḓ** | LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **1E25** | **ḥ** | LATIN SMALL LETTER H WITH DOT BELOW | Kabyle (5) | [155], [156] |
|  |  | **1E37** | **ḷ** | LATIN SMALL LETTER L WITH DOT BELOW | Marshallese (1)Mundari (5) | [213], [214], [215], [216], [165] |
|  | **Included in IP's proposal for MSR3.** | **1E3D** | **ḽ** | LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **1E43** | **ṃ** | LATIN SMALL LETTER M WITH DOT BELOW | Marshallese (1) | [213], [136], [215], [216] |
|  | **Only languages with good evidence are kept.** | **1E45** | **ṅ** | LATIN SMALL LETTER N WITH DOT ABOVE | Venda (1)Tswa (5) | [164], [217] |
|  |  | **1E47** | **ṇ** | LATIN SMALL LETTER N WITH DOT BELOW | Mundari (5)Marshallese (1) | [165], [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] |
|  |  | **1E5B** | **ṛ** | LATIN SMALL LETTER R WITH DOT BELOW | Kabyle (5) | [155], [156] |
|  |  | **1E63** | **ṣ** | LATIN SMALL LETTER S WITH DOT BELOW | Yoruba (2)Kabyle (5) | [181], [155], [156] |
|  |  | **1E6D** | **ṭ** | LATIN SMALL LETTER T WITH DOT BELOW | Mizo (4)Mundari (5)Kabyle (5) | [242], [165], [155], [156] |
|  |  | **1E71** | **ṱ** | LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **1E8D** | **ẍ** | LATIN SMALL LETTER X WITH DIAERESIS | Mam (4) | [248], [249] |
|  |  | **1E91** | **ẑ** | LATIN SMALL LETTER Z WITH CIRCUMFLEX | Tswa (5) | [217] |
|  |  | **1E93** | **ẓ** | LATIN SMALL LETTER Z WITH DOT BELOW | Kabyle (5) | [155] |
|  |  | **1EA1** | **ạ** | LATIN SMALL LETTER A WITH DOT BELOW | Vietnamese (1) | [109] |
|  |  | **1EA3** | **ả** | LATIN SMALL LETTER A WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **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) | [118] |
|  |  | **1EAB** | **ẫ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EAD** | **ậ** | LATIN SMALL LETTER A WITH CIRCUMFLEXAND 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) | [118] |
|  |  | **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) | [118] |
|  |  | **1EC1** | **ề** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EC3** | **ể** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EC5** | **ễ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EC7** | **ệ** | LATIN SMALL LETTER E WITH CIRCUMFLEXAND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EC9** | **ỉ** | LATIN SMALL LETTER I WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **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) | [118] |
|  |  | **1ED1** | **ố** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1ED3** | **ồ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1ED5** | **ổ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1ED7** | **ỗ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1ED9** | **ộ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EDB** | **ớ** | LATIN SMALL LETTER O WITH HORN AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1EDD** | **ờ** | LATIN SMALL LETTER O WITH HORN AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EDF** | **ở** | LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EE1** | **ỡ** | LATIN SMALL LETTER O WITH HORN AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EE3** | **ợ** | LATIN SMALL LETTER O WITH HORN AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EE5** | **ụ** | LATIN SMALL LETTER U WITH DOT BELOW | Igbo (2) | [204], [205] |
|  |  | **1EE7** | **ủ** | LATIN SMALL LETTER U WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EE9** | **ứ** | LATIN SMALL LETTER U WITH HORN AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1EEB** | **ừ** | LATIN SMALL LETTER U WITH HORN AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EED** | **ử** | LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EEF** | **ữ** | LATIN SMALL LETTER U WITH HORN AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EF1** | **ự** | LATIN SMALL LETTER U WITH HORN AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EF3** | **ỳ** | LATIN SMALL LETTER Y WITH GRAVE | Vietnamese (1) | [118] |
|  |  | **1EF5** | **ỵ** | LATIN SMALL LETTER Y WITH DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EF7** | **ỷ** | LATIN SMALL LETTER Y WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EF9** | **ỹ** | LATIN SMALL LETTER Y WITH TILDE | Vietnamese (1) Guarani (1) | [118] [142] |

## Code points proposed to be added to MSR-3

Latin GP has found support for six Code Points not in MSR2. The Code Points are listed in the table below with supporting language and link to reference material.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Unicode** | **Glyph** | **Unicode name** | **Languages**  | **Reference supporting inclusion** |
| **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> |
| **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> |
| **01C0** | **ǀ** | LATIN LETTER DENTAL CLICK  | Khoekhoe(4) | <https://www.britannica.com/topic/Khoisan-languages><https://en.wikipedia.org/wiki/Khoe_languages><https://www.newera.com.na/tag/khoekhoegowab/><http://www.omniglot.com/writing/khoekhoe.htm> |
| **01C1** | **ǁ** | LATIN LETTER LATERAL CLICK | Khoekhoe(4) | <https://www.britannica.com/topic/Khoisan-languages><https://en.wikipedia.org/wiki/Khoe_languages><https://www.newera.com.na/tag/khoekhoegowab/><http://www.omniglot.com/writing/khoekhoe.htm> |
| **01C2** | **ǂ** | LATIN LETTER ALVEOLAR CLICK | Khoekhoe(4) | <https://www.britannica.com/topic/Khoisan-languages><https://en.wikipedia.org/wiki/Khoe_languages><https://www.newera.com.na/tag/khoekhoegowab/><http://www.omniglot.com/writing/khoekhoe.htm> |
| 1**E3D** | **ḽ** | LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW | Venda (1) | http://www.omniglot.com/writing/venda.htm |

## Code points excluded

The Internet Architecture Board (IAB) has mandated that punctuation marks not 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* Latin letters followed by punctuation marks. Accordingly, the following letters from various languages using the Latin script have been excluded from the repertoire

### Punctuation marks

| **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.htmhttp://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

| **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> |
| **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> |
| **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 these letters in MSR-3 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 summarizes all variant relationships discovered which are encoded in the LGR. There are two principle types of variants:

* in-script variants
* cross-script variants.

In-script variants are such where a variant relationship was established among two or more code points which form part of the repertoire of the LGR.

Cross-script variants are such where a variant relationship was established in between code points which form part of the repertoire of the LGR and code points from other scripts part of MSR-3.

Before presenting the two types of variants in sections 6.2 and 6.4, the principles which were employed for developing variant relationship, as well as the types of variants relationships and possible dispositions in the LGR are discussed in sections 6.1 and 6.3.

## Principles for developing in-script Variants

WORK IN PROGRESS!

## In-script Latin variants

WORK IN PROGRESS!

## Principles for developing Cross script variants

WORK IN PROGRESS!

## Cross-script variants

WORK IN PROGRESS!

### Armenian variants

The Latin script has the following cross-script variants with Armenian script:

WORK IN PROGRESS!

### Cyrillic variants

The Latin script has the following cross-script variants with Cyrillic script:

WORK IN PROGRESS!

### Greek variants

The Latin script has the following cross-script variants with Greek script:

WORK IN PROGRESS!

### Other Cross-script variants

The Latin script has the following cross-script variants with the following additional script.

WORK IN PROGRESS!

# Whole Label Evaluation Rules (WLE)

WLE rules for the Latin script.

WORK IN PROGRESS!

# Contributors

# References

## References for code points inclusion

[100] https://www.icann.org/sites/default/files/packages/lgr/lgr-second-level-spanish-30aug16-en.html

[101] http://www.omniglot.com/writing/czech.htm

[102] http://www.omniglot.com/writing/icelandic.htm

[103] http://www.omniglot.com/writing/faroese.htm

[104] https://en.wikipedia.org/wiki/Burundi\_Bwacu#Kirundi\_.28with\_tonal\_diacritics\_.E2.80.94\_utw.C3.A2tuzo.29

[105] http://www.omniglot.com/writing/chuukese.htm

[106] http://www.webcitation.org/6siTI8ieQ

[107] http://www.omniglot.com/writing/lulesami.htm

[108] https://en.wikipedia.org/wiki/Northern\_Sami

[109] http://www.omniglot.com/writing/vietnamese.htm

[110] http://www.omniglot.com/writing/romanian.htm

[111] https://www.omniglot.com/writing/bavarian.htm

[112] https://bar.wikipedia.org/wiki/Wikipedia:Boarische\_Umschrift

[113] http://www.omniglot.com/writing/skoltsami.htm

[114] http://omniglot.com/writing/french.htm

[115] http://www.omniglot.com/writing/westfrisian.htm

[116] http://www.omniglot.com/writing/friulian.htm

[117] http://www.silbrasil.org.br/resources/archives/17019

[118] https://www.omniglot.com/writing/vietnamese.htm

[119] http://www.omniglot.com/writing/german.htm

[120] http://www.omniglot.com/writing/finnish.htm

[121] http://www.omniglot.com/writing/turkmen.htm

[122] http://www.omniglot.com/writing/estonian.htm

[123] http://www.omniglot.com/writing/swedish.htm

[124] http://www.omniglot.com/writing/yapese.htm

[125] https://www.omniglot.com/writing/dinka.php

[126] http://www.omniglot.com/writing/kaqchikel.htm

[127] http://www.omniglot.com/writing/bashkir.htm

[128] https://www.omniglot.com/writing/alsatian.htm

[129] https://en.wikipedia.org/wiki/Nuer\_language

[130] http://www.omniglot.com/writing/italian.htm

[131] https://en.wikipedia.org/wiki/Italian\_orthography

[132] http://www.omniglot.com/writing/wolof.htm

[133] http://www.omniglot.com/writing/latvian.htm

[134] http://www.omniglot.com/writing/tongan.htm

[135] http://www.omniglot.com/writing/hawaiian.htm

[136] http://www.omniglot.com/writing/marshallese.php

[137] http://www.omniglot.com/writing/polish.htm

[138] http://www.omniglot.com/writing/lithuanian.htm

[139] http://www.omniglot.com/writing/danish.htm

[140] http://www.omniglot.com/writing/chamorro.htm

[141] http://www.omniglot.com/writing/umbundu.htm

[142] http://www.omniglot.com/writing/guarani.htm

[143] https://en.wikipedia.org/wiki/Guarani\_alphabet

[144] http://www.omniglot.com/writing/nauruan.htm

[145] https://www.omniglot.com/writing/khoekhoe.htm

[146] https://www.omniglot.com/writing/nuer.htm

[147] http://www.omniglot.com/writing/hausa.htm

[148] http://www.omniglot.com/writing/dagaare.htm

[149] http://www.omniglot.com/writing/fula.htm

[150] http://www.omniglot.com/writing/croatian.htm

[151] http://www.omniglot.com/writing/serbian.htm

[152] https://en.wikipedia.org/wiki/Polish\_language

[153] http://www.omniglot.com/writing/slovak.htm

[154] http://www.evertype.com/alphabets/lithuanian.pdf

[155] http://www.omniglot.com/writing/kabyle.php

[156] https://en.wikipedia.org/wiki/Kabyle\_language

[157] http://www.omniglot.com/writing/turkish.htm

[158] http://www.omniglot.com/writing/kurdish.htm

[159] http://www.omniglot.com/writing/azeri.htm

[160] http://www.omniglot.com/writing/basque.htm

[161] https://en.wikipedia.org/wiki/Basque\_language#Writing\_system

[162] http://scriptsource.org/cms/scripts/page.php?item\_id=wrSys\_detail\_sym&key=gl

[163] http://www.omniglot.com/writing/maltese.htm

[164] http://www.omniglot.com/writing/venda.htm

[165] https://www.omniglot.com/writing/mundari.htm

[166] https://en.wikipedia.org/wiki/Hausa\_language

[167] http://phoible.org/inventories/view/809#tsource

[168] https://www.omniglot.com/writing/brahui.htm

[169] https://en.wikipedia.org/wiki/Fon\_language

[170] http://www.omniglot.com/writing/ewe.htm

[171] http://www.geonames.de/alphfj.html

[172] https://www.omniglot.com/writing/sorbian.htm

[173] http://files.peacecorps.gov/multimedia/audio/languagelessons/botswana/Bw\_Setswana\_Language\_Lessons.pdf

[174] http://omniglot.com/writing/tswana.php

[175] https://en.wikipedia.org/wiki/Afrikaans

[176] http://www.omniglot.com/writing/albanian.htm

[177] https://en.wikipedia.org/wiki/Albanian\_alphabet

[178] http://www.jesuitvolunteers.org/wp-content/uploads/2015/08/So\_you\_want\_to\_learn\_chuukese\_-\_only\_for\_Chuuk\_JVs.pdf

[179] https://en.wikipedia.org/wiki/Uyghur\_Latin\_alphabet

[180] http://www.omniglot.com/writing/drehu.php

[181] http://www.omniglot.com/writing/yoruba.htm

[182] http://www.omniglot.com/writing/haitiancreole.htm

[183] https://en.wikipedia.org/wiki/Haitian\_Creole#Orthography

[184] http://www.omniglot.com/writing/minangkabau.htm

[185] http://www.omniglot.com/writing/palauan.htm

[186] http://www.omniglot.com/writing/cubeo.htm

[187] https://www.sil.org/system/files/reapdata/10/58/27/10582785843693992331766506069073895620/40337\_01.pdf

[188] http://www.omniglot.com/writing/inarisami.htm

[189] http://www.omniglot.com/charts/dagbani.pdf

[190] http://www.omniglot.com/writing/ewondo.php

[191] http://www.omniglot.com/writing/ganda.php

[192] http://www.omniglot.com/writing/adzera.htm

[193] http://www.omniglot.com/writing/ga.htm

[194] http://www.omniglot.com/writing/duala.php

[195] http://www.omniglot.com/writing/soga.htm

[196] http://www.omniglot.com/writing/alur.htm

[197] http://www.omniglot.com/writing/mandinka.htm

[198] https://www.omniglot.com/writing/acholi.htm

[199] http://www.omniglot.com/writing/bambara.htm

[200] http://www.omniglot.com/writing/raga.htm

[201] http://www.omniglot.com/writing/tatar.htm

[202] https://www.omniglot.com/writing/zazaki.htm

[203] https://en.wikipedia.org/wiki/Turkish\_alphabet

[204] https://www.degruyter.com/downloadpdf/j/psicl.2007.43.issue-1/v10010-007-0009-0/v10010-007-0009-0.pdf

[205] http://www.omniglot.com/writing/igbo.htm

[206] https://www.italianpod101.com/italian-accents

[207] http://www.affaritaliani.it/blog/monica-la-pensa-cosi

[208] http://dictionary.reverso.net/italian-english/venerd%C3%AC

[209] http://www.omniglot.com/writing/kikuyu.htm

[210] http://www.omniglot.com/writing/hixkaryana.htm

[211] http://www.omniglot.com/writing/maasai.htm

[212] http://www.omniglot.com/writing/mossi.htm

[213] http://www.omniglot.com/babel/marshallese.htm

[214] https://en.wikipedia.org/wiki/Cedilla#Marshallese

[215] https://en.wikipedia.org/wiki/Marshallese\_language#Display\_issues

[216] http://www.trussel2.com/MOD/

[217] http://www.omniglot.com/writing/tswa.htm

[218] https://www.omniglot.com/writing/susu.htm

[219] https://www.omniglot.com/writing/zarma.htm

[220] https://www.omniglot.com/writing/pitjantjatjara.htm

[221] http://www.omniglot.com/writing/spanish.htm

[222] http://www.omniglot.com/writing/filipino.htm

[223] http://www.omniglot.com/writing/chavacano.php

[224] https://en.wikipedia.org/wiki/Ilocano\_language#Modern\_alphabet

[225] http://www.omniglot.com/writing/quechua.htm

[226] https://en.wikipedia.org/wiki/Quechua\_alphabet

[227] http://www.omniglot.com/writing/kriol.php

[228] http://www.omniglot.com/writing/waray.php

[229] http://www.omniglot.com/writing/lozi.htm

[230] http://africanlanguages.com/northern\_sotho/

[231] https://www.omniglot.com/writing/lowgerman.htm

[232] https://en.wikipedia.org/wiki/Chechen\_language

[233] http://www.omniglot.com/writing/hungarian.htm

[234] https://en.wikipedia.org/wiki/Hungarian\_alphabet

[235] http://www.omniglot.com/writing/khoekhoe.htm

[236] http://www.omniglot.com/writing/lingala.htm

[237] https://www.omniglot.com/writing/akan.htm

[238] https://en.wikipedia.org/wiki/Mossi\_language

[239] https://www.sil.org/system/files/reapdata/10/06/46/100646256099282892829790816212446104791/OPSL\_9.pdf (p. 75)

[240] http://www.omniglot.com/writing/kanuri.htm

[241] http://www.omniglot.com/writing/bugis.htm

[242] http://www.omniglot.com/writing/mizo.htm

[243] http://www.omniglot.com/writing/miskito.htm

[244] http://www.omniglot.com/writing/zazaki.htm

[245] https://en.wikipedia.org/wiki/Papiamento

[246] http://www.omniglot.com/writing/papiamento.php

[247] http://www.omniglot.com/writing/chichewa.php

[248] http://www.native-languages.org/mam\_words.htm

[249] http://www.omniglot.com/writing/mam.htm

[250] https://en.wikipedia.org/wiki/Pulaar\_language

[251] https://en.wikipedia.org/wiki/Fula\_language#Writing\_systems

[252] https://en.wikipedia.org/wiki/Polish\_alphabet

[253] https://en.wikipedia.org/wiki/French\_orthography

[254] https://www.omniglot.com/writing/yoruba.htm

## 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). <http://www.icann.org/en/resources/idn/variant-tlds/draft-lgrprocedure-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-)

[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.

Appendix A: Table of processed languages used to develop Latin Script Repertoire

Data about population, used in creating this table, are collected in the year 2015.

|  | **Language** | **ISO 639-3** | **Population** | **EGIDS** |
| --- | --- | --- | --- | --- |
|  | [Afrikaans,](http://www.omniglot.com/writing/afrikaans.htm) | [afr](http://www-01.sil.org/iso639-3/documentation.asp?id=afr) | 7,096,810 | 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) | 5,367,000 | 1 |
|  | [Azeri,Azerbaijani](http://www.omniglot.com/writing/azeri.htm) | [azj](http://www-01.sil.org/iso639-3/documentation.asp?id=azj) | 24,226,940 | 1 |
|  | [Chamorro,Chamorru Tjamoro](http://www.omniglot.com/writing/chamorro.htm) | [cha](http://www-01.sil.org/iso639-3/documentation.asp?id=cha) | 94,700 | 1 |
|  | [Croatian,Hrvatski](http://www.omniglot.com/writing/croatian.htm) | [hrv](http://www-01.sil.org/iso639-3/documentation.asp?id=hrv) | 5,609,290 | 1 |
|  | [Czech Bohemian Cestina](http://www.omniglot.com/writing/czech.htm) | [ces](http://www-01.sil.org/iso639-3/documentation.asp?id=ces) | 10,619,340 | 1 |
|  | [DanishDansk Rigsdansk](http://www.omniglot.com/writing/danish.htm) | [dan](http://www-01.sil.org/iso639-3/documentation.asp?id=dan) | 5,520,860 | 1 |
|  | [Dutch Hollands Nederlands](http://www.omniglot.com/writing/dutch.htm) | [nld](http://www-01.sil.org/iso639-3/documentation.asp?id=nld) | 21,944,690 | 1 |
|  | [English](http://www.omniglot.com/writing/english.htm) | [eng](http://www-01.sil.org/iso639-3/documentation.asp?id=eng) | 341,000,000 | 1 |
|  | [Estonian Eesti keel](http://www.omniglot.com/writing/estonian.htm) | [ekk](http://www-01.sil.org/iso639-3/documentation.asp?id=ekk) | 1,132,500 | 1 |
|  | [Filipino](http://www.omniglot.com/writing/filipino.htm) | [fil](http://www-01.sil.org/iso639-3/documentation.asp?id=fil) | 45,000,000 | 1 |
|  | [Finnish, Suomi](http://www.omniglot.com/writing/finnish.htm) | [fin](http://www-01.sil.org/iso639-3/documentation.asp?id=fin) | 5,392,180 | 1 |
|  | [French,Français](http://www.omniglot.com/writing/french.htm) | [fra](http://www-01.sil.org/iso639-3/documentation.asp?id=fra) | 87,000,000 | 1 |
|  | [German Deutsch Tedesco](http://www.omniglot.com/writing/german.htm) | [deu](http://www-01.sil.org/iso639-3/documentation.asp?id=deu) | 69,800,000 | 1 |
|  | [Greenlandic Kalaallisut, Inuktitut,](http://www.omniglot.com/writing/greenlandic.htm) | [kal](http://www-01.sil.org/iso639-3/documentation.asp?id=kal) | 57,000 | 1 |
|  | [Guarani Avañe’e Paraguayan](http://www.omniglot.com/writing/guarani.htm) | grn | 4,939,180 | 1 |
|  | [Haitian CreoleCreole Haitian Creole Western Caribbean Creole](http://www.omniglot.com/writing/haitiancreole.htm) | [hat](http://www-01.sil.org/iso639-3/documentation.asp?id=hat) | 7,731,240 | 1 |
|  | [Hungarian Magyar](http://www.omniglot.com/writing/hungarian.htm) | [hun](http://www-01.sil.org/iso639-3/documentation.asp?id=hun) | 9,840,000 | 1 |
|  | [Icelandic Íslenska](http://www.omniglot.com/writing/icelandic.htm) | [isl](http://www-01.sil.org/iso639-3/documentation.asp?id=isl) | 300,000 | 1 |
|  | [Indonesian](http://www.omniglot.com/writing/indonesian.htm) | [ind](http://www-01.sil.org/iso639-3/documentation.asp?id=ind) | 22,800,000 | 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) | 1,000,000 | 1 |
|  | [Italian Italiano](http://www.omniglot.com/writing/italian.htm) | [ita](http://www-01.sil.org/iso639-3/documentation.asp?id=ita) | 63,738,247 | 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) | 12,881,960 | 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) | 7,189,900 | 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) | 116,280 | 1 |
|  | [Kirundi, Rundi Urundi,](http://www.omniglot.com/writing/kirundi.php) | [run](http://www-01.sil.org/iso639-3/documentation.asp?id=run) | 10,000,000 | 1 |
|  | [Latvian,  “Lettisch” (pej,), “Lettish” (pej,)](http://www.omniglot.com/writing/latvian.htm) | lav | 1,552,260 | 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) | 3,001,430 | 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) | 7,529,640 | 1 |
|  | [Malay,](http://www.omniglot.com/writing/malay.htm) | msa | 60,475,586 | 1 |
|  | [Maltese, Malti](http://www.omniglot.com/writing/maltese.htm) | [mlt](http://www-01.sil.org/iso639-3/documentation.asp?id=mlt) | 522,000 | 1 |
|  | [Marshallese,  Ebon](http://www.omniglot.com/writing/marshallese.php) | [mah](http://www-01.sil.org/iso639-3/documentation.asp?id=mah) | 54,600 | 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) | 1,090,000 | 1 |
|  | [Niuean,Niue, “Niuefekai” (pej,)](http://www.omniglot.com/writing/niuean.php) | [niu](http://www-01.sil.org/iso639-3/documentation.asp?id=niu) | 7,767 | 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) | 4,631,000 | 1 |
|  | [Norwegian,Norsk](http://www.omniglot.com/writing/norwegian.htm) | [nor](http://www-01.sil.org/iso639-3/documentation.asp?id=nor) | 4,741,780 | 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) | 263,200 | 1 |
|  | [Polish, Polnisch, Polski](http://www.omniglot.com/writing/polish.htm) | [pol](http://www-01.sil.org/iso639-3/documentation.asp?id=pol) | 38,636,480 | 1 |
|  | [Portuguese,](http://www.omniglot.com/writing/portuguese.htm) | [por](http://www-01.sil.org/iso639-3/documentation.asp?id=por) | 203,352,100 | 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) | 23,681,610 | 1 |
|  | [Samoan,](http://www.omniglot.com/writing/samoan.htm) | [smo](http://www-01.sil.org/iso639-3/documentation.asp?id=smo) | 413,257 | 1 |
|  | [Sango, Sangho](http://www.omniglot.com/writing/sango.php) | [sag](http://www-01.sil.org/iso639-3/documentation.asp?id=sag) | 404,000 | 1 |
|  | [Serbian, srpski,српски,](http://www.omniglot.com/writing/serbian.htm) | [srp](http://www-01.sil.org/iso639-3/documentation.asp?id=srp) | 8,638,906 | 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) | 72,700 | 1 |
|  | [Slovak, Slovakian, Slovencina](http://www.omniglot.com/writing/slovak.htm) | [slk](http://www-01.sil.org/iso639-3/documentation.asp?id=slk) | 4,750,000 | 1 |
|  | [Slovenian, Slovenscina,Slovene](http://www.omniglot.com/writing/slovene.htm) | [slv](http://www-01.sil.org/iso639-3/documentation.asp?id=slv) | 2,085,160 | 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) | 14,762,900 | 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) | 5,634,000 | 1 |
|  | [Spanish, Castellano, Castilian, Español](http://www.omniglot.com/writing/spanish.htm) | [spa](http://www-01.sil.org/iso639-3/documentation.asp?id=spa) | 398,931,840 | 1 |
|  | [Swahili, Kisuaheli, Kiswahili](http://www.omniglot.com/writing/swahili.htm) | [swh](http://www-01.sil.org/iso639-3/documentation.asp?id=swh) | 15,437,390 | 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) | 2,324,200 | 1 |
|  | [Swedish, Ruotsi, Svenska](http://www.omniglot.com/writing/swedish.htm) | [swe](http://www-01.sil.org/iso639-3/documentation.asp?id=swe) | 9,197,090 | 1 |
|  | [Tahitian,](http://www.omniglot.com/writing/tahitian.htm) | [tah](http://www-01.sil.org/iso639-3/documentation.asp?id=tah) | 68,260 | 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) | 122,000 | 1 |
|  | [Tongan, Tonga](http://www.omniglot.com/writing/tongan.htm) | [ton](http://www-01.sil.org/iso639-3/documentation.asp?id=ton) | 169,120 | 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) | 4,009,000 | 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) | 5,181,700 | 1 |
|  | [Turkish, Anatolian, Türkçe, Türkisch](http://www.omniglot.com/writing/turkish.htm) | [tur](http://www-01.sil.org/iso639-3/documentation.asp?id=tur) | 70,890,130 | 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) | 3,470,000 | 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) | 27,001,230 | 1 |
|  | [Venda, Chivenda, Tshivenda](http://www.omniglot.com/writing/venda.htm) | [ven](http://www-01.sil.org/iso639-3/documentation.asp?id=ven) | 1,294,000 | 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) | 73,600,000 | 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) | 8,177,300 | 1 |
|  | [Zulu, Isizulu, Zunda](http://www.omniglot.com/writing/zulu.htm) | [zul](http://www-01.sil.org/iso639-3/documentation.asp?id=zul) | 15,700,000 | 1 |
|  | [BasqueEuskara Euskera Vascuense](http://www.omniglot.com/writing/basque.htm) | [eus](http://www-01.sil.org/iso639-3/documentation.asp?id=eus) | 2,000,000 | 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) | 4,079,420 | 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) | 1,430,000 | 2 |
|  | [ChuukeseChuuk Lagoon Chuukese Ruk Truk Trukese](http://www.omniglot.com/writing/chuukese.htm) | [chk](http://www-01.sil.org/iso639-3/documentation.asp?id=chk) | 45,900 | 2 |
|  | [FaroeseFøroyskt](http://www.omniglot.com/writing/faroese.htm) | [fao](http://www-01.sil.org/iso639-3/documentation.asp?id=fao) | 66,150 | 2 |
|  | [FrisianFries Frysk](http://www.omniglot.com/writing/westfrisian.htm) | [fry](http://www-01.sil.org/iso639-3/documentation.asp?id=fry) | 467,000 | 2 |
|  | [Galician Galego Gallego](http://www.omniglot.com/writing/galician.htm) | [glg](http://www-01.sil.org/iso639-3/documentation.asp?id=glg) | 2,340,000 | 2 |
|  | [GaroGarrow Mande Mandi](http://www.omniglot.com/writing/garo.htm) | [grt](http://www-01.sil.org/iso639-3/documentation.asp?id=grt) | 1,009,000 | 2 |
|  | [HausaAbakwariga Habe Haoussa Hausawa Kado Mgbakpa](http://www.omniglot.com/writing/hausa.htm) | [hau](http://www-01.sil.org/iso639-3/documentation.asp?id=hau) | 25,109,000 | 2 |
|  | [HawaiianOlelo 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) | 336,000 | 2 |
|  | [Igbo](http://www.omniglot.com/writing/igbo.htm) | [ibo](http://www-01.sil.org/iso639-3/documentation.asp?id=ibo) | 18,000,000 | 2 |
|  | [Inari Sámi Anarâškielâ Anar “Finnish Lapp” (pej,) “Inari Lappish” (pej,) “Lapp” (pej,) Saam Saame Sámi Samic](http://www.omniglot.com/writing/inarisami.htm) | [smn](http://www-01.sil.org/iso639-3/documentation.asp?id=smn) | 4,700 | 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) | 2,423,540 | 2 |
|  | [Kurdish,](http://www.omniglot.com/writing/kurdish.htm) | [kur](http://www-01.sil.org/iso639-3/documentation.asp?id=kur) | 29,972,890 | 2 |
|  | [Lingala, Ngala](http://www.omniglot.com/writing/lingala.htm) | [lin](http://www-01.sil.org/iso639-3/documentation.asp?id=lin) | 2,141,300 | 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) | 6,000 | 2 |
|  | [Mirandese, Mirandês](http://www.omniglot.com/writing/mirandese.htm) | [mwl](http://www-01.sil.org/iso639-3/documentation.asp?id=mwl) | 15,000 | 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) | 183,000 | 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) | 40,000 | 2 |
|  | [Palauan, Belauan, Palau](http://www.omniglot.com/writing/palauan.htm) | [pau](http://www-01.sil.org/iso639-3/documentation.asp?id=pau) | 14,950 | 2 |
|  | [Pohnpeian, Ponapean](http://www.omniglot.com/writing/pohnpeian.htm) | [pon](http://www-01.sil.org/iso639-3/documentation.asp?id=pon) | 31,350 | 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) | 600 | 2 |
|  | [Tatar, Tartar](http://www.omniglot.com/writing/tatar.htm) | [tat](http://www-01.sil.org/iso639-3/documentation.asp?id=tat) | 5,310,000 | 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) | 6,300,000 | 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) | 10,416,910 | 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) | 805,700 | 2 |
|  | [Welsh, Cymraeg](http://www.omniglot.com/writing/welsh.htm) | [cym](http://www-01.sil.org/iso639-3/documentation.asp?id=cym) | 536,890 | 2 |
|  | [West Frisian, Fries, Frysk](http://www.omniglot.com/writing/westfrisian.htm) | [fry](http://www-01.sil.org/iso639-3/documentation.asp?id=fry) | 467,000 | 2 |
|  | [Yapese,](http://www.omniglot.com/writing/yapese.htm) | [yap](http://www-01.sil.org/iso639-3/documentation.asp?id=yap) | 7,390 | 2 |
|  | [Yoruba, Yariba, Yooba](http://www.omniglot.com/writing/yoruba.htm) | [yor](http://www-01.sil.org/iso639-3/documentation.asp?id=yor) | 19,380,800 | 2 |
|  | Akan, Twi, Ajan Twi | [aka](http://www-01.sil.org/iso639-3/documentation.asp?id=aka) | 8,314,600 | 3 |
|  | Bislama, Bichelamar | [bis](http://www-01.sil.org/iso639-3/documentation.asp?id=bis) | 10,000 | 3 |
|  | [BugisBasa 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) | 5,017,800 | 3 |
|  | Cebuano, Binisaya Bisayan Sebuano Sugbuanon Sugbuhanon Visayan | [ceb](http://www-01.sil.org/iso639-3/documentation.asp?id=ceb) | 15,810,000 | 3 |
|  | [ChichewaChewa Chinyanja Nyanja Nyanja-Chewa](http://www.omniglot.com/writing/chichewa.php) | [nya](http://www-01.sil.org/iso639-3/documentation.asp?id=nya) | 7,000,000 | 3 |
|  | Cubeo Cuveo Hehenawa Hipnwa Kobeua Kobewa Kubwa Pamiwa | [cub](http://www-01.sil.org/iso639-3/documentation.asp?id=cub) | 6,260 | 3 |
|  | [DualaDiwala Douala Dualla Dwala Dwela Sawa](http://www.omniglot.com/writing/duala.php) | [dua](http://www-01.sil.org/iso639-3/documentation.asp?id=dua) | 2,000,000 | 3 |
|  | [EweEbwe Efe Eibe Eue Eve Gbe Krepe Krepi Popo VheEʋegbe](http://www.omniglot.com/writing/ewe.htm) | [ewe](http://www-01.sil.org/iso639-3/documentation.asp?id=ewe) | 3,112,000 | 3 |
|  | [EwondoEwundu Jaunde Yaounde Yaunde](http://www.omniglot.com/writing/ewondo.php) | [ewo](http://www-01.sil.org/iso639-3/documentation.asp?id=ewo) | 578,000 | 3 |
|  | [FanagaloFanakaloPidgin Zulu Fanekolo Isikula Lololo or Isilololo Piki or Isipiki Silunguboi,ChilapalapaCikabanga](http://www.omniglot.com/writing/fanagalo.htm) | [fng](http://www-01.sil.org/iso639-3/documentation.asp?id=fng) | 200,000 | 3 |
|  | [FonDahomeen Fongbe](http://www.omniglot.com/writing/fon.htm) | [fon](http://www-01.sil.org/iso639-3/documentation.asp?id=fon) | 1,435,500 | 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) | 13,000,000 | 3 |
|  | [Ganda Luganda](http://www.omniglot.com/writing/ganda.php) | [lug](http://www-01.sil.org/iso639-3/documentation.asp?id=lug) | 4,130,000 | 3 |
|  | [Hiligaynon Hiligainon Illogo Ilonggo](http://www.omniglot.com/writing/hiligaynon.htm) | [hil](http://www-01.sil.org/iso639-3/documentation.asp?id=hil) | 5,770,000 | 3 |
|  | [Iban Dayak](http://www.omniglot.com/writing/iban.htm) | [iba](http://www-01.sil.org/iso639-3/documentation.asp?id=iba) | 788,400 | 3 |
|  | [IlokoIlokano Ilocano](http://www.omniglot.com/writing/ilocano.htm) | [ilo](http://www-01.sil.org/iso639-3/documentation.asp?id=ilo) | 6,920,000 | 3 |
|  | [Kanuri,](http://www.omniglot.com/writing/kanuri.htm) | [kau](http://www-01.sil.org/iso639-3/documentation.asp?id=kau) | 3,760,500 | 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) | 1,905,430 | 3 |
|  | [Latin,Latina](http://www.omniglot.com/writing/latin2.htm) | [lat](http://www-01.sil.org/iso639-3/documentation.asp?id=lat) | 0 | 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) | 850,000 | 3 |
|  | [Masbateño, Masbatenyo, Minasbate](http://www.omniglot.com/writing/masbateno.htm) | [msb](http://www-01.sil.org/iso639-3/documentation.asp?id=msb) | 700,000 | 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) | 5,075,300 | 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) | 30,000 | 3 |
|  | [Nauruan,](http://www.omniglot.com/writing/nauruan.htm) | [nau](http://www-01.sil.org/iso639-3/documentation.asp?id=nau) | 6,000 | 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) | 668,000 | 3 |
|  | [Pangasinan,](http://www.omniglot.com/writing/pangasinan.htm) | [pag](http://www-01.sil.org/iso639-3/documentation.asp?id=pag) | 1,162,140 | 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) | 24,400 | 3 |
|  | [Quechua, Runasimi ,Qhichwa simi](http://www.omniglot.com/writing/quechua.htm) | [que](http://www-01.sil.org/iso639-3/documentation.asp?id=que) | 8,912,820 | 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) | 6,500 | 3 |
|  | Roviana,Robiana, Rubiana, Ruviana | [rug](http://www-01.sil.org/iso639-3/documentation.asp?id=rug) | 9,870 | 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) | 10,741,700 | 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) | 126,400 | 3 |
|  | [Tagalog,](http://www.omniglot.com/writing/tagalog.htm) | [tgl](http://www-01.sil.org/iso639-3/documentation.asp?id=tgl) | 24,310,000 | 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) | 1,062,000 | 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) | 6,040 | 3 |
|  | [Tuvaluan, Ellice, Ellicean, Tuvalu](http://www.omniglot.com/writing/tuvaluan.htm) | [tvl](http://www-01.sil.org/iso639-3/documentation.asp?id=tvl) | 12,380 | 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) | 6,000,000 | 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) | 2,560,000 | 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) | 1,710,000 | 3 |
|  | [Zhuang,Nong](http://www.omniglot.com/writing/zhuang.htm) | [zha](http://www-01.sil.org/iso639-3/documentation.asp?id=zha) | 14,936,200 | 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) | 30,000 | 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) | 460,000 | 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) | 2,380 | 4 |
|  | [Bambara,Bamanankan](http://www.omniglot.com/writing/bambara.htm) | [bam](http://www-01.sil.org/iso639-3/documentation.asp?id=bam) | 4,072,040 | 4 |
|  | [BashkirBashkir Bashqort Basquort](http://www.omniglot.com/writing/bashkir.htm) | [bak](http://www-01.sil.org/iso639-3/documentation.asp?id=bak) | 1,590,000 | 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) | 1,200,200 | 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) | 105,000 | 4 |
|  | [Chavacano,Chabacano Chabakano Zamboangueño](http://www.omniglot.com/writing/chavacano.php) | [cbk](http://www-01.sil.org/iso639-3/documentation.asp?id=cbk) | 1,200,000 | 4 |
|  | [CorsicanCorse Corsi Corso Corsu](http://www.omniglot.com/writing/corsican.htm) | [cos](http://www-01.sil.org/iso639-3/documentation.asp?id=cos) | 31,000 | 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) | 1,000,000 | 4 |
|  | [DagbaniDagbamba Dagbane Dagomba](http://www.omniglot.com/writing/dagbani.htm) | [dag](http://www-01.sil.org/iso639-3/documentation.asp?id=dag) | 800,000 | 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) | 1,365,900 | 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) | 15,586 | 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) | 337,000 | 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) | 300,000 | 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) | 600,000 | 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) | 680 | 4 |
|  | [Ifugao, Ifugaw, Mayaoyaw, Mayoyao](http://www.omniglot.com/writing/ifugao.htm) | [ifu](http://www-01.sil.org/iso639-3/documentation.asp?id=ifu) | 30,000 | 4 |
|  | [Ixil](http://www.omniglot.com/writing/ixil.htm) | [ixl](http://www-01.sil.org/iso639-3/documentation.asp?id=ixl) | 69,000 | 4 |
|  | [JavaneseDjawa Jawa](http://www.omniglot.com/writing/javanese.htm) | [jav](http://www-01.sil.org/iso639-3/documentation.asp?id=jav) | 84,308,740 | 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) | 30,000 | 4 |
|  | [Kaqchikel, Cakchiquel, Kaqchikel, Kaqchiquel](http://www.omniglot.com/writing/kaqchikel.htm) | [cak](http://www-01.sil.org/iso639-3/documentation.asp?id=cak) | 451,000 | 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) | 251,000 | 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) | 2,330,000 | 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) | 724,600 | 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) | 336,710 | 4 |
|  | [Mam, Huehuetenango Mam](http://www.omniglot.com/writing/mam.htm) | [mam](http://www-01.sil.org/iso639-3/documentation.asp?id=mam) | 537,980 | 4 |
|  | [Maranao, Maranaw, Ranao](http://www.omniglot.com/writing/maranao.htm) | [mrw](http://www-01.sil.org/iso639-3/documentation.asp?id=mrw) | 776,000 | 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) | 4,500 | 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) | 687,750 | 4 |
|  | [Nuer, Naadh, Naath](http://www.omniglot.com/writing/nuer.htm) | [nus](http://www-01.sil.org/iso639-3/documentation.asp?id=nus) | 891,000 | 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) | 2,000,000 | 4 |
|  | [Pitjantjatjara, Pitjantjara](http://www.omniglot.com/writing/pitjantjatjara.htm) | [pjt](http://www-01.sil.org/iso639-3/documentation.asp?id=pjt) | 2,660 | 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) | 823,500 | 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) | 40,039 | 4 |
|  | [Scottish Gaelic,Gaelic-Scotish](http://www.omniglot.com/writing/gaelic.htm) | [gla](http://www-01.sil.org/iso639-3/documentation.asp?id=gla) | 63,130 | 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) | 9,600 | 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) | 45,000 | 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) | 1,060,280 | 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) | 43,000 | 4 |
|  | [Talysh, Talesh, Talish, Talyshi](http://www.omniglot.com/writing/talysh.htm) | [tly](http://www-01.sil.org/iso639-3/documentation.asp?id=tly) | 915,400 | 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) | 2,566,000 | 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) | 268,000 | 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) | 3,930,000 | 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) | 2,440,100 | 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) | 140,000 | 4 |
|  | Acehnese, Achehnese AchineseAceh | [ace](http://www-01.sil.org/iso639-3/documentation.asp?id=ace) | 3,500,032 | 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) | 1,197,000 | 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) | 30,000,000 | 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) | 1,379,200 | 5 |
|  | Alsatian, Elsässerdeutsche Alsacien Alemanic Alemannisch Schwyzerdütsch | [gsw](http://www-01.sil.org/iso639-3/documentation.asp?id=gsw) | 6,319,000 | 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) | 1,367,000 | 5 |
|  | Bavarian, Bairisch Bavarian Austrian Bayerisch Ost-Oberdeutsch | [bar](http://www-01.sil.org/iso639-3/documentation.asp?id=bar) | 7,830,000 | 5 |
|  | Brahui, Birahui Brahuidi Brahuigi Kur Galli | [brh](http://www-01.sil.org/iso639-3/documentation.asp?id=brh) | 4,220,000 | 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) | 4,044,000 | 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) | 3,205,000 | 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) | 1,005,000 | 5 |
|  | [Kikuyu, Gĩkũyũ, Gekoyo, Gigikuyu,](http://www.omniglot.com/writing/kikuyu.htm) | [kik](http://www-01.sil.org/iso639-3/documentation.asp?id=kik) | 6,623,000 | 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) | 3,000,000 | 5 |
|  | [Maasai, Maa, Masai](http://www.omniglot.com/writing/maasai.htm) | [mas](http://www-01.sil.org/iso639-3/documentation.asp?id=mas) | 1,297,000 | 5 |
|  | [Madurese,Madura, Basa Mathura](http://www.omniglot.com/writing/madurese.htm) | [mad](http://www-01.sil.org/iso639-3/documentation.asp?id=mad) | 6,770,900 | 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) | 3,090,000 | 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) | 1,346,000 | 5 |
|  | [Minangkabau, Minang, Padang](http://www.omniglot.com/writing/minangkabau.htm) | [min](http://www-01.sil.org/iso639-3/documentation.asp?id=min) | 5,530,000 | 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) | 1,120,280 | 5 |
|  | [Neapolitan,Napoletano, Neapolitan-Calabrese](http://www.omniglot.com/writing/neapolitan.php) | [nap](http://www-01.sil.org/iso639-3/documentation.asp?id=nap) | 5,700,000 | 5 |
|  | [Piedmontese,Piemontese, Piemontèis](http://www.omniglot.com/writing/piedmontese.htm) | [pms](http://www-01.sil.org/iso639-3/documentation.asp?id=pms) | 1,600,000 | 5 |
|  | [Romany,](http://www.omniglot.com/writing/romany.htm) | [rom](http://www-01.sil.org/iso639-3/documentation.asp?id=rom) | 2,600,180 | 5 |
|  | [Sasak, Lombok](http://www.omniglot.com/writing/sasak.htm) | [sas](http://www-01.sil.org/iso639-3/documentation.asp?id=sas) | 2,100,000 | 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) | 4,700,000 | 5 |
|  | [Soga, Lusoga, Olusoga](http://www.omniglot.com/writing/soga.htm) | [xog](http://www-01.sil.org/iso639-3/documentation.asp?id=xog) | 2,060,000 | 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) | 2,114,000 | 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) | 1,180,000 | 5 |
|  | [Venetian, Talian, Venet](http://www.omniglot.com/writing/venetian.htm) | [vec](http://www-01.sil.org/iso639-3/documentation.asp?id=vec) | 3,852,500 | 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) | 3,000,000 | 5 |
|  |   |   | **2,324,515,350** |  |

# Appendix B: Statistics about processed languages using Latin script

Data about population, used in creating following tables, are collected in the year 2015. Total world population in 2015 was 7.3 billion according to [Population Reference Bureau](http://www.prb.org/pdf15/2015-world-population-data-sheet_eng.pdf).

## B.1 Number of languages processed per EGIDS

Processed Languages using Latin script per Expanded Graded Intergenerational Disruption Scale - EGIDS

|  |  |  |
| --- | --- | --- |
| **EGIDS** | **Number of languages** | **Population** |
| 1 | 64 | 1,792,271,753 |
| 2 | 30 | 132,940,490 |
| 3 | 39 | 158,668,240 |
| 4 | 47 | 118,884,775 |
| 5 | 29 | 121,750,092 |
| **In total** | **209** | **2,324,515,350** |

## B.2 Number of languages processed per continent

|  |  |  |
| --- | --- | --- |
| **Continent** | **Number of languages** | **Population** |
| Africa | 62 | 319,264,920 |
| Asia | 49 | 530,793,372 |
| Central America | 9 | 15,593,920 |
| Europe | 49 | 430,511,142 |
| North America | 2 | 393,000 |
| Oceania | 28 | 70,680,116 |
| South America | 6 | 13,994,940 |
| English, Portuguese, Spanish | 3 | 943,283,940 |
| Latin | 1 | 0 |
|  **In total:** | **209** | **2,324,515,350** |

## B.3 Number of processed languages per language family

| **Language classification** | **Number of languages** | **Population** |
| --- | --- | --- |
| Afro-Asiatic | 6 | 73,483,100 |
| Australian | 2 | 5,040 |
| Austro-Asiatic, | 3 | 75,525,980 |
| Austronesian Malayo-Polynesian | 46 | 307,246,078 |
| Cariban Waiwai | 1 | 680 |
| Creole | 14 | 15,245,180 |
| Dravidian Northern | 1 | 4,220,000 |
| Eskimo-Aleut | 1 | 57,000 |
| Indo-European | 50 | 1,398,405,742 |
| Jean, Central | 1 | 9,600 |
| Khoe-Kwadi | 1 | 251,000 |
| Language isolate(Basque/Euskara) | 1 | 2,000,000 |
| Mayan Yucatecan | 5 | 4,211,480 |
| Misumalpan | 1 | 183,000 |
| Niger-Congo | 43 | 220,716,610 |
| Nilo-Saharan | 8 | 16,362,500 |
| North Caucasian | 1 | 1,430,000 |
| Pidgin, Zulu based | 1 | 200,000 |
| Quechuan | 1 | 8,912,820 |
| Sino-Tibetan | 3 | 3,696,750 |
| Tai-Kadai | 1 | 14,936,200 |
| Tucanoan | 1 | 6,260 |
| Tupian | 1 | 4,939,180 |
| Turkic | 9 | 156,055,170 |
| Uralic | 7 | 16,415,980 |
|  **In total:** | **209** | **2,324,515,350** |

## B.4 Number of states represented per continent

All countries speaking English, Spanish and Portuguese languages are not counted in this table. These languages are counted only in United Kingdom, Spain and Portugal.

|  |  |
| --- | --- |
| **Continent** | **Number of states** |
| Africa | 41 |
| Asia | 22 |
| Central America | 9 |
| Europe | 35 |
| North America | 0 |
| Oceania | 20 |
| South America | 7 |
| **In total** | **134** |

# Appendix C: Repertoire table grouped by Glyph

| # | Comment on inclusion | Unicode | Glyph | Unicode name | Languages using the code point (EGIDS) | Reference supporting inclusion (URL etc.) |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **0061** | **a** | LATIN SMALL LETTER A | Many languages | **ASCII a-z** |
|  |  | **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 (1)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)Bavarian ( 5) | [109], [110], [111], [112] |
|  |  | **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) | [118] |
|  |  | **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) | [118] |
|  |  | **1EAB** | **ẫ** | LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EAD** | **ậ** | LATIN SMALL LETTER A WITH CIRCUMFLEXAND 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 | Many languages | **ASCII a-z** |
|  |  | **0253** | **ɓ** | LATIN SMALL LETTER B WITH HOOK | Hausa (2)Dagaare - Burkina Faso (4) Pulaar, (3) | [147], [148], [149] |
|  |  | **0063** | **c** | LATIN SMALL LETTER C | Many languages | **ASCII a-z** |
|  |  | **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], [162], [116], [127] |
|  |  | **0107** | **ć** | LATIN SMALL LETTER C WITH ACUTE | Croatian (1)Serbian (1)Polish (1) | [150], [151], [152] |
|  |  | **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)Kabyle (5) | [150], [151], [152], [133], [153], [108], [154], [155], [156] |
|  |  | **0064** | **d** | LATIN SMALL LETTER D | Many languages | **ASCII a-z** |
|  |  | **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 SámiBrahui (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], [167] |
|  |  | **1E0D** | **ḍ** | LATIN SMALL LETTER D WITH DOT BELOW | Mundari (5)Kabyle (5) | [165], [155], [156]  |
|  |  | **1E13** | **ḓ** | LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **0065** | **e** | LATIN SMALL LETTER E | Many languages | **ASCII a-z** |
|  |  | **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], [162], [132], [117], [171] |
|  |  | **00EA** | **ê** | LATIN SMALL LETTER E WITH CIRCUMFLEX | French (1)Tswana (1)Afrikaans (1)Vietnamese (1)Kurdish (2)Kirundi (1)West Frisian (1)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], [171], [129] |
|  |  | **0113** | **ē** | LATIN SMALL LETTER E WITH MACRON | Latvian (1)Hawaiian (2)Tongan (1)Minangkabau (5) | [133], [135], [134], [184] |
|  |  | **0115** | **ĕ** | LATIN SMALL LETTER E WITH BREVE | Bavarian(5) | [111], [112] |
|  |  | **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] |
|  | **Note complexity of "schwa" and "turned E" and that those are homeglyphs in lower case. 2017-12-31/MD** | **01DD** | **ǝ** | LATIN SMALL LETTER TURNED E | Kanuri (3) | [240] |
|  | **Note complexity of "schwa" and "turned E", and that those are homeglyphs in lower case. 2017-12-31/MD** | **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) | [118] |
|  |  | **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) | [118] |
|  |  | **1EC1** | **ề** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EC3** | **ể** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EC5** | **ễ** | LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EC7** | **ệ** | LATIN SMALL LETTER E WITH CIRCUMFLEXAND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **0066** | **f** | LATIN SMALL LETTER F | Many languages | **ASCII a-z** |
|  |  | **0192** | **ƒ** | LATIN SMALL LETTER F WITH HOOK | Ewe(3) | [170] |
|  |  | **0067** | **g** | LATIN SMALL LETTER G | Many languages | **ASCII a-z** |
|  |  | **0067 + 0303** | **g̃** | LATIN SMALL LETTER G + COMBING TILDE | Guarani (1) | [142], [143] |
|  |  | **0067 + 0304** | **ḡ** | LATIN SMALL LETTER G + COMBINING MACRON | Raga (Hano) (3) | [200] |
|  |  | **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 | Many languages | **ASCII a-z** |
|  |  | **0127** | **ħ** | LATIN SMALL LETTER H WITH STROKE | Maltese(1) | [163] |
|  |  | **1E25** | **ḥ** | LATIN SMALL LETTER H WITH DOT BELOW | Kabyle (5) | [155], [156] |
|  |  | **0069** | **i** | LATIN SMALL LETTER I | Many languages | **ASCII a-z** |
|  |  | **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], [207], [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], [162], [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], [171] |
|  |  | **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] |
|  |  | **01D0** | **ǐ** | LATIN SMALL LETTER I WITH CARON | Kirundi (1) | [104] |
|  | **Included in IP's proposal for MSR3.** | **0268** | **ɨ** | LATIN SMALL LETTER I WITH STROKE | Cubeo (3)Dagbani (Dagomba) (4)HIxkaryána (4)Maasai (5) | [186], [189], [210], [211] |
|  | **U+0268 included in IP's proposal for MSR3.** | **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) | [118] |
|  |  | **1ECB** | **ị** | LATIN SMALL LETTER I WITH DOT BELOW | Igbo (2) | [205] |
|  |  | **006A** | **j** | LATIN SMALL LETTER J | Many languages | **ASCII a-z** |
|  |  | **0269** | **ɩ** | LATIN SMALL LETTER IOTA | Dagaare - Burkina Faso (4) Mossi (3) | [148], [212] |
|  |  | **006B** | **k** | LATIN SMALL LETTER K | Many languages | **ASCII a-z** |
|  |  | **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 | Many languages | **ASCII a-z** |
|  |  | **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)Mundari (5) | [213], [214], [215], [216], [165] |
|  | **Included in IP's proposal for MSR3.** | **1E3D** | **ḽ** | LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **006D** | **m** | LATIN SMALL LETTER M | Many languages | **ASCII a-z** |
|  |  | **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 | Many languages | **ASCII a-z** |
|  |  | **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)Chomorro (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], [222], [142], [143], [223], [160], [162], [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] |
|  | **Included in IP's proposal for MSR3.** | **0272** | **ɲ** | LATIN SMALL LETTER N WITH LEFT HOOK | Susu (4)Zarma (4)Bambara (4) | [218], [219], [199] |
|  | **Only languages with good evidence are kept.** | **1E45** | **ṅ** | LATIN SMALL LETTER N WITH DOT ABOVE | Venda (1)Tswa (5) | [164], [217] |
|  |  | **1E47** | **ṇ** | LATIN SMALL LETTER N WITH DOT BELOW | Mundari (5)Marshallese (1) | [165], [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] |
|  |  | **006F** | **o** | LATIN SMALL LETTER O | Many languages | **ASCII a-z** |
|  |  | **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], [162], [132] |
|  |  | **00F4** | **ô** | LATIN SMALL LETTER O WITH CIRCUMFLEX | Tswana (1)Afrikaans (1)Vietnamese (1)Kirundi (1)French (1)Northern Sotho(1)West Frisian (1)Galician (2)Friulian (4)Xavante(4) | [173], [174], [175], [109], [104], [114], [230], [115], [162], [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 VersionWest Frisian (2)Nuer (4) | [119], [120], [175], [157], [123], [179], [124], [180], [126], [125], [127], [231], [232], [171], [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) | [118] |
|  |  | **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) | [118] |
|  |  | **1ED1** | **ố** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1ED3** | **ồ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1ED5** | **ổ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1ED7** | **ỗ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE | Vietnamese (1) | [118] |
|  |  | **1ED9** | **ộ** | LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EDB** | **ớ** | LATIN SMALL LETTER O WITH HORN AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1EDD** | **ờ** | LATIN SMALL LETTER O WITH HORN AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EDF** | **ở** | LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EE1** | **ỡ** | LATIN SMALL LETTER O WITH HORN AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EE3** | **ợ** | LATIN SMALL LETTER O WITH HORN AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **0070** | **p** | LATIN SMALL LETTER P  | Many languages | **ASCII a-z** |
|  |  | **00FE** | **þ** | LATIN SMALL LETTER THORN | Icelandic (1) | [102] |
|  |  | **0071** | **q** | LATIN SMALL LETTER Q | Many languages | **ASCII a-z** |
|  |  | **0072** | **r** | LATIN SMALL LETTER R | Many languages | **ASCII a-z** |
|  |  | **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] |
|  |  | **1E5B** | **ṛ** | LATIN SMALL LETTER R WITH DOT BELOW | Kabyle (5) | [155], [156] |
|  |  | **0073** | **s** | LATIN SMALL LETTER S | Many languages | **ASCII a-z** |
|  |  | **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 | Tswa (5) | [217] |
|  |  | **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)Nothert 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)Kabyle (5) | [181], [155], [156] |
|  |  | **0074** | **t** | LATIN SMALL LETTER T | Many languages | **ASCII a-z** |
|  |  | **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)Mundari (5)Kabyle (5) | [242], [165], [155], [156] |
|  |  | **1E71** | **ṱ** | LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW | Venda (1) | [164] |
|  |  | **0075** | **u** | LATIN SMALL LETTER U | Many languages | **ASCII a-z** |
|  |  | **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 (1)Galician (2) | [100], [101], [102], [103], [104], [105], [115], [162] |
|  |  | **00FB** | **û** | LATIN SMALL LETTER U WITH CIRCUMFLEX | Afrikaans (1) Kurdish (2)Kirundi (1)French (1)Miskito (2)West Frisian (1)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)Low German (5) | [119], [100], [175], [157], [123], [114], [159], [161], [162], [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] |
|  |  | **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 | Igbo (2) | [204], [205] |
|  |  | **1EE7** | **ủ** | LATIN SMALL LETTER U WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EE9** | **ứ** | LATIN SMALL LETTER U WITH HORN AND ACUTE | Vietnamese (1) | [118] |
|  |  | **1EEB** | **ừ** | LATIN SMALL LETTER U WITH HORN AND GRAVE | Vietnamese (1) | [118] |
|  |  | **1EED** | **ử** | LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EEF** | **ữ** | LATIN SMALL LETTER U WITH HORN AND TILDE | Vietnamese (1) | [118] |
|  |  | **1EF1** | **ự** | LATIN SMALL LETTER U WITH HORN AND DOT BELOW | Vietnamese (1) | [118] |
|  |  | **0076** | **v** | LATIN SMALL LETTER V | Many languages | **ASCII a-z** |
|  |  | **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 | Many languages | **ASCII a-z** |
|  |  | **0175** | **ŵ** | LATIN SMALL LETTER W WITH CIRCUMFLEX | Chichewa (3) | [247] |
|  |  | **0078** | **x** | LATIN SMALL LETTER X | Many languages | **ASCII a-z** |
|  |  | **1E8D** | **ẍ** | LATIN SMALL LETTER X WITH DIAERESIS | Mam (4) | [248], [249] |
|  |  | **0079** | **y** | LATIN SMALL LETTER Y | Many languages | **ASCII a-z** |
|  |  | **00FD** | **ý** | LATIN SMALL LETTER Y WITH ACUTE | Turkmen (1)Czech (1)Icelandic (1)Faroese (2)Guarani (1) | [121], [101], [102], [103], [142], [143] |
|  |  | **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) | [118] |
|  |  | **1EF5** | **ỵ** | LATIN SMALL LETTER Y WITH DOT BELOW | Vietnamese (1) | [118] |
|  |  | **1EF7** | **ỷ** | LATIN SMALL LETTER Y WITH HOOK ABOVE | Vietnamese (1) | [118] |
|  |  | **1EF9** | **ỹ** | LATIN SMALL LETTER Y WITH TILDE | Vietnamese (1) Guarani (1) | [118] [142] |
|  |  | **007A** | **z** | LATIN SMALL LETTER Z | Many languages | **ASCII a-z** |
|  |  | **017A** | **ź** | LATIN SMALL LETTER Z WITH ACUTE | Polish (1)Brahui (5)(Lower) 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] |
|  |  | **1E91** | **ẑ** | LATIN SMALL LETTER Z WITH CIRCUMFLEX | Tswa (5) | [217] |
|  |  | **1E93** | **ẓ** | LATIN SMALL LETTER Z WITH DOT BELOW | Kabyle (5) | [155] |

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)