**Considerations in the identification and management of**
**variant elements in Latin script tables for IDN registration**

**Cary Karp**
**Swedish Museum of Natural History**
**for .SE and .MUSEUM**

**This Draft Support Brief is presented to further ICANN's VIP initiative by**
**the host of its Latin script study, the Internet Infrastructure Foundation.**

The code points available for use in IDNs are all taken from the Unicode Character Code Charts. The Latin script is divided there into nine blocks. The one headed "Basic Latin" restates the ASCII repertoire and therefore includes the familiar letter-digit-hyphen (LDH) array to which TLD registries previously restricted all second-level domain names. The TLD labels, themselves, were further restricted to only the letters in that repertoire.

Latin letters other than the 'a–z' encoded in ASCII, as well as diacritically marked and otherwise decorated forms of them all are presented in the supplemental and extended Latin blocks, with further Latin letters in blocks under the heading "Phonetic Symbols".

Many of the marked letters can be represented with differing series of code points. Other letters that are intrinsically different and have different code points may share the same glyph. Protocol constraint renders the first of these situations tractable. Contextual restriction on the use of certain code points is necessary for the second.

Basic concepts and considerations associated with the protocol and contextual management of these conditions are discussed below, with specific regard to the local collation of permissible IDN character repertoires and the identification of variant relationships among the listed characters.

A specimen character table is presented that includes almost the entire available Latin repertoire and can be used to illustrate all of the issues with variant characters that may require consideration in the management of the root and TLD zones.

*       *       *

The IDNA2008 protocol disallows the Latin block headed 'Fullwidth Latin Letters' (which IDNA2003 remapped to ASCII) but permits a large number of code points from all the other Latin blocks, as well as Latin letters placed in the phonetic blocks. IDN registries typically assemble smaller repertoires of code points from the full array, appropriate to the communities they serve. These repertoires are documented in tabular form aggregated by script or language, again as appropriate to a specific target community.

There was no counterpart to this selectivity in prior practice. The basic Latin alphabet was the only available script and its twenty-six letters could be supplemented with nothing more than the ten digits, '0-9', and a hyphen. These letters could be used without distinction between upper- and lowercase. The DNS protocol treats both forms as fully equivalent and processes the labels 'oldnews' and 'OldNews' identically. The visually similarities between 'OldNews' and the separate '01dNews' were noted, as was the prevalent use of fonts that obscured their visual differentiation. However, no attempt was made at limiting either the confusion this might cause or the potential for its deceptive exploitation.

Similarly, differences such as those between American and British spelling were noted but not reflected in registration policy – 'organization' and 'organisation' were two separate labels, as were 'theater' and 'theatre'. The registration of one label in such a pair conferred no rights to the other. Any name holder wishing to obviate risk of user confusion was free to register both forms but had no recourse if the one was already taken. One of the reasons why such things were not perceived as a source of concern was that labels were regarded as mnemonic conveniences and any word-like properties they displayed were of secondary interest.

The vast extension of the available repertoire provided by IDNA2003 was the outcome of a protracted effort highlighting the restriction to the Latin alphabet as a barrier to the participation of language communities that do not use it. The linguistic and cultural relevance that attach to domain names rose during the ensuing dialog, from being effectively nonexistent, into a sensitive matter of language rights. The ability to use dictionary words and proper names in localized form became a bannerhead issue. When the scripts needed for this ultimately became available, concern with orthographic rigor suffused the discussion of localized implementation. Orthographic variation of the type that was recognized only in passing in LDH days became a serious issue, as did differentiating characters that appear similarly but are not equivalent, and the equation of abstract characters that appear differently but are the same.

*        *        *

Various terms have been used to designate this concept of equivalence but none has yet been provided with a definition that is adequate in all the contexts where it is needed. This terminological intractability is likely, at least in part, to be a manifestation of differing perceptions and needs from one language community to another, even among those that share a common script. This is also a concern with the Latin script, which is used for writing a larger number of languages than are all other scripts combined.

The most obvious variant relationship between two Latin letters is between their upper-case and lower-case forms. As noted, the DNS obviates any need for concern with it, as long as the repertoire is restricted to ASCII. The form of equality that it conveys to those characters – absolutely identical behavior in every regard – is taken to be somehow extensible into other forms of variant relationships, even though a correspondingly broader descriptive framework would be needed. Similarly greater complexity would attach to to the technical implementation of whatever that framework might require. In fact, however, the ASCII case insensitivity is supported by a trivial mathematical operation. Devising means for the algorithmic enforcement of any other form of variant relationship – from the straightforward upper- and lower-case convertibility of non-ASCII Latin letters, to the equivalence between characters that are consider similar in some other regard – is by no means a straightforward undertaking. If nothing else, characters that are used in the writing systems of multiple languages frequently have variant relationships that differ from one language to another, which would somehow also need invariable quantification.

As will be seen below, the IDNA protocols (using that term to designate both IDNA2003 and IDNA2008) address central concerns with variant relationships in the Latin script and deal, for example, directly with case sensitivity (albeit in what might be an unexpected manner). Other variant issues are so utterly dependent on detailed understanding of local considerations that the expectation of general protocol relief is likely to remain unmet. Since the Latin alphabet was the only one used for domain names prior to the advent of IDNA it provides a grateful point of departure in the consideration of the changes to earlier conditions that characterize IDNs. It must, however, be noted that other scripts have properties that differ from those of the Latin alphabet, which is the sole focus of the following remarks.

*        *        *

Three types of variant relationship are regularly noted in the discussion of IDNs: graphemic (same squiggles), phonemic (same sounds), and semantic (same meanings). The first of them is assessed character by character, and a character will invariably have graphemic properties. The other two are primarily properties of entire labels but are generally applicable only to those that are words, or short form constructs that are treated as words. An example of a label lacking any such attribute would be "x5z". The IDNA protocols operate on the (Unicode) properties of individual characters and the remaining discussion here will be focused primarily on the character level, as well.

The variant issues appearing on this level derive largely from the orthographic conventions that determine whether such things as ligatures, say, 'æ', decompose to nonligated digraphs, in this case, 'ae', or if and how letters bearing diacritical marks can be alternately represented. Typical approached involve the decomposition of a marked single letter into a digraph – 'ö' to 'oe' – or the optional removal of the mark – 'ö' to 'o' – or its equation with another composite form – 'ö' and 'ø'. Such rules tend to be language specific and can differ among countries that use the same language.

The other variant situations that need to be considered in the IDN context are the shared use of the same glyph for different abstract characters, and the representation of a given character by multiple sequences of code points.

* * *

The most immediate difference between the presentation of a string of Latin letters in the LDH and IDN environments can be illustrated by returning to mixed-case labels. The IDNABIS protocol does not operate on LDH labels and only accepts lower case letters in the labels that it does process. This means that the labels 'clubamerica' and 'ClubAmerica' remain equivalent but 'ClubAmérica' is invalid and can only be used as 'clubamérica'. In this case, the presence or absence of the accent is unlikely to change any meaning the label is intended to convey and the name holder might ascribe greater value to case insensitivity than to the diacritical mark.

There is less latitude in other cases involving no more than a single non-ASCII letter. Someone with the Swedish given name, Östen, wishing to use it as a label in a domain name would almost certainly accept the restriction to lowercase in preference to forgoing the umlaut and using 'osten', which is the Swedish word for cheese. In the German language expanding the 'ö' to 'oe' is an established alternative (albeit of limited applicability to proper names) but Swedish uses a 29-letter alphabet in which 'ö' is the final letter and does not decompose to a digraph. Nor is it regarded as a diacritically marked 'o' (even if sometimes treated so in analytic discourse) and the two characters are not variants of each other. In text entry contexts where an 'ö' is not available the fall-back alternative is an 'o', still affording Östen no relief. Providing a converse example using the same letters, the Swedish word "norden" means "The North" – the region in which the country is found (and for which a regional TLD is readily conceivable). Replacing the 'o' with an 'ö' gives "nörden" – "the nerd" (presumably also a candidate gTLD label).

In cases where diacritical marks clarify pronunciation but do not change the meaning of a word – 'naivete' and 'näiveté' – their use is "decorative". Where they also change or disambiguate meaning – 'resume' and 'résumé' – their use is "contrastive". All diacritically marked letters are, however, commonly referred to as decorated regardless of the function of the marks. The same term is also applied to letters that appear in modified shapes ( n ɳ ɲ ), whether or not they are marked.

* * *

An extensive illustration of the decorated Latin letters available for use in IDNs, together with base letters that are not used in English and therefore not included in ASCII, is given in the script table

appended to this text (referred to in continuation simply as "the table"). This includes almost all of the code points that are both labeled as letters in the Latin script blocks and valid under the IDNABIS protocol. The phonetic blocks also include Latin elements that are used in writing systems that appear in the IDN space or for which future need can reasonably be expected. The table currently includes the code points from the phonetic blocks specifically needed to accommodate the "African Reference Alphabet" ("ARA"). This is used repeatedly in the illustrations both because it is well suited for that purpose and to emphasize that concern with variant characters in the Latin script is not primarily anchored in European languages. Many of the code points included on its basis also appear in other alphabets.

The table does not provide an exhaustive listing and is subject to the addition of further code points from the blocks already mentioned. In particular, it excludes letters that cannot be visually distinguished from punctuation marks and may even appear with the same glyphs. They can readily be included in local repertoires by communities that use them in their writing systems, understand how to embed them in strings that disambiguate their function – a clear instance of label-based means for resolving character-level concerns – and are able to articulate the contextual constraints needed to prevent their inappropriate use elsewhere. Such characters do, however, appear in the examples below. Characters requiring multiple code points are similarly not included in the table but are used for exemplification.

The Latin blocks have been considered through the middle of Latin Extended-C (stopping at U+2C74). It is unclear if higher code points are needed for IDNs (although one is used in a following example) and many of them have yet to be included in production fonts. Regardless of any changes that may be made to it, the current version of the table exemplifies the issues relating to variant characters that attach to the Latin script.

*       *       *

The Unicode Latin blocks include a prodigious number of single code points that represent a base character combined with one or more diacritical marks or appear in other decorated forms. Further marks can, however, be applied to these precombined characters. No separate "combining marks" are included in the table but the marks that appear in precombined characters are also protocol valid in their combining forms. Anyone collating a local character table would therefore be free to include any of the following forms of the same character.

The precombined,

LATIN SMALL LETTER C WITH CEDILLA AND ACUTE                                    ; ǵ

can also be represented by combining characters at two separate code points (explicitly noted further on in this text),

LATIN SMALL LETTER C WITH CEDILLA + COMBINING ACUTE ACCENT          ; ḉ
LATIN SMALL LETTER C WITH ACUTE + COMBINING CEDILLA                       ; ḉ

or at three separate code points,

LATIN SMALL LETTER C + COMBINING ACUTE ACCENT + COMBINING CEDILLA    ; ḉ
LATIN SMALL LETTER C + COMBINING CEDILLA + COMBINING ACUTE ACCENT    ; ḉ

These one-, two-, and three code point representations are equivalent in every regard and thus an unambiguous instantiation of a variant character collection. The precombined characters in the second and third of these examples can similarly be represented in both pre- and post-combined forms,

providing two further collections. The number of variant character collections that can be culled recursively from the table in this manner is truly overwhelming.

*       *       *

As long as the separate combining marks are excluded from the table there is no need to indicate the plethora of variant forms they can generate. There are, however, writing systems that use characters which the Unicode chart does not include in precombined form. The ARA includes several underlined consonants of which some are available in precombined form, such as,

U+1E6F                 ; t       # LATIN SMALL LETTER T WITH LINE BELOW

but others require the separate combining mark, such as,

U+0063 U+0332          ; c       # LATIN SMALL LETTER C + COMBINING LOW LINE

The Unicode glossary refers to a composite character that does not exist in precombined form as a "grapheme cluster":

"A grapheme cluster represents a horizontally segmentable unit of text, consisting of some grapheme base (which may consist of a Korean syllable) together with any number of nonspacing marks applied to it."

There are also situations where similar marks, such as the COMBINING LINE BELOW and the COMBINING MACRON BELOW, can be used interchangeably, with graphemically more distinct marks such the COMBINING DOT BELOW also being elements in what might be treated as a variant collection. There is, however, rarely any intrinsic variant relationship among the marks, themselves. That property is modulated by the base character with which they are combined, and further modulated by the writing system in which the combined characters are used. The treatment of marks as variants of each other in one writing system in no way precludes their being used independently of each other in another.

Combined characters can easily be placed in a table without requiring the separate listing of the combining marks, indicating variant relationships where appropriate. There must, however, be some form of contextual constraint placed on the character(s) to which a given mark may be applied.

This is a good illustration of a situation where a massive problem in variant management can be averted on the protocol level. The initiation of such action is, however, not a present concern. The IDNA protocols recognize the need for reducing variant forms of this type to a single canonical representation and algorithmically normalize all multi-code point alternatives to the precombined form – if one exists – policy constraint is still necessary where they do not.

The five marked forms of the 'c' in the previous example, when included in a candidate U-label, appear indistinguishably as:

U+0061 U+1E09 U+0065                          ; aće
U+0061 U+00E7 U+0301 U+0065                   ; aće
U+0061 U+0107 U+0327 U+0065                   ; aće
U+0061 U+0063 U+0301 U+0327 U+0065           ; aće
U+0061 U+0063 U+0327 U+0301 U+0065           ; aće

Although only the first of these is valid, the protocol processes them all and generates the same Alabel: xn--ae-ess.

It is, however, important to note that this normalization is not performed by the encoding algorithm, itself. Each of the five candidate U-labels, valid and invalid, has its own Punycode form:

```
U+0061 U+1E09 U+0065                      ; aće   # ae-ess
U+0061 U+00E7 U+0301 U+0065               ; aće   # ae-4ia66s
U+0061 U+0107 U+0327 U+0065               ; aće   # ae-vla09s
U+0061 U+0063 U+0301 U+0327 U+0065        ; aće   # ace-ldc1qd
U+0061 U+0063 U+0327 U+0301 U+0065        ; aće   # ace-ldc3q
```

[Digression: There are online "IDN conversion" facilities that simply perform the Punycode encoding and prefix the result with 'xn--'. It is therefore to be expected that some prospective name holders will submit requests with the A-label presented in an invalid form. As a further caveat, precombined letters are often more stable typographically than are their multi-code point variants. Typical errors in the display of the latter are the misplacement of combining marks, for example in the aće case, with one of the combining marks shifting from the 'c' to the 'e', without changing the actual code points. It is therefore also to be expected that requests for U-labels may also be submitted in incorrect form. It is fully conceivable that the simultaneous appearance of both incorrect A- and U-labels will make it impossible to identify the actual label being requested.]

*       *       *

The single punctuation mark permitted in the LDH repertoire is the,

```
U+002D        ; -        # HYPHEN-MINUS
```

The IDNA2008 repertoire includes a number of characters that are not punctuation marks but use glyphs that cannot be readily distinguished from them (and have Unicode names that are letter/punctuation hybrids). The use of such characters obviously requires extraordinary care. One example is,

```
U+02BB        ; ʻ        # MODIFIER LETTER TURNED COMMA
```

This is used in numerous contexts of which one is as the " ʻokina ", that indicates the glottal stop in, for example, "Hawaiʻi".

Another example of a character used to indicate a glottal stop but which is a Latin letter and not a punctuation mark, is,

```
U+A78C        ; ꞌ        # LATIN SMALL LETTER SALTILLO
```

Its Unicode annotation states that, "it is widely used in many languages in Mexico and other regions, including Izere in Nigeria", and it therefore may appear sooner or later in localized Latin script tables (but lies beyond the cut-off point applied to the table here). Similar expectations apply to the ʻokina in a Polynesian language context.

The only basis for assembling these characters into a variant character collection is through their graphemic similarity. They are otherwise used in separate writing systems and are not mutually recognized across those boundaries. In any case, what would be the obvious preferred variant,

U+0027          ; '        # APOSTROPHE

is disallowed. The problem is that it is a shared fallback representation for the other forms, which are rarely indicated directly on stock keyboards and for which users commonly substitute an apostrophe. This all but guarantees a high failure rate when a label containing an apostrophe-like character is transcribed from ink on paper. The latitude for deceptive substitution in a born-digital situation is similarly high.

Using the ARA once again, it represents the glottal stop as:

U+0294          ; ʔ        # LATIN LETTER GLOTTAL STOP

This resembles the protocol invalid,

U+003F          ; ?        # QUESTION MARK

Taking one of the four characters used to represent the click consonants used in many languages spoken in Southern Africa (but not included in the ARA),

U+01C3          ; !        # LATIN LETTER RETROFLEX CLICK

This similarly resembles the protocol invalid,

U+0021          ; !        # EXCLAMATION MARK

For obvious reasons these characters, together with all other valid characters that resemble invalid punctuation marks, must not be permitted for use as surrogates for them.

*          *          *

Similar concern is due to letters that reside at different code points, are used in unrelated writing systems, and are commonly represented with similar – if not identical – glyphs. One illustration of this is,

U+01DD          ; ə        # LATIN SMALL LETTER TURNED E
U+0259          ; ə        # LATIN SMALL LETTER SCHWA

The use of the same glyph for both in many widely deployed fonts does not imply that the characters are variants forms of each other, nor is any writing system likely to include more than the one of them. Unforeseen problems might result, nonetheless, if a repertoire including one of these code points were to be compiled without awareness of the other. Therefore, in addition to any column in a script table that indicates variant relationships between code points, a separate column could indicate code points that should not appear in the same localized repertoire without careful consideration. If both are indeed required, contextual rules for their use should be clearly articulated and enforced.

The converse also applies to characters within a single writing system. The ARA includes both,

U+0061          ; a        # LATIN SMALL LETTER A

and

U+0251        ; ɑ      # LATIN SMALL LETTER ALPHA

which cannot be taken as variants of each other, but the second of them clearly needs to be restricted to use in labels written with an alphabet that includes both.

There is a daunting challenge in balancing need for minimizing the opportunity for abuse generated by the simultaneous availability of two characters such as these, against legitimate orthographic expectation – if not to say requirement – in linguistic contexts where the differences between the two letters are recognized and understood. This applies to many of the other illustrations given here.

The expertise needed to deal with such situations without any semblance of cultural discrimination extends beyond detailed first-hand knowledge of the respective language community's writing system. It requires in-depth familiarity with localized working environments, particularly in regard to the ability of commonly used rendering engines to display decorated characters correctly in widely deployed fonts, and awareness of any discrepancies between the code point for a given character as it is generated by a stock keyboard and its Unicode code point. There is further need for close enough proximity to the community to be able to track developments in the user environment and ensure the rapid adjustment of IDN systems to reflect changes as they are made.

If there is any variation within a language community in such things as the code point(s) produced for a given character on keyboard text entry, the preparation of a single IDN character repertoire for that community would also need to take those differences into account. (This is an important issue in variant management but not of major concern with the Latin script.) Similar understanding of local conditions is necessary when assessing potential need for special treatment of characters at different code points that, at least in some contexts, are regularly displayed with glyphs that are indistinguishable to even the most erudite members of the speech communities using them.

Beyond familiarity with local conditions, assessing and managing these risks requires understanding the effects they can have on users external to a community who, on the one hand are comfortably able to read the basic script it uses for IDN labels, but on the other, will not recognize crucial code point distinctions projected onto familiar graphemes.

20110706 ver. 1.0

APPENDIX: List of Latin Script Code Points

```
U+0061      ; a     # LATIN SMALL LETTER A
U+00E0      ; à     # LATIN SMALL LETTER A WITH GRAVE
U+00E1      ; á     # LATIN SMALL LETTER A WITH ACUTE
U+00E2      ; â     # LATIN SMALL LETTER A WITH CIRCUMFLEX
U+00E3      ; ã     # LATIN SMALL LETTER A WITH TILDE
U+00E4      ; ä     # LATIN SMALL LETTER A WITH DIAERESIS
U+00E5      ; å     # LATIN SMALL LETTER A WITH RING ABOVE
U+0101      ; ā     # LATIN SMALL LETTER A WITH MACRON
U+0103      ; ă     # LATIN SMALL LETTER A WITH BREVE
U+0105      ; ą     # LATIN SMALL LETTER A WITH OGONEK
U+01CE      ; ǎ     # LATIN SMALL LETTER A WITH CARON
U+01DF      ; ǟ     # LATIN SMALL LETTER A WITH DIAERESIS AND MACRON
U+01E1      ; ǡ     # LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON
U+01FB      ; ǻ     # LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE
U+0201      ; ȁ     # LATIN SMALL LETTER A WITH DOUBLE GRAVE
U+0203      ; ȃ     # LATIN SMALL LETTER A WITH INVERTED BREVE
U+1E01      ; ḁ     # LATIN SMALL LETTER A WITH RING BELOW
U+1EA1      ; ạ     # LATIN SMALL LETTER A WITH DOT BELOW
U+1EA3      ; ả     # LATIN SMALL LETTER A WITH HOOK ABOVE
U+1EA5      ; ấ     # LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE
U+1EA7      ; ầ     # LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE
U+1EA9      ; ẩ     # LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK
U+1EAB      ; ẫ     # LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE
U+1EAD      ; ậ     # LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW
U+1EAF      ; ắ     # LATIN SMALL LETTER A WITH BREVE AND ACUTE
U+1EB1      ; ằ     # LATIN SMALL LETTER A WITH BREVE AND GRAVE
U+1EB3      ; ẳ     # LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE
U+1EB5      ; ẵ     # LATIN SMALL LETTER A WITH BREVE AND TILDE
U+1EB7      ; ặ     # LATIN SMALL LETTER A WITH BREVE AND DOT BELOW
U+0227      ; ȧ     # LATIN SMALL LETTER A WITH DOT ABOVE
U+2C65      ; ⱥ     # LATIN SMALL LETTER A WITH STROKE
U+0251      ; ɑ     # LATIN SMALL LETTER ALPHA
U+00E6      ; æ     # LATIN SMALL LETTER AE
U+0062      ; b     # LATIN SMALL LETTER B
U+0180      ; ƀ     # LATIN SMALL LETTER B WITH STROKE
U+0183      ; ƃ     # LATIN SMALL LETTER B WITH TOPBAR
U+1E03      ; ḃ     # LATIN SMALL LETTER B WITH DOT ABOVE
U+1E05      ; ḅ     # LATIN SMALL LETTER B WITH DOT BELOW
U+1E07      ; ḇ     # LATIN SMALL LETTER B WITH LINE BELOW
U+0253      ; ɓ     # LATIN SMALL LETTER B WITH HOOK
U+0063      ; c     # LATIN SMALL LETTER C
U+00E7      ; ç     # LATIN SMALL LETTER C WITH CEDILLA
U+0107      ; ć     # LATIN SMALL LETTER C WITH ACUTE
U+0109      ; ĉ     # LATIN SMALL LETTER C WITH CIRCUMFLEX
```

```
U+010B     ; ċ     # LATIN SMALL LETTER C WITH DOT ABOVE
U+010D     ; č     # LATIN SMALL LETTER C WITH CARON
U+0188     ; ƈ     # LATIN SMALL LETTER C WITH HOOK
U+023C     ; ȼ     # LATIN SMALL LETTER C WITH STROKE
U+0255     ; ɕ     # LATIN SMALL LETTER C WITH CURL
U+1E09     ; ḉ     # LATIN SMALL LETTER C WITH CEDILLA AND ACUTE
U+0064     ; d     # LATIN SMALL LETTER D
U+010F     ; ď     # LATIN SMALL LETTER D WITH CARON
U+0111     ; đ     # LATIN SMALL LETTER D WITH STROKE
U+018C     ; ƌ     # LATIN SMALL LETTER D WITH TOPBAR
U+0221     ; ȡ     # LATIN SMALL LETTER D WITH CURL
U+1E0D     ; ḍ     # LATIN SMALL LETTER D WITH DOT BELOW
U+1E0F     ; ḏ     # LATIN SMALL LETTER D WITH LINE BELOW
U+1E11     ; ḑ     # LATIN SMALL LETTER D WITH CEDILLA
U+1E13     ; ḓ     # LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW
U+1E0B     ; ḋ     # LATIN SMALL LETTER D WITH DOT ABOVE
U+0256     ; ɖ     # LATIN SMALL LETTER D WITH TAIL
U+0257     ; ɗ     # LATIN SMALL LETTER D WITH HOOK
U+00F0     ; ð     # LATIN SMALL LETTER ETH
U+0065     ; e     # LATIN SMALL LETTER E
U+00E8     ; è     # LATIN SMALL LETTER E WITH GRAVE
U+00E9     ; é     # LATIN SMALL LETTER E WITH ACUTE
U+00EA     ; ê     # LATIN SMALL LETTER E WITH CIRCUMFLEX
U+00EB     ; ë     # LATIN SMALL LETTER E WITH DIAERESIS
U+0113     ; ē     # LATIN SMALL LETTER E WITH MACRON
U+0115     ; ĕ     # LATIN SMALL LETTER E WITH BREVE
U+0117     ; ė     # LATIN SMALL LETTER E WITH DOT ABOVE
U+0119     ; ę     # LATIN SMALL LETTER E WITH OGONEK
U+011B     ; ě     # LATIN SMALL LETTER E WITH CARON
U+0205     ; ȅ     # LATIN SMALL LETTER E WITH DOUBLE GRAVE
U+0207     ; ȇ     # LATIN SMALL LETTER E WITH INVERTED BREVE
U+1E15     ; ḕ     # LATIN SMALL LETTER E WITH MACRON AND GRAVE
U+1E17     ; ḗ     # LATIN SMALL LETTER E WITH MACRON AND ACUTE
U+1E19     ; ḙ     # LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW
U+1E1B     ; ḛ     # LATIN SMALL LETTER E WITH TILDE BELOW
U+1E1D     ; ḝ     # LATIN SMALL LETTER E WITH CEDILLA AND BREVE
U+1EB9     ; ẹ     # LATIN SMALL LETTER E WITH DOT BELOW
U+1EBB     ; ẻ     # LATIN SMALL LETTER E WITH HOOK ABOVE
U+1EBD     ; ẽ     # LATIN SMALL LETTER E WITH TILDE
U+1EBF     ; ế     # LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE
U+1EC1     ; ề     # LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE
U+1EC3     ; ể     # LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE
U+1EC5     ; ễ     # LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE
U+1EC7     ; ệ     # LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW
U+0229     ; ȩ     # LATIN SMALL LETTER E WITH CEDILLA
U+0247     ; ɇ     # LATIN SMALL LETTER E WITH STROKE
```

```
U+025B      ; ɛ       # LATIN SMALL LETTER OPEN E
U+01DD      ; ə       # LATIN SMALL LETTER TURNED E
U+0259      ; ə       # LATIN SMALL LETTER SCHWA
U+0066      ; f       # LATIN SMALL LETTER F
U+0192      ; ƒ       # LATIN SMALL LETTER F WITH HOOK
U+1E1F      ; ḟ       # LATIN SMALL LETTER F WITH DOT ABOVE
U+0067      ; g       # LATIN SMALL LETTER G
U+011D      ; ĝ       # LATIN SMALL LETTER G WITH CIRCUMFLEX
U+011F      ; ğ       # LATIN SMALL LETTER G WITH BREVE
U+0121      ; ġ       # LATIN SMALL LETTER G WITH DOT ABOVE
U+0123      ; ģ       # LATIN SMALL LETTER G WITH CEDILLA
U+01E5      ; ǥ       # LATIN SMALL LETTER G WITH STROKE
U+01E7      ; ǧ       # LATIN SMALL LETTER G WITH CARON
U+01F5      ; ǵ       # LATIN SMALL LETTER G WITH ACUTE
U+0260      ; ɠ       # LATIN SMALL LETTER G WITH HOOK
U+1E21      ; ḡ       # LATIN SMALL LETTER G WITH MACRON
U+1D77      ; ᵷ       # LATIN SMALL LETTER TURNED G
U+0263      ; ɣ       # LATIN SMALL LETTER GAMMA
U+0068      ; h       # LATIN SMALL LETTER H
U+0125      ; ĥ       # LATIN SMALL LETTER H WITH CIRCUMFLEX
U+0127      ; ħ       # LATIN SMALL LETTER H WITH STROKE
U+021F      ; ȟ       # LATIN SMALL LETTER H WITH CARON
U+1E23      ; ḣ       # LATIN SMALL LETTER H WITH DOT ABOVE
U+1E25      ; ḥ       # LATIN SMALL LETTER H WITH DOT BELOW
U+1E27      ; ḧ       # LATIN SMALL LETTER H WITH DIAERESIS
U+1E29      ; ḩ       # LATIN SMALL LETTER H WITH CEDILLA
U+1E2B      ; ḫ       # LATIN SMALL LETTER H WITH BREVE BELOW
U+2C68      ; ⱨ       # LATIN SMALL LETTER H WITH DESCENDER
U+1E96      ; ẖ       # LATIN SMALL LETTER H WITH LINE BELOW
U+0266      ; ɦ       # LATIN SMALL LETTER H WITH HOOK
U+02AE      ; ʮ       # LATIN SMALL LETTER TURNED H WITH FISHHOOK
U+02AF      ; ʯ       # LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
U+0069      ; i       # LATIN SMALL LETTER I
U+00EC      ; ì       # LATIN SMALL LETTER I WITH GRAVE
U+00ED      ; í       # LATIN SMALL LETTER I WITH ACUTE
U+00EE      ; î       # LATIN SMALL LETTER I WITH CIRCUMFLEX
U+00EF      ; ï       # LATIN SMALL LETTER I WITH DIAERESIS
U+0129      ; ĩ       # LATIN SMALL LETTER I WITH TILDE
U+012B      ; ī       # LATIN SMALL LETTER I WITH MACRON
U+012D      ; ĭ       # LATIN SMALL LETTER I WITH BREVE
U+012F      ; į       # LATIN SMALL LETTER I WITH OGONEK
U+0131      ; ı       # LATIN SMALL LETTER DOTLESS I
U+01D0      ; ǐ       # LATIN SMALL LETTER I WITH CARON
U+0209      ; ȉ       # LATIN SMALL LETTER I WITH DOUBLE GRAVE
U+020B      ; ȋ       # LATIN SMALL LETTER I WITH INVERTED BREVE
U+1E2D      ; ḭ       # LATIN SMALL LETTER I WITH TILDE BELOW
```

```
U+1E2F    ; ḯ    # LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE
U+1EC9    ; ỉ    # LATIN SMALL LETTER I WITH HOOK ABOVE
U+1ECB    ; ị    # LATIN SMALL LETTER I WITH DOT BELOW
U+0269    ; ɩ    # LATIN SMALL LETTER IOTA
U+006A    ; j    # LATIN SMALL LETTER J
U+0135    ; ĵ    # LATIN SMALL LETTER J WITH CIRCUMFLEX
U+01F0    ; ǰ    # LATIN SMALL LETTER J WITH CARON
U+0237    ; ȷ    # LATIN SMALL LETTER DOTLESS J
U+0249    ; ɉ    # LATIN SMALL LETTER J WITH STROKE
U+025F    ; ɟ    # LATIN SMALL LETTER DOTLESS J WITH STROKE
U+006B    ; k    # LATIN SMALL LETTER K
U+0137    ; ķ    # LATIN SMALL LETTER K WITH CEDILLA
U+0199    ; ƙ    # LATIN SMALL LETTER K WITH HOOK
U+01E9    ; ǩ    # LATIN SMALL LETTER K WITH CARON
U+1E31    ; ḱ    # LATIN SMALL LETTER K WITH ACUTE
U+1E33    ; ḳ    # LATIN SMALL LETTER K WITH DOT BELOW
U+1E35    ; ḵ    # LATIN SMALL LETTER K WITH LINE BELOW
U+2C6A    ; ⱪ    # LATIN SMALL LETTER K WITH DESCENDER
U+0138    ; ĸ    # LATIN SMALL LETTER KRA
U+006C    ; l    # LATIN SMALL LETTER L
U+013A    ; ĺ    # LATIN SMALL LETTER L WITH ACUTE
U+013C    ; ļ    # LATIN SMALL LETTER L WITH CEDILLA
U+013E    ; ľ    # LATIN SMALL LETTER L WITH CARON
U+0142    ; ł    # LATIN SMALL LETTER L WITH STROKE
U+019A    ; ƚ    # LATIN SMALL LETTER L WITH BAR
U+0234    ; ȴ    # LATIN SMALL LETTER L WITH CURL
U+1E37    ; ḷ    # LATIN SMALL LETTER L WITH DOT BELOW
U+1E39    ; ḹ    # LATIN SMALL LETTER L WITH DOT BELOW AND MACRON
U+1E3B    ; ḻ    # LATIN SMALL LETTER L WITH LINE BELOW
U+1E3D    ; ḽ    # LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW
U+2C61    ; ⱡ    # LATIN SMALL LETTER L WITH DOUBLE BAR
U+006D    ; m    # LATIN SMALL LETTER M
U+1E3F    ; ḿ    # LATIN SMALL LETTER M WITH ACUTE
U+1E41    ; ṁ    # LATIN SMALL LETTER M WITH DOT ABOVE
U+1E43    ; ṃ    # LATIN SMALL LETTER M WITH DOT BELOW
U+006E    ; n    # LATIN SMALL LETTER N
U+00F1    ; ñ    # LATIN SMALL LETTER N WITH TILDE
U+0144    ; ń    # LATIN SMALL LETTER N WITH ACUTE
U+0146    ; ņ    # LATIN SMALL LETTER N WITH CEDILLA
U+0148    ; ň    # LATIN SMALL LETTER N WITH CARON
U+019E    ; ŋ    # LATIN SMALL LETTER N WITH LONG RIGHT LEG
U+01F9    ; ǹ    # LATIN SMALL LETTER N WITH GRAVE
U+0235    ; ȵ    # LATIN SMALL LETTER N WITH CURL
U+1E45    ; ṅ    # LATIN SMALL LETTER N WITH DOT ABOVE
U+1E47    ; ṇ    # LATIN SMALL LETTER N WITH DOT BELOW
U+1E49    ; ṉ    # LATIN SMALL LETTER N WITH LINE BELOW
```

```
U+1E4B    ; ṋ     # LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW
U+0272    ; ɲ     # LATIN SMALL LETTER N WITH LEFT HOOK
U+014B    ; ŋ     # LATIN SMALL LETTER ENG
U+006F    ; o     # LATIN SMALL LETTER O
U+00F2    ; ò     # LATIN SMALL LETTER O WITH GRAVE
U+00F3    ; ó     # LATIN SMALL LETTER O WITH ACUTE
U+00F4    ; ô     # LATIN SMALL LETTER O WITH CIRCUMFLEX
U+00F5    ; õ     # LATIN SMALL LETTER O WITH TILDE
U+00F6    ; ö     # LATIN SMALL LETTER O WITH DIAERESIS
U+00F8    ; ø     # LATIN SMALL LETTER O WITH STROKE
U+014D    ; ō     # LATIN SMALL LETTER O WITH MACRON
U+014F    ; ŏ     # LATIN SMALL LETTER O WITH BREVE
U+0151    ; ő     # LATIN SMALL LETTER O WITH DOUBLE ACUTE
U+01A1    ; ơ     # LATIN SMALL LETTER O WITH HORN
U+01D2    ; ǒ     # LATIN SMALL LETTER O WITH CARON
U+01EB    ; ǫ     # LATIN SMALL LETTER O WITH OGONEK
U+01ED    ; ǭ     # LATIN SMALL LETTER O WITH OGONEK AND MACRON
U+01FF    ; ǿ     # LATIN SMALL LETTER O WITH STROKE AND ACUTE
U+020D    ; ȍ     # LATIN SMALL LETTER O WITH DOUBLE GRAVE
U+020F    ; ȏ     # LATIN SMALL LETTER O WITH INVERTED BREVE
U+022B    ; ȫ     # LATIN SMALL LETTER O WITH DIAERESIS AND MACRON
U+022D    ; ȭ     # LATIN SMALL LETTER O WITH TILDE AND MACRON
U+022F    ; ȯ     # LATIN SMALL LETTER O WITH DOT ABOVE
U+0231    ; ȱ     # LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON
U+1E4D    ; ṍ     # LATIN SMALL LETTER O WITH TILDE AND ACUTE
U+1E4F    ; ṏ     # LATIN SMALL LETTER O WITH TILDE AND DIAERESIS
U+1E51    ; ṑ     # LATIN SMALL LETTER O WITH MACRON AND GRAVE
U+1E53    ; ṓ     # LATIN SMALL LETTER O WITH MACRON AND ACUTE
U+1ECD    ; ọ     # LATIN SMALL LETTER O WITH DOT BELOW
U+1ECF    ; ỏ     # LATIN SMALL LETTER O WITH HOOK ABOVE
U+1ED1    ; ố     # LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE
U+1ED3    ; ồ     # LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE
U+1ED5    ; ổ     # LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE
U+1ED7    ; ỗ     # LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE
U+1ED9    ; ộ     # LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW
U+1EDB    ; ớ     # LATIN SMALL LETTER O WITH HORN AND ACUTE
U+1EDD    ; ờ     # LATIN SMALL LETTER O WITH HORN AND GRAVE
U+1EDF    ; ở     # LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE
U+1EE1    ; ỡ     # LATIN SMALL LETTER O WITH HORN AND TILDE
U+1EE3    ; ợ     # LATIN SMALL LETTER O WITH HORN AND DOT BELOW
U+0254    ; ɔ     # LATIN SMALL LETTER OPEN O
U+0275    ; ɵ     # LATIN SMALL LETTER BARRED O
U+0153    ; œ     # LATIN SMALL LIGATURE OE
U+0070    ; p     # LATIN SMALL LETTER P
U+01A5    ; ƥ     # LATIN SMALL LETTER P WITH HOOK
U+1E55    ; ṕ     # LATIN SMALL LETTER P WITH ACUTE
```

```
U+1E57    ; ṗ    # LATIN SMALL LETTER P WITH DOT ABOVE
U+0071    ; q    # LATIN SMALL LETTER Q
U+024B    ; ɋ    # LATIN SMALL LETTER Q WITH HOOK TAIL
U+0072    ; r    # LATIN SMALL LETTER R
U+0155    ; ŕ    # LATIN SMALL LETTER R WITH ACUTE
U+0157    ; ŗ    # LATIN SMALL LETTER R WITH CEDILLA
U+0159    ; ř    # LATIN SMALL LETTER R WITH CARON
U+0211    ; ȑ    # LATIN SMALL LETTER R WITH DOUBLE GRAVE
U+0213    ; ȓ    # LATIN SMALL LETTER R WITH INVERTED BREVE
U+1E59    ; ṙ    # LATIN SMALL LETTER R WITH DOT ABOVE
U+1E5B    ; ṛ    # LATIN SMALL LETTER R WITH DOT BELOW
U+1E5D    ; ṝ    # LATIN SMALL LETTER R WITH DOT BELOW AND MACRON
U+1E5F    ; ṟ    # LATIN SMALL LETTER R WITH LINE BELOW
U+024D    ; ɍ    # LATIN SMALL LETTER R WITH STROKE
U+027D    ; ɽ    # LATIN SMALL LETTER R WITH TAIL
U+027F    ; ɿ    # LATIN SMALL LETTER REVERSED R WITH FISHHOOK
U+0073    ; s    # LATIN SMALL LETTER S
U+015B    ; ś    # LATIN SMALL LETTER S WITH ACUTE
U+015D    ; ŝ    # LATIN SMALL LETTER S WITH CIRCUMFLEX
U+015F    ; ş    # LATIN SMALL LETTER S WITH CEDILLA
U+0161    ; š    # LATIN SMALL LETTER S WITH CARON
U+0219    ; □    # LATIN SMALL LETTER S WITH COMMA BELOW
U+1E61    ; ṡ    # LATIN SMALL LETTER S WITH DOT ABOVE
U+1E63    ; ṣ    # LATIN SMALL LETTER S WITH DOT BELOW
U+1E65    ; ṥ    # LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE
U+1E67    ; ṧ    # LATIN SMALL LETTER S WITH CARON AND DOT ABOVE
U+1E69    ; ṩ    # LATIN SMALL LETTER S WITH DOT BELOW AND DOT
U+023F    ; ȿ    # LATIN SMALL LETTER S WITH SWASH TAIL
U+00DF    ; ß    # LATIN SMALL LETTER SHARP S
U+1E9C    ; ẜ    # LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE
U+1E9D    ; ẝ    # LATIN SMALL LETTER LONG S WITH HIGH STROKE
U+0283    ; ʃ    # LATIN SMALL LETTER ESH
U+0074    ; t    # LATIN SMALL LETTER T
U+0163    ; ţ    # LATIN SMALL LETTER T WITH CEDILLA
U+0165    ; ť    # LATIN SMALL LETTER T WITH CARON
U+0167    ; ŧ    # LATIN SMALL LETTER T WITH STROKE
U+01AB    ; ƫ    # LATIN LETTER T WITH PALATAL HOOK
U+01AD    ; ƭ    # LATIN SMALL LETTER T WITH HOOK
U+1E6B    ; ṫ    # LATIN SMALL LETTER T WITH DOT ABOVE
U+1E6D    ; ṭ    # LATIN SMALL LETTER T WITH DOT BELOW
U+1E6F    ; ṯ    # LATIN SMALL LETTER T WITH LINE BELOW
U+1E71    ; ṱ    # LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW
U+0236    ; ȶ    # LATIN SMALL LETTER T WITH CURL
U+021B    ; □    # LATIN SMALL LETTER T WITH COMMA BELOW
U+1E97    ; ẗ    # LATIN SMALL LETTER T WITH DIAERESIS
U+2C66    ; ⱦ    # LATIN SMALL LETTER T WITH DIAGONAL STROKE
```

```
U+0288      ; ʈ      # LATIN SMALL LETTER T WITH RETROFLEX HOOK
U+00FE      ; þ      # LATIN SMALL LETTER THORN
U+0075      ; u      # LATIN SMALL LETTER U
U+00F9      ; ù      # LATIN SMALL LETTER U WITH GRAVE
U+00FA      ; ú      # LATIN SMALL LETTER U WITH ACUTE
U+00FB      ; û      # LATIN SMALL LETTER U WITH CIRCUMFLEX
U+00FC      ; ü      # LATIN SMALL LETTER U WITH DIAERESIS
U+0169      ; ũ      # LATIN SMALL LETTER U WITH TILDE
U+016B      ; ū      # LATIN SMALL LETTER U WITH MACRON
U+016D      ; ŭ      # LATIN SMALL LETTER U WITH BREVE
U+016F      ; ů      # LATIN SMALL LETTER U WITH RING ABOVE
U+0171      ; ű      # LATIN SMALL LETTER U WITH DOUBLE ACUTE
U+0173      ; ų      # LATIN SMALL LETTER U WITH OGONEK
U+01B0      ; ư      # LATIN SMALL LETTER U WITH HORN
U+01D4      ; ǔ      # LATIN SMALL LETTER U WITH CARON
U+01D6      ; ǖ      # LATIN SMALL LETTER U WITH DIAERESIS AND MACRON
U+01D8      ; ǘ      # LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE
U+01DA      ; ǚ      # LATIN SMALL LETTER U WITH DIAERESIS AND CARON
U+01DC      ; ǜ      # LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE
U+0215      ; ȕ      # LATIN SMALL LETTER U WITH DOUBLE GRAVE
U+0217      ; ȗ      # LATIN SMALL LETTER U WITH INVERTED BREVE
U+1E73      ; ṳ      # LATIN SMALL LETTER U WITH DIAERESIS BELOW
U+1E75      ; ṵ      # LATIN SMALL LETTER U WITH TILDE BELOW
U+1E77      ; ṷ      # LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW
U+1E79      ; ṹ      # LATIN SMALL LETTER U WITH TILDE AND ACUTE
U+1E7B      ; ṻ      # LATIN SMALL LETTER U WITH MACRON AND DIAERESIS
U+1EE5      ; ụ      # LATIN SMALL LETTER U WITH DOT BELOW
U+1EE7      ; ủ      # LATIN SMALL LETTER U WITH HOOK ABOVE
U+1EE9      ; ứ      # LATIN SMALL LETTER U WITH HORN AND ACUTE
U+1EEB      ; ừ      # LATIN SMALL LETTER U WITH HORN AND GRAVE
U+1EED      ; ử      # LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE
U+1EEF      ; ữ      # LATIN SMALL LETTER U WITH HORN AND TILDE
U+1EF1      ; ự      # LATIN SMALL LETTER U WITH HORN AND DOT BELOW
U+0076      ; v      # LATIN SMALL LETTER V
U+1E7D      ; ṽ      # LATIN SMALL LETTER V WITH TILDE
U+1E7F      ; ṿ      # LATIN SMALL LETTER V WITH DOT BELOW
U+1EFD      ; ỽ      # LATIN SMALL LETTER MIDDLE-WELSH V
U+2C71      ; ⱱ      # LATIN SMALL LETTER V WITH RIGHT HOOK
U+2C74      ; ⱴ      # LATIN SMALL LETTER V WITH CURL
U+028B      ; ʋ      # LATIN SMALL LETTER V WITH HOOK
U+028C      ; ʌ      # LATIN SMALL LETTER TURNED V
U+0077      ; w      # LATIN SMALL LETTER W
U+0175      ; ŵ      # LATIN SMALL LETTER W WITH CIRCUMFLEX
U+1E81      ; ẁ      # LATIN SMALL LETTER W WITH GRAVE
U+1E83      ; ẃ      # LATIN SMALL LETTER W WITH ACUTE
U+1E85      ; ẅ      # LATIN SMALL LETTER W WITH DIAERESIS
```

```
U+1E87     ; ẇ     # LATIN SMALL LETTER W WITH DOT ABOVE
U+1E89     ; ẉ     # LATIN SMALL LETTER W WITH DOT BELOW
U+1E98     ; ẘ     # LATIN SMALL LETTER W WITH RING ABOVE
U+2C73     ; ⱳ     # LATIN SMALL LETTER W WITH HOOK
U+01BF     ; ƿ     # LATIN LETTER WYNN
U+0078     ; x     # LATIN SMALL LETTER X
U+1E8B     ; ẋ     # LATIN SMALL LETTER X WITH DOT ABOVE
U+1E8D     ; ẍ     # LATIN SMALL LETTER X WITH DIAERESIS
U+0079     ; y     # LATIN SMALL LETTER Y
U+00FD     ; ý     # LATIN SMALL LETTER Y WITH ACUTE
U+00FF     ; ÿ     # LATIN SMALL LETTER Y WITH DIAERESIS
U+0177     ; ŷ     # LATIN SMALL LETTER Y WITH CIRCUMFLEX
U+01B4     ; ƴ     # LATIN SMALL LETTER Y WITH HOOK
U+0233     ; ȳ     # LATIN SMALL LETTER Y WITH MACRON
U+024F     ; ɏ     # LATIN SMALL LETTER Y WITH STROKE
U+1E99     ; ẙ     # LATIN SMALL LETTER Y WITH RING ABOVE
U+1EF3     ; ỳ     # LATIN SMALL LETTER Y WITH GRAVE
U+1EF5     ; ỵ     # LATIN SMALL LETTER Y WITH DOT BELOW
U+1EF7     ; ỷ     # LATIN SMALL LETTER Y WITH HOOK ABOVE
U+1EF9     ; ỹ     # LATIN SMALL LETTER Y WITH TILDE
U+1EFF     ; ỿ     # LATIN SMALL LETTER Y WITH LOOP
U+1E8F     ; ẏ     # LATIN SMALL LETTER Y WITH DOT ABOVE
U+021D     ; ȝ     # LATIN SMALL LETTER YOGH
U+007A     ; z     # LATIN SMALL LETTER Z
U+017A     ; ź     # LATIN SMALL LETTER Z WITH ACUTE
U+017C     ; ż     # LATIN SMALL LETTER Z WITH DOT ABOVE
U+017E     ; ž     # LATIN SMALL LETTER Z WITH CARON
U+01B6     ; ƶ     # LATIN SMALL LETTER Z WITH STROKE
U+0225     ; ȥ     # LATIN SMALL LETTER Z WITH HOOK
U+0240     ; ɀ     # LATIN SMALL LETTER Z WITH SWASH TAIL
U+1E91     ; ẑ     # LATIN SMALL LETTER Z WITH CIRCUMFLEX
U+1E93     ; ẓ     # LATIN SMALL LETTER Z WITH DOT BELOW
U+1E95     ; ẕ     # LATIN SMALL LETTER Z WITH LINE BELOW
U+2C6C     ; ⱬ     # LATIN SMALL LETTER Z WITH DESCENDER
U+0292     ; ʒ     # LATIN SMALL LETTER EZH
U+01B9     ; ƹ     # LATIN SMALL LETTER EZH REVERSED
U+01BA     ; ƺ     # LATIN SMALL LETTER EZH WITH TAIL
U+01EF     ; ǯ     # LATIN SMALL LETTER EZH WITH CARON
U+0294     ; ʔ     # LATIN LETTER GLOTTAL STOP
```