Proposal to Adjust Identifier Properties

Date: 2019-10-07
Authors: Asmus Freytag, Mark Davis and Michel Suignard

1   Status

This is an updated document. Appendices 3 and 4 are new, and for each section the proposed changes have been summarized.

2   Overview

UTS#39 "Unicode Security Measures" provides a breakdown of code points by several overlapping Identifier Types. For example, "Exclusion" is based on UAX#31 "Unicode Identifier and Pattern Syntax" Table 4 and defines code points that should be excluded from identifiers as belonging to scripts and blocks of special use, archaic, liturgical and other uncommon or problematic characters.

Since 2013, ICANN has been engaged in a process of defining the repertoire for IDN top level domain names resulting in a specification called the Root Zone Label Generation Rules (RZ-LGR). As first step in that process, a Maximal Starting Repertoire (MSR) has been created, that similar to the Identifier Types attempts to set an outer boundary of code points, within which the LGRs for specific scripts can be designed.

The scope of the RZ-LGR and therefore the MSR in principle encompasses all modern writing systems that are in general, everyday use. (See the [Procedure] document establishing the parameters and goals of the project). The development of the MSR took into account the information from UAX#31 on recommended scripts, but excluded Bopomofo, which was seen as not sufficiently general.

Where code points weren't associated with known orthographies, the developers consulted the original character proposals (where available) to understand the nature of the proposed character and whether they were for modern orthographies of more specialized purposes. The list of orthographies consulted was limited to languages classified as having some institutional support (e.g. use in education), for details see [MSR-4].

The list of recommended scripts in UAX#31 excludes some scripts, such as Cherokee, Canadian Syllabics, Ol Chiki and some African scripts that have user communities that are in principle not much different from those of some languages that can be written with extensions to some more widely used scripts. Developing a RZ-LGR for a script requires sustained commitment from the community involved; so far that has not materialized for any script not listed as Recommended in UAX#31 Table 5.

The latest published version of the Root Zone LGR [RZ-LGR-3] covers the majority of these scripts (with drafts for almost all the remaining ones in various stages of development). Most of the scripts do not fully exhaust their maximal repertoire, rejecting some code points as too uncommon or too specialized.

The effort has reached a stage where it is possible to compare some of it conclusions on the use of characters in identifiers with the recommendations made in UTS#39. This proposal document presents some of the issues and suggests some possible adjustments.

3 Unified Ideographs

In the following analysis, Unified ideographs are ignored. The MSR does not include the entire Han script, but creates a subset of 19,855 everyday common use ideographs that is informed by a number of widely supported standards as well as the IICORE subset. The subset includes the repertoires supported by .jp, .cn, .tw and .asia registries and only insignificantly exceeds it.

We believe that including ideographs outside this subset provides diminishing return from the perspective of identifiers and primarily opens up additional avenues for spoofing attacks.

Proposed: Our recommendation for UTS#39 would be to assign "Uncommon_Use" or a similar identifier type to all CJK Unified Ideographs outside the subset of CJK unified ideographs contained in the [MSR-4]. (We can provide soft-copy of the set in a suitable format).

4 21 characters not recommended in UTS#39 but part of the Root Zone

There are 21 characters that are included in the Root Zone (or pending drafts for which the repertoire development has been completed), but that are not listed as Recommended in UTS#39. For the most part, the detailed research by the script community panels for the RZ-LGR uncovered their use in modern orthographies that met the standards for widespread everyday use.

For domain names, RFC 6912 singles out the Root Zone as the most restrictive; consequently, we feel that there should be little reason for Unicode to recommend against these characters for general identifiers.

For languages that use these code points see Appendix 4.

Proposed: change the identifier type of these 21 character to "Recommended"

Code Point	Glyph	Script	Name	IdentifierType
U+0192	ƒ	Latin	LATIN SMALL LETTER F WITH HOOK	Uncommon_Use
U+0199	ƙ	Latin	LATIN SMALL LETTER K WITH HOOK	Uncommon_Use
U+01B4	ƴ	Latin	LATIN SMALL LETTER Y WITH HOOK	Uncommon_Use
U+01DD	ə	Latin	LATIN SMALL LETTER TURNED E	Uncommon_Use
U+024D	ɍ	Latin	LATIN SMALL LETTER R WITH STROKE	Uncommon_Use
U+0253..U+0254	ɓ..ɔ	Latin	LATIN SMALL LETTER B WITH HOOK..LATIN SMALL LETTER OPEN O	Uncommon_Use
U+0256..U+0257	ɖ..ɗ	Latin	LATIN SMALL LETTER D WITH TAIL..LATIN	Uncommon_Use

Code Point	Glyph	Script	Name	Tags
			SMALL LETTER D WITH HOOK	
U+025B	ɛ	Latin	LATIN SMALL LETTER OPEN E	Uncommon_Use
U+0263	ɣ	Latin	LATIN SMALL LETTER GAMMA	Uncommon_Use
U+0268..U+0269	ɨ..ɩ	Latin	LATIN SMALL LETTER I WITH STROKE..LATIN SMALL LETTER IOTA	Uncommon_Use
U+0272	ɲ	Latin	LATIN SMALL LETTER N WITH LEFT HOOK	Uncommon_Use
U+0289	ʉ	Latin	LATIN SMALL LETTER U BAR	Technical, Uncommon_Use
U+0292	ʒ	Latin	LATIN SMALL LETTER EZH	Uncommon_Use
U+068E	ڎ	Arabic	ARABIC LETTER DUL	Obsolete
U+17CB..U+17CD	់..៍	Khmer	KHMER SIGN BANTOC..KHMER SIGN TOANDAKHIAT	Technical
U+17D0	័	Khmer	KHMER SIGN SAMYOK SANNYA	Technical

5   90 Characters from the MSR not picked up by RZ-LGR scripts

The following 90 characters were included in [MSR-4] but have not been picked up by their respective RZ-LGR scripts. A comparison shows that they all fall outside the Recommended range in UTS#39. In essence this confirms the IdentifierType assignments in UTS#39 (if perhaps not the particular breakdown between Uncommon_Use, Obsolete and Technical).

Proposed: No change. The detailed breakdown between Uncommon_Use, Obsolete and Technical is informative and ICANN analysis does not lay claim to being more authoritative.

Code Point	Glyph	Script	Name	Tags
U+0180	ƀ	Latin	LATIN SMALL LETTER B WITH STROKE	Technical
U+0188	ƈ	Latin	LATIN SMALL LETTER C WITH HOOK	Uncommon_Use
U+01A3	ƣ	Latin	LATIN SMALL LETTER GHA	Uncommon_Use
U+01A5	ƥ	Latin	LATIN SMALL LETTER P WITH HOOK	Uncommon_Use
U+01AD	ƭ	Latin	LATIN SMALL LETTER T WITH HOOK	Uncommon_Use
U+01B6	ƶ	Latin	LATIN SMALL LETTER Z WITH STROKE	Uncommon_Use
U+01E5	ǥ	Latin	LATIN SMALL LETTER G WITH STROKE	Uncommon_Use
U+0242	ɂ	Latin	LATIN SMALL LETTER GLOTTAL STOP	Uncommon_Use
U+0247	ɇ	Latin	LATIN SMALL LETTER E WITH STROKE	Uncommon_Use
U+0249	ɉ	Latin	LATIN SMALL LETTER J WITH STROKE	Uncommon_Use
U+024F	ɏ	Latin	LATIN SMALL LETTER Y WITH STROKE	Uncommon_Use
U+0251	ɑ	Latin	LATIN SMALL LETTER ALPHA	Technical
U+0260	ɠ	Latin	LATIN SMALL LETTER G WITH HOOK	Uncommon_Use
U+0265..U+0266	ɥ..ɦ	Latin	LATIN SMALL LETTER TURNED H..LATIN	Technical

			SMALL LETTER H WITH HOOK	
U+026A..U+026B	ɪ..ɫ	Latin	LATIN LETTER SMALL CAPITAL I..LATIN SMALL LETTER L WITH MIDDLE TILDE	Technical
U+0275	ɵ	Latin	LATIN SMALL LETTER BARRED O	Uncommon_Use
U+027D	ɽ	Latin	LATIN SMALL LETTER R WITH TAIL	Technical
U+0283	ʃ	Latin	LATIN SMALL LETTER ESH	Uncommon_Use
U+028A..U+028B	ʊ..ʋ	Latin	LATIN SMALL LETTER UPSILON..LATIN SMALL LETTER V WITH HOOK	Uncommon_Use
U+028C	ʌ	Latin	LATIN SMALL LETTER TURNED V	Technical
U+0294	ʔ	Latin	LATIN LETTER GLOTTAL STOP	Uncommon_Use
U+0329	̩	Inherited	COMBINING VERTICAL LINE BELOW	Technical
U+0358	̘	Inherited	COMBINING DOT ABOVE RIGHT	Uncommon_Use
U+05B0..U+05B3	ֹ..ֳ	Hebrew	HEBREW POINT SHEVA..HEBREW POINT HATAF QAMATS	Uncommon_Use
U+05B5..U+05B9	ֹ..ֹ	Hebrew	HEBREW POINT TSERE..HEBREW POINT HOLAM	Uncommon_Use
U+05BB..U+05BC	ֹ..ּ	Hebrew	HEBREW POINT QUBUTS..HEBREW POINT DAGESH OR MAPIQ	Uncommon_Use
U+05BF	ֿ	Hebrew	HEBREW POINT RAFE	Uncommon_Use
U+05C1..U+05C2	ׁ..ׂ	Hebrew	HEBREW POINT SHIN DOT..HEBREW POINT SIN DOT	Uncommon_Use
U+0656..U+0658	ٖ..٘	Arabic	ARABIC SUBSCRIPT ALEF..ARABIC MARK NOON GHUNNA	Uncommon_Use
U+0659..U+065E	ٙ..ٞ	Arabic	ARABIC ZWARAKAY..ARABIC FATHA WITH TWO DOTS	Uncommon_Use
U+065F		Arabic	ARABIC WAVY HAMZA BELOW	Uncommon_Use
U+08E4..U+08EF	٠..	Arabic	ARABIC CURLY FATHA..ARABIC TONE LOOP BELOW	Uncommon_Use
U+08F4..U+08FE	٠..	Arabic	ARABIC FATHA WITH RING..ARABIC DAMMA WITH DOT	Uncommon_Use
U+0A51	ੑ	Gurmukhi	GURMUKHI SIGN UDAAT	Uncommon_Use
U+0A75	ੵ	Gurmukhi	GURMUKHI SIGN YAKASH	Uncommon_Use
U+0B44	ୄ	Oriya	ORIYA VOWEL SIGN VOCALIC RR	Uncommon_Use
U+0D44	ൄ	Malayalam	MALAYALAM VOWEL SIGN VOCALIC RR	Uncommon_Use
U+0D8F..U+0D90	ඏ..ඐ	Sinhala	SINHALA LETTER ILUYANNA..SINHALA LETTER ILUUYANNA	Technical, Uncommon_Use
U+0DA6	ඦ	Sinhala	SINHALA LETTER SANYAKA JAYANNA	Technical,

				Uncommon_Use
U+0DDF	◌ඟ	Sinhala	SINHALA VOWEL SIGN GAYANUKITTA	Technical, Uncommon_Use
U+0DF3	◌ඳ	Sinhala	SINHALA VOWEL SIGN DIGA GAYANUKITTA	Technical, Uncommon_Use
U+10F4	ჴ	Georgian	GEORGIAN LETTER HAR	Obsolete
U+10F6	ჶ	Georgian	GEORGIAN LETTER FI	Obsolete
U+17CE..U+17CF	◌̇..◌̇	Khmer	KHMER SIGN KAKABAT..KHMER SIGN AHSDA	Technical
U+1DCA	◌̗	Inherited	COMBINING LATIN SMALL LETTER R BELOW	Technical
U+2C61	ɫ	Latin	LATIN SMALL LETTER L WITH DOUBLE BAR	Technical
U+2C73	ⱳ	Latin	LATIN SMALL LETTER W WITH HOOK	Obsolete
U+FB1E	◌ֆ	Hebrew	HEBREW POINT JUDEO-SPANISH VARIKA	Technical, Uncommon_Use

6   658 allowed characters not included in MSR

This comparison set covers 658 characters that UTS#39 considers "Allowed" but that are **excluded** from the MSR for a variety of reasons. Code points not in IDNA2008 as well as the CONTEXTJ and CONTEXTO classifications are categorically excluded for use in the Root Zone. The first two are considered "Inclusion" in UTS#39 marking them as available for exceptional inclusion, as are the two middle dots and Hebrew punctuation. The Greek character should be considered Obsolete to match the designation in MSR.

Note that the Root Zone may not contain HYPHEN-MINUS or digits; these code points would be acceptable for domain names in other zones (such as the second level). Therefore, we have excluded them from the analysis here to make it more relevant in other contexts. (In other words, even though [MSR-4] does not contain them, we believe they are appropriate for a generic profile for other public zones).

A special note on the Greek script. For domain names, modern practice is limited to the monotonic orthography; registries such as ".gr" have taken an explicit stand on this issue. We therefore feel justified in recommending that classical Greek (polytonic) should be considered "obsolete" from the point of view of making recommendations for identifiers. There are a number of Greek code points in the Extended Greek block that are DISALLOWED in IDNA 2008 because of the way they casefold. Those also have no place among the "Recommended" set. (Perhaps they could be given some other IdentifierType such as "Not_NFKC_CF").

A special note on the Bopomofo script: it is set here to "educational" based on the note in UAX#31 and the fact that the MSR "identifie[s] Bopomofo as a modern use script that is limited to educational use and therefore not eligible for the root zone".

A special note on UPPERCASE: IDNA2008 identifiers are all lowercase. It would be useful to mark any "Allowed" characters with their own "Uppercase" IdentifierType, so that profiles can more easily be created that are in IDNA 2008 (and without having to rely on multiple UTSs) – the affected code points are not listed here as they are readily identified by GC.

Proposed: Except for the code points covered by "Inclusion" in UTS#39 we recommend changing the IdentifierType to the nearest equivalent of the MSR Tags listed here. (With special attention to the Greek, Bopomofo and Uppercase subsets as noted above).

Code Point	Glyph	Script	Name	MSR Tags	Comment
U+0027	'	Common	APOSTROPHE		Not IDN2008
U+002E	.	Common	FULL STOP		Not IDN2008
U+003A	:	Common	COLON		Not IDN2008
U+005F	_	Common	LOW LINE		Not IDN2008
U+00B7	·	Common	MIDDLE DOT	context-other	IDNA2008 CONTEXTO
U+0138	ĸ	Latin	LATIN SMALL LETTER KRA	obsolete	(Greenlandic)
U+0201	ȁ	Latin	LATIN SMALL LETTER A WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+0203	ȃ	Latin	LATIN SMALL LETTER A WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+0205	ȅ	Latin	LATIN SMALL LETTER E WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+0207	ȇ	Latin	LATIN SMALL LETTER E WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+0209	ȉ	Latin	LATIN SMALL LETTER I WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+020B	ȋ	Latin	LATIN SMALL LETTER I WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+020D	ȍ	Latin	LATIN SMALL LETTER O WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+020F	ȏ	Latin	LATIN SMALL LETTER O WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+0211	ȑ	Latin	LATIN SMALL LETTER R WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+0213	ȓ	Latin	LATIN SMALL LETTER R WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+0215	ȕ	Latin	LATIN SMALL LETTER U WITH DOUBLE GRAVE	poetic, technical	(tone in poetry)
U+0217	ȗ	Latin	LATIN SMALL LETTER U WITH INVERTED BREVE	poetic, technical	(tone in poetry)
U+02BB	ʻ	Common	MODIFIER LETTER TURNED COMMA	punctuation	
U+02BC	ʼ	Common	MODIFIER LETTER APOSTROPHE	technical	
U+02EC	ˬ	Common	MODIFIER LETTER VOICING	technical	

U+030F		Inherited	COMBINING DOUBLE GRAVE ACCENT	technical	(Serbian and Croatian poetics)
U+0310		Inherited	COMBINING CANDRABINDU	technical	
U+0311		Inherited	COMBINING INVERTED BREVE	technical	(Serbian and Croatian poetics)
U+0313		Inherited	COMBINING COMMA ABOVE	technical	
U+0314		Inherited	COMBINING REVERSED COMMA ABOVE	technical	
U+0324		Inherited	COMBINING DIAERESIS BELOW	technical	
U+0325		Inherited	COMBINING RING BELOW	technical	
U+032D		Inherited	COMBINING CIRCUMFLEX ACCENT BELOW	technical	
U+032E		Inherited	COMBINING BREVE BELOW	technical	
U+0330		Inherited	COMBINING TILDE BELOW	technical	
U+0335	-	Inherited	COMBINING SHORT STROKE OVERLAY	technical	
U+0338	⁄	Inherited	COMBINING LONG SOLIDUS OVERLAY	technical	
U+0339		Inherited	COMBINING RIGHT HALF RING BELOW	technical	
U+0342		Inherited	COMBINING GREEK PERISPOMENI	technical	
U+0345		Inherited	COMBINING GREEK YPOGEGRAMMENI		NOT IDNA2008
U+0375	͵	Greek	GREEK LOWER NUMERAL SIGN	obsolete, context-other	IDNA2008 CONTEXTO
U+037B..U+037D	ͻ..ͽ	Greek	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL	obsolete	
U+03FC	ϼ	Greek	GREEK RHO WITH STROKE SYMBOL	obsolete, symbol	
U+048B	й	Cyrillic	CYRILLIC SMALL LETTER SHORT I WITH TAIL	uncommon_use	(nearly extinct)
U+048D	ҍ	Cyrillic	CYRILLIC SMALL LETTER SEMISOFT SIGN	uncommon_use	(nearly extinct)
U+048F	ҏ	Cyrillic	CYRILLIC SMALL LETTER	uncommon_use	(nearly extinct)

			ER WITH TICK		
U+049D	қ	Cyrillic	CYRILLIC SMALL LETTER KA WITH VERTICAL STROKE	obsolete	(Azerbaijani)
U+04A7	ҧ	Cyrillic	CYRILLIC SMALL LETTER PE WITH MIDDLE HOOK	obsolete	(Abkhazian)
U+04B9	ҹ	Cyrillic	CYRILLIC SMALL LETTER CHE WITH VERTICAL STROKE	obsolete	(Azerbaijani)
U+04C4	ӄ	Cyrillic	CYRILLIC SMALL LETTER KA WITH HOOK	uncommon_use	(threatened)
U+04C6	ӆ	Cyrillic	CYRILLIC SMALL LETTER EL WITH TAIL	uncommon_use	(nearly extinct)
U+04C8	ӈ	Cyrillic	CYRILLIC SMALL LETTER EN WITH HOOK	uncommon_use	(threatened)
U+04CA	ӊ	Cyrillic	CYRILLIC SMALL LETTER EN WITH TAIL	uncommon_use	(nearly extinct)
U+04CE	ӎ	Cyrillic	CYRILLIC SMALL LETTER EM WITH TAIL	uncommon_use	(nearly extinct)
U+04F7	ӷ	Cyrillic	CYRILLIC SMALL LETTER GHE WITH DESCENDER	uncommon_use	(educational)
U+04FB	ӻ	Cyrillic	CYRILLIC SMALL LETTER GHE WITH STROKE AND HOOK	uncommon_use	(nearly extinct)
U+04FD	ӽ	Cyrillic	CYRILLIC SMALL LETTER HA WITH HOOK	uncommon_use	(nearly extinct)
U+04FF	ӿ	Cyrillic	CYRILLIC SMALL LETTER HA WITH STROKE	uncommon_use	(nearly extinct)
U+0501	ԁ	Cyrillic	CYRILLIC SMALL LETTER KOMI DE	obsolete	(Komi, Mordvin)
U+0503	ԃ	Cyrillic	CYRILLIC SMALL LETTER KOMI DJE	obsolete	(Komi, Mordvin)
U+0505	ԅ	Cyrillic	CYRILLIC SMALL LETTER KOMI ZJE	obsolete	(Komi, Mordvin)
U+0507	ԇ	Cyrillic	CYRILLIC SMALL LETTER KOMI DZJE	obsolete	(Komi, Mordvin)
U+0509	ԉ	Cyrillic	CYRILLIC SMALL LETTER KOMI LJE	obsolete	(Komi, Mordvin)
U+050B	ԋ	Cyrillic	CYRILLIC SMALL LETTER KOMI NJE	obsolete	(Komi, Mordvin)
U+050D	ԍ	Cyrillic	CYRILLIC SMALL LETTER KOMI SJE	obsolete	(Komi, Mordvin)
U+050F	ԏ	Cyrillic	CYRILLIC SMALL LETTER KOMI TJE	obsolete	(Komi, Mordvin)

U+0511	ε	Cyrillic	CYRILLIC SMALL LETTER REVERSED ZE	uncommon_use	(threatened)
U+0513	ӆ	Cyrillic	CYRILLIC SMALL LETTER EL WITH HOOK	uncommon_use	(threatened)
U+0515		Cyrillic	CYRILLIC SMALL LETTER LHA	obsolete	(Mordvin)
U+0517		Cyrillic	CYRILLIC SMALL LETTER RHA	obsolete	(Mordvin)
U+0519		Cyrillic	CYRILLIC SMALL LETTER YAE	obsolete	(Mordvin)
U+051B	q	Cyrillic	CYRILLIC SMALL LETTER QA	obsolete	(Abkhaz, Kurdish)
U+051D	w	Cyrillic	CYRILLIC SMALL LETTER WE	obsolete	(Abkhaz, Kurdish)
U+051F		Cyrillic	CYRILLIC SMALL LETTER ALEUT KA	obsolete	(Aleut)
U+0521		Cyrillic	CYRILLIC SMALL LETTER EL WITH MIDDLE HOOK	obsolete	(Abkhaz, Chuvash)
U+0523		Cyrillic	CYRILLIC SMALL LETTER EN WITH MIDDLE HOOK	obsolete	(Chuvash)
U+0527		Cyrillic	CYRILLIC SMALL LETTER SHHA WITH DESCENDER	obsolete	(Azerbaijani)
U+0529		Cyrillic	CYRILLIC SMALL LETTER EN WITH LEFT HOOK	obsolete	(Orok) (7.0)
U+052F		Cyrillic	CYRILLIC SMALL LETTER EL WITH DESCENDER	obsolete	(Khanty, Nenets) (7.0)
U+0559		Armenian	ARMENIAN MODIFIER LETTER LEFT HALF RING	technical	
U+0560		Armenian	ARMENIAN SMALL LETTER TURNED AYB	uncommon_use	(11.0 17/032)
U+0588		Armenian	ARMENIAN SMALL LETTER YI WITH STROKE	uncommon_use	(11.0 17/032)
U+058A	֊	Armenian	ARMENIAN HYPHEN		NOT IDNA2008
U+05EF		Hebrew	HEBREW YOD TRIANGLE	uncommon_use	(11.0 16/305)
U+05F3	׳	Hebrew	HEBREW PUNCTUATION GERESH	context-other	IDNA2008 CONTEXTO
U+05F4	״	Hebrew	HEBREW PUNCTUATION GERSHAYIM	context-other	IDNA2008 CONTEXTO
U+063B..U+063F	..	Arabic	ARABIC LETTER KEHEH WITH TWO DOTS ABOVE..ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE	obsolete	(historic)

U+0653	◌ٓ	Inherited	ARABIC MADDAH ABOVE	religious_use	
U+06AC	ڬ	Arabic	ARABIC LETTER KAF WITH DOT ABOVE	obsolete	(old Malay-Jawi)*use 0762 instead
U+06E5..U+06E6	ۥ ۦ	Arabic	ARABIC SMALL WAW..ARABIC SMALL YEH	technical	(religious annotation)
U+06FD	۽	Arabic	ARABIC SIGN SINDHI AMPERSAND	punctuation	
U+06FE	۾	Arabic	ARABIC SIGN SINDHI POSTPOSITION MEN	punctuation	
U+077E	ݾ	Arabic	ARABIC LETTER SEEN WITH INVERTED V	obsolete	(early Persian)
U+077F	ݿ	Arabic	ARABIC LETTER KAF WITH TWO DOTS ABOVE	obsolete	(early Persian)
U+08B2		Arabic	ARABIC LETTER ZAIN WITH INVERTED V ABOVE	uncommon_use	(Berber) (7.0)
U+08B6..U+08BA	..	Arabic	ARABIC LETTER BEH WITH SMALL MEEM ABOVE..ARABIC LETTER YEH WITH TWO DOTS BELOW AND SMALL NOON ABOVE	uncommon_use	(Bravanese) (9.0 13/178)
U+093D	ऽ	Devanagari	DEVANAGARI SIGN AVAGRAHA	obsolete	(Sanskrit)
U+0950	ॐ	Devanagari	DEVANAGARI OM	religious_use, symbol	
U+0960..U+0963	ॠ ◌ॣ	Devanagari	DEVANAGARI LETTER VOCALIC RR..DEVANAGARI VOWEL SIGN VOCALIC LL	obsolete	(Sanskrit)
U+0971	ॱ	Devanagari	DEVANAGARI SIGN HIGH SPACING DOT	obsolete	(Sanskrit)
U+097D	ॽ	Devanagari	DEVANAGARI LETTER GLOTTAL STOP	technical	(Limbu)
U+09BD	ঽ	Bengali	BENGALI SIGN AVAGRAHA	obsolete	(Sanskrit)
U+09E0	ৠ	Bengali	BENGALI LETTER VOCALIC RR	obsolete	(Sanskrit)
U+09E1	ৡ	Bengali	BENGALI LETTER VOCALIC LL	obsolete	(Sanskrit)
U+09E2	◌ৢ	Bengali	BENGALI VOWEL SIGN	obsolete	(Sanskrit)

			VOCALIC L		
U+09E3	◌ৣ	Bengali	BENGALI VOWEL SIGN VOCALIC LL	obsolete	(Sanskrit)
U+09FC		Bengali	BENGALI LETTER VEDIC ANUSVARA	uncommon_use	(Vedic) (10.0 15/161R)
U+09FE		Bengali	BENGALI SANDHI MARK	uncommon_use	(Sanskrit) (11.0 16/322)
U+0A01	◌ਂ	Gurmukhi	GURMUKHI SIGN ADAK BINDI	uncommon_use	
U+0A74	ੴ	Gurmukhi	GURMUKHI EK ONKAR	religious_use, symbol	
U+0ABD	ऽ	Gujarati	GUJARATI SIGN AVAGRAHA	obsolete	(Sanskrit)
U+0AD0	ॐ	Gujarati	GUJARATI OM	religious_use, symbol	
U+0AE0	ૠ	Gujarati	GUJARATI LETTER VOCALIC RR	obsolete	(Sanskrit)
U+0AE1..U+0AE3	ૡ..◌ૣ	Gujarati	GUJARATI LETTER VOCALIC LL..GUJARATI VOWEL SIGN VOCALIC LL	obsolete	(Sanskrit)
U+0AFA..U+0AFF	..	Gujarati	GUJARATI SIGN SUKUN..GUJARATI SIGN TWO-CIRCLE NUKTA ABOVE	uncommon_use	(6) (Arabic transliteration) (10.0 13/143)
U+0B3D	ଽ	Oriya	ORIYA SIGN AVAGRAHA	obsolete	(Sanskrit)
U+0B60..U+0B61	ୠ..ୡ	Oriya	ORIYA LETTER VOCALIC RR..ORIYA LETTER VOCALIC LL	obsolete	(Sanskrit)
U+0B82	◌ஂ	Tamil	TAMIL SIGN ANUSVARA	technical	(not used in Tamil
U+0BD0	ௐ	Tamil	TAMIL OM	religious_use, symbol	
U+0C01	◌ఁ	Telugu	TELUGU SIGN CANDRABINDU	uncommon_use	
U+0C04		Telugu	TELUGU SIGN COMBINING ANUSVARA ABOVE	uncommon_use	(Prakrit) (11.0 16/285)
U+0C3D	ఽ	Telugu	TELUGU SIGN AVAGRAHA	obsolete	(Sanskrit)

U+0C60..U+0C61	ఋు.. ౡ	Telugu	TELUGU LETTER VOCALIC RR..TELUGU LETTER VOCALIC LL	obsolete	(Sanskrit)
U+0C80		Kannada	KANNADA SIGN SPACING CANDRABINDU	uncommon_use	(Badaga) (9.0 14/153)
U+0CBD	ಽ	Kannada	KANNADA SIGN AVAGRAHA	obsolete	(Sanskrit)
U+0CE0..U+0CE1	ಋ.. ೡ	Kannada	KANNADA LETTER VOCALIC RR..KANNADA LETTER VOCALIC LL	obsolete	(Sanskrit)
U+0CE2..U+0CE3	ಿ	Kannada	KANNADA VOWEL SIGN VOCALIC L..KANNADA VOWEL SIGN VOCALIC LL	obsolete	(Sanskrit)
U+0CF1..U+0CF2	Ⴝ..೦೦	Kannada	KANNADA SIGN JIHVAMULIYA..KANNADA SIGN UPADHMANIYA	obsolete	(Sankrit)*
U+0D00		Malayalam	MALAYALAM SIGN COMBINING ANUSVARA ABOVE	uncommon_use	(Prakrit) (10.0 14/003)
U+0D3A		Malayalam	MALAYALAM LETTER TTTA	obsolete	(Historic)
U+0D3B		Malayalam	MALAYALAM SIGN VERTICAL BAR VIRAMA	uncommon_use	(10.0 14/015R)
U+0D3C		Malayalam	MALAYALAM SIGN CIRCULAR VIRAMA	uncommon_use	(10.0 14/014R)
U+0D3D	∫	Malayalam	MALAYALAM SIGN AVAGRAHA	obsolete	(Sanskrit)
U+0D4C	�⊙ൗ	Malayalam	MALAYALAM VOWEL SIGN AU	obsolete	(Archaic)
U+0D4E		Malayalam	MALAYALAM LETTER DOT REPH	obsolete	(Historic)
U+0D54..U+0D56	..	Malayalam	MALAYALAM LETTER CHILLU M..MALAYALAM LETTER CHILLU LLL	uncommon_use	(Chillu) (9.0 14/013)
U+0D60..U+0D61	ൠ.. ൡ	Malayalam	MALAYALAM LETTER VOCALIC RR..MALAYALAM LETTER VOCALIC LL	obsolete	(Sanskrit)
U+0E2F	ฯ	Thai	THAI CHARACTER PAIYANNOI	symbol	

U+0E86			LAO LETTER PALI GHA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0E89			LAO LETTER PALI CHA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0E8C			LAO LETTER PALI JHA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0E8E..U+0E93	..		LAO LETTER PALI NYA..LAO LETTER PALI NNA	uncommon_use	(6) (Pali/Sanskrit) (12.0 17/106)
U+0E98			LAO LETTER PALI DHA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0EA0			LAO LETTER PALI BHA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0EA8..U+0EA9	..		LAO LETTER SANSKRIT SHA..LAO LETTER SANSKRIT SSA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0EAC			LAO LETTER PALI LLA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0EAF	ๅ	Lao	LAO ELLIPSIS	symbol	
U+0EBA			LAO SIGN PALI VIRAMA	uncommon_use	(Pali/Sanskrit) (12.0 17/106)
U+0F00	ༀ	Tibetan	TIBETAN SYLLABLE OM	religious_use, symbol	
U+0F35	༵	Tibetan	TIBETAN MARK NGAS BZUNG NYI ZLA	symbol	(honorific, emphasis)
U+0F37	༷	Tibetan	TIBETAN MARK NGAS BZUNG SGOR RTAGS	symbol	(emphasis)
U+0F3E	༾	Tibetan	TIBETAN SIGN YAR TSHES	numeric	(almanacs)
U+0F3F	༿	Tibetan	TIBETAN SIGN MAR TSHES	numeric	(almanacs)
U+0F6A	ཪ	Tibetan	TIBETAN LETTER FIXED-FORM RA	uncommon_use, obsolete	(Sanskrit)
U+0F6B..U+0F6C	··	Tibetan	TIBETAN LETTER KKA..TIBETAN LETTER RRA	uncommon_use	(Balti)
U+0F7B	ཻ	Tibetan	TIBETAN VOWEL SIGN EE	duplicate	*homoglyph (digraph of 0F7A 0F7A)
U+0F7D	ཽ	Tibetan	TIBETAN VOWEL SIGN OO	duplicate	*homoglyph (digraph of 0F7C 0F7C)
U+0F82..U+0F83	ྂ ..	Tibetan	TIBETAN SIGN NYI ZLA NAA DA..TIBETAN SIGN	uncommon_use, obsolete	(Sanskrit)

			SNA LDAN		
U+0F86..U+0F8B		Tibetan	TIBETAN SIGN LCI RTAGS..TIBETAN SIGN GRU MED RGYINGS	obsolete	(historic)
U+0F8C..U+0F8F	··	Tibetan	TIBETAN SIGN INVERTED MCHU CAN..TIBETAN SUBJOINED SIGN INVERTED MCHU CAN	obsolete	(historic)
U+0FAE..U+0FAF		Tibetan	TIBETAN SUBJOINED LETTER ZHA..TIBETAN SUBJOINED LETTER ZA	uncommon_use	
U+0FB0		Tibetan	TIBETAN SUBJOINED LETTER -A	uncommon_use	
U+0FC6		Tibetan	TIBETAN SYMBOL PADMA GDAN	uncommon_use, symbol	
U+1050..U+1059		Myanmar	MYANMAR LETTER SHA..MYANMAR VOWEL SIGN VOCALIC LL	uncommon_use, obsolete	(Pali) (Sanskrit)
U+1065..U+106D		Myanmar	MYANMAR LETTER WESTERN PWO KAREN THA..MYANMAR SIGN WESTERN PWO KAREN TONE-5	uncommon_use	(Western Pwo Karen)
U+106E..U+1070		Myanmar	MYANMAR LETTER EASTERN PWO KAREN NNA..MYANMAR LETTER EASTERN PWO KAREN GHWA	uncommon_use	(Eastern Pwo Karen)
U+1071		Myanmar	MYANMAR VOWEL SIGN GEBA KAREN I	uncommon_use	(Geba Karen)
U+1072..U+1074		Myanmar	MYANMAR VOWEL SIGN KAYAH OE..MYANMAR VOWEL SIGN KAYAH EE	uncommon_use	(Kayah)
U+108E	Ⲑ	Myanmar	MYANMAR LETTER RUMAI PALAUNG FA	uncommon_use	(Rumai Palaung)
U+109A..U+109B		Myanmar	MYANMAR SIGN KHAMTI TONE-1..MYANMAR SIGN KHAMTI TONE-3	uncommon_use	(Kamti Shan)
U+109C..U+109D		Myanmar	MYANMAR VOWEL SIGN AITON A..MYANMAR VOWEL SIGN AITON AI	uncommon_use	(Aiton, Phake)
U+10F9		Georgian	GEORGIAN LETTER	uncommon_use	(educational)

			TURNED GAN	, technical	
U+10FA	ჺ	Georgian	GEORGIAN LETTER AIN	uncommon_use, technical	(threatened)
U+10FD		Georgian	GEORGIAN LETTER AEN	uncommon_use	(Ossetian, Abkhaz)
U+10FE..U+10FF	..	Georgian	GEORGIAN LETTER HARD SIGN..GEORGIAN LETTER LABIAL SIGN	uncommon_use	(Ossetian, Abkhaz)
U+1380..U+138F	ᎀᎁᎏ	Ethiopic	ETHIOPIC SYLLABLE SEBATBEIT MWA..ETHIOPIC SYLLABLE PWE	uncommon_use	(Sebatbeit)
U+17DC	ៜ	Khmer	KHMER SIGN AVAKRAHASANYA	obsolete	(Sanskrit)
U+1C80..U+1C88	..	Cyrillic	CYRILLIC SMALL LETTER ROUNDED VE..CYRILLIC SMALL LETTER UNBLENDED UK	Allowed	??? post 6.3
U+1E01	ḁ	Latin	LATIN SMALL LETTER A WITH RING BELOW	technical	
U+1E19	ḙ	Latin	LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW	technical	
U+1E1B	ḛ	Latin	LATIN SMALL LETTER E WITH TILDE BELOW	technical	
U+1E2B	ḫ	Latin	LATIN SMALL LETTER H WITH BREVE BELOW	technical	(Semitic transliteration)
U+1E2D	ḭ	Latin	LATIN SMALL LETTER I WITH TILDE BELOW	technical	
U+1E73	ṳ	Latin	LATIN SMALL LETTER U WITH DIAERESIS BELOW	technical	
U+1E75	ṵ	Latin	LATIN SMALL LETTER U WITH TILDE BELOW	technical	
U+1E77	ṷ	Latin	LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW	technical	
U+1F00..U+1F07	ἀ..ἇ	Greek	GREEK SMALL LETTER ALPHA WITH PSILI..GREEK SMALL LETTER ALPHA WITH DASIA AND PERISPOMENI	historic	*polytoniko orthography
U+1F10..U+1F15	ἐ..ἕ	Greek	GREEK SMALL LETTER EPSILON WITH PSILI..GREEK SMALL LETTER EPSILON WITH	historic	*polytoniko orthography

			DASIA AND OXIA		
U+1F20..U+1F27	ἠ..ἧ	Greek	GREEK SMALL LETTER ETA WITH PSILI..GREEK SMALL LETTER ETA WITH DASIA AND PERISPOMENI	historic	*polytoniko orthography
U+1F30..U+1F37	ἰ..ἷ	Greek	GREEK SMALL LETTER IOTA WITH PSILI..GREEK SMALL LETTER IOTA WITH DASIA AND PERISPOMENI	historic	*polytoniko orthography
U+1F40..U+1F45	ὀ..ὅ	Greek	GREEK SMALL LETTER OMICRON WITH PSILI..GREEK SMALL LETTER OMICRON WITH DASIA AND OXIA	historic	(6) *polytoniko orthography
U+1F50..U+1F57	ὐ..ὗ	Greek	GREEK SMALL LETTER UPSILON WITH PSILI..GREEK SMALL LETTER UPSILON WITH DASIA AND PERISPOMENI	historic	*polytoniko orthography
U+1F60..U+1F67	ὠ..ὧ	Greek	GREEK SMALL LETTER OMEGA WITH PSILI..GREEK SMALL LETTER OMEGA WITH DASIA AND PERISPOMENI	historic	*polytoniko orthography
U+1F70	ὰ	Greek	GREEK SMALL LETTER ALPHA WITH VARIA	historic	*polytoniko orthography
U+1F72	ὲ	Greek	GREEK SMALL LETTER EPSILON WITH VARIA	historic	*polytoniko orthography
U+1F74	ὴ	Greek	GREEK SMALL LETTER ETA WITH VARIA	historic	*polytoniko orthography
U+1F76	ὶ	Greek	GREEK SMALL LETTER IOTA WITH VARIA	historic	*polytoniko orthography
U+1F78	ὸ	Greek	GREEK SMALL LETTER OMICRON WITH VARIA	historic	*polytoniko orthography
U+1F7A	ὺ	Greek	GREEK SMALL LETTER UPSILON WITH VARIA	historic	*polytoniko orthography
U+1F7C	ὼ	Greek	GREEK SMALL LETTER OMEGA WITH VARIA	historic	*polytoniko orthography
U+1F80..U+1FAF	ᾀ..Ὧ	Greek	GREEK SMALL LETTER ALPHA WITH PSILI AND YPOGEGRAMMENI..GREEK CAPITAL LETTER		NOT IDNA2008

			OMEGA WITH DASIA AND PERISPOMENI AND PROSGEGRAMMENI		
U+1FB0..U+1FB1	ᾰ..ᾱ	Greek	GREEK SMALL LETTER ALPHA WITH VRACHY..GREEK SMALL LETTER ALPHA WITH MACRON	technical	
U+1FB2..U+1FB4	ᾲ..ᾴ	Greek	GREEK SMALL LETTER ALPHA WITH VARIA AND YPOGEGRAMMENI..GREEK SMALL LETTER ALPHA WITH OXIA AND YPOGEGRAMMENI		NOT IDNA208
U+1FB6	ᾶ	Greek	GREEK SMALL LETTER ALPHA WITH PERISPOMENI	historic	*polytoniko orthography
U+1FB7	ᾷ	Greek	GREEK SMALL LETTER ALPHA WITH PERISPOMENI AND YPOGEGRAMMENI		NOT IDNA2008
U+1FBC	ᾼ	Greek	GREEK CAPITAL LETTER ALPHA WITH PROSGEGRAMMENI		NOT IDNA2008
U+1FC2..U+1FC4	ῂ..ῄ	Greek	GREEK SMALL LETTER ETA WITH VARIA AND YPOGEGRAMMENI..GREEK SMALL LETTER ETA WITH OXIA AND YPOGEGRAMMENI		NOT IDNA2008
U+1FC6	ῆ	Greek	GREEK SMALL LETTER ETA WITH PERISPOMENI	historic	*polytoniko orthography
U+1FC7	ῇ	Greek	GREEK SMALL LETTER ETA WITH PERISPOMENI AND YPOGEGRAMMENI		NOT IDNA2008
U+1FCC	ῌ	Greek	GREEK CAPITAL LETTER ETA WITH PROSGEGRAMMENI		NOT IDNA2008
U+1FD0..U+1FD2	ῐ..ῒ	Greek	GREEK SMALL LETTER IOTA WITH VRACHY..GREEK SMALL LETTER IOTA WITH DIALYTIKA AND VARIA	historic	*polytoniko orthography
U+1FD6..U+1FD7	ῖ..ῗ	Greek	GREEK SMALL LETTER IOTA WITH PERISPOMENI..GREEK	historic	*polytoniko orthography

			SMALL LETTER IOTA WITH DIALYTIKA AND PERISPOMENI		
U+1FE0..U+1FE2	ῠ..ῢ	Greek	GREEK SMALL LETTER UPSILON WITH VRACHY..GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND VARIA	historic	*polytoniko orthography
U+1FE4..U+1FE7	ῤ..ῧ	Greek	GREEK SMALL LETTER RHO WITH PSILI..GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND PERISPOMENI	historic	*polytoniko orthography
U+1FF2..U+1FF4	ῲ..ῴ	Greek	GREEK SMALL LETTER OMEGA WITH VARIA AND YPOGEGRAMMENI..GREEK SMALL LETTER OMEGA WITH OXIA AND YPOGEGRAMMENI		NOT IDNA2008
U+1FF6	ῶ	Greek	GREEK SMALL LETTER OMEGA WITH PERISPOMENI	historic	*polytoniko orthography
U+1FF7	ῷ	Greek	GREEK SMALL LETTER OMEGA WITH PERISPOMENI AND YPOGEGRAMMENI		NOT IDNA2008
U+1FFC	ῼ	Greek	GREEK CAPITAL LETTER OMEGA WITH PROSGEGRAMMENI		NOT IDNA2008
U+200C..U+200D	..	Inherited	ZERO WIDTH NON-JOINER..ZERO WIDTH JOINER	joiner	IDNA2008 CONTEXTJ
U+2010	‐	Common	HYPHEN		NOT IDNA2008
U+2019	'	Common	RIGHT SINGLE QUOTATION MARK		NOT IDNA2008
U+2027	·	Common	HYPHENATION POINT		NOT IDNA2008
U+2D27		Georgian	GEORGIAN SMALL LETTER YN	religious_use	(Khutsuri)
U+2D2D		Georgian	GEORGIAN SMALL LETTER AEN	religious_use	(Khutsuri)
U+2DA0..U+2DA6	ⶠ..ⶦ	Ethiopic	ETHIOPIC SYLLABLE SSA..ETHIOPIC SYLLABLE SSO	uncommon_use	(Sebatbeit)
U+2DA8..U+2DAE	ⶨ..ⶮ	Ethiopic	ETHIOPIC SYLLABLE	uncommon_use	(Sebatbeit)

			CCA..ETHIOPIC SYLLABLE CCO		
U+2DB0..U+2DB6	ዠ..ዦ	Ethiopic	ETHIOPIC SYLLABLE ZZA..ETHIOPIC SYLLABLE ZZO	uncommon_use	(Sebatbeit)
U+2DB8..U+2DBE	ዸ..ዾ	Ethiopic	ETHIOPIC SYLLABLE CCHA..ETHIOPIC SYLLABLE CCHO	uncommon_use	(Sebatbeit)
U+2DC0..U+2DC6	ቀ..ቆ	Ethiopic	ETHIOPIC SYLLABLE QYA..ETHIOPIC SYLLABLE QYO	uncommon_use	(Sebatbeit)
U+2DC8..U+2DCE	ከ..ኮ	Ethiopic	ETHIOPIC SYLLABLE KYA..ETHIOPIC SYLLABLE KYO	uncommon_use	(Sebatbeit)
U+2DD0..U+2DD6	ዀ..ዄ	Ethiopic	ETHIOPIC SYLLABLE XYA..ETHIOPIC SYLLABLE XYO	uncommon_use	(Sebatbeit)
U+2DD8..U+2DDE	ጘ..ጞ	Ethiopic	ETHIOPIC SYLLABLE GYA..ETHIOPIC SYLLABLE GYO	uncommon_use	(Sebatbeit)
U+3099	゙	Inherited	COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK	punctuation	
U+309A	゚	Inherited	COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK	punctuation	
U+30A0	=	Common	KATAKANA-HIRAGANA DOUBLE HYPHEN	Allowed	
U+30FB	・	Common	KATAKANA MIDDLE DOT	context-other	IDNA2008 CONTEXTO
U+3105..U+312C	ㄅ..□	Bopomofo	BOPOMOFO LETTER B..BOPOMOFO LETTER GN	limited_use	* educational only
U+312D	□	Bopomofo	BOPOMOFO LETTER IH	limited_use	* educational only
U+312E	□	Bopomofo	BOPOMOFO LETTER O WITH DOT ABOVE	limited_use	* educational only
U+312F	□	Bopomofo	BOPOMOFO LETTER NN	limited_use	* educational only
U+31A0..U+31B7	ㆠ..ㆷ	Bopomofo	BOPOMOFO LETTER BU..BOPOMOFO FINAL LETTER H	limited_use	* educational only
U+31B8..U+31BA	..	Bopomofo	BOPOMOFO LETTER GH..BOPOMOFO LETTER ZY	limited_use	* educational only
U+A661		Cyrillic	CYRILLIC SMALL LETTER REVERSED TSE	obsolete	

U+A674..U+A67B	..	Cyrillic	COMBINING CYRILLIC LETTER UKRAINIAN IE..COMBINING CYRILLIC LETTER OMEGA	obsolete	
U+A67F		Cyrillic	CYRILLIC PAYEROK	obsolete	
U+A69F		Cyrillic	COMBINING CYRILLIC LETTER IOTIFIED E	obsolete	
U+A717..U+A71A	˙..˗	Common	MODIFIER LETTER DOT VERTICAL BAR..MODIFIER LETTER LOWER RIGHT CORNER ANGLE	modifier	*no modifiers
U+A71B..U+A71F	↑..ᵎ	Common	MODIFIER LETTER RAISED UP ARROW..MODIFIER LETTER LOW INVERTED EXCLAMATION MARK	modifier	*no modifiers
U+A788	ˆ	Common	MODIFIER LETTER LOW CIRCUMFLEX ACCENT	technical	
U+A78E		Latin	LATIN SMALL LETTER L WITH RETROFLEX HOOK AND BELT	phonetic	
U+A791		Latin	LATIN SMALL LETTER N WITH DESCENDER	limited_use	*Janalif
U+A793		Latin	LATIN SMALL LETTER C WITH BAR	limited_use	*Nanai
U+A7A1		Latin	LATIN SMALL LETTER G WITH OBLIQUE STROKE	obsolete	
U+A7A3		Latin	LATIN SMALL LETTER K WITH OBLIQUE STROKE	obsolete	
U+A7A5		Latin	LATIN SMALL LETTER N WITH OBLIQUE STROKE	obsolete	
U+A7A7		Latin	LATIN SMALL LETTER R WITH OBLIQUE STROKE	obsolete	
U+A7A9		Latin	LATIN SMALL LETTER S WITH OBLIQUE STROKE	obsolete	
U+A7AF		Latin	LATIN LETTER SMALL CAPITAL Q	uncommon_use	(Japanese dialectology) (11.0 15/241)
U+A7BB			LATIN SMALL LETTER GLOTTAL A	uncommon_use	(Egyptian transliteration/Ugaritic) (12.0 17/076R2 17/362)
U+A7BD			LATIN SMALL LETTER GLOTTAL I	uncommon_use	(Egyptian transliteration/Ugari

					tic) (12.0 17/076R2 17/362)
U+A7BF			LATIN SMALL LETTER GLOTTAL U	uncommon_use	(Egyptian transliteration/Ugaritic) (12.0 17/076R2 17/362)
U+A7C3			LATIN SMALL LETTER ANGLICANA W	uncommon_use	(medieval English/Cornish) (12.0 17/238)
U+A7FA		Latin	LATIN LETTER SMALL CAPITAL TURNED M	technical	
U+A9E7..U+A9EF	..	Myanmar	MYANMAR LETTER TAI LAING NYA..MYANMAR LETTER TAI LAING NNA	uncommon_use	(Tai Laing) (7.0 11/130R)
U+A9FA..U+A9FE	..	Myanmar	MYANMAR LETTER TAI LAING LLA..MYANMAR LETTER TAI LAING BHA	uncommon_use	(Tai Laing) (7.0 11/130R)
U+AA60..U+AA76	..	Myanmar	MYANMAR LETTER KHAMTI GA..MYANMAR LOGOGRAM KHAMTI HM	uncommon_use	(Khamti Shan)
U+AA7A		Myanmar	MYANMAR LETTER AITON RA	uncommon_use	(Aiton)
U+AA7C..U+AA7D	..	Myanmar	MYANMAR SIGN TAI LAING TONE-2..MYANMAR SIGN TAI LAING TONE-5	uncommon_use	(Tai Laing) (7.0 11/130R)
U+AA7E..U+AA7F	..	Myanmar	MYANMAR LETTER SHWE PALAUNG CHA..MYANMAR LETTER SHWE PALAUNG SHA	uncommon_use	(Shwe Palaung) (7.0 11/130R)
U+AB11..U+AB16	..	Ethiopic	ETHIOPIC SYLLABLE DZU..ETHIOPIC SYLLABLE DZO	uncommon_use	(Gamo-Gofa-Dawro)
U+AB20..U+AB26	..	Ethiopic	ETHIOPIC SYLLABLE CCHHA..ETHIOPIC SYLLABLE CCHHO	uncommon_use	(Gumuz)
U+AB28..U+AB2E	..	Ethiopic	ETHIOPIC SYLLABLE BBA..ETHIOPIC SYLLABLE BBO	uncommon_use	(Gumuz)
U+AB66..U+AB67	..		LATIN SMALL LETTER DZ DIGRAPH WITH RETROFLEX HOOK..LATIN SMALL LETTER TS DIGRAPH WITH RETROFLEX HOOK	uncommon_use	(Sinology) (12.0 17/299 17/367)

U+1133B		Inherited	COMBINING BINDU BELOW		*excluded
U+1B150..U+1B152	..		HIRAGANA LETTER SMALL WI..HIRAGANA LETTER SMALL WO	obsolete	(12.0 16/354 16/385R)
U+1B164..U+1B167	..		KATAKANA LETTER SMALL WI..KATAKANA LETTER SMALL N	obsolete	(12.0 16/354 16/385R)

Proposed: TBD add proposed changes

7   359 Characters Allowed and in MSR but not included in Root Zone

The following 359 characters have IdentifierType "Allowed"; in addition, they are not excluded from the MSR. Yet none of the Root Zone script LGRs (or pending drafts) has chosen to include them, casting doubts as to whether they should properly be considered "Recommended".

For the particular Rationale behind not including certain code points, see the LGR Proposal document for the given script (for a list of published ones, see [LGR-3]). For example, the Arabic LGR categorically excludes combining marks as unsuitable for IDN TLDs; this follows RFC 5564 where similar recommendations were made for Arabic domain names in general. We think this should be reason enough to not make these characters "Recommended".

Note that there are no available proposed repertoires for Thaana and Tibetan, therefore these code points are removed from the following list. (Also, all CJK characters are ignored in this section).

Proposed: With exception of a few of the post Unicode 6.3 Arabic characters, we recommend that the IdentifierType be changed to "Uncommon_Use" for most of these (perhaps "technical" for the diacritics). We deem the precise IdentifierType as less important than removing them from the "Recommended set".

Code Point	Glyph	Script	Name	Comment
U+0115	ĕ	Latin	LATIN SMALL LETTER E WITH BREVE	
U+012D	ĭ	Latin	LATIN SMALL LETTER I WITH BREVE	
U+014F	ŏ	Latin	LATIN SMALL LETTER O WITH BREVE	
U+0157	ŗ	Latin	LATIN SMALL LETTER R WITH CEDILLA	
U+0163	ţ	Latin	LATIN SMALL LETTER T WITH CEDILLA	
U+01D6	ǖ	Latin	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON	
U+01D8	ǘ	Latin	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE	
U+01DA	ǚ	Latin	LATIN SMALL LETTER U WITH DIAERESIS AND CARON	
U+01DC	ǜ	Latin	LATIN SMALL LETTER U WITH DIAERESIS AND

			GRAVE	
U+01DF	ǟ	Latin	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	
U+01E1	ǡ	Latin	LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON	
U+01E3	ǣ	Latin	LATIN SMALL LETTER AE WITH MACRON	
U+01EB	ǫ	Latin	LATIN SMALL LETTER O WITH OGONEK	
U+01ED	ǭ	Latin	LATIN SMALL LETTER O WITH OGONEK AND MACRON	
U+01F0	ǰ	Latin	LATIN SMALL LETTER J WITH CARON	
U+01F5	ǵ	Latin	LATIN SMALL LETTER G WITH ACUTE	
U+01F9	ǹ	Latin	LATIN SMALL LETTER N WITH GRAVE	
U+01FB	ǻ	Latin	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE	
U+01FD	ǽ	Latin	LATIN SMALL LETTER AE WITH ACUTE	
U+01FF	ǿ	Latin	LATIN SMALL LETTER O WITH STROKE AND ACUTE	
U+021F	ȟ	Latin	LATIN SMALL LETTER H WITH CARON	
U+0227	ȧ	Latin	LATIN SMALL LETTER A WITH DOT ABOVE	
U+0229	ȩ	Latin	LATIN SMALL LETTER E WITH CEDILLA	
U+022B	ȫ	Latin	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON	
U+022D	ȭ	Latin	LATIN SMALL LETTER O WITH TILDE AND MACRON	
U+022F	ȯ	Latin	LATIN SMALL LETTER O WITH DOT ABOVE	
U+0231	ȱ	Latin	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON	
U+0233	ȳ	Latin	LATIN SMALL LETTER Y WITH MACRON	
U+0300..U+0304	̈	Inherited	COMBINING GRAVE ACCENT..COMBINING MACRON	
U+0306..U+030C	̈	Inherited	COMBINING BREVE..COMBINING CARON	
U+031B	̛	Inherited	COMBINING HORN	
U+0323		Inherited	COMBINING DOT BELOW	
U+0326..U+0328	̧	Inherited	COMBINING COMMA BELOW..COMBINING OGONEK	
U+0331		Inherited	COMBINING MACRON BELOW	
U+0450	ѐ	Cyrillic	CYRILLIC SMALL LETTER IE WITH GRAVE	
U+045D	ѝ	Cyrillic	CYRILLIC SMALL LETTER I WITH GRAVE	
U+04C2	ӂ	Cyrillic	CYRILLIC SMALL LETTER ZHE WITH BREVE	
U+04CC	ӌ	Cyrillic	CYRILLIC SMALL LETTER KHAKASSIAN CHE	
U+04DB	ӛ	Cyrillic	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS	
U+04EB	ӫ	Cyrillic	CYRILLIC SMALL LETTER BARRED O WITH	

			DIAERESIS	
U+04ED	ӭ	Cyrillic	CYRILLIC SMALL LETTER E WITH DIAERESIS	
U+05B4	ִ◌	Hebrew	HEBREW POINT HIRIQ	
U+05F0..U+05F2	ﬡ..ײ	Hebrew	HEBREW LIGATURE YIDDISH DOUBLE VAV..HEBREW LIGATURE YIDDISH DOUBLE YOD	
U+064B..U+0652	◌ْ..◌ً	Inherited	ARABIC FATHATAN..ARABIC SUKUN	
U+0654..U+0655	◌ٕ..◌ٔ	Inherited	ARABIC HAMZA ABOVE..ARABIC HAMZA BELOW	
U+0670	◌ٰ	Inherited	ARABIC LETTER SUPERSCRIPT ALEF	
U+0671	ٱ	Arabic	ARABIC LETTER ALEF WASLA	
U+0674	ٴ	Arabic	ARABIC LETTER HIGH HAMZA	
U+0682	ڂ	Arabic	ARABIC LETTER HAH WITH TWO DOTS VERTICAL ABOVE	
U+0690	ڐ	Arabic	ARABIC LETTER DAL WITH FOUR DOTS ABOVE	
U+0692	ڒ	Arabic	ARABIC LETTER REH WITH SMALL V	
U+0694	ڔ	Arabic	ARABIC LETTER REH WITH DOT BELOW	
U+069B..U+069E	ڛ..ڞ	Arabic	ARABIC LETTER SEEN WITH THREE DOTS BELOW..ARABIC LETTER SAD WITH THREE DOTS ABOVE	
U+06A1	ڡ	Arabic	ARABIC LETTER DOTLESS FEH	
U+06A3	ڣ	Arabic	ARABIC LETTER FEH WITH DOT BELOW	
U+06A5	ڥ	Arabic	ARABIC LETTER FEH WITH THREE DOTS BELOW	
U+06B2	ڲ	Arabic	ARABIC LETTER GAF WITH TWO DOTS BELOW	
U+06B4	ڴ	Arabic	ARABIC LETTER GAF WITH THREE DOTS ABOVE	
U+06B6..U+06B7	ڶ..ڷ	Arabic	ARABIC LETTER LAM WITH DOT ABOVE..ARABIC LETTER LAM WITH THREE DOTS ABOVE	
U+06B8..U+06B9	ڸ..ڹ	Arabic	ARABIC LETTER LAM WITH THREE DOTS BELOW..ARABIC LETTER NOON WITH DOT BELOW	
U+06BF	ڿ	Arabic	ARABIC LETTER TCHEH WITH DOT ABOVE	
U+06C5	ۅ	Arabic	ARABIC LETTER KIRGHIZ OE	
U+06C7..U+06CA	ۇ..ۊ	Arabic	ARABIC LETTER U..ARABIC LETTER WAW WITH TWO DOTS ABOVE	

U+06D3	ئے	Arabic	ARABIC LETTER YEH BARREE WITH HAMZA ABOVE	
U+06EE..U+06EF	ڎ..گ	Arabic	ARABIC LETTER DAL WITH INVERTED V..ARABIC LETTER REH WITH INVERTED V	
U+06FA..U+06FC	ں..ڼ	Arabic	ARABIC LETTER SHEEN WITH DOT BELOW..ARABIC LETTER GHAIN WITH DOT BELOW	
U+06FF	ۿ	Arabic	ARABIC LETTER HEH WITH INVERTED V	
U+0750	ݐ	Arabic	ARABIC LETTER BEH WITH THREE DOTS HORIZONTALLY BELOW	
U+0753..U+0755	ݓ..ݕ	Arabic	ARABIC LETTER BEH WITH THREE DOTS POINTING UPWARDS BELOW AND TWO DOTS ABOVE..ARABIC LETTER BEH WITH INVERTED SMALL V BELOW	
U+0757..U+075F	ݗ..ݟ	Arabic	ARABIC LETTER HAH WITH TWO DOTS ABOVE..ARABIC LETTER AIN WITH TWO DOTS VERTICALLY ABOVE	
U+0761	ݡ	Arabic	ARABIC LETTER FEH WITH THREE DOTS POINTING UPWARDS BELOW	
U+0764..U+0765	ݤ..ݥ	Arabic	ARABIC LETTER KEHEH WITH THREE DOTS POINTING UPWARDS BELOW..ARABIC LETTER MEEM WITH DOT ABOVE	
U+0769	ݩ	Arabic	ARABIC LETTER NOON WITH SMALL V	
U+076B..U+076D	ݫ..ݭ	Arabic	ARABIC LETTER REH WITH TWO DOTS VERTICALLY ABOVE..ARABIC LETTER SEEN WITH TWO DOTS VERTICALLY ABOVE	
U+0772..U+077D	..	Arabic	ARABIC LETTER HAH WITH SMALL ARABIC LETTER TAH ABOVE..ARABIC LETTER SEEN WITH EXTENDED ARABIC-INDIC DIGIT FOUR ABOVE	
U+08A1		Arabic	ARABIC LETTER BEH WITH HAMZA ABOVE	(7.0 10/288)
U+08AA..U+08AC	..	Arabic	ARABIC LETTER REH WITH LOOP..ARABIC LETTER ROHINGYA YEH	
U+08BB..U+08BD	..	Arabic	ARABIC LETTER AFRICAN FEH..ARABIC LETTER AFRICAN NOON	(Warsh 9.0 14/211)
U+0904	ऄ	Devanagari	DEVANAGARI LETTER SHORT A	
U+090C	ऌ	Devanagari	DEVANAGARI LETTER VOCALIC L	
U+0929	ऩ	Devanagari	DEVANAGARI LETTER NNNA	

U+0931	ऱ	Devanagari	DEVANAGARI LETTER RRA	
U+0934	ऴ	Devanagari	DEVANAGARI LETTER LLLA	
U+0944	ृ	Devanagari	DEVANAGARI VOWEL SIGN VOCALIC RR	
U+0979..U+097A	..	Devanagari	DEVANAGARI LETTER ZHA..DEVANAGARI LETTER HEAVY YA	
U+0981..U+0983	ঁ..ঃ	Bengali	BENGALI SIGN CANDRABINDU..BENGALI SIGN VISARGA	
U+0985..U+098C	অ..ঌ	Bengali	BENGALI LETTER A..BENGALI LETTER VOCALIC L	
U+098F..U+0990	এ..ঐ	Bengali	BENGALI LETTER E..BENGALI LETTER AI	
U+0993..U+09A8	ও..ন	Bengali	BENGALI LETTER O..BENGALI LETTER NA	
U+09AA..U+09B0	প..র	Bengali	BENGALI LETTER PA..BENGALI LETTER RA	
U+09B2	ল	Bengali	BENGALI LETTER LA	
U+09B6..U+09B9	শ..হ	Bengali	BENGALI LETTER SHA..BENGALI LETTER HA	
U+09BC	়	Bengali	BENGALI SIGN NUKTA	
U+09BE..U+09C4	া..ৄ	Bengali	BENGALI VOWEL SIGN AA..BENGALI VOWEL SIGN VOCALIC RR	
U+09C7..U+09C8	ে..ৈ	Bengali	BENGALI VOWEL SIGN E..BENGALI VOWEL SIGN AI	
U+09CB..U+09CD	ো..্	Bengali	BENGALI VOWEL SIGN O..BENGALI SIGN VIRAMA	
U+09CE	ৎ	Bengali	BENGALI LETTER KHANDA TA	
U+09D7	ৗ	Bengali	BENGALI AU LENGTH MARK	
U+09F0..U+09F1	ৰ..ৱ	Bengali	BENGALI LETTER RA WITH MIDDLE DIAGONAL..BENGALI LETTER RA WITH LOWER DIAGONAL	
U+0A03	ਃ	Gurmukhi	GURMUKHI SIGN VISARGA	
U+0A72..U+0A73	ੲ..ੳ	Gurmukhi	GURMUKHI IRI..GURMUKHI URA	

U+0A81	ઁ	Gujarati	GUJARATI SIGN CANDRABINDU	
U+0B0C	ଌ	Oriya	ORIYA LETTER VOCALIC L	
U+0B35	ଵ	Oriya	ORIYA LETTER VA	
U+0B57	ୗ	Oriya	ORIYA AU LENGTH MARK	
U+0BD7	ௗ	Tamil	TAMIL AU LENGTH MARK	
U+0C0C	ఌ	Telugu	TELUGU LETTER VOCALIC L	
U+0C31	ఱ	Telugu	TELUGU LETTER RRA	
U+0C55..U+0C56	ౕ..ౖ	Telugu	TELUGU LENGTH MARK..TELUGU AI LENGTH MARK	
U+0C8C	ಌ	Kannada	KANNADA LETTER VOCALIC L	
U+0CB1	ಱ	Kannada	KANNADA LETTER RRA	
U+0CBC	಼	Kannada	KANNADA SIGN NUKTA	
U+0CC4	ೄ	Kannada	KANNADA VOWEL SIGN VOCALIC RR	
U+0CD5..U+0CD6	ೕ..ೖ	Kannada	KANNADA LENGTH MARK..KANNADA AI LENGTH MARK	
U+0D0C	ഌ	Malayalam	MALAYALAM LETTER VOCALIC L	
U+0D29		Malayalam	MALAYALAM LETTER NNNA	
U+0D8E	ඎ	Sinhala	SINHALA LETTER IRUUYANNA	
U+0D9E	ඞ	Sinhala	SINHALA LETTER KANTAJA NAASIKYAYA	
U+0E45..U+0E46	ๅ..ๆ	Thai	THAI CHARACTER LAKKHANGYAO..THAI CHARACTER MAIYAMOK	
U+0E4E	๎	Thai	THAI CHARACTER YAMAKKAN	
U+0EDE..U+0EDF	..	Lao	LAO LETTER KHMU GO..LAO LETTER KHMU NYO	

U+1063	◌ၣ	Myanmar	MYANMAR TONE MARK SGAW KAREN HATHI	
U+108B..U+108D	◌ႋ..◌ႍ	Myanmar	MYANMAR SIGN SHAN COUNCIL TONE-2..MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE	
U+10F7..U+10F8	ჷ..ჸ	Georgian	GEORGIAN LETTER YN..GEORGIAN LETTER ELIFI	
U+1207	ሇ	Ethiopic	ETHIOPIC SYLLABLE HOA	
U+1287	ኇ	Ethiopic	ETHIOPIC SYLLABLE XOA	
U+12AF	ኯ	Ethiopic	ETHIOPIC SYLLABLE KOA	
U+12F8..U+12FF	ዸ..ዿ	Ethiopic	ETHIOPIC SYLLABLE DDA..ETHIOPIC SYLLABLE DDWA	
U+130F	ጏ	Ethiopic	ETHIOPIC SYLLABLE GOA	
U+131F	ጟ	Ethiopic	ETHIOPIC SYLLABLE GGWAA	
U+1347	ፇ	Ethiopic	ETHIOPIC SYLLABLE TZOA	
U+135A	ፚ	Ethiopic	ETHIOPIC SYLLABLE FYA	
U+135D..U+135E	◌፝..◌፞	Ethiopic	ETHIOPIC COMBINING GEMINATION AND VOWEL LENGTH MARK..ETHIOPIC COMBINING VOWEL LENGTH MARK	
U+135F		Ethiopic	ETHIOPIC COMBINING GEMINATION MARK	
U+179D..U+179E	ឝ..ឞ	Khmer	KHMER LETTER SHA..KHMER LETTER SSO	
U+17A9	ឩ	Khmer	KHMER INDEPENDENT VOWEL QUU	
U+17B2	ឲ	Khmer	KHMER INDEPENDENT VOWEL QOO TYPE TWO	
U+17D7	ៗ	Khmer	KHMER SIGN LEK TOO	
U+1E03	ḃ	Latin	LATIN SMALL LETTER B WITH DOT ABOVE	
U+1E05	ḅ	Latin	LATIN SMALL LETTER B WITH DOT BELOW	
U+1E07	ḇ	Latin	LATIN SMALL LETTER B WITH LINE BELOW	
U+1E09	ḉ	Latin	LATIN SMALL LETTER C WITH CEDILLA AND ACUTE	
U+1E0B	ḋ	Latin	LATIN SMALL LETTER D WITH DOT ABOVE	
U+1E0D	ḍ	Latin	LATIN SMALL LETTER D WITH DOT BELOW	
U+1E0F	ḏ	Latin	LATIN SMALL LETTER D WITH LINE BELOW	
U+1E11	ḑ	Latin	LATIN SMALL LETTER D WITH CEDILLA	
U+1E15	ḕ	Latin	LATIN SMALL LETTER E WITH MACRON AND GRAVE	
U+1E17	ḗ	Latin	LATIN SMALL LETTER E WITH MACRON AND ACUTE	
U+1E1D	ḝ	Latin	LATIN SMALL LETTER E WITH CEDILLA AND BREVE	
U+1E1F	ḟ	Latin	LATIN SMALL LETTER F WITH DOT ABOVE	

U+1E23	ḣ	Latin	LATIN SMALL LETTER H WITH DOT ABOVE	
U+1E25	ḥ	Latin	LATIN SMALL LETTER H WITH DOT BELOW	
U+1E27	ḧ	Latin	LATIN SMALL LETTER H WITH DIAERESIS	
U+1E29	ḩ	Latin	LATIN SMALL LETTER H WITH CEDILLA	
U+1E2F	ḯ	Latin	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE	
U+1E31	ḱ	Latin	LATIN SMALL LETTER K WITH ACUTE	
U+1E33	ḳ	Latin	LATIN SMALL LETTER K WITH DOT BELOW	
U+1E35	ḵ	Latin	LATIN SMALL LETTER K WITH LINE BELOW	
U+1E39	ḹ	Latin	LATIN SMALL LETTER L WITH DOT BELOW AND MACRON	
U+1E3B	ḻ	Latin	LATIN SMALL LETTER L WITH LINE BELOW	
U+1E3F	ḿ	Latin	LATIN SMALL LETTER M WITH ACUTE	
U+1E41	ṁ	Latin	LATIN SMALL LETTER M WITH DOT ABOVE	
U+1E4D	ṍ	Latin	LATIN SMALL LETTER O WITH TILDE AND ACUTE	
U+1E4F	ṏ	Latin	LATIN SMALL LETTER O WITH TILDE AND DIAERESIS	
U+1E51	ṑ	Latin	LATIN SMALL LETTER O WITH MACRON AND GRAVE	
U+1E53	ṓ	Latin	LATIN SMALL LETTER O WITH MACRON AND ACUTE	
U+1E55	ṕ	Latin	LATIN SMALL LETTER P WITH ACUTE	
U+1E57	ṗ	Latin	LATIN SMALL LETTER P WITH DOT ABOVE	
U+1E59	ṙ	Latin	LATIN SMALL LETTER R WITH DOT ABOVE	
U+1E5B	ṛ	Latin	LATIN SMALL LETTER R WITH DOT BELOW	
U+1E5D	ṝ	Latin	LATIN SMALL LETTER R WITH DOT BELOW AND MACRON	
U+1E5F	ṟ	Latin	LATIN SMALL LETTER R WITH LINE BELOW	
U+1E61	ṡ	Latin	LATIN SMALL LETTER S WITH DOT ABOVE	
U+1E65	ṥ	Latin	LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE	
U+1E67	ṧ	Latin	LATIN SMALL LETTER S WITH CARON AND DOT ABOVE	
U+1E69	ṩ	Latin	LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE	
U+1E6B	ṫ	Latin	LATIN SMALL LETTER T WITH DOT ABOVE	
U+1E6F	ṯ	Latin	LATIN SMALL LETTER T WITH LINE BELOW	
U+1E79	ṹ	Latin	LATIN SMALL LETTER U WITH TILDE AND ACUTE	
U+1E7B	ṻ	Latin	LATIN SMALL LETTER U WITH MACRON AND DIAERESIS	
U+1E7D	ṽ	Latin	LATIN SMALL LETTER V WITH TILDE	

U+1E7F	ṿ	Latin	LATIN SMALL LETTER V WITH DOT BELOW	
U+1E81	ẁ	Latin	LATIN SMALL LETTER W WITH GRAVE	
U+1E83	ẃ	Latin	LATIN SMALL LETTER W WITH ACUTE	
U+1E85	ẅ	Latin	LATIN SMALL LETTER W WITH DIAERESIS	
U+1E87	ẇ	Latin	LATIN SMALL LETTER W WITH DOT ABOVE	
U+1E89	ẉ	Latin	LATIN SMALL LETTER W WITH DOT BELOW	
U+1E8B	ẋ	Latin	LATIN SMALL LETTER X WITH DOT ABOVE	
U+1E8F	ẏ	Latin	LATIN SMALL LETTER Y WITH DOT ABOVE	
U+1E91	ẑ	Latin	LATIN SMALL LETTER Z WITH CIRCUMFLEX	
U+1E93	ẓ	Latin	LATIN SMALL LETTER Z WITH DOT BELOW	
U+1E95..U+1E99	ẕ..ẙ	Latin	LATIN SMALL LETTER Z WITH LINE BELOW..LATIN SMALL LETTER Y WITH RING ABOVE	
U+2D80..U+2D96	ⶀ..ⶖ	Ethiopic	ETHIOPIC SYLLABLE LOA..ETHIOPIC SYLLABLE GGWE	
U+A7B9		Latin	LATIN SMALL LETTER U WITH STROKE	(Mazahua EGIDS 6a) *deferred repertoire (11.0 16/032)
U+AB01..U+AB06	..	Ethiopic	ETHIOPIC SYLLABLE TTHU..ETHIOPIC SYLLABLE TTHO	
U+AB09..U+AB0E	..	Ethiopic	ETHIOPIC SYLLABLE DDHU..ETHIOPIC SYLLABLE DDHO	

Proposed: TBD add proposed changes

8   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-lgr-procedure-20mar13-en.pdf

[RFC5564] El-Sherbiny, A., Farah, M., Oueichek, I., and A. Al-Zoman, "Linguistic Guidelines for the Use of the Arabic Language in Internet Domains", RFC 5564, DOI 10.17487/RFC5564, February 2010, http://www.rfc-editor.org/info/rfc5564

[RFC6912] Sullivan, A., *et al.,* "Principles for Unicode Code Point Inclusion in Labels in the DNS", RFC 6912, April 2013. = IABCP

[MSR-4] Integration Panel, "Maximal Starting Repertoire — MSR-4 Overview and Rationale", 7 February 2019 https://www.icann.org/en/system/files/files/msr-4-overview-25jan19-en.pdf [PDF, 0.8 MB]

[LGR-3] Integration Panel, "Integration Panel: Root Zone Label Generation Rules — LGR-2", 10 July 2019, https://www.icann.org/sites/default/files/lgr/lgr-3-overview-10jul19-en.pdf

[UAX31] UAX #31: *Unicode Identifier and Pattern Syntax.* An integral part of The Unicode Standard. Most recent version available from http://www.unicode.org/reports/tr31

[UTS31] UTS #39: *Unicode Security Mechanisms.* Most recent version available from http://www.unicode.org/reports/tr39/

Appendix 1: Documenting the Identifier Types

Looking at UTS#39 identifier category "Exclusion" there appears to be a discrepancy between the stated definition and the actual assignment of values in the IdentifierTypes.txt data file. We believe that the cases identified represent a shortcoming of the definitional language rather than an incorrect assignment of identifier type. In that sense, we view the issue as editorial.

The category "Exclusion" is supposed to be based on UAX#31 table of Exclusion candidates; UTS#39 gives the following definition:

"Exclusion: Characters from Table 4, *Candidate Characters for Exclusion from Identifiers* from [UAX31]"

However, the IdentifierType date file for 12.0.0 lists, for example

A9CF JAVANESE PANGRANGKEP [sc:Common]

which is not part of any of the listed scripts in table 4 and does not fit any of the property-based derivations given at the end of the table:

[\p{Extender=True} & \p{Joining_Type=Join_Causing}]	0640 (‑) ARABIC TATWEEL
	07FA ( ) NKO LAJANYALAN
\p{Default_Ignorable_Code_Point}	Default Ignorable Code Points See *Section 2.3, Layout and Format Control Characters*

\p{block=Combining_Diacritical_Marks_for_Symbols}
    \p{block=Musical_Symbols}
    \p{block=Ancient_Greek_Musical_Notation}
    \p{block=Phaistos_Disc}	

JAVANESE PANGRANGKEP is neither "Join-causing" nor "Default-ignorable", and also not member of any of the blocks listed.

Other characters that appear equally not covered by the stated definition of "Exclusion":

U+1735..U+1736	..	Common	PHILIPPINE SINGLE PUNCTUATION..PHILIPPINE DOUBLE PUNCTUATION	[5]	**Exclusion**, Not_XID
U+1802..U+1803	᠂ ᠃	Common	MONGOLIAN COMMA..MONGOLIAN FULL STOP	[3]	**Exclusion, Not_XID**
U+1805	᠅	Common	MONGOLIAN FOUR DOTS	[3]	**Exclusion, Not_XID**
U+1CFA		Common	VEDIC SIGN DOUBLE ANUSVARA ANTARGOMUKHA	[20]	**Exclusion, Obsolete**

U+10100..U+10102	..	Common	AEGEAN WORD SEPARATOR LINE..AEGEAN CHECK MARK	[6]	**Exclusion**, Not_XID
U+10107..U+10133	..	Common	AEGEAN NUMBER ONE..AEGEAN NUMBER NINETY THOUSAND	[6]	**Exclusion**, Not_XID
U+10137..U+1013F	..	Common	AEGEAN WEIGHT BASE UNIT..AEGEAN MEASURE THIRD SUBUNIT	[6]	**Exclusion**, Not_XID

For the last three items it looks like a \p{block=Agean_Numbers} is missing from Table 4 in UAX#31, while the remainder could be covered by an update to the definition of "Exclusion". The definitions for Limited_Use and Recommended would be reworded to match.

Proposed: add the Agean Number blocks to the list of exclusion candidates in Table 4 in UX#31.

Proposed updated definitions:

Exclusion:

> Characters from Table 4[1], Candidate Characters for Exclusion from Identifiers from [UAX31], and those that do not have an explicit script extensions value in either Table 5. Recommended Scripts or Table 7, Limited Use Scripts
>
> This tag is suppressed where characters have *Type=Not_Character*

Limited_Use:

> Characters that have an explicit script extensions value in Table 7, Limited Use Scripts in [UAX31], and no explicit script extensions value in Table 5. Recommended Scripts

Recommended:

> Characters with an explicit script extensions value int Table 5, Recommended Scripts in [UAX31], except for those characters that are Restricted above.

Other suggested changes in the text:

Below table #1 in UTS#39 it says:

> The distinctions among the **Type** values is not strict; if there are multiple Types for restricting a character only one is given.

This is no longer the case.

---

[1] Or "Tables 4 and 4a" if modified as suggested further below.

Proposed Replacement:

There may be multiple reasons for restricting a character. For some, such as the qualifiers on usage, Obsolete, Uncommon_Use and Technical, the distinctions among the **Type** values is not strict and only one is given. For others, multiple values may be given unless suppressed.

9   Appendix 2: Additional notes on UAX#31

From recent initiatives in both IETF and ICANN it would be seen as detrimental if the discussion in UAX#31 was simply reduced to scripts. The problem is that the cautions given for the scripts in Table 4,

"Some characters are not in modern customary use, and thus implementations may want to exclude them from identifiers. These include characters in historic and obsolete scripts, scripts used mostly liturgically, and regional scripts used only in very small communities or with very limited current usage. Some scripts also have unresolved architectural issues that make them currently unsuitable for identifiers. The set of characters in *Table 4, Candidate Characters for Exclusion from Identifiers* provides candidates of these, plus some inappropriate technical blocks.

apply equally well to specific characters in otherwise recommended scripts. That would argue for leaving table 4 largely as it is, except for properties that are neither blocks nor scripts.

Perhaps it would be worthwhile to a similar statement below table 4.

"Some characters used with recommended scripts may still be problematic for identifiers, for example because they are part of extensions that are not in modern customary use, and thus implementations may want to exclude them from identifiers. These include characters for historic and obsolete orthographies, characters used mostly liturgically, and in orthographies for languages used only in very small communities or with very limited current or declining usage. Some characters also have architectural issues that may make them unsuitable for identifiers. The set of characters in *Table 6, Characters Problematic in Identifiers* lists some subsets of these that are defined by property. For additional suggested sets of characters that are based on usage, see UTS#39."

and create a Table 4a. The new table would list property expressions for

(1) default ignorable

(2) deprecated <-- those seem an omission from current table 4

(3) joining extenders

(4) other?

This would focus UAX#31 on issues related to existing prorperties and relegate to UTS#39 the task of making the judgment calls about "Technical", "Obsolete", "Uncommon_Use" on an individual character in recommended scripts, which is a task that presents different maintenance characteristics than a property-based derivation. Importantly, UAX#31 would retain a clear statement that stopping at "recommended scripts" is not what this is about.

Proposed: clarify that exclusion candidates are not limited to whole scripts and ensure that the full list of them breaks down by Unicode property (leaving one-off analysis to UTS#39).

10 Appendix 3: 21 code points: break-down by CLDR exemplars

The following shows which languages' exemplars in CLDR support a given character from the list of 21, some of which have been supplied by SIL and not reviewed by vetters. Also, there CLDR languages are not selected by the same cutoff criteria that were used for the root zone, so may include threatened or extinct languages. However, each of them is also cited for at least one language above the cutoff used for the Root Zone.

For

ƒ => [avn, ee, wci] (← The ICANN analysis identifies Ewe)

ƙ => [anc, ank, ckl, ha, hia, ikx, kai, mbu, nin, pip, tal, tan, wja, wji] (← Hausa)

y => [dgi, ff, ff_Latn, ffm, fub_Latn, fue, fuf, fuh, fuq, fuv, ha, kai, kzr, sav, sok, srr]
        (←Dagaare - Burkina Faso)

ə => [ckl, dop, hbb, hia, jen, jgk, kai, kby_Latn, kpo, kr, ksf, las, mbu, mua, nmg]
        (←Kanuri)

ɍ => [kr] (← Kanuri)

ɓ => [anc, ank, asg, bas, bcn, bjt, bkc, bmq, bsq_Latn, bwr, bys, ckl, cky, cla, dbq, dgh, dgi, dnj,
    dow, dua, enn, ff, ff_Latn, ffm, fub_Latn, fue, fuf, fuh, fuq, fuv, gba, gby, gde, gkp, gmm,
    gnd, ha, hbb, hia, hig, ikx, jgk, kai, kdl, kkj, kpe, kzr, mbo, mbu, mif, mua, mzm, nin, nmg,
    pbi, pip, sav, sld, sok, srr, sur, tal, tan, tik, tsw, ttr, vai_Latn, vut, wja, wji, yay, yer]

    (←Hausa, Dagaare - Burkina Faso, Pulaar)

ɔ => [abi, acd, ada, ade, adj, agc, agq, aha, ahl, ajg, ak, any, avn, ayb, bas, bav, bba, bci, bcn, bet,
    bex, bfd, bhy, bib, bim, bkc, bm, bqc, bqp, bsq_Latn, bss, bud, bum, buu, bza, cbj, cko,
    cme, daf, dag, ddn, dgi, dnj, dop, dow, dua, dyu, dzg, ee, etu, ewo, fan, fmp, fod, fon,
    gaa, gba, gej, gkp, gmm, gng, god, gud, gur, guw, gux, ica, ich, idd, idu, ife, ijj, jen, jgo,
    kbp, kdh_Latn, ken, kez, kkj, kmw, knp, kpe, kpo, kqs, kri, ksf, kye, kzr, las, ldb, lee, lem,
    lgg, lia, lig, ln, lok, lor, lu, mas, maw, mbo, mcu, mda, mdj, men, mfo, mfq, mgo, mhi,
    mkl, mmu, mnf, moa, mql, mur, mzm, mzw, ncu, neb, nfr, nga, ngb, nhb, niy, nko, nmg,
    nmz, nnh, ntm, ntr, nus, nuv, nwb, ozm, pil, saf, sba, sef, sig, sil, sld, sok, soy, sus, sxw,
    tbz, ted, tem, tfi, tik, tpm, tuq, tvu, utr, vag, vai_Latn, vut, wan, wci, wib, wob, wwa, xon,
    xrb, xsm, xwe, yas, yat, yav, yba, ybb, yre]

    (←Dagaare - Burkina Faso, Dagbani (Dagomba), Lingala, Akan, Ewondo, Fon, Ga, Duala,
        EWE, Nuer)

ɖ => [ahl, ajg, avn, bsq_Latn, ee, fon, gej, ife, kbp, kdh_Latn, las, sxw, tfi, wci, xwe] (←Fon, Ewe)

ɗ => [anc, ank, asg, ayb, bcn, bkc, bwr, bys, ckl, cky, cla, dbq, dow, dua, enn, ff, ff_Latn, ffm,
    fub_Latn, fue, fuf, fuh, fuq, fuv, gba, gby, gde, gmm, gnd, guw, ha, hbb, hia, hig, ikx, jgk,
    kai, kdl, kkj, kzr, mbu, mif, mua, mzm, pbi, pip, sav, sok, srr, sur, tal, tan, tik, tsw, ttr,
    vai_Latn, vut, wja, wji, yay, yer] (←Hausa, Pulaar)

ɛ => [abi, acd, ada, ade, adj, agc, agq, aha, ahl, ajg, ak, any, avn, ayb, bas, bba, bci, bet, bfd, bhy,
    bib, bkc, bm, bqc, bqp, bsq_Latn, bss, buu, bza, bzw, cbj, ckl, cko, cme, daf, dag, ddn,
    dga, dgi, dnj, dop, dow, dua, dyu, dzg, ee, etu, ewo, fod, fon, gaa, gba, gej, gkp, gmm,
    gng, god, gud, gur, guw, ica, ich, idd, idu, ife, ijj, jen, jgo, kab, kbp, kdh_Latn, ken, kez,
    kkj, kmw, knp, kpe, kpo, kqs, kri, ksf, kye, las, ldb, lee, lem, lgg, lia, lig, lmp, ln, lok, lor, lu,
    mas, maw, mbo, mcp, mda, mdj, men, mfo, mfq, mhi, mkl, mmu, moa, mos, mql, mur,
    mzm, mzw, ncu, neb, nfr, nga, ngb, nhb, niy, nko, nmg, nmz, nnh, ntm, ntr, nus, nuv,
    nwb, ozm, pil, saf, sba, sef, shi_Latn, sig, sil, sld, sok, soy, sus, sxw, tbz, ted, tem, tfi, tik,
    tuq, tvu, tzm, utr, vag, vai_Latn, wan, wci, wib, wob, wwa, xrb, xsm, xwe, yam, yas, yat,
    yav, yba, ybb, yre]

    (←Dagaare - Burkina Faso, Lingala, Akan, Ewondo, Dagbani (Dagomba), Fon, Mossi, Ga,
        Ewe, Duala, Bambara, Nuer)

ɣ => [aha, ajg, bza, dag, dop, ee, gej, kab, kbp, kpe, kpo, nus, pil, shi_Latn, sxw, tzm, wci, xwe]

(←Dagbani (Dagomba), Dinka, Ewe, Nuer)

ɨ => [aak, agg, agm, ago, agq, apz, avt, bav, bio, bkm, buu, bye, byr, bzf, can, cky, dia, etu, gdr, geb, ian, iws, jen, kbx, ken, knp, lag, las, led, mas, mcp, mdj, mnf, mtf, nif, nin, niy, ozm, pbi, ptp, rao, sur, tgu, van, vut, yll, yuj, yut]
(←Cubeo, Dagbani (Dagomba), HIxkaryána, Maasai)

ɩ => [abi, ade, aha, any, bet, bib, dgi, dop, fod, god, gud, kbp, kdh_Latn, kpo, kye, las, lgg, lor, mhi, mos, neb, nko, nuv, nwb, sig, sil, sld, ted, wob, xsm, yre]

(←Dagaare - Burkina Faso, Mossi)

ŋ => [bm, dje, fuf, fuv, gkp, ikx, khq, kpe, kqs, ses, sus, twq]

(← Susu, Zarma, Bambara)

ʉ => [agq, arh, buu, etu, fmp, jgo, ken, knp, lag, led, mas, mci, mcp, mdj, med, niy, nnh, ozm, yam, ybb] (←Cubeo, Maasai)

ʒ => [ajg, dag, sms] (←Skolt Sami, Dagbani (Dagomba))

ڽ => []   (← the ICANN analysis identifies Malay, Information technology - Jawi Coded Character Set for Information Interchange MS 2443:2012, Department of Standards, Malaysia. http://www.standardsmalaysia.gov.my; Omniglot lists this also for Buruskashi, but this language is below the cutoff for the Root Zone)

̍ => [km] (← The ICANN analysis supports standard use in Khmer, not "technical")

̎ => [km] (← The ICANN analysis supports standard use in Khmer, not "technical")

̐ => [km] (← The ICANN analysis supports standard use in Khmer, not "technical")

11 Appendix 4: Criteria and Parameters for including Code Points in the Root Zone

The analysis of code points for MSR and Root Zone processes proceeded based on orthographies (languages) used with a given script. For the purposes of the Root Zone development languages were considered only if written in a given script for "everyday widespread use". As basis of the classification the process used the EGIDS scale from Ethnologue as a starting criteria, generally excluding languages that do have institutional support (such as use in education, administration etc.) but occasionally also considering those that are: "in vigorous use, with literature in a standardized form", especially in contexts where written use of that language was expect to develop further.

For a detailed description of these parameters, see [MSR-4]

For example, for the Latin script, the responsible panel investigated over 180 distinct orthographies.

For those scripts already published in [LGR-3], the supporting documentation cites at least one orthography supporting the inclusion for each code point.