L2/09-222

Source: Mark Davis
Date: June 11, 2009
Subject: Formalized invariants

This document provides a set of machine-readable data for testing Unicode invariants. These can applied in verifying that successive versions of Unicode do not violate those invariants, and can also be used by implementations for testing, if desired.

The proposal is to produce a proposed draft UAX (for post-5.2) that explains the structure and usage of this file.

Many of those invariants are stability constraints from the Unicode Stability Policies. Those are marked with "Stability" in the preceding comment. Others are property constraints established by other standards, such as the Regex properties alpha, alphanum, etc. [Note: these should also be marked with a special flag.] Others are "red flag" invariants, which are simply used to detect when a change in property might be problematic. Typically those have a set of exceptions (inclusions or exclusions) that are modified with each release. Note that some of the "red flag" invariants might be candidates for stability policies in the future.

FAQ

Q. how will invariant tests be maintained: When does a rule failure indicate a problem in properties versus a problem in the rule?

A. Failures in the tests marked with Stablity are true problems in the properties. Other failures need to be examined to see where the problem lies.

Q. How are the exception lists to be checked and maintained?

A. Currently the exception lists are maintained by the editorial committee. When a problem is detected in a beta version, a determination is made as to whether there is a problem in the rule or in the property. Rule issues are fixed; property issues are handled like others; for new characters the property value is adjusted in the beta; for characters from the previous version the issue is raised to the UTC.

Q. Why make this a UAX?

A. The fact that the exception lists are maintained and updated with each release implies that the invariant rules need to be updated in lockstep with each release of the Unicode Standard itself. That makes them a clear candidate for versioning in synch with the UCD, and hence the logical vehicle for documentation is a UAX. There would be two additional files in the UCD (perhaps in the auxiliary folder?): the data file of invariant tests (a plaintext file), and the generated html version of results for that version (such as document XYZ).

In essence, the guarantee would be that when a version of the UCD is released, it will go out with a set of invariant tests that are known to pass for *that* release, once the editorial committee gets done correcting it for the edge cases that may have failed starting from the invariant tests for the prior release.

#Unicode Invariant Results

Invariance Tests
#
This file provides a set of machine-readable invariance tests for Unicode Properties.
#

Format

Let <$variable> = <unicodeSet>
Assign a variable to a value. The variable must start with $.
#
<unicodeSet> is a boolean combinations of properties and character ranges, as defined in LDML,
with the following extensions.
#
Example:
[\p{General_Category=Unassigned}-[a-zA-Z]]
#
Property Name:
<propertyName> can be the short or long form as in the PropertyAliases.txt
<propertyName> can be prefixed with "U<version>:"
A version of -1 indicates that the property is the previous released version.
That is, if the version is 4.0.1, then the U-1 version is 4.0.0
#
Example: U5.1:Whitespace
#
Property Value:
If the propertyValue is missing, it is defaulted to true
If the value is of the form /.../, then the ... is interpreted as a regular expression
The value (if enumerated) can be the short or long form as in PropertyValueAliases.txt

---

Show <unicodeSet>
List any set on the console, for viewing and debugging.

Test <unicodeSet> <relation> <unicodeSet>
#
Tests that the relation is true for the two sets. The "Test" keyword is optional.
#
relation := '=' // has identical contents to
:= '⊃' // is proper superset of
:= '⊇' // is superset of
:= '⊂' // is proper subset of
:= '⊆' // is subset of
:= ' ∥ ' // has no intersection
:= '≠' // none of the above (they overlap, and neither contains the other)
#
When this file is parsed, a parse error message may contain <@>
to indicate the location of an error in the input line.
#
If there is an error in the test, a comparison listing of the two sides of the relation is generated.

In <unicodeSet> <props> (=|≠) <props>
#

For each character in <unicodeSet>, verify that the result of applying the left <props>
is (=|≠) the result of applying the right <props>.
<props> is of the form (<unicodeSet> | <prop>) ("*" (<unicodeSet> | <prop>))?
It is the functional composition of the properties applied to strings, whereby
<unicodeSet> is used to filter the result.
<prop> for a string property is applied to each character, and the result concatenated
That is, cf("A1") is cf("A")+cf("1") = "a1"
<prop> for an enumerated property, is applied to each character, and the result is a concatenated set.
That is, gc("A1") is gc("A")+gc("1") = "Uppercase_LetterDecimal_Number"
#
Example: for <props> of bc * \P{bc=NSM} * cf * dm, the result applied to Å (angstrom sign) are:
bc * \P{bc=NSM} * cf * dm ("Å")
bc * \P{bc=NSM} * cf ("A" + umlaut)
bc * \P{bc=NSM} ("a" + umlaut)
bc ("a")
"Left"
#
Example: In \p{dt=canonical} bc * \P{bc=NSM} * dm = bc * \P{bc=NSM}
This examines only those characters that have canonical compositions. For each such character X
it gets the decomposition mapping of X, then filters out all NSM characters, then gets the Bidi_Class.
It then tests that against the result of filtering out NSM characters from X, then getting the BIDI_Class.
#

General Constants

Let $gcAllPunctuation = \p{gc=/_Punctuation/}

0021..0023	#	Po	[3]	(!..#)	EXCLAMATION MARK..NUMBER SIGN
0025..0027	#	Po	[3]	(%..')	PERCENT SIGN..APOSTROPHE
0028	#	Ps		(()	LEFT PARENTHESIS
0029	#	Pe		())	RIGHT PARENTHESIS
002A	#	Po		(*)	ASTERISK
002C	#	Po		(,)	COMMA
002D	#	Pd		(-)	HYPHEN-MINUS
002E..002F	#	Po	[2]	(.../)	FULL STOP..SOLIDUS
003A..003B	#	Po	[2]	(:..;)	COLON..SEMICOLON
003F..0040	#	Po	[2]	(?..@)	QUESTION MARK..COMMERCIAL AT
005B	#	Ps		([)	LEFT SQUARE BRACKET
005C	#	Po		(\)	REVERSE SOLIDUS
005D	#	Pe		(])	RIGHT SQUARE BRACKET
005F	#	Pc		(_)	LOW LINE
007B	#	Ps		({)	LEFT CURLY BRACKET
007D	#	Pe		(})	RIGHT CURLY BRACKET
00A1	#	Po		(¡)	INVERTED EXCLAMATION MARK
00AB	#	Pi		(«)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00B7	#	Po		(·)	MIDDLE DOT
00BB	#	Pf		(»)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00BF	#	Po		(¿)	INVERTED QUESTION MARK
037E	#	Po		(;)	GREEK QUESTION MARK

0387	#	Po		(·)	GREEK ANO TELEIA
055A..055F	#	Po	[6]	(՚..՟)	ARMENIAN APOSTROPHE..ARMENIAN ABBREVIATION MARK
0589	#	Po		(։)	ARMENIAN FULL STOP
058A	#	Pd		(֊)	ARMENIAN HYPHEN
05BE	#	Pd		(־)	HEBREW PUNCTUATION MAQAF
05C0	#	Po		(׀)	HEBREW PUNCTUATION PASEQ

Total: 40 ...(omitting 543 from listing)...

Let $gcAllSymbols = \p{gc=/_Symbol/}

0024	#	Sc		($)	DOLLAR SIGN
002B	#	Sm		(+)	PLUS SIGN
003C..003E	#	Sm	[3]	(<..>)	LESS-THAN SIGN..GREATER-THAN SIGN
005E	#	Sk		(^)	CIRCUMFLEX ACCENT
0060	#	Sk		(`)	GRAVE ACCENT
007C	#	Sm		(|)	VERTICAL LINE
007E	#	Sm		(~)	TILDE
00A2..00A5	#	Sc	[4]	(¢..¥)	CENT SIGN..YEN SIGN
00A6..00A7	#	So	[2]	(¦..§)	BROKEN BAR..SECTION SIGN
00A8	#	Sk		(¨)	DIAERESIS
00A9	#	So		(©)	COPYRIGHT SIGN
00AC	#	Sm		(¬)	NOT SIGN
00AE	#	So		(®)	REGISTERED SIGN
00AF	#	Sk		(¯)	MACRON
00B0	#	So		(°)	DEGREE SIGN
00B1	#	Sm		(±)	PLUS-MINUS SIGN
00B4	#	Sk		(´)	ACUTE ACCENT
00B6	#	So		(¶)	PILCROW SIGN
00B8	#	Sk		(¸)	CEDILLA
00D7	#	Sm		(×)	MULTIPLICATION SIGN
00F7	#	Sm		(÷)	DIVISION SIGN
02C2..02C5	#	Sk	[4]	(˂..˅)	MODIFIER LETTER LEFT ARROWHEAD..MODIFIER LETTER DOWN ARROWHEAD
02D2..02DF	#	Sk	[14]	(˒..□)	MODIFIER LETTER CENTRED RIGHT HALF RING..MODIFIER LETTER CROSS ACCENT
02E5..02EB	#	Sk	[7]	(˥..□)	MODIFIER LETTER EXTRA-HIGH TONE BAR..MODIFIER LETTER YANG DEPARTING TONE MARK
02ED	#	Sk		(□)	MODIFIER LETTER UNASPIRATED
02EF..02FF	#	Sk	[17]	(□..□)	MODIFIER LETTER LOW DOWN ARROWHEAD..MODIFIER LETTER LOW LEFT ARROW

Total: 70 ...(omitting 4434 from listing)...

Let $gcAllMarks = \p{gc=/_Mark/}

0300..036F	#	Mn	[112]	(◌̀..◌ͯ)	COMBINING GRAVE ACCENT..COMBINING LATIN SMALL LETTER X
0483..0487	#	Mn	[5]	(◌҃..◌҇)	COMBINING CYRILLIC TITLO..COMBINING CYRILLIC POKRYTIE
0488..0489	#	Me	[2]	(◌҈..◌҉)	COMBINING CYRILLIC HUNDRED THOUSANDS SIGN..COMBINING CYRILLIC MILLIONS SIGN
0591..05BD	#	Mn	[45]	(◌֑..◌ֽ)	HEBREW ACCENT ETNAHTA..HEBREW POINT METEG
05BF	#	Mn		(◌ֿ)	HEBREW POINT RAFE
05C1..05C2	#	Mn	[2]	(◌ׁ..◌ׂ)	HEBREW POINT SHIN DOT..HEBREW POINT SIN DOT
05C4..05C5	#	Mn	[2]	(◌ׄ..◌ׅ)	HEBREW MARK UPPER DOT..HEBREW MARK LOWER DOT
05C7	#	Mn		(◌ׇ)	HEBREW POINT QAMATS QATAN
0610..061A	#	Mn	[11]	(◌؀..◌؊)	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM..ARABIC SMALL KASRA
064B..065E	#	Mn	[20]	(◌ً..◌ٞ)	ARABIC FATHATAN..ARABIC FATHA WITH TWO DOTS
0670	#	Mn		(◌ٰ)	ARABIC LETTER SUPERSCRIPT ALEF
06D6..06DC	#	Mn	[7]	(◌ۖ..◌ۜ)	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA..ARABIC SMALL HIGH SEEN
06DE	#	Me		(◌۞)	ARABIC START OF RUB EL HIZB
06DF..06E4	#	Mn	[6]	(◌۟..◌۠)	ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL HIGH MADDA
06E7..06E8	#	Mn	[2]	(◌ۧ..◌ۨ)	ARABIC SMALL HIGH YEH..ARABIC SMALL HIGH NOON
06EA..06ED	#	Mn	[4]	(◌۪..◌ۭ)	ARABIC EMPTY CENTRE LOW STOP..ARABIC SMALL LOW MEEM
0711	#	Mn		(◌ܑ)	SYRIAC LETTER SUPERSCRIPT ALAPH
0730..074A	#	Mn	[27]	(◌ܰ..◌݊)	SYRIAC PTHAHA ABOVE..SYRIAC BARREKH
07A6..07B0	#	Mn	[11]	(◌ަ..◌ް)	THAANA ABAFILI..THAANA SUKUN
07EB..07F3	#	Mn	[9]	(◌߫..◌߳)	NKO COMBINING SHORT HIGH TONE..NKO COMBINING DOUBLE DOT ABOVE
0816..0819	#	Mn	[4]	(◌ࠖ..◌࠙)	SAMARITAN MARK IN..SAMARITAN MARK DAGESH
081B..0823	#	Mn	[9]	(◌ࠛ..◌ࠣ)	SAMARITAN MARK EPENTHETIC YUT..SAMARITAN VOWEL SIGN A

Total: 283 ...(omitting 1168 from listing)...

Main Stability Policies
http://www.unicode.org/policies/property_value_stability_table.html

---

TODO: Formal Name Alias Stability, Named Character Sequence Stability, Name Uniqueness,
TODO: Identity Stability, Property Stability, Alias Stability, Property Alias Uniqueness

Encoding Stability: Once a character is encoded, it will not be moved or removed.
\p{GC=unassigned} ⊆ \p{U-1:GC=unassigned}

Name Stability: The Unicode Name property value for any non-reserved code point will not be changed. In particular, once a character is encoded, its name will not be changed.
In \P{U-1:GC=Cn} name=U-1:name

Formal Name Alias Stability
TODO

Named Character Sequence Stability
TODO

Name Uniqueness
TODO

Strong Normalization Stability (decomposition mapping, Canonical Combining Class don't change)
In Property Section

Identity Stability
Can't be tested

Property Stability: Normative and informative properties, once defined in the Unicode Character Database, will never be removed.
TODO

Alias Stability: Property aliases and property value aliases, once defined in the Unicode Character Database, will never be removed.
TODO

Property Alias Uniqueness: All property aliases constitute a single namespace. Property aliases are guaranteed to be unique within this namespace. For each property, all of its property value aliases constitute a separate namespace, one per property. Within each of these property value alias namespaces, property value aliases are guaranteed to be unique.
TODO

Identifier Stability: All strings that are valid default Unicode identifiers will continue to be valid default Unicode identifiers in all subsequent versions of Unicode. Furthermore, default identifiers never contain characters with the Pattern_Syntax or Pattern_White_Space properties.
Covered in Property Stability Section

Case Folding Stability: Caseless matching of Unicode strings used for identifiers is stable.
TODO

Case Pair Stability: If two characters form a case pair in a version of Unicode, they will remain a case pair in each subsequent version of Unicode. If two characters do not form a case pair in a version of Unicode, they will never become a case pair in any subsequent version of Unicode.
TODO

Property Stability Policies
http://www.unicode.org/policies/property_value_stability_table.html

---

BIDI

Stability: The Bidi_Class property values will not be further subdivided.
\p{bc=/^(AL|AN|B|BN|CS|EN|ES|ET|L|LRE|LRO|NSM|ON|PDF|R|RLE|RLO|S|WS)$/} = [\u0000-

\U0010FFFF]

Stability: The property values for the bidirectional properties Bidi_Class and Bidi_Mirrored preserve canonical equivalence.
This test utilizes the fact that bc=NSM inherit behavior in the algorithm, so these are just filtered
In \p{dt=canonical} bc * \P{bc=NSM} * dm = bc * \P{bc=NSM}

Stability: The property values for the bidirectional properties Bidi_Class and Bidi_Mirrored preserve canonical equivalence.
This test utilizes the fact that bc=NSM inherit behavior in the algorithm, so these are just filtered
There are 5 special cases:
Let $BMExclusions =[≠ ∤ ∦ ≢ \u2ADC]

2224	#	Sm	(∤)	DOES NOT DIVIDE
2226	#	Sm	(∦)	NOT PARALLEL TO
2260	#	Sm	(≠)	NOT EQUAL TO
2262	#	Sm	(≢)	NOT IDENTICAL TO
2ADC	#	Sm	(□)	FORKING

Total: 5

In [\p{dt=canonical}-$BMExclusions] Bidi_M * \P{bc=NSM} * dm = Bidi_M * \P{bc=NSM}

Additional BIDI invariant constants
Let $R_blocks = [\u0590-\u05FF \u07C0-\u08FF \uFB1D-\uFB4F \U00010800-\U00010FFF \U0001E800-\U0001EFFF]

0590	#	Cn		(□)	
0591..05BD	#	Mn	[45]	(□..ֽ)	HEBREW ACCENT ETNAHTA..HEBREW POINT METEG
05BE	#	Pd		(־)	HEBREW PUNCTUATION MAQAF
05BF	#	Mn		(ֿ)	HEBREW POINT RAFE
05C0	#	Po		(׀)	HEBREW PUNCTUATION PASEQ
05C1..05C2	#	Mn	[2]	(ׁ.. ׂ)	HEBREW POINT SHIN DOT..HEBREW POINT SIN DOT
05C3	#	Po		(׃)	HEBREW PUNCTUATION SOF PASUQ
05C4..05C5	#	Mn	[2]	(□..□)	HEBREW MARK UPPER DOT..HEBREW MARK LOWER DOT
05C6	#	Po		(□)	HEBREW PUNCTUATION NUN HAFUKHA
05C7	#	Mn		(□)	HEBREW POINT QAMATS QATAN
05C8..05CF	#	Cn	[8]	(□..□)	..
05D0..05EA	#	Lo	[27]	(א..ת)	HEBREW LETTER ALEF..HEBREW LETTER TAV
05EB..05EF	#	Cn	[5]	(□..□)	..
05F0..05F2	#	Lo	[3]	(ױ..ײ)	HEBREW LIGATURE YIDDISH DOUBLE VAV..HEBREW LIGATURE YIDDISH DOUBLE YOD
05F3..05F4	#	Po	[2]	(׳..״)	HEBREW PUNCTUATION GERESH..HEBREW PUNCTUATION GERSHAYIM
05F5..05FF	#	Cn	[11]	(□..□)	..
07C0..07C9	#	Nd	[10]	(□..□)	NKO DIGIT ZERO..NKO DIGIT NINE
07CA..07EA	#	Lo	[33]	(□..□)	NKO LETTER A..NKO LETTER JONA RA
07EB..07F3	#	Mn	[9]	(□..□)	NKO COMBINING SHORT HIGH TONE..NKO COMBINING DOUBLE DOT ABOVE
07F4..07F5	#	Lm	[2]	(□..□)	NKO HIGH TONE APOSTROPHE..NKO LOW TONE

					APOSTROPHE
07F6	#	So		(□)	NKO SYMBOL OO DENNEN
07F7..07F9	#	Po	[3]	(□..□)	NKO SYMBOL GBAKURUNEN..NKO EXCLAMATION MARK
07FA	#	Lm		(□)	NKO LAJANYALAN
07FB..07FF	#	Cn	[5]	(□..□)	..
0800..0815	#	Lo	[22]	(□..□)	SAMARITAN LETTER ALAF..SAMARITAN LETTER TAAF
0816..0819	#	Mn	[4]	(□..□)	SAMARITAN MARK IN..SAMARITAN MARK DAGESH
081A	#	Lm		(□)	SAMARITAN MODIFIER LETTER EPENTHETIC YUT
081B..0823	#	Mn	[9]	(□..□)	SAMARITAN MARK EPENTHETIC YUT..SAMARITAN VOWEL SIGN A
0824	#	Lm		(□)	SAMARITAN MODIFIER LETTER SHORT A
0825..0827	#	Mn	[3]	(□..□)	SAMARITAN VOWEL SIGN SHORT A..SAMARITAN VOWEL SIGN U
0828	#	Lm		(□)	SAMARITAN MODIFIER LETTER I
0829..082D	#	Mn	[5]	(□..□)	SAMARITAN VOWEL SIGN LONG I..SAMARITAN MARK NEQUDAA
082E..082F	#	Cn	[2]	(□..□)	..
0830..083E	#	Po	[15]	(□..□)	SAMARITAN PUNCTUATION NEQUDAA..SAMARITAN PUNCTUATION ANNAAU
083F..08FF	#	Cn	[193]	(□..□)	..
FB1D	#	Lo		(יִ)	HEBREW LETTER YOD WITH HIRIQ
FB1E	#	Mn		(ֽ)	HEBREW POINT JUDEO-SPANISH VARIKA
FB1F..FB28	#	Lo	[10]	(□..תּ)	HEBREW LIGATURE YIDDISH YOD YOD PATAH..HEBREW LETTER WIDE TAV
FB29	#	Sm		(□)	HEBREW LETTER ALTERNATIVE PLUS SIGN
FB2A..FB36	#	Lo	[13]	(שׁ..זּ)	HEBREW LETTER SHIN WITH SHIN DOT..HEBREW LETTER ZAYIN WITH DAGESH
FB37	#	Cn		(□)	
FB38..FB3C	#	Lo	[5]	(טּ..לּ)	HEBREW LETTER TET WITH DAGESH..HEBREW LETTER LAMED WITH DAGESH
FB3D	#	Cn		(□)	
FB3E	#	Lo		(מּ)	HEBREW LETTER MEM WITH DAGESH
FB3F	#	Cn		(□)	
FB40..FB41	#	Lo	[2]	(נּ..סּ)	HEBREW LETTER NUN WITH DAGESH..HEBREW LETTER SAMEKH WITH DAGESH
FB42	#	Cn		(□)	
FB43..FB44	#	Lo	[2]	(פּ..פּ)	HEBREW LETTER FINAL PE WITH DAGESH..HEBREW LETTER PE WITH DAGESH
FB45	#	Cn		(□)	
FB46..FB4F	#	Lo	[10]	(צּ..אל)	HEBREW LETTER TSADI WITH DAGESH..HEBREW LIGATURE ALEF LAMED
10800..10805	#	Lo	[6]	(□..□)	CYPRIOT SYLLABLE A..CYPRIOT SYLLABLE JA
10806..10807	#	Cn	[2]	(□..□)	..
10808	#	Lo		(□)	CYPRIOT SYLLABLE JO
10809	#	Cn		(□)	
1080A..10835	#	Lo	[44]	(□..□)	CYPRIOT SYLLABLE KA..CYPRIOT SYLLABLE WO

10836	#	Cn		(□)	
10837..10838	#	Lo	[2]	(□..□)	CYPRIOT SYLLABLE XA..CYPRIOT SYLLABLE XE
10839..1083B	#	Cn	[3]	(□..□)	..
1083C	#	Lo		(□)	CYPRIOT SYLLABLE ZA
1083D..1083E	#	Cn	[2]	(□..□)	..
1083F..10855	#	Lo	[23]	(□..□)	CYPRIOT SYLLABLE ZO..IMPERIAL ARAMAIC LETTER TAW
10856	#	Cn		(□)	
10857	#	Po		(□)	IMPERIAL ARAMAIC SECTION SIGN
10858..1085F	#	No	[8]	(□..□)	IMPERIAL ARAMAIC NUMBER ONE..IMPERIAL ARAMAIC NUMBER TEN THOUSAND
10860..108FF	#	Cn	[160]	(□..□)	..
10900..10915	#	Lo	[22]	(□..□)	PHOENICIAN LETTER ALF..PHOENICIAN LETTER TAU
10916..1091B	#	No	[6]	(□..□)	PHOENICIAN NUMBER ONE..PHOENICIAN NUMBER THREE
1091C..1091E	#	Cn	[3]	(□..□)	..
1091F	#	Po		(□)	PHOENICIAN WORD SEPARATOR
10920..10939	#	Lo	[26]	(□..□)	LYDIAN LETTER A..LYDIAN LETTER C
1093A..1093E	#	Cn	[5]	(□..□)	..
1093F	#	Po		(□)	LYDIAN TRIANGULAR MARK
10940..109FF	#	Cn	[192]	(□..□)	..
10A00	#	Lo		(□)	KHAROSHTHI LETTER A
10A01..10A03	#	Mn	[3]	(□..□)	KHAROSHTHI VOWEL SIGN I..KHAROSHTHI VOWEL SIGN VOCALIC R
10A04	#	Cn		(□)	
10A05..10A06	#	Mn	[2]	(□..□)	KHAROSHTHI VOWEL SIGN E..KHAROSHTHI VOWEL SIGN O
10A07..10A0B	#	Cn	[5]	(□..□)	..
10A0C..10A0F	#	Mn	[4]	(□..□)	KHAROSHTHI VOWEL LENGTH MARK..KHAROSHTHI SIGN VISARGA
10A10..10A13	#	Lo	[4]	(□..□)	KHAROSHTHI LETTER KA..KHAROSHTHI LETTER GHA
10A14	#	Cn		(□)	
10A15..10A17	#	Lo	[3]	(□..□)	KHAROSHTHI LETTER CA..KHAROSHTHI LETTER JA
10A18	#	Cn		(□)	
10A19..10A33	#	Lo	[27]	(□..□)	KHAROSHTHI LETTER NYA..KHAROSHTHI LETTER TTTHA
10A34..10A37	#	Cn	[4]	(□..□)	..
10A38..10A3A	#	Mn	[3]	(□..□)	KHAROSHTHI SIGN BAR ABOVE..KHAROSHTHI SIGN DOT BELOW
10A3B..10A3E	#	Cn	[4]	(□..□)	..
10A3F	#	Mn		(□)	KHAROSHTHI VIRAMA
10A40..10A47	#	No	[8]	(□..□)	KHAROSHTHI DIGIT ONE..KHAROSHTHI NUMBER ONE THOUSAND
10A48..10A4F	#	Cn	[8]	(□..□)	..
10A50..10A58	#	Po	[9]	(□..□)	KHAROSHTHI PUNCTUATION DOT..KHAROSHTHI PUNCTUATION LINES
10A59..10A5F	#	Cn	[7]	(□..□)	..
10A60..10A7C	#	Lo	[29]	(□..□)	OLD SOUTH ARABIAN LETTER HE..OLD SOUTH ARABIAN LETTER THETH
10A7D..10A7E	#	No	[2]	(□..□)	OLD SOUTH ARABIAN NUMBER ONE..OLD SOUTH ARABIAN

					NUMBER FIFTY
10A7F	#	Po		(□)	OLD SOUTH ARABIAN NUMERIC INDICATOR
10A80..10AFF	#	Cn	[128]	(□..□)	..
10B00..10B35	#	Lo	[54]	(□..□)	AVESTAN LETTER A..AVESTAN LETTER HE
10B36..10B38	#	Cn	[3]	(□..□)	..
10B39..10B3F	#	Po	[7]	(□..□)	AVESTAN ABBREVIATION MARK..LARGE ONE RING OVER TWO RINGS PUNCTUATION
10B40..10B55	#	Lo	[22]	(□..□)	INSCRIPTIONAL PARTHIAN LETTER ALEPH..INSCRIPTIONAL PARTHIAN LETTER TAW
10B56..10B57	#	Cn	[2]	(□..□)	..
10B58..10B5F	#	No	[8]	(□..□)	INSCRIPTIONAL PARTHIAN NUMBER ONE..INSCRIPTIONAL PARTHIAN NUMBER ONE THOUSAND
10B60..10B72	#	Lo	[19]	(□..□)	INSCRIPTIONAL PAHLAVI LETTER ALEPH..INSCRIPTIONAL PAHLAVI LETTER TAW
10B73..10B77	#	Cn	[5]	(□..□)	..
10B78..10B7F	#	No	[8]	(□..□)	INSCRIPTIONAL PAHLAVI NUMBER ONE..INSCRIPTIONAL PAHLAVI NUMBER ONE THOUSAND
10B80..10BFF	#	Cn	[128]	(□..□)	..
10C00..10C48	#	Lo	[73]	(□..□)	OLD TURKIC LETTER ORKHON A..OLD TURKIC LETTER ORKHON BASH
10C49..10E5F	#	Cn	[535]	(□..□)	..
10E60..10E7E	#	No	[31]	(□..□)	RUMI DIGIT ONE..RUMI FRACTION TWO THIRDS
10E7F..10FFF	#	Cn	[385]	(□..□)	..
1E800..1EFFF	#	Cn	[2048]	(□..□)	..

Total: 4579

Let $AL_blocks = [\u0600-\u07BF \uFB50-\uFDCF \uFDF0-\uFDFF \uFE70-\uFEFF]

0600..0603	#	Cf	[4]	(□..□)	ARABIC NUMBER SIGN..ARABIC SIGN SAFHA
0604..0605	#	Cn	[2]	(□..□)	..
0606..0608	#	Sm	[3]	(□..□)	ARABIC-INDIC CUBE ROOT..ARABIC RAY
0609..060A	#	Po	[2]	(□..□)	ARABIC-INDIC PER MILLE SIGN..ARABIC-INDIC PER TEN THOUSAND SIGN
060B	#	Sc		(□)	AFGHANI SIGN
060C..060D	#	Po	[2]	(،..□)	ARABIC COMMA..ARABIC DATE SEPARATOR
060E..060F	#	So	[2]	(□..□)	ARABIC POETIC VERSE SIGN..ARABIC SIGN MISRA
0610..061A	#	Mn	[11]	(□..□)	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM..ARABIC SMALL KASRA
061B	#	Po		(؛)	ARABIC SEMICOLON
061C..061D	#	Cn	[2]	(□..□)	..
061E..061F	#	Po	[2]	(؞..□)	ARABIC TRIPLE DOT PUNCTUATION MARK..ARABIC QUESTION MARK
0620	#	Cn		(□)	
0621..063F	#	Lo	[31]	(□..ؿ)	ARABIC LETTER HAMZA..ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE
0640	#	Lm		(-)	ARABIC TATWEEL

0641..064A	#	Lo	[10]	(ف..ي)	ARABIC LETTER FEH..ARABIC LETTER YEH
064B..065E	#	Mn	[20]	(ً..□)	ARABIC FATHATAN..ARABIC FATHA WITH TWO DOTS
065F	#	Cn		(□)	
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
066A..066D	#	Po	[4]	(٪..*)	ARABIC PERCENT SIGN..ARABIC FIVE POINTED STAR
066E..066F	#	Lo	[2]	(ب..ق)	ARABIC LETTER DOTLESS BEH..ARABIC LETTER DOTLESS QAF
0670	#	Mn		(ٰ)	ARABIC LETTER SUPERSCRIPT ALEF
0671..06D3	#	Lo	[99]	(ٱ..ے)	ARABIC LETTER ALEF WASLA..ARABIC LETTER YEH BARREE WITH HAMZA ABOVE
06D4	#	Po		(۔)	ARABIC FULL STOP
06D5	#	Lo		(ە)	ARABIC LETTER AE
06D6..06DC	#	Mn	[7]	(ۖ..ۜ)	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA..ARABIC SMALL HIGH SEEN
06DD	#	Cf		(۝)	ARABIC END OF AYAH
06DE	#	Me		(۞)	ARABIC START OF RUB EL HIZB
06DF..06E4	#	Mn	[6]	(۟..ۤ)	ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL HIGH MADDA
06E5..06E6	#	Lm	[2]	(ۥ..ۦ)	ARABIC SMALL WAW..ARABIC SMALL YEH
06E7..06E8	#	Mn	[2]	(ۧ..ۨ)	ARABIC SMALL HIGH YEH..ARABIC SMALL HIGH NOON
06E9	#	So		(۩)	ARABIC PLACE OF SAJDAH
06EA..06ED	#	Mn	[4]	(ۭ..ۭ)	ARABIC EMPTY CENTRE LOW STOP..ARABIC SMALL LOW MEEM
06EE..06EF	#	Lo	[2]	(□..□)	ARABIC LETTER DAL WITH INVERTED V..ARABIC LETTER REH WITH INVERTED V
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
06FA..06FC	#	Lo	[3]	(ﺶ..ﻍ)	ARABIC LETTER SHEEN WITH DOT BELOW..ARABIC LETTER GHAIN WITH DOT BELOW
06FD..06FE	#	So	[2]	(ﷺ..ﷻ)	ARABIC SIGN SINDHI AMPERSAND..ARABIC SIGN SINDHI POSTPOSITION MEN
06FF	#	Lo		(□)	ARABIC LETTER HEH WITH INVERTED V
0700..070D	#	Po	[14]	(܀..܍)	SYRIAC END OF PARAGRAPH..SYRIAC HARKLEAN ASTERISCUS
070E	#	Cn		(□)	
070F	#	Cf		(܏)	SYRIAC ABBREVIATION MARK
0710	#	Lo		(ܐ)	SYRIAC LETTER ALAPH
0711	#	Mn		(◌ܑ)	SYRIAC LETTER SUPERSCRIPT ALAPH
0712..072F	#	Lo	[30]	(ܒ..ܯ)	SYRIAC LETTER BETH..SYRIAC LETTER PERSIAN DHALATH
0730..074A	#	Mn	[27]	(◌ܰ..◌ܺ)	SYRIAC PTHAHA ABOVE..SYRIAC BARREKH
074B..074C	#	Cn	[2]	(□..□)	..
074D..07A5	#	Lo	[89]	(ܭ..ޥ)	SYRIAC LETTER SOGDIAN ZHAIN..THAANA LETTER WAAVU
07A6..07B0	#	Mn	[11]	(ަ..ް)	THAANA ABAFILI..THAANA SUKUN
07B1	#	Lo		(ޱ)	THAANA LETTER NAA
07B2..07BF	#	Cn	[14]	(□..□)	..
FB50..FBB1	#	Lo	[98]	(ﭐ..ﮮ)	ARABIC LETTER ALEF WASLA ISOLATED FORM..ARABIC LETTER YEH BARREE WITH HAMZA ABOVE FINAL FORM

FBB2..FBD2	#	Cn	[33]	(□..□)	..
FBD3..FD3D	#	Lo	[363]	(□..ﴽ)	ARABIC LETTER NG ISOLATED FORM..ARABIC LIGATURE ALEF WITH FATHATAN ISOLATED FORM
FD3E	#	Ps		(()	ORNATE LEFT PARENTHESIS
FD3F	#	Pe		())	ORNATE RIGHT PARENTHESIS
FD40..FD4F	#	Cn	[16]	(□..□)	..
FD50..FD8F	#	Lo	[64]	(□..□)	ARABIC LIGATURE TEH WITH JEEM WITH MEEM INITIAL FORM..ARABIC LIGATURE MEEM WITH KHAH WITH MEEM INITIAL FORM
FD90..FD91	#	Cn	[2]	(□..□)	..
FD92..FDC7	#	Lo	[54]	(□..□)	ARABIC LIGATURE MEEM WITH JEEM WITH KHAH INITIAL FORM..ARABIC LIGATURE NOON WITH JEEM WITH YEH FINAL FORM
FDC8..FDCF	#	Cn	[8]	(□..□)	..
FDF0..FDFB	#	Lo	[12]	(□..□)	ARABIC LIGATURE SALLA USED AS KORANIC STOP SIGN ISOLATED FORM..ARABIC LIGATURE JALLAJALALOUHOU
FDFC	#	Sc		(□)	RIAL SIGN
FDFD	#	So		(□)	ARABIC LIGATURE BISMILLAH AR-RAHMAN AR-RAHEEM
FDFE..FDFF	#	Cn	[2]	(□..□)	..
FE70..FE74	#	Lo	[5]	(□..□)	ARABIC FATHATAN ISOLATED FORM..ARABIC KASRATAN ISOLATED FORM
FE75	#	Cn		(□)	
FE76..FEFC	#	Lo	[135]	(ﻵ..□)	ARABIC FATHA ISOLATED FORM..ARABIC LIGATURE LAM WITH ALEF FINAL FORM
FEFD..FEFE	#	Cn	[2]	(□..□)	..
FEFF	#	Cf		()	ZERO WIDTH NO-BREAK SPACE

Total: 1248

Unassigned characters in these blocks have R or AL respectively
\p{Bidi_Class=R} ⊇ [$R_blocks & \p{gc=Cn}]
\p{Bidi_Class=AL} ⊇ [$AL_blocks & \p{gc=Cn}]

There are no strong characters of the other directionalities (out of L, AL, R) in these blocks,
and anything R or L is in the block (or RLM)
$R_blocks ‖ [\p{Bidi_Class=L} \p{Bidi_Class=AL}]
$AL_blocks ‖ [\p{Bidi_Class=L} \p{Bidi_Class=R}]
[$R_blocks $AL_blocks \N{RIGHT-TO-LEFT MARK}] ⊇ [\p{Bidi_Class=AL} \p{Bidi_Class=R}] #200f

---

Case

Stability: The Case_Folding property value is limited so that no string when case folded expands to more than 3× in length (measured in code units).
TODO

Stability: All characters with the Lowercase property and all characters with the Uppercase property have the Alphabetic property.

\p{Alphabetic} ⊃ [\p{Uppercase} \p{Lowercase}]

General

Stability: The General_Category property values will not be further subdivided.
\p{gc=/^(Cc|Cf|Cn|Co|Cs|Ll|Lm|Lo|Lt|Lu|Mc|Me|Mn|Nd|Nl|No|Pc|Pd|Pe|Pf|Pi|Po|Ps|Sc|Sk|Sm|So|Zl|Zp|Zs)$/} = [\u0000-\U0010FFFF]

Stability: The General_Category property value Control (Cc) is immutable: the set of code points with that value will never change.
\p{GC=Cc} = \p{U-1:GC=Cc}

Stability: The General_Category property value Private_Use (Co) is immutable: the set of code points with that value will never change.
\p{GC=Co} = \p{U-1:GC=Co}

Stability: The General_Category property value Surrogate (Cs) is immutable: the set of code points with that value will never change.
\p{GC=Cs} = \p{U-1:GC=Cs}

Stability: The set of characters having General_Category=Nd will always be the same as the set of characters having Numeric_Type=de.
\p{General_Category=Decimal_Number} = \p{Numeric_Type=Decimal}

Stability: Once a character is assigned, both its Name and its Jamo_Short_Name will never change.
Name is covered in Main policies
TODO: Short Name

Stability: The Noncharacter_Code_Point property is an immutable code point property, which means that its property values for all Unicode code points will never change.
\p{NChar} = \p{U-1:NChar}

Identifier Stability

Stability: Once a character is ID_Continue, it must continue to be so in all future versions.
\p{ID_Continue} ⊇ \p{U-1:ID_Continue}

Stability: If a character is ID_Start then it must also be ID_Continue.
\p{ID_Continue} ⊇ \p{ID_Start}

Stability: Once a character is ID_Start, it must continue to be so in all future versions.
\p{ID_Start} ⊇ \p{U-1:ID_Start}

Stability: Once a character is XID_Continue, it must continue to be so in all future versions.
\p{XID_Continue} ⊇ \p{U-1:XID_Continue}

Stability: If a character is XID_Start then it must also be XID_Continue.
\p{XID_Continue} ⊇ \p{XID_Start}

Stability: If a character is XID_Start then it must also be XID_Continue.

\p{XID_Start} ⊇ \p{U-1:XID_Start}

Stability: The Pattern_Syntax and Pattern_Whitespace properties are immutable code point properties, which means that their property values for all Unicode code points will never change.
\p{Pattern_Whitespace} = \p{U-1:Pattern_Whitespace}
\p{Pattern_Syntax} = \p{U-1:Pattern_Syntax}

Stability: If a character has the Pattern_Syntax or Pattern_White_Space property, then it cannot have the ID_Continue or XID_Continue property.
(Also tests that Pattern_Syntax is disjoint from Pattern_White_Space)

\p{ID_Continue} ∥ [\p{Pattern_Whitespace} \p{Pattern_Syntax}]
\p{Pattern_Whitespace} ∥ [\p{ID_Continue} \p{Pattern_Syntax}]
\p{Pattern_Syntax} ∥ [\p{ID_Continue} \p{Pattern_Whitespace}]

\p{XID_Continue} ∥ [\p{Pattern_Whitespace} \p{Pattern_Syntax}]
\p{Pattern_Whitespace} ∥ [\p{XID_Continue} \p{Pattern_Syntax}]
\p{Pattern_Syntax} ∥ [\p{XID_Continue} \p{Pattern_Whitespace}]

The X versions are subsets of the the plain versions
Should add as stability provision
\p{ID_Continue} ⊇ \p{XID_Continue}
\p{ID_Start} ⊇ \p{XID_Start}

Normalization

Stability: The Canonical_Combining_Class property values are limited to the values 0 to 255.
\p{CCC=/^([0-9][0-9]?|1[0-9][0-9]|2[0-4][0-9]|25[0-5])$/} = [\u0000-\U0010FFFF]

Stability: Once a character is assigned, its Canonical_Combining_Class will never change.
In \P{U-1:GC=Cn} ccc=U-1:ccc

Canonical decompositions (minus exclusions) must be identical across releases (also required by strong normalization stability)
Should be stability policy
[\p{Decomposition_Type=Canonical} - \p{Full_Composition_Exclusion}] = [\p{U-1:Decomposition_Type=Canonical} - \p{U-1:Full_Composition_Exclusion}]

Stability: All characters other than those with General_Category property values Spacing_Mark (Mc) and Nonspacing_Mark (Mn) have the Canonical_Combining_Class property value 0.
\p{CCC=0} ⊇ [^ \p{GC=Mc} \p{GC=Mn}]

Stability: Canonical and compatibility mappings (Decomposition_Mapping property values) are always in canonical order, and the resulting recursive decomposition will also be in canonical order.
TODO

Stability: Canonical mappings (Decomposition_Mapping property values) are always limited either to a single value or to a pair. The second character in the pair cannot itself have a canonical mapping.
TODO

Stability: Canonical mappings (Decomposition_Mapping property values) are always limited so that no string when normalized to NFC expands to more than 3× in length (measured in code units).

TODO

Stability: Once a character is assigned, its Decomposition_Mapping will never change.
In \P{U-1:GC=Cn} dm=U-1:dm

Other Invariant Tests, not in Stability Policies

Numbers

Decimals are 0-9

Let $decimalValue = [\p{Numeric_Value=/[0-9]+(.0)?/}]

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
00B2..00B3	#	No	[2]	(²..³)	SUPERSCRIPT TWO..SUPERSCRIPT THREE
00B9	#	No		(¹)	SUPERSCRIPT ONE
00BC..00BE	#	No	[3]	(¼..¾)	VULGAR FRACTION ONE QUARTER..VULGAR FRACTION THREE QUARTERS
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
07C0..07C9	#	Nd	[10]	(□..□)	NKO DIGIT ZERO..NKO DIGIT NINE
0966..096F	#	Nd	[10]	(०..९)	DEVANAGARI DIGIT ZERO..DEVANAGARI DIGIT NINE
09E6..09EF	#	Nd	[10]	(০..৯)	BENGALI DIGIT ZERO..BENGALI DIGIT NINE
09F4..09F9	#	No	[6]	(৴..৹)	BENGALI CURRENCY NUMERATOR ONE..BENGALI CURRENCY DENOMINATOR SIXTEEN
0A66..0A6F	#	Nd	[10]	(੦..੯)	GURMUKHI DIGIT ZERO..GURMUKHI DIGIT NINE
0AE6..0AEF	#	Nd	[10]	(૦..૯)	GUJARATI DIGIT ZERO..GUJARATI DIGIT NINE
0B66..0B6F	#	Nd	[10]	(□..□)	ORIYA DIGIT ZERO..ORIYA DIGIT NINE
0BE6..0BEF	#	Nd	[10]	(..௯)	TAMIL DIGIT ZERO..TAMIL DIGIT NINE
0BF0..0BF2	#	No	[3]	(൰..൲)	TAMIL NUMBER TEN..TAMIL NUMBER ONE THOUSAND
0C66..0C6F	#	Nd	[10]	(౦..౯)	TELUGU DIGIT ZERO..TELUGU DIGIT NINE
0C78..0C7E	#	No	[7]	(౸..౾)	TELUGU FRACTION DIGIT ZERO FOR ODD POWERS OF FOUR..TELUGU FRACTION DIGIT THREE FOR EVEN POWERS OF FOUR
0CE6..0CEF	#	Nd	[10]	(೦..೯)	KANNADA DIGIT ZERO..KANNADA DIGIT NINE
0D66..0D6F	#	Nd	[10]	(൦..൯)	MALAYALAM DIGIT ZERO..MALAYALAM DIGIT NINE
0D70..0D75	#	No	[6]	(൰..൵)	MALAYALAM NUMBER TEN..MALAYALAM FRACTION THREE QUARTERS
0E50..0E59	#	Nd	[10]	(๐..๙)	THAI DIGIT ZERO..THAI DIGIT NINE
0ED0..0ED9	#	Nd	[10]	(□..□)	LAO DIGIT ZERO..LAO DIGIT NINE

Total: 178 ...(omitting 963 from listing)...

$decimalValue ⊇ \p{General_Category=Decimal_Number}

All and only those items with numeric types have numeric values

Let $anyNumericValue = \p{Numeric_Value=/-?[0-9]+.[0-9]+/}

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
00B2..00B3	#	No	[2]	(²..³)	SUPERSCRIPT TWO..SUPERSCRIPT THREE
00B9	#	No		(¹)	SUPERSCRIPT ONE
00BC..00BE	#	No	[3]	(¼..¾)	VULGAR FRACTION ONE QUARTER..VULGAR FRACTION THREE QUARTERS
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
07C0..07C9	#	Nd	[10]	(□..□)	NKO DIGIT ZERO..NKO DIGIT NINE
0966..096F	#	Nd	[10]	(०..९)	DEVANAGARI DIGIT ZERO..DEVANAGARI DIGIT NINE
09E6..09EF	#	Nd	[10]	(০..৯)	BENGALI DIGIT ZERO..BENGALI DIGIT NINE
09F4..09F9	#	No	[6]	(৴..৹)	BENGALI CURRENCY NUMERATOR ONE..BENGALI CURRENCY DENOMINATOR SIXTEEN
0A66..0A6F	#	Nd	[10]	(੦..੯)	GURMUKHI DIGIT ZERO..GURMUKHI DIGIT NINE
0AE6..0AEF	#	Nd	[10]	(૦..૯)	GUJARATI DIGIT ZERO..GUJARATI DIGIT NINE
0B66..0B6F	#	Nd	[10]	(□..□)	ORIYA DIGIT ZERO..ORIYA DIGIT NINE
0BE6..0BEF	#	Nd	[10]	(..௯)	TAMIL DIGIT ZERO..TAMIL DIGIT NINE
0BF0..0BF2	#	No	[3]	(൘..൲)	TAMIL NUMBER TEN..TAMIL NUMBER ONE THOUSAND
0C66..0C6F	#	Nd	[10]	(౦..౯)	TELUGU DIGIT ZERO..TELUGU DIGIT NINE
0C78..0C7E	#	No	[7]	(□..□)	TELUGU FRACTION DIGIT ZERO FOR ODD POWERS OF FOUR..TELUGU FRACTION DIGIT THREE FOR EVEN POWERS OF FOUR
0CE6..0CEF	#	Nd	[10]	(೦..೯)	KANNADA DIGIT ZERO..KANNADA DIGIT NINE
0D66..0D6F	#	Nd	[10]	(൦..൯)	MALAYALAM DIGIT ZERO..MALAYALAM DIGIT NINE
0D70..0D75	#	No	[6]	(□..□)	MALAYALAM NUMBER TEN..MALAYALAM FRACTION THREE QUARTERS
0E50..0E59	#	Nd	[10]	(๐..๙)	THAI DIGIT ZERO..THAI DIGIT NINE
0ED0..0ED9	#	Nd	[10]	(□..□)	LAO DIGIT ZERO..LAO DIGIT NINE

Total: 178 ...(omitting 963 from listing)...

[\p{Numeric_Type=Decimal} \p{Numeric_Type=Digit} \p{Numeric_Type=Numeric}] = $anyNumericValue

---

Misc Properties

Musical symbol combining marks, other oddities

Let $AlphaExclusions = [\u0F3E\u0F3F\u1063\u1064\u1069-\u106D\u1087-\u108C\u108F\u109A\u109B\u1B44\u1BAA\u1CE1\uA953\uA9C0\uAA7B\uABEC\U0001D165\U0001D166\U0001D16D-\U0001D172]

0F3E..0F3F	#	Mc	[2]	(◌ཾ..◌ཿ)	TIBETAN SIGN YAR TSHES..TIBETAN SIGN MAR TSHES
1063..1064	#	Mc	[2]	(◌ ..	MYANMAR TONE MARK SGAW KAREN HATHI..MYANMAR

				)	TONE MARK SGAW KAREN KE PHO
1069..106D	#	Mc	[5]	(..)	MYANMAR SIGN WESTERN PWO KAREN TONE-1..MYANMAR SIGN WESTERN PWO KAREN TONE-5
1087..108C	#	Mc	[6]	(..)	MYANMAR SIGN SHAN TONE-2..MYANMAR SIGN SHAN COUNCIL TONE-3
108F	#	Mc		( )	MYANMAR SIGN RUMAI PALAUNG TONE-5
109A..109B	#	Mc	[2]	(□..□)	MYANMAR SIGN KHAMTI TONE-1..MYANMAR SIGN KHAMTI TONE-3
1B44	#	Mc		(□)	BALINESE ADEG ADEG
1BAA	#	Mc		(□)	SUNDANESE SIGN PAMAAEH
1CE1	#	Mc		(□)	VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA
A953	#	Mc		(□)	REJANG VIRAMA
A9C0	#	Mc		(□)	JAVANESE PANGKON
AA7B	#	Mc		(□)	MYANMAR SIGN PAO KAREN TONE
ABEC	#	Mc		(□)	MEETEI MAYEK LUM IYEK
1D165..1D166	#	Mc	[2]	(□..□)	MUSICAL SYMBOL COMBINING STEM..MUSICAL SYMBOL COMBINING SPRECHGESANG STEM
1D16D..1D172	#	Mc	[6]	(□..□)	MUSICAL SYMBOL COMBINING AUGMENTATION DOT..MUSICAL SYMBOL COMBINING FLAG-5

Total: 33

\p{Alphabetic} ⊇ [\p{GC=Lu} \p{GC=Ll} \p{GC=Lt} \p{GC=Lm} \p{GC=Lo} \p{GC=Nl} \p{GC=Mc} - $AlphaExclusions]

Show [\p{GC=Mc} - \p{alphabetic}]
Show [\p{GC=Mc} & \p{alphabetic}]

\p{Whitespace} ⊃ [\p{GC=Zs} \p{GC=Zp} \p{GC=Zl}]
\p{GC=Zs} ≠ \p{Name=/SPACE/}

\p{Dash} ⊃ [\p{GC=Pd}]

\p{Script=Common} ‖ [\p{GC=Mn} \p{GC=Me} \p{Join_Control}]
\p{Script=Inherited} ⊆ [\p{GC=Mn} \p{GC=Me} \p{Join_Control}]
\p{Script=Unknown} = [\p{GC=Cn} \p{GC=Co} \p{GC=Cs}]

[\p{Alphabetic}] ‖ \p{Script=Common}
& [\p{Decomposition_Type=None} \p{Decomposition_Type=Canonical}]

LineBreak property

Let $IDInclusions = [[:block=/Ideographs/:][\U00020000-\U0003FFFF] & [:gc=Cn:] - [:NChar:]]

4DB6..4DBF	#	Cn	[10]	(□..□)	..
9FCC..9FFF	#	Cn	[52]	(□..□)	..

FA2E..FA2F	#	Cn	[2]	(□..□)	..
FA6E..FA6F	#	Cn	[2]	(□..□)	..
FADA..FAFF	#	Cn	[38]	(□..□)	..
2A6D7..2A6FF	#	Cn	[41]	(□..□)	..
2B735..2F7FF	#	Cn	[16587]	(□..□)	..
2FA1E..2FFFD	#	Cn	[1504]	(□..□)	..
30000..3FFFD	#	Cn	[65534]	(□..□)	..

Total: 83770

\p{LB=ID} ⊃ $IDInclusions
\p{Line_Break=Unknown} = [\p{General_Category=Unassigned} \p{GeneralCategory=PrivateUse} - $IDInclusions]

Let $OPInclusions = [\u00A1\u00BF\u2E18\U00013258-\U0001325A\U00013286\U00013288\U00013379]

00A1	#	Po		(¡)	INVERTED EXCLAMATION MARK
00BF	#	Po		(¿)	INVERTED QUESTION MARK
2E18	#	Po		(□)	INVERTED INTERROBANG
13258..1325A	#	Lo	[3]	(□..□)	EGYPTIAN HIEROGLYPH O006A..EGYPTIAN HIEROGLYPH O006C
13286	#	Lo		(□)	EGYPTIAN HIEROGLYPH O036A
13288	#	Lo		(□)	EGYPTIAN HIEROGLYPH O036C
13379	#	Lo		(□)	EGYPTIAN HIEROGLYPH V011A

Total: 9

\p{LB=OP} = [\p{GC=Ps} $OPInclusions]
\p{LB=CL} ⊃ \p{GC=Pe}
\p{LB=CM} = [\p{GC=Mn} \p{GC=Me} \p{GC=Mc} \p{GC=Cc} \p{GC=Cf} -\p{LB=SA} -\p{LB=WJ} -\p{LB=ZW} -\p{LB=BA} -\p{LB=LF} -\p{LB=BK} -\p{LB=CR} -\p{LB=NL} -\p{LB=GL} -\p{LB=AL}]

Let $NUInclusions = [\u066B\u066C]

066B..066C	#	Po	[2]	(٫..٬)	ARABIC DECIMAL SEPARATOR..ARABIC THOUSANDS SEPARATOR

Total: 2

\p{LB=NU} = [\p{GC=Nd} $NUInclusions - \p{EA=F}]

Let $PRInclusions = [\u002b\u005c\u00b1\u2116\u2212\u2213]

002B	#	Sm		(+)	PLUS SIGN
005C	#	Po		(\)	REVERSE SOLIDUS
00B1	#	Sm		(±)	PLUS-MINUS SIGN
2116	#	So		(№)	NUMERO SIGN
2212..2213	#	Sm	[2]	(−..∓)	MINUS SIGN..MINUS-OR-PLUS SIGN

Total: 6

\p{LB=PR} = [\p{GC=Sc} $PRInclusions - \p{LB=PO}]

Let $QUInclusions = [\u0022 \u0027 \u275B-\u275E \u2E00-\u2E01 \u2E06-\u2E08 \u2E0B]

0022	#	Po		(")	QUOTATION MARK
0027	#	Po		(')	APOSTROPHE
275B..275E	#	So	[4]	(❛ .. ❞)	HEAVY SINGLE TURNED COMMA QUOTATION MARK ORNAMENT..HEAVY DOUBLE COMMA QUOTATION MARK ORNAMENT
2E00..2E01	#	Po	[2]	(□..□)	RIGHT ANGLE SUBSTITUTION MARKER..RIGHT ANGLE DOTTED SUBSTITUTION MARKER
2E06..2E08	#	Po	[3]	(□..□)	RAISED INTERPOLATION MARKER..DOTTED TRANSPOSITION MARKER
2E0B	#	Po		(□)	RAISED SQUARE

Total: 12

\p{LB=QU} = [\p{GC=Pf} \p{GC=Pi} $QUInclusions]
\p{LB=SG} = \p{GC=Cs}
\p{LB=SP} = \N{SPACE}
\p{LB=SY} = \N{SOLIDUS}
\p{LB=WJ} = [\N{WORD JOINER} \N{ZERO WIDTH NO-BREAK SPACE}]
\p{LB=ZW} = \N{ZERO WIDTH SPACE}

SA are limited to certain scripts:
Let $SAScripts = [\p{script=thai} \p{script=lao} \p{script=myanmar} \p{script=khmer} \p{script=Tai_Le} \p{script=New_Tai_Lue} \p{script=Tai_Tham} \p{script=Tai_Viet}]

0E01..0E30	#	Lo	[48]	(ก..ะ)	THAI CHARACTER KO KAI..THAI CHARACTER SARA A
0E31	#	Mn		(◌ั)	THAI CHARACTER MAI HAN-AKAT
0E32..0E33	#	Lo	[2]	(า..ำ)	THAI CHARACTER SARA AA..THAI CHARACTER SARA AM
0E34..0E3A	#	Mn	[7]	(◌ิ..◌ฺ)	THAI CHARACTER SARA I..THAI CHARACTER PHINTHU
0E40..0E45	#	Lo	[6]	(เ..ๅ)	THAI CHARACTER SARA E..THAI CHARACTER LAKKHANGYAO
0E46	#	Lm		(ๆ)	THAI CHARACTER MAIYAMOK
0E47..0E4E	#	Mn	[8]	(◌็..◌๎)	THAI CHARACTER MAITAIKHU..THAI CHARACTER YAMAKKAN
0E4F	#	Po		(๏)	THAI CHARACTER FONGMAN
0E50..0E59	#	Nd	[10]	(๐..๙)	THAI DIGIT ZERO..THAI DIGIT NINE
0E5A..0E5B	#	Po	[2]	(๚..๛)	THAI CHARACTER ANGKHANKHU..THAI CHARACTER KHOMUT
0E81..0E82	#	Lo	[2]	(□..□)	LAO LETTER KO..LAO LETTER KHO SUNG
0E84	#	Lo		(□)	LAO LETTER KHO TAM
0E87..0E88	#	Lo	[2]	(□..□)	LAO LETTER NGO..LAO LETTER CO
0E8A	#	Lo		(□)	LAO LETTER SO TAM
0E8D	#	Lo		(□)	LAO LETTER NYO
0E94..0E97	#	Lo	[4]	(□..□)	LAO LETTER DO..LAO LETTER THO TAM
0E99..0E9F	#	Lo	[7]	(□..□)	LAO LETTER NO..LAO LETTER FO SUNG
0EA1..0EA3	#	Lo	[3]	(□..□)	LAO LETTER MO..LAO LETTER LO LING
0EA5	#	Lo		(□)	LAO LETTER LO LOOT
0EA7	#	Lo		(□)	LAO LETTER WO

0EAA..0EAB	#	Lo	[2]	(□..□)	LAO LETTER SO SUNG..LAO LETTER HO SUNG
0EAD..0EB0	#	Lo	[4]	(□..□)	LAO LETTER O..LAO VOWEL SIGN A
0EB1	#	Mn		(□)	LAO VOWEL SIGN MAI KAN
0EB2..0EB3	#	Lo	[2]	(□..□)	LAO VOWEL SIGN AA..LAO VOWEL SIGN AM
0EB4..0EB9	#	Mn	[6]	(□..□)	LAO VOWEL SIGN I..LAO VOWEL SIGN UU
0EBB..0EBC	#	Mn	[2]	(□..□)	LAO VOWEL SIGN MAI KON..LAO SEMIVOWEL SIGN LO
0EBD	#	Lo		(□)	LAO SEMIVOWEL SIGN NYO
0EC0..0EC4	#	Lo	[5]	(□..□)	LAO VOWEL SIGN E..LAO VOWEL SIGN AI
0EC6	#	Lm		(□)	LAO KO LA
0EC8..0ECD	#	Mn	[6]	(□..□)	LAO TONE MAI EK..LAO NIGGAHITA
0ED0..0ED9	#	Nd	[10]	(□..□)	LAO DIGIT ZERO..LAO DIGIT NINE
0EDC..0EDD	#	Lo	[2]	(□..□)	LAO HO NO..LAO HO MO

Total: 151 ...(omitting 651 from listing)...

$SAScripts ⊇ \p{LineBreak=SA}

And in those scripts, they are all the alphabetic spacing characters, plus some odd Cf & Mn
Let $SAScriptExceptions = [\u1063\u1064\u1069-\u106D\u1087-\u108C\u108F\u109A\u109B\u109E\u109F\u19DE\u19DF\u1AA0-\u1AA6\u1AA8-\u1AAD\uAA77-\uAA79\uAA7B\uAADB-\uAADF]

1063..1064	#	Mc	[2]	(..)	MYANMAR TONE MARK SGAW KAREN HATHI..MYANMAR TONE MARK SGAW KAREN KE PHO
1069..106D	#	Mc	[5]	(..)	MYANMAR SIGN WESTERN PWO KAREN TONE-1..MYANMAR SIGN WESTERN PWO KAREN TONE-5
1087..108C	#	Mc	[6]	(..)	MYANMAR SIGN SHAN TONE-2..MYANMAR SIGN SHAN COUNCIL TONE-3
108F	#	Mc		( )	MYANMAR SIGN RUMAI PALAUNG TONE-5
109A..109B	#	Mc	[2]	(□..□)	MYANMAR SIGN KHAMTI TONE-1..MYANMAR SIGN KHAMTI TONE-3
109E..109F	#	So	[2]	(..)	MYANMAR SYMBOL SHAN ONE..MYANMAR SYMBOL SHAN EXCLAMATION
19DE..19DF	#	Po	[2]	(□..□)	NEW TAI LUE SIGN LAE..NEW TAI LUE SIGN LAEV
1AA0..1AA6	#	Po	[7]	(□..□)	TAI THAM SIGN WIANG..TAI THAM SIGN REVERSED ROTATED RANA
1AA8..1AAD	#	Po	[6]	(□..□)	TAI THAM SIGN KAAN..TAI THAM SIGN CAANG
AA77..AA79	#	So	[3]	(□..□)	MYANMAR SYMBOL AITON EXCLAMATION..MYANMAR SYMBOL AITON TWO
AA7B	#	Mc		(□)	MYANMAR SIGN PAO KAREN TONE
AADB..AADF	#	So	[5]	(□..□)	TAI VIET SYMBOL KON..TAI VIET SYMBOL KOI KOI

Total: 42

[$SAScripts & [\p{Alphabetic} \p{gc=cf} \p{gc=Mn} $SAScriptExceptions]] = [$SAScripts & [\p{LineBreak=SA} \p{LineBreak=CM}]]

\p{Math} = [\p{Other_Math} \p{GC=Sm}]
\p{Alphabetic} = [\p{Other_Alphabetic} \p{GC=Lu} \p{GC=Ll} \p{GC=Lt} \p{GC=Lm} \p{GC=Lo} \p{GC=Nl}]
\p{Lowercase} = [\p{Other_Lowercase} \p{GC=Ll}]
\p{Uppercase} = [\p{Other_Uppercase} \p{GC=Lu}]
\p{ID_Start} = [\p{Other_ID_Start} \p{GC=Lu} \p{GC=Ll} \p{GC=Lt} \p{GC=Lm} \p{GC=Lo} \p{GC=Nl} - \p{Pattern_Syntax} - \p{Pattern_White_Space}]
\p{ID_Continue} = [\p{Other_ID_Continue} \p{ID_Start} \p{GC=Mn} \p{GC=Mc} \p{GC=Nd} \p{GC=Pc} - \p{Pattern_Syntax} - \p{Pattern_White_Space}]

Let $DIExclusions = [\u0600-\u0603\u06DD\u070F\uFFF9-\uFFFB\U000110BD]

0600..0603	#	Cf	[4]	(□..□)	ARABIC NUMBER SIGN..ARABIC SIGN SAFHA
06DD	#	Cf		(◌)	ARABIC END OF AYAH
070F	#	Cf		(▯)	SYRIAC ABBREVIATION MARK
FFF9..FFFB	#	Cf	[3]	(▯..▯)	INTERLINEAR ANNOTATION ANCHOR..INTERLINEAR ANNOTATION TERMINATOR
110BD	#	Cf		(□)	KAITHI NUMBER SIGN

Total: 10

\p{Default_Ignorable_Code_Point} = [\p{Other_Default_Ignorable_Code_Point} \p{GC=Cf} \p{Variation_Selector} - [\p{White_Space} $DIExclusions]]

\p{Grapheme_Extend} = [\p{Other_Grapheme_Extend} \p{GC=Me} \p{GC=Mn}]

\p{Grapheme_Base} = [^\p{GC=Cc} \p{GC=Cf} \p{GC=Cs} \p{GC=Co} \p{GC=Cn} \p{GC=Zl} \p{GC=Zp} \p{Grapheme_Extend}]
\p{Grapheme_Link} = \p{CCC=Virama}

"Minimal" Other_: NOT hard requirements; just if we want to be minimal
(Should add way to make these warnings, not errors)

\p{Other_Math} = [\p{Math} - \p{GC=Sm}]
\p{Other_Alphabetic} = [\p{Alphabetic} - [\p{GC=Lu} \p{GC=Ll} \p{GC=Lt} \p{GC=Lm} \p{GC=Lo} \p{GC=Nl}]]
\p{Other_Lowercase} = [\p{Lowercase} - \p{GC=Ll}]
\p{Other_Uppercase} = [\p{Uppercase} - \p{GC=Lu}]
\p{Other_ID_Start} = [\p{ID_Start} - [\p{GC=Lu} \p{GC=Ll} \p{GC=Lt} \p{GC=Lm} \p{GC=Lo} \p{GC=Nl} - \p{Pattern_Syntax} - \p{Pattern_White_Space}]]
\p{Other_ID_Continue} = [\p{ID_Continue} - [\p{ID_Start} \p{GC=Mn} \p{GC=Mc} \p{GC=Nd} \p{GC=Pc} - \p{Pattern_Syntax} - \p{Pattern_White_Space}]]

Let $Annotations = [\uFFF9-\uFFFB]

FFF9..FFFB	#	Cf	[3]	(▯..▯)	INTERLINEAR ANNOTATION ANCHOR..INTERLINEAR ANNOTATION TERMINATOR

Total: 3

\p{Other_Default_Ignorable_Code_Point} = [\p{Default_Ignorable_Code_Point} - [\p{GC=Cf} \p{Variation_Selector} - [\p{White_Space} $Annotations]]]
\p{Other_Grapheme_Extend} = [\p{Grapheme_Extend} - [\p{GC=Me} \p{GC=Mn}]]

POSIX Compatibility Properties (UTS#18)
http://www.opengroup.org/onlinepubs/007904975/basedefs/xbd_chap07.html

---

constants

Let $SP = [\u0020] # [\N{space}]

0020	#	Zs	( )	SPACE

Total: 1

Let $TAB = [\u0009] # [\N{CHARACTER TABULATION}]

0009	#	Cc	(\t)	

Total: 1

Let $LF = [\u000A] # \N{linefeed}

000A	#	Cc	(\n)	

Total: 1

Let $VTAB = [\u000B] # [\N{LINE TABULATION}]

000B	#	Cc	(\v)	

Total: 1

Let $FF = [\u000C] # [\N{formfeed}]

000C	#	Cc	(\f)	

Total: 1

Let $CR = [\u000D] # \N{carriage return}

000D	#	Cc	(\r)	

Total: 1

Let $NEL = [\u0085] # \N{next line}

0085	#	Cc	( )	

Total: 1

#Let $ZWNJ = [\u200C] # [\N{ZERO WIDTH NON-JOINER}]
#Let $ZWJ = [\u200D] # [\N{ZERO WIDTH JOINER}]

Let $CircledAsciiLetters = [\u24B6-\u24E9]

24B6..24E9	#	So	[52]	(Ⓐ..ⓩ)	CIRCLED LATIN CAPITAL LETTER A..CIRCLED LATIN SMALL LETTER Z

Total: 52

Unassigned, Control, Format, Private_Use, Surrogate,
Uppercase_Letter, Lowercase_Letter, Titlecase_Letter, Modifier_Letter, Other_Letter,
Nonspacing_Mark, Enclosing_Mark, Spacing_Mark,
Decimal_Number, Letter_Number, Other_Number,
Space_Separator, Line_Separator, Paragraph_Separator,
Dash_Punctuation, Open_Punctuation, Close_Punctuation, Connector_Punctuation, Other_Punctuation, Initial_Punctuation, Final_Punctuation
Math_Symbol, Currency_Symbol, Modifier_Symbol, Other_Symbol

UTS Rules

Let $alpha = [\p{Alphabetic} $CircledAsciiLetters]

0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00B5	#	L&		(µ)	MICRO SIGN
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL LETTER O WITH DIAERESIS
00D8..00F6	#	L&	[31]	(Ø..ö)	LATIN CAPITAL LETTER O WITH STROKE..LATIN SMALL LETTER O WITH DIAERESIS
00F8..01BA	#	L&	[195]	(ø..ƺ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER EZH WITH TAIL
01BB	#	Lo		(ƻ)	LATIN LETTER TWO WITH STROKE
01BC..01BF	#	L&	[4]	(Ƽ..ƿ)	LATIN CAPITAL LETTER TONE FIVE..LATIN LETTER WYNN
01C0..01C3	#	Lo	[4]	(ǀ..ǃ)	LATIN LETTER DENTAL CLICK..LATIN LETTER RETROFLEX CLICK
01C4..0293	#	L&	[208]	(DŽ..ʓ)	LATIN CAPITAL LETTER DZ WITH CARON..LATIN SMALL LETTER EZH WITH CURL
0294	#	Lo		(ʔ)	LATIN LETTER GLOTTAL STOP
0295..02AF	#	L&	[27]	(ʕ..ʯ)	LATIN LETTER PHARYNGEAL VOICED FRICATIVE..LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
02B0..02C1	#	Lm	[18]	(ʰ..ˁ)	MODIFIER LETTER SMALL H..MODIFIER LETTER REVERSED GLOTTAL STOP
02C6..02D1	#	Lm	[12]	(ˆ..ˑ)	MODIFIER LETTER CIRCUMFLEX ACCENT..MODIFIER LETTER HALF TRIANGULAR COLON
02E0..02E4	#	Lm	[5]	(ˠ..ˤ)	MODIFIER LETTER SMALL GAMMA..MODIFIER LETTER SMALL REVERSED GLOTTAL STOP

02EC	#	Lm		(□)	MODIFIER LETTER VOICING
02EE	#	Lm		(□)	MODIFIER LETTER DOUBLE APOSTROPHE
0345	#	Mn		(□)	COMBINING GREEK YPOGEGRAMMENI
0370..0373	#	L&	[4]	(□..□)	GREEK CAPITAL LETTER HETA..GREEK SMALL LETTER ARCHAIC SAMPI
0374	#	Lm		(′)	GREEK NUMERAL SIGN
0376..0377	#	L&	[2]	(□..□)	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA..GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
037A	#	Lm		(ͺ)	GREEK YPOGEGRAMMENI
037B..037D	#	L&	[3]	(□..□)	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
0386	#	L&		(Ά)	GREEK CAPITAL LETTER ALPHA WITH TONOS
0388..038A	#	L&	[3]	(Έ..Ί)	GREEK CAPITAL LETTER EPSILON WITH TONOS..GREEK CAPITAL LETTER IOTA WITH TONOS
038C	#	L&		(Ό)	GREEK CAPITAL LETTER OMICRON WITH TONOS
038E..03A1	#	L&	[20]	(Ύ..Ρ)	GREEK CAPITAL LETTER UPSILON WITH TONOS..GREEK CAPITAL LETTER RHO

Total: 623 ...(omitting 99894 from listing)...

Let $lower = \p{Lowercase}

0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00B5	#	L&		(µ)	MICRO SIGN
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00DF..00F6	#	L&	[24]	(ß..ö)	LATIN SMALL LETTER SHARP S..LATIN SMALL LETTER O WITH DIAERESIS
00F8..00FF	#	L&	[8]	(ø..ÿ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER Y WITH DIAERESIS
0101	#	L&		(ā)	LATIN SMALL LETTER A WITH MACRON
0103	#	L&		(ă)	LATIN SMALL LETTER A WITH BREVE
0105	#	L&		(ą)	LATIN SMALL LETTER A WITH OGONEK
0107	#	L&		(ć)	LATIN SMALL LETTER C WITH ACUTE
0109	#	L&		(ĉ)	LATIN SMALL LETTER C WITH CIRCUMFLEX
010B	#	L&		(ċ)	LATIN SMALL LETTER C WITH DOT ABOVE
010D	#	L&		(č)	LATIN SMALL LETTER C WITH CARON
010F	#	L&		(ď)	LATIN SMALL LETTER D WITH CARON
0111	#	L&		(đ)	LATIN SMALL LETTER D WITH STROKE
0113	#	L&		(ē)	LATIN SMALL LETTER E WITH MACRON
0115	#	L&		(ĕ)	LATIN SMALL LETTER E WITH BREVE
0117	#	L&		(ė)	LATIN SMALL LETTER E WITH DOT ABOVE
0119	#	L&		(ę)	LATIN SMALL LETTER E WITH OGONEK
011B	#	L&		(ě)	LATIN SMALL LETTER E WITH CARON

Total: 75 ...(omitting 1833 from listing)...

Let $upper = [\p{Uppercase}]

0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL LETTER O WITH DIAERESIS
00D8..00DE	#	L&	[7]	(Ø..Þ)	LATIN CAPITAL LETTER O WITH STROKE..LATIN CAPITAL LETTER THORN
0100	#	L&		(Ā)	LATIN CAPITAL LETTER A WITH MACRON
0102	#	L&		(Ă)	LATIN CAPITAL LETTER A WITH BREVE
0104	#	L&		(Ą)	LATIN CAPITAL LETTER A WITH OGONEK
0106	#	L&		(Ć)	LATIN CAPITAL LETTER C WITH ACUTE
0108	#	L&		(Ĉ)	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
010A	#	L&		(Ċ)	LATIN CAPITAL LETTER C WITH DOT ABOVE
010C	#	L&		(Č)	LATIN CAPITAL LETTER C WITH CARON
010E	#	L&		(Ď)	LATIN CAPITAL LETTER D WITH CARON
0110	#	L&		(Đ)	LATIN CAPITAL LETTER D WITH STROKE
0112	#	L&		(Ē)	LATIN CAPITAL LETTER E WITH MACRON
0114	#	L&		(Ĕ)	LATIN CAPITAL LETTER E WITH BREVE
0116	#	L&		(Ė)	LATIN CAPITAL LETTER E WITH DOT ABOVE
0118	#	L&		(Ę)	LATIN CAPITAL LETTER E WITH OGONEK
011A	#	L&		(Ě)	LATIN CAPITAL LETTER E WITH CARON
011C	#	L&		(Ĝ)	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
011E	#	L&		(Ğ)	LATIN CAPITAL LETTER G WITH BREVE
0120	#	L&		(Ġ)	LATIN CAPITAL LETTER G WITH DOT ABOVE

Total: 73 ...(omitting 1396 from listing)...

Let $punct = [$gcAllPunctuation $gcAllSymbols - $alpha]

0021..0023	#	Po	[3]	(!..#)	EXCLAMATION MARK..NUMBER SIGN
0024	#	Sc		($)	DOLLAR SIGN
0025..0027	#	Po	[3]	(%..')	PERCENT SIGN..APOSTROPHE
0028	#	Ps		(()	LEFT PARENTHESIS
0029	#	Pe		())	RIGHT PARENTHESIS
002A	#	Po		(*)	ASTERISK
002B	#	Sm		(+)	PLUS SIGN
002C	#	Po		(,)	COMMA
002D	#	Pd		(-)	HYPHEN-MINUS
002E..002F	#	Po	[2]	(../)	FULL STOP..SOLIDUS
003A..003B	#	Po	[2]	(:..;)	COLON..SEMICOLON
003C..003E	#	Sm	[3]	(<..>)	LESS-THAN SIGN..GREATER-THAN SIGN
003F..0040	#	Po	[2]	(?..@)	QUESTION MARK..COMMERCIAL AT
005B	#	Ps		([)	LEFT SQUARE BRACKET
005C	#	Po		(\)	REVERSE SOLIDUS
005D	#	Pe		(])	RIGHT SQUARE BRACKET
005E	#	Sk		(^)	CIRCUMFLEX ACCENT

005F	#	Pc		(_)	LOW LINE
0060	#	Sk		(`)	GRAVE ACCENT
007B	#	Ps		({)	LEFT CURLY BRACKET
007C	#	Sm		(|)	VERTICAL LINE
007D	#	Pe		(})	RIGHT CURLY BRACKET
007E	#	Sm		(~)	TILDE
00A1	#	Po		(¡)	INVERTED EXCLAMATION MARK
00A2..00A5	#	Sc	[4]	(¢..¥)	CENT SIGN..YEN SIGN
00A6..00A7	#	So	[2]	(¦..§)	BROKEN BAR..SECTION SIGN
00A8	#	Sk		(¨)	DIAERESIS
00A9	#	So		(©)	COPYRIGHT SIGN
00AB	#	Pi		(«)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00AC	#	Sm		(¬)	NOT SIGN
00AE	#	So		(®)	REGISTERED SIGN
00AF	#	Sk		(¯)	MACRON
00B0	#	So		(°)	DEGREE SIGN
00B1	#	Sm		(±)	PLUS-MINUS SIGN
00B4	#	Sk		(´)	ACUTE ACCENT
00B6	#	So		(¶)	PILCROW SIGN
00B7	#	Po		(·)	MIDDLE DOT
00B8	#	Sk		(¸)	CEDILLA
00BB	#	Pf		(»)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00BF	#	Po		(¿)	INVERTED QUESTION MARK
00D7	#	Sm		(×)	MULTIPLICATION SIGN
00F7	#	Sm		(÷)	DIVISION SIGN
02C2..02C5	#	Sk	[4]	(˂..˅)	MODIFIER LETTER LEFT ARROWHEAD..MODIFIER LETTER DOWN ARROWHEAD
02D2..02DF	#	Sk	[14]	(˒..□)	MODIFIER LETTER CENTRED RIGHT HALF RING..MODIFIER LETTER CROSS ACCENT
02E5..02EB	#	Sk	[7]	(˥..□)	MODIFIER LETTER EXTRA-HIGH TONE BAR..MODIFIER LETTER YANG DEPARTING TONE MARK
02ED	#	Sk		(□)	MODIFIER LETTER UNASPIRATED
02EF..02FF	#	Sk	[17]	(□..□)	MODIFIER LETTER LOW DOWN ARROWHEAD..MODIFIER LETTER LOW LEFT ARROW
0375	#	Sk		(ʻ)	GREEK LOWER NUMERAL SIGN
037E	#	Po		(;)	GREEK QUESTION MARK

Total: 100 ...(omitting 4935 from listing)...

Let $digit = \p{gc=Decimal_Number}

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
07C0..07C9	#	Nd	[10]	(□..□)	NKO DIGIT ZERO..NKO DIGIT NINE

0966..096F	#	Nd	[10]	(०..९)	DEVANAGARI DIGIT ZERO..DEVANAGARI DIGIT NINE
09E6..09EF	#	Nd	[10]	(০..৯)	BENGALI DIGIT ZERO..BENGALI DIGIT NINE
0A66..0A6F	#	Nd	[10]	(੦..੯)	GURMUKHI DIGIT ZERO..GURMUKHI DIGIT NINE
0AE6..0AEF	#	Nd	[10]	(૦..૯)	GUJARATI DIGIT ZERO..GUJARATI DIGIT NINE
0B66..0B6F	#	Nd	[10]	(□..□)	ORIYA DIGIT ZERO..ORIYA DIGIT NINE
0BE6..0BEF	#	Nd	[10]	(..௯)	TAMIL DIGIT ZERO..TAMIL DIGIT NINE
0C66..0C6F	#	Nd	[10]	(౦..౯)	TELUGU DIGIT ZERO..TELUGU DIGIT NINE
0CE6..0CEF	#	Nd	[10]	(೦..೯)	KANNADA DIGIT ZERO..KANNADA DIGIT NINE
0D66..0D6F	#	Nd	[10]	(൦..൯)	MALAYALAM DIGIT ZERO..MALAYALAM DIGIT NINE
0E50..0E59	#	Nd	[10]	(๐..๙)	THAI DIGIT ZERO..THAI DIGIT NINE
0ED0..0ED9	#	Nd	[10]	(□..□)	LAO DIGIT ZERO..LAO DIGIT NINE
0F20..0F29	#	Nd	[10]	(༠..༩)	TIBETAN DIGIT ZERO..TIBETAN DIGIT NINE
1040..1049	#	Nd	[10]	(F ..F)	MYANMAR DIGIT ZERO..MYANMAR DIGIT NINE
1090..1099	#	Nd	[10]	(..)	MYANMAR SHAN DIGIT ZERO..MYANMAR SHAN DIGIT NINE
17E0..17E9	#	Nd	[10]	(□..□)	KHMER DIGIT ZERO..KHMER DIGIT NINE
1810..1819	#	Nd	[10]	(᠐..᠙)	MONGOLIAN DIGIT ZERO..MONGOLIAN DIGIT NINE

Total: 200 ...(omitting 211 from listing)...

Let $xdigit = [\p{gc=Decimal_Number} \p{Hex_Digit}] # in both!

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
0041..0046	#	L&	[6]	(A..F)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER F
0061..0066	#	L&	[6]	(a..f)	LATIN SMALL LETTER A..LATIN SMALL LETTER F
0660..0669	#	Nd	[10]	(٠..٩)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
07C0..07C9	#	Nd	[10]	(□..□)	NKO DIGIT ZERO..NKO DIGIT NINE
0966..096F	#	Nd	[10]	(०..९)	DEVANAGARI DIGIT ZERO..DEVANAGARI DIGIT NINE
09E6..09EF	#	Nd	[10]	(০..৯)	BENGALI DIGIT ZERO..BENGALI DIGIT NINE
0A66..0A6F	#	Nd	[10]	(੦..੯)	GURMUKHI DIGIT ZERO..GURMUKHI DIGIT NINE
0AE6..0AEF	#	Nd	[10]	(૦..૯)	GUJARATI DIGIT ZERO..GUJARATI DIGIT NINE
0B66..0B6F	#	Nd	[10]	(□..□)	ORIYA DIGIT ZERO..ORIYA DIGIT NINE
0BE6..0BEF	#	Nd	[10]	(..௯)	TAMIL DIGIT ZERO..TAMIL DIGIT NINE
0C66..0C6F	#	Nd	[10]	(౦..౯)	TELUGU DIGIT ZERO..TELUGU DIGIT NINE
0CE6..0CEF	#	Nd	[10]	(೦..೯)	KANNADA DIGIT ZERO..KANNADA DIGIT NINE
0D66..0D6F	#	Nd	[10]	(൦..൯)	MALAYALAM DIGIT ZERO..MALAYALAM DIGIT NINE
0E50..0E59	#	Nd	[10]	(๐..๙)	THAI DIGIT ZERO..THAI DIGIT NINE
0ED0..0ED9	#	Nd	[10]	(□..□)	LAO DIGIT ZERO..LAO DIGIT NINE
0F20..0F29	#	Nd	[10]	(༠..༩)	TIBETAN DIGIT ZERO..TIBETAN DIGIT NINE
1040..1049	#	Nd	[10]	(F ..F)	MYANMAR DIGIT ZERO..MYANMAR DIGIT NINE
1090..1099	#	Nd	[10]	(..)	MYANMAR SHAN DIGIT ZERO..MYANMAR SHAN DIGIT NINE

Total: 192 ...(omitting 243 from listing)...

Let $alnum = [$alpha $digit]

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00B5	#	L&		(µ)	MICRO SIGN
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL LETTER O WITH DIAERESIS
00D8..00F6	#	L&	[31]	(Ø..ö)	LATIN CAPITAL LETTER O WITH STROKE..LATIN SMALL LETTER O WITH DIAERESIS
00F8..01BA	#	L&	[195]	(ø..ƺ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER EZH WITH TAIL
01BB	#	Lo		(ƻ)	LATIN LETTER TWO WITH STROKE
01BC..01BF	#	L&	[4]	(Ƽ..ƿ)	LATIN CAPITAL LETTER TONE FIVE..LATIN LETTER WYNN
01C0..01C3	#	Lo	[4]	(ǀ..ǃ)	LATIN LETTER DENTAL CLICK..LATIN LETTER RETROFLEX CLICK
01C4..0293	#	L&	[208]	(DŽ..ʓ)	LATIN CAPITAL LETTER DZ WITH CARON..LATIN SMALL LETTER EZH WITH CURL
0294	#	Lo		(ʔ)	LATIN LETTER GLOTTAL STOP
0295..02AF	#	L&	[27]	(ʕ..ʯ)	LATIN LETTER PHARYNGEAL VOICED FRICATIVE..LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
02B0..02C1	#	Lm	[18]	(ʰ..ˁ)	MODIFIER LETTER SMALL H..MODIFIER LETTER REVERSED GLOTTAL STOP
02C6..02D1	#	Lm	[12]	(ˆ..ˑ)	MODIFIER LETTER CIRCUMFLEX ACCENT..MODIFIER LETTER HALF TRIANGULAR COLON
02E0..02E4	#	Lm	[5]	(ˠ..ˤ)	MODIFIER LETTER SMALL GAMMA..MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
02EC	#	Lm		(˭)	MODIFIER LETTER VOICING
02EE	#	Lm		(ˮ)	MODIFIER LETTER DOUBLE APOSTROPHE
0345	#	Mn		(◌ͅ)	COMBINING GREEK YPOGEGRAMMENI
0370..0373	#	L&	[4]	(Ͱ..ͳ)	GREEK CAPITAL LETTER HETA..GREEK SMALL LETTER ARCHAIC SAMPI
0374	#	Lm		(ʹ)	GREEK NUMERAL SIGN
0376..0377	#	L&	[2]	(Ͷ..ͷ)	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA..GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
037A	#	Lm		(ͺ)	GREEK YPOGEGRAMMENI
037B..037D	#	L&	[3]	(ͻ..ͽ)	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
0386	#	L&		(Ά)	GREEK CAPITAL LETTER ALPHA WITH TONOS
0388..038A	#	L&	[3]	(Έ..Ί)	GREEK CAPITAL LETTER EPSILON WITH TONOS..GREEK CAPITAL LETTER IOTA WITH TONOS
038C	#	L&		(Ό)	GREEK CAPITAL LETTER OMICRON WITH TONOS

Total: 613 ...(omitting 100315 from listing)...

Let $space = \p{Whitespace}

0009..000D	#	Cc	[5]	(..)	..
0020	#	Zs		( )	SPACE
0085	#	Cc		( )	
00A0	#	Zs		( )	NO-BREAK SPACE
1680	#	Zs		(─)	OGHAM SPACE MARK
180E	#	Zs		(᠎)	MONGOLIAN VOWEL SEPARATOR
2000..200A	#	Zs	[11]	(..)	EN QUAD..HAIR SPACE
2028	#	Zl		( )	LINE SEPARATOR
2029	#	Zp		( )	PARAGRAPH SEPARATOR
202F	#	Zs		(□)	NARROW NO-BREAK SPACE
205F	#	Zs		(▪)	MEDIUM MATHEMATICAL SPACE
3000	#	Zs		( )	IDEOGRAPHIC SPACE

Total: 26

Let $blank = [\p{Whitespace} - [$LF $VTAB $FF $CR $NEL \p{gc=Line_Separator} \p{gc=Paragraph_Separator}]]

0009	#	Cc		( )	
0020	#	Zs		( )	SPACE
00A0	#	Zs		( )	NO-BREAK SPACE
1680	#	Zs		(─)	OGHAM SPACE MARK
180E	#	Zs		(᠎)	MONGOLIAN VOWEL SEPARATOR
2000..200A	#	Zs	[11]	(..)	EN QUAD..HAIR SPACE
202F	#	Zs		(□)	NARROW NO-BREAK SPACE
205F	#	Zs		(▪)	MEDIUM MATHEMATICAL SPACE
3000	#	Zs		( )	IDEOGRAPHIC SPACE

Total: 19

Let $cntrl = \p{gc=Control}

0000..001F	#	Cc	[32]	(□..□)	..
007F..009F	#	Cc	[33]	(□..□)	..

Total: 65

Let $graph = [^$space \p{gc=Control} \p{gc=Surrogate} \p{gc=Unassigned}]

0021..0023	#	Po	[3]	(!..#)	EXCLAMATION MARK..NUMBER SIGN
0024	#	Sc		($)	DOLLAR SIGN
0025..0027	#	Po	[3]	(%..')	PERCENT SIGN..APOSTROPHE
0028	#	Ps		(()	LEFT PARENTHESIS
0029	#	Pe		())	RIGHT PARENTHESIS
002A	#	Po		(*)	ASTERISK
002B	#	Sm		(+)	PLUS SIGN

002C	#	Po		(,)	COMMA
002D	#	Pd		(-)	HYPHEN-MINUS
002E..002F	#	Po	[2]	(../)	FULL STOP..SOLIDUS
0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
003A..003B	#	Po	[2]	(:..;)	COLON..SEMICOLON
003C..003E	#	Sm	[3]	(<..>)	LESS-THAN SIGN..GREATER-THAN SIGN
003F..0040	#	Po	[2]	(?..@)	QUESTION MARK..COMMERCIAL AT
0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
005B	#	Ps		([)	LEFT SQUARE BRACKET
005C	#	Po		(\)	REVERSE SOLIDUS
005D	#	Pe		(])	RIGHT SQUARE BRACKET
005E	#	Sk		(^)	CIRCUMFLEX ACCENT
005F	#	Pc		(_)	LOW LINE
0060	#	Sk		(`)	GRAVE ACCENT
0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
007B	#	Ps		({)	LEFT CURLY BRACKET
007C	#	Sm		(|)	VERTICAL LINE
007D	#	Pe		(})	RIGHT CURLY BRACKET
007E	#	Sm		(~)	TILDE
00A1	#	Po		(¡)	INVERTED EXCLAMATION MARK
00A2..00A5	#	Sc	[4]	(¢..¥)	CENT SIGN..YEN SIGN
00A6..00A7	#	So	[2]	(¦..§)	BROKEN BAR..SECTION SIGN
00A8	#	Sk		(¨)	DIAERESIS
00A9	#	So		(©)	COPYRIGHT SIGN
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00AB	#	Pi		(«)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00AC	#	Sm		(¬)	NOT SIGN
00AD	#	Cf		()	SOFT HYPHEN
00AE	#	So		(®)	REGISTERED SIGN
00AF	#	Sk		(¯)	MACRON
00B0	#	So		(°)	DEGREE SIGN
00B1	#	Sm		(±)	PLUS-MINUS SIGN
00B2..00B3	#	No	[2]	(²..³)	SUPERSCRIPT TWO..SUPERSCRIPT THREE
00B4	#	Sk		(´)	ACUTE ACCENT
00B5	#	L&		(µ)	MICRO SIGN
00B6	#	So		(¶)	PILCROW SIGN
00B7	#	Po		(·)	MIDDLE DOT
00B8	#	Sk		(¸)	CEDILLA
00B9	#	No		(¹)	SUPERSCRIPT ONE
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00BB	#	Pf		(»)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00BC..00BE	#	No	[3]	(¼..¾)	VULGAR FRACTION ONE QUARTER..VULGAR FRACTION THREE QUARTERS
00BF	#	Po		(¿)	INVERTED QUESTION MARK
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL

					LETTER O WITH DIAERESIS
00D7	#	Sm		(×)	MULTIPLICATION SIGN
00D8..00F6	#	L&	[31]	(Ø..ö)	LATIN CAPITAL LETTER O WITH STROKE..LATIN SMALL LETTER O WITH DIAERESIS
00F7	#	Sm		(÷)	DIVISION SIGN
00F8..01BA	#	L&	[195]	(ø..ƺ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER EZH WITH TAIL
01BB	#	Lo		(ƻ)	LATIN LETTER TWO WITH STROKE
01BC..01BF	#	L&	[4]	(Ƽ..ƿ)	LATIN CAPITAL LETTER TONE FIVE..LATIN LETTER WYNN
01C0..01C3	#	Lo	[4]	(ǀ..ǃ)	LATIN LETTER DENTAL CLICK..LATIN LETTER RETROFLEX CLICK
01C4..0293	#	L&	[208]	(DŽ..ʓ)	LATIN CAPITAL LETTER DZ WITH CARON..LATIN SMALL LETTER EZH WITH CURL
0294	#	Lo		(ʔ)	LATIN LETTER GLOTTAL STOP
0295..02AF	#	L&	[27]	(ʕ..ʯ)	LATIN LETTER PHARYNGEAL VOICED FRICATIVE..LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
02B0..02C1	#	Lm	[18]	(ʰ..ˁ)	MODIFIER LETTER SMALL H..MODIFIER LETTER REVERSED GLOTTAL STOP
02C2..02C5	#	Sk	[4]	(˂..˅)	MODIFIER LETTER LEFT ARROWHEAD..MODIFIER LETTER DOWN ARROWHEAD
02C6..02D1	#	Lm	[12]	(ˆ..˒)	MODIFIER LETTER CIRCUMFLEX ACCENT..MODIFIER LETTER HALF TRIANGULAR COLON
02D2..02DF	#	Sk	[14]	(˓..˟)	MODIFIER LETTER CENTRED RIGHT HALF RING..MODIFIER LETTER CROSS ACCENT
02E0..02E4	#	Lm	[5]	(ˠ..ˤ)	MODIFIER LETTER SMALL GAMMA..MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
02E5..02EB	#	Sk	[7]	(˥..˫)	MODIFIER LETTER EXTRA-HIGH TONE BAR..MODIFIER LETTER YANG DEPARTING TONE MARK
02EC	#	Lm		(ˬ)	MODIFIER LETTER VOICING
02ED	#	Sk		(˭)	MODIFIER LETTER UNASPIRATED
02EE	#	Lm		(ˮ)	MODIFIER LETTER DOUBLE APOSTROPHE
02EF..02FF	#	Sk	[17]	(˯..˿)	MODIFIER LETTER LOW DOWN ARROWHEAD..MODIFIER LETTER LOW LEFT ARROW
0300..036F	#	Mn	[112]	(◌̀..◌ͯ)	COMBINING GRAVE ACCENT..COMBINING LATIN SMALL LETTER X
0370..0373	#	L&	[4]	(Ͱ..ͳ)	GREEK CAPITAL LETTER HETA..GREEK SMALL LETTER ARCHAIC SAMPI
0374	#	Lm		(ʹ)	GREEK NUMERAL SIGN
0375	#	Sk		(͵)	GREEK LOWER NUMERAL SIGN
0376..0377	#	L&	[2]	(Ͷ..ͷ)	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA..GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
037A	#	Lm		(◌ͺ)	GREEK YPOGEGRAMMENI
037B..037D	#	L&	[3]	(ͻ..ͽ)	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
037E	#	Po		(;)	GREEK QUESTION MARK

0384..0385	#	Sk	[2]	(΄..΅)	GREEK TONOS..GREEK DIALYTIKA TONOS
0386	#	L&		(Ά)	GREEK CAPITAL LETTER ALPHA WITH TONOS
0387	#	Po		(·)	GREEK ANO TELEIA
0388..038A	#	L&	[3]	(Έ..Ί)	GREEK CAPITAL LETTER EPSILON WITH TONOS..GREEK CAPITAL LETTER IOTA WITH TONOS
038C	#	L&		(Ό)	GREEK CAPITAL LETTER OMICRON WITH TONOS
038E..03A1	#	L&	[20]	(Ύ..Ρ)	GREEK CAPITAL LETTER UPSILON WITH TONOS..GREEK CAPITAL LETTER RHO
03A3..03F5	#	L&	[83]	(Σ..϶)	GREEK CAPITAL LETTER SIGMA..GREEK LUNATE EPSILON SYMBOL
03F6	#	Sm		(϶)	GREEK REVERSED LUNATE EPSILON SYMBOL
03F7..0481	#	L&	[139]	(Ϸ..ҁ)	GREEK CAPITAL LETTER SHO..CYRILLIC SMALL LETTER KOPPA
0482	#	So		(҂)	CYRILLIC THOUSANDS SIGN
0483..0487	#	Mn	[5]	(◌҃..◌҇)	COMBINING CYRILLIC TITLO..COMBINING CYRILLIC POKRYTIE
0488..0489	#	Me	[2]	(◌҈..◌҉)	COMBINING CYRILLIC HUNDRED THOUSANDS SIGN..COMBINING CYRILLIC MILLIONS SIGN
048A..0525	#	L&	[156]	(Ҋ..ԥ)	CYRILLIC CAPITAL LETTER SHORT I WITH TAIL..CYRILLIC SMALL LETTER PE WITH DESCENDER
0531..0556	#	L&	[38]	(Ա..Ֆ)	ARMENIAN CAPITAL LETTER AYB..ARMENIAN CAPITAL LETTER FEH
0559	#	Lm		(ՙ)	ARMENIAN MODIFIER LETTER LEFT HALF RING
055A..055F	#	Po	[6]	(՚..՟)	ARMENIAN APOSTROPHE..ARMENIAN ABBREVIATION MARK
0561..0587	#	L&	[39]	(ա..և)	ARMENIAN SMALL LETTER AYB..ARMENIAN SMALL LIGATURE ECH YIWN
0589	#	Po		(։)	ARMENIAN FULL STOP
058A	#	Pd		(֊)	ARMENIAN HYPHEN
0591..05BD	#	Mn	[45]	(◌֑..◌ֽ)	HEBREW ACCENT ETNAHTA..HEBREW POINT METEG
05BE	#	Pd		(־)	HEBREW PUNCTUATION MAQAF
05BF	#	Mn		(◌ֿ)	HEBREW POINT RAFE
05C0	#	Po		(׀)	HEBREW PUNCTUATION PASEQ
05C1..05C2	#	Mn	[2]	(◌ׁ..◌ׂ)	HEBREW POINT SHIN DOT..HEBREW POINT SIN DOT
05C3	#	Po		(׃)	HEBREW PUNCTUATION SOF PASUQ
05C4..05C5	#	Mn	[2]	(◌ׄ..◌ׅ)	HEBREW MARK UPPER DOT..HEBREW MARK LOWER DOT
05C6	#	Po		(׆)	HEBREW PUNCTUATION NUN HAFUKHA
05C7	#	Mn		(◌ׇ)	HEBREW POINT QAMATS QATAN
05D0..05EA	#	Lo	[27]	(א..ת)	HEBREW LETTER ALEF..HEBREW LETTER TAV
05F0..05F2	#	Lo	[3]	(ױ..ײ)	HEBREW LIGATURE YIDDISH DOUBLE VAV..HEBREW LIGATURE YIDDISH DOUBLE YOD
05F3..05F4	#	Po	[2]	(׳..״)	HEBREW PUNCTUATION GERESH..HEBREW PUNCTUATION GERSHAYIM
0600..0603	#	Cf	[4]	(◌؀..◌؃)	ARABIC NUMBER SIGN..ARABIC SIGN SAFHA
0606..0608	#	Sm	[3]	(◌؆..◌؈)	ARABIC-INDIC CUBE ROOT..ARABIC RAY
0609..060A	#	Po	[2]	(◌؉..◌؊)	ARABIC-INDIC PER MILLE SIGN..ARABIC-INDIC PER TEN THOUSAND SIGN

060B	#	Sc		(□)	AFGHANI SIGN
060C..060D	#	Po	[2]	(،..□)	ARABIC COMMA..ARABIC DATE SEPARATOR
060E..060F	#	So	[2]	(□..□)	ARABIC POETIC VERSE SIGN..ARABIC SIGN MISRA
0610..061A	#	Mn	[11]	(□..□)	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM..ARABIC SMALL KASRA
061B	#	Po		(؛)	ARABIC SEMICOLON
061E..061F	#	Po	[2]	(؞..□)	ARABIC TRIPLE DOT PUNCTUATION MARK..ARABIC QUESTION MARK
0621..063F	#	Lo	[31]	(□..ع)	ARABIC LETTER HAMZA..ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE
0640	#	Lm		(-)	ARABIC TATWEEL
0641..064A	#	Lo	[10]	(ف..ي)	ARABIC LETTER FEH..ARABIC LETTER YEH
064B..065E	#	Mn	[20]	(ً..□)	ARABIC FATHATAN..ARABIC FATHA WITH TWO DOTS
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
066A..066D	#	Po	[4]	(٪..*)	ARABIC PERCENT SIGN..ARABIC FIVE POINTED STAR
066E..066F	#	Lo	[2]	(ب..ق)	ARABIC LETTER DOTLESS BEH..ARABIC LETTER DOTLESS QAF
0670	#	Mn		(◌ٰ)	ARABIC LETTER SUPERSCRIPT ALEF
0671..06D3	#	Lo	[99]	(ٱ..ے)	ARABIC LETTER ALEF WASLA..ARABIC LETTER YEH BARREE WITH HAMZA ABOVE
06D4	#	Po		(۔)	ARABIC FULL STOP
06D5	#	Lo		(ە)	ARABIC LETTER AE
06D6..06DC	#	Mn	[7]	(◌ۖ..◌ۜ)	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA..ARABIC SMALL HIGH SEEN
06DD	#	Cf		(◌۝)	ARABIC END OF AYAH
06DE	#	Me		(◌۞)	ARABIC START OF RUB EL HIZB
06DF..06E4	#	Mn	[6]	(◌..◌)	ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL HIGH MADDA
06E5..06E6	#	Lm	[2]	(ۥ..ۦ)	ARABIC SMALL WAW..ARABIC SMALL YEH
06E7..06E8	#	Mn	[2]	(◌ۧ..◌ۨ)	ARABIC SMALL HIGH YEH..ARABIC SMALL HIGH NOON
06E9	#	So		(۩)	ARABIC PLACE OF SAJDAH
06EA..06ED	#	Mn	[4]	(◌ۭ..◌ۭ)	ARABIC EMPTY CENTRE LOW STOP..ARABIC SMALL LOW MEEM
06EE..06EF	#	Lo	[2]	(□..□)	ARABIC LETTER DAL WITH INVERTED V..ARABIC LETTER REH WITH INVERTED V
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
06FA..06FC	#	Lo	[3]	(ﺵ..ﻍ)	ARABIC LETTER SHEEN WITH DOT BELOW..ARABIC LETTER GHAIN WITH DOT BELOW
06FD..06FE	#	So	[2]	(ﻡ..ﻉ)	ARABIC SIGN SINDHI AMPERSAND..ARABIC SIGN SINDHI POSTPOSITION MEN
06FF	#	Lo		(□)	ARABIC LETTER HEH WITH INVERTED V
0700..070D	#	Po	[14]	(܀..܍)	SYRIAC END OF PARAGRAPH..SYRIAC HARKLEAN ASTERISCUS
070F	#	Cf		(◌)	SYRIAC ABBREVIATION MARK
0710	#	Lo		(ܐ)	SYRIAC LETTER ALAPH
0711	#	Mn		(◌ܑ)	SYRIAC LETTER SUPERSCRIPT ALAPH

0712..072F	#	Lo	[30]	(ܒ..ܨ)	SYRIAC LETTER BETH..SYRIAC LETTER PERSIAN DHALATH
0730..074A	#	Mn	[27]	(◌ܰ..◌ܺ)	SYRIAC PTHAHA ABOVE..SYRIAC BARREKH

Total: 1738 ...(omitting 243006 from listing)...

Let $print = [$graph $blank - $cntrl]

0020	#	Zs		( )	SPACE
0021..0023	#	Po	[3]	(!..#)	EXCLAMATION MARK..NUMBER SIGN
0024	#	Sc		($)	DOLLAR SIGN
0025..0027	#	Po	[3]	(%..')	PERCENT SIGN..APOSTROPHE
0028	#	Ps		(()	LEFT PARENTHESIS
0029	#	Pe		())	RIGHT PARENTHESIS
002A	#	Po		(*)	ASTERISK
002B	#	Sm		(+)	PLUS SIGN
002C	#	Po		(,)	COMMA
002D	#	Pd		(-)	HYPHEN-MINUS
002E..002F	#	Po	[2]	(../)	FULL STOP..SOLIDUS
0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
003A..003B	#	Po	[2]	(:..;)	COLON..SEMICOLON
003C..003E	#	Sm	[3]	(<..>)	LESS-THAN SIGN..GREATER-THAN SIGN
003F..0040	#	Po	[2]	(?..@)	QUESTION MARK..COMMERCIAL AT
0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
005B	#	Ps		([)	LEFT SQUARE BRACKET
005C	#	Po		(\)	REVERSE SOLIDUS
005D	#	Pe		(])	RIGHT SQUARE BRACKET
005E	#	Sk		(^)	CIRCUMFLEX ACCENT
005F	#	Pc		(_)	LOW LINE
0060	#	Sk		(`)	GRAVE ACCENT
0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
007B	#	Ps		({)	LEFT CURLY BRACKET
007C	#	Sm		(|)	VERTICAL LINE
007D	#	Pe		(})	RIGHT CURLY BRACKET
007E	#	Sm		(~)	TILDE
00A0	#	Zs		( )	NO-BREAK SPACE
00A1	#	Po		(¡)	INVERTED EXCLAMATION MARK
00A2..00A5	#	Sc	[4]	(¢..¥)	CENT SIGN..YEN SIGN
00A6..00A7	#	So	[2]	(¦..§)	BROKEN BAR..SECTION SIGN
00A8	#	Sk		(¨)	DIAERESIS
00A9	#	So		(©)	COPYRIGHT SIGN
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00AB	#	Pi		(«)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00AC	#	Sm		(¬)	NOT SIGN
00AD	#	Cf		()	SOFT HYPHEN
00AE	#	So		(®)	REGISTERED SIGN
00AF	#	Sk		(¯)	MACRON

00B0	#	So		(°)	DEGREE SIGN
00B1	#	Sm		(±)	PLUS-MINUS SIGN
00B2..00B3	#	No	[2]	(²..³)	SUPERSCRIPT TWO..SUPERSCRIPT THREE
00B4	#	Sk		(´)	ACUTE ACCENT
00B5	#	L&		(µ)	MICRO SIGN
00B6	#	So		(¶)	PILCROW SIGN
00B7	#	Po		(·)	MIDDLE DOT
00B8	#	Sk		(¸)	CEDILLA
00B9	#	No		(¹)	SUPERSCRIPT ONE
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00BB	#	Pf		(»)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00BC..00BE	#	No	[3]	(¼..¾)	VULGAR FRACTION ONE QUARTER..VULGAR FRACTION THREE QUARTERS
00BF	#	Po		(¿)	INVERTED QUESTION MARK
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL LETTER O WITH DIAERESIS
00D7	#	Sm		(×)	MULTIPLICATION SIGN
00D8..00F6	#	L&	[31]	(Ø..ö)	LATIN CAPITAL LETTER O WITH STROKE..LATIN SMALL LETTER O WITH DIAERESIS
00F7	#	Sm		(÷)	DIVISION SIGN
00F8..01BA	#	L&	[195]	(ø..ƺ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER EZH WITH TAIL
01BB	#	Lo		(ƻ)	LATIN LETTER TWO WITH STROKE
01BC..01BF	#	L&	[4]	(Ƽ..ƿ)	LATIN CAPITAL LETTER TONE FIVE..LATIN LETTER WYNN
01C0..01C3	#	Lo	[4]	(ǀ..ǃ)	LATIN LETTER DENTAL CLICK..LATIN LETTER RETROFLEX CLICK
01C4..0293	#	L&	[208]	(DŽ..ʓ)	LATIN CAPITAL LETTER DZ WITH CARON..LATIN SMALL LETTER EZH WITH CURL
0294	#	Lo		(ʔ)	LATIN LETTER GLOTTAL STOP
0295..02AF	#	L&	[27]	(ʕ..ʯ)	LATIN LETTER PHARYNGEAL VOICED FRICATIVE..LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
02B0..02C1	#	Lm	[18]	(ʰ..ˁ)	MODIFIER LETTER SMALL H..MODIFIER LETTER REVERSED GLOTTAL STOP
02C2..02C5	#	Sk	[4]	(˂..˅)	MODIFIER LETTER LEFT ARROWHEAD..MODIFIER LETTER DOWN ARROWHEAD
02C6..02D1	#	Lm	[12]	(ˆ..˒)	MODIFIER LETTER CIRCUMFLEX ACCENT..MODIFIER LETTER HALF TRIANGULAR COLON
02D2..02DF	#	Sk	[14]	(˓..˟)	MODIFIER LETTER CENTRED RIGHT HALF RING..MODIFIER LETTER CROSS ACCENT
02E0..02E4	#	Lm	[5]	(ˠ..ˤ)	MODIFIER LETTER SMALL GAMMA..MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
02E5..02EB	#	Sk	[7]	(˥..˫)	MODIFIER LETTER EXTRA-HIGH TONE BAR..MODIFIER LETTER YANG DEPARTING TONE MARK
02EC	#	Lm		(ˬ)	MODIFIER LETTER VOICING
02ED	#	Sk		(˭)	MODIFIER LETTER UNASPIRATED
02EE	#	Lm		(ˮ)	MODIFIER LETTER DOUBLE APOSTROPHE

02EF..02FF	#	Sk	[17]	(□..□)	MODIFIER LETTER LOW DOWN ARROWHEAD..MODIFIER LETTER LOW LEFT ARROW
0300..036F	#	Mn	[112]	(◌̀..◌ͯ)	COMBINING GRAVE ACCENT..COMBINING LATIN SMALL LETTER X
0370..0373	#	L&	[4]	(□..□)	GREEK CAPITAL LETTER HETA..GREEK SMALL LETTER ARCHAIC SAMPI
0374	#	Lm		(ʹ)	GREEK NUMERAL SIGN
0375	#	Sk		(͵)	GREEK LOWER NUMERAL SIGN
0376..0377	#	L&	[2]	(□..□)	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA..GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
037A	#	Lm		(◌ͺ)	GREEK YPOGEGRAMMENI
037B..037D	#	L&	[3]	(□..□)	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
037E	#	Po		(;)	GREEK QUESTION MARK
0384..0385	#	Sk	[2]	(΄..΅)	GREEK TONOS..GREEK DIALYTIKA TONOS
0386	#	L&		(Ά)	GREEK CAPITAL LETTER ALPHA WITH TONOS
0387	#	Po		(·)	GREEK ANO TELEIA
0388..038A	#	L&	[3]	(Έ..Ί)	GREEK CAPITAL LETTER EPSILON WITH TONOS..GREEK CAPITAL LETTER IOTA WITH TONOS
038C	#	L&		(Ό)	GREEK CAPITAL LETTER OMICRON WITH TONOS
038E..03A1	#	L&	[20]	(Ύ..Ρ)	GREEK CAPITAL LETTER UPSILON WITH TONOS..GREEK CAPITAL LETTER RHO
03A3..03F5	#	L&	[83]	(Σ..϶)	GREEK CAPITAL LETTER SIGMA..GREEK LUNATE EPSILON SYMBOL
03F6	#	Sm		(϶)	GREEK REVERSED LUNATE EPSILON SYMBOL
03F7..0481	#	L&	[139]	(Ϸ..ҁ)	GREEK CAPITAL LETTER SHO..CYRILLIC SMALL LETTER KOPPA
0482	#	So		(҂)	CYRILLIC THOUSANDS SIGN
0483..0487	#	Mn	[5]	(◌҃..◌҇)	COMBINING CYRILLIC TITLO..COMBINING CYRILLIC POKRYTIE
0488..0489	#	Me	[2]	(◌҈..◌҉)	COMBINING CYRILLIC HUNDRED THOUSANDS SIGN..COMBINING CYRILLIC MILLIONS SIGN
048A..0525	#	L&	[156]	(Ҋ..□)	CYRILLIC CAPITAL LETTER SHORT I WITH TAIL..CYRILLIC SMALL LETTER PE WITH DESCENDER
0531..0556	#	L&	[38]	(Ա..Ֆ)	ARMENIAN CAPITAL LETTER AYB..ARMENIAN CAPITAL LETTER FEH
0559	#	Lm		(◌ՙ)	ARMENIAN MODIFIER LETTER LEFT HALF RING
055A..055F	#	Po	[6]	(◌՚..◌՟)	ARMENIAN APOSTROPHE..ARMENIAN ABBREVIATION MARK
0561..0587	#	L&	[39]	(ա..և)	ARMENIAN SMALL LETTER AYB..ARMENIAN SMALL LIGATURE ECH YIWN
0589	#	Po		(։)	ARMENIAN FULL STOP
058A	#	Pd		(֊)	ARMENIAN HYPHEN
0591..05BD	#	Mn	[45]	(◌֑..◌ֽ)	HEBREW ACCENT ETNAHTA..HEBREW POINT METEG
05BE	#	Pd		(־)	HEBREW PUNCTUATION MAQAF
05BF	#	Mn		(◌ֿ)	HEBREW POINT RAFE

05C0	#	Po		(׀)	HEBREW PUNCTUATION PASEQ
05C1..05C2	#	Mn	[2]	(◌ׁ)	HEBREW POINT SHIN DOT..HEBREW POINT SIN DOT
05C3	#	Po		(׃)	HEBREW PUNCTUATION SOF PASUQ
05C4..05C5	#	Mn	[2]	(◌..◌)	HEBREW MARK UPPER DOT..HEBREW MARK LOWER DOT
05C6	#	Po		(◌)	HEBREW PUNCTUATION NUN HAFUKHA
05C7	#	Mn		(◌)	HEBREW POINT QAMATS QATAN
05D0..05EA	#	Lo	[27]	(א..ת)	HEBREW LETTER ALEF..HEBREW LETTER TAV
05F0..05F2	#	Lo	[3]	(וו..יי)	HEBREW LIGATURE YIDDISH DOUBLE VAV..HEBREW LIGATURE YIDDISH DOUBLE YOD
05F3..05F4	#	Po	[2]	(״..׳)	HEBREW PUNCTUATION GERESH..HEBREW PUNCTUATION GERSHAYIM
0600..0603	#	Cf	[4]	(◌..◌)	ARABIC NUMBER SIGN..ARABIC SIGN SAFHA
0606..0608	#	Sm	[3]	(◌..◌)	ARABIC-INDIC CUBE ROOT..ARABIC RAY
0609..060A	#	Po	[2]	(◌..◌)	ARABIC-INDIC PER MILLE SIGN..ARABIC-INDIC PER TEN THOUSAND SIGN
060B	#	Sc		(◌)	AFGHANI SIGN
060C..060D	#	Po	[2]	(،..◌)	ARABIC COMMA..ARABIC DATE SEPARATOR
060E..060F	#	So	[2]	(◌..◌)	ARABIC POETIC VERSE SIGN..ARABIC SIGN MISRA
0610..061A	#	Mn	[11]	(◌..◌)	ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM..ARABIC SMALL KASRA
061B	#	Po		(؛)	ARABIC SEMICOLON
061E..061F	#	Po	[2]	(؞..◌)	ARABIC TRIPLE DOT PUNCTUATION MARK..ARABIC QUESTION MARK
0621..063F	#	Lo	[31]	(◌..ء)	ARABIC LETTER HAMZA..ARABIC LETTER FARSI YEH WITH THREE DOTS ABOVE
0640	#	Lm		(ـ)	ARABIC TATWEEL
0641..064A	#	Lo	[10]	(ف..ي)	ARABIC LETTER FEH..ARABIC LETTER YEH
064B..065E	#	Mn	[20]	(◌ً..◌)	ARABIC FATHATAN..ARABIC FATHA WITH TWO DOTS
0660..0669	#	Nd	[10]	(٩..٠)	ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT NINE
066A..066D	#	Po	[4]	(٪..٭)	ARABIC PERCENT SIGN..ARABIC FIVE POINTED STAR
066E..066F	#	Lo	[2]	(ب..ق)	ARABIC LETTER DOTLESS BEH..ARABIC LETTER DOTLESS QAF
0670	#	Mn		(◌ٰ)	ARABIC LETTER SUPERSCRIPT ALEF
0671..06D3	#	Lo	[99]	(ٱ..ۓ)	ARABIC LETTER ALEF WASLA..ARABIC LETTER YEH BARREE WITH HAMZA ABOVE
06D4	#	Po		(۔)	ARABIC FULL STOP
06D5	#	Lo		(ە)	ARABIC LETTER AE
06D6..06DC	#	Mn	[7]	(◌ۖ..◌ۜ)	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA..ARABIC SMALL HIGH SEEN
06DD	#	Cf		(◌۝)	ARABIC END OF AYAH
06DE	#	Me		(◌۞)	ARABIC START OF RUB EL HIZB
06DF..06E4	#	Mn	[6]	(◌۟..◌۠)	ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL HIGH MADDA
06E5..06E6	#	Lm	[2]	(◌ۥ..◌ۦ)	ARABIC SMALL WAW..ARABIC SMALL YEH
06E7..06E8	#	Mn	[2]	(◌ۧ..◌ۨ)	ARABIC SMALL HIGH YEH..ARABIC SMALL HIGH NOON
06E9	#	So		(۩)	ARABIC PLACE OF SAJDAH

06EA..06ED	#	Mn	[4]	(◌۪..◌۫)	ARABIC EMPTY CENTRE LOW STOP..ARABIC SMALL LOW MEEM
06EE..06EF	#	Lo	[2]	(□..□)	ARABIC LETTER DAL WITH INVERTED V..ARABIC LETTER REH WITH INVERTED V
06F0..06F9	#	Nd	[10]	(۰..۹)	EXTENDED ARABIC-INDIC DIGIT ZERO..EXTENDED ARABIC-INDIC DIGIT NINE
06FA..06FC	#	Lo	[3]	(غ..ﺷ)	ARABIC LETTER SHEEN WITH DOT BELOW..ARABIC LETTER GHAIN WITH DOT BELOW
06FD..06FE	#	So	[2]	(ﷺ..ﷻ)	ARABIC SIGN SINDHI AMPERSAND..ARABIC SIGN SINDHI POSTPOSITION MEN
06FF	#	Lo		(□)	ARABIC LETTER HEH WITH INVERTED V
0700..070D	#	Po	[14]	(܀..܍)	SYRIAC END OF PARAGRAPH..SYRIAC HARKLEAN ASTERISCUS
070F	#	Cf		(◌)	SYRIAC ABBREVIATION MARK
0710	#	Lo		(ܐ)	SYRIAC LETTER ALAPH
0711	#	Mn		(◌ܑ)	SYRIAC LETTER SUPERSCRIPT ALAPH
0712..072F	#	Lo	[30]	(ܒ..ܯ)	SYRIAC LETTER BETH..SYRIAC LETTER PERSIAN DHALATH
0730..074A	#	Mn	[27]	(◌ܰ..◌ܺ)	SYRIAC PTHAHA ABOVE..SYRIAC BARREKH

Total: 1740 ...(omitting 243022 from listing)...

Let $word = [$alpha $gcAllMarks $digit \p{gc=Connector_Punctuation}]

0030..0039	#	Nd	[10]	(0..9)	DIGIT ZERO..DIGIT NINE
0041..005A	#	L&	[26]	(A..Z)	LATIN CAPITAL LETTER A..LATIN CAPITAL LETTER Z
005F	#	Pc		(_)	LOW LINE
0061..007A	#	L&	[26]	(a..z)	LATIN SMALL LETTER A..LATIN SMALL LETTER Z
00AA	#	L&		(ª)	FEMININE ORDINAL INDICATOR
00B5	#	L&		(µ)	MICRO SIGN
00BA	#	L&		(º)	MASCULINE ORDINAL INDICATOR
00C0..00D6	#	L&	[23]	(À..Ö)	LATIN CAPITAL LETTER A WITH GRAVE..LATIN CAPITAL LETTER O WITH DIAERESIS
00D8..00F6	#	L&	[31]	(Ø..ö)	LATIN CAPITAL LETTER O WITH STROKE..LATIN SMALL LETTER O WITH DIAERESIS
00F8..01BA	#	L&	[195]	(ø..ƺ)	LATIN SMALL LETTER O WITH STROKE..LATIN SMALL LETTER EZH WITH TAIL
01BB	#	Lo		(ƻ)	LATIN LETTER TWO WITH STROKE
01BC..01BF	#	L&	[4]	(Ƽ..ƿ)	LATIN CAPITAL LETTER TONE FIVE..LATIN LETTER WYNN
01C0..01C3	#	Lo	[4]	(ǀ..ǃ)	LATIN LETTER DENTAL CLICK..LATIN LETTER RETROFLEX CLICK
01C4..0293	#	L&	[208]	(DŽ..ʓ)	LATIN CAPITAL LETTER DZ WITH CARON..LATIN SMALL LETTER EZH WITH CURL
0294	#	Lo		(ʔ)	LATIN LETTER GLOTTAL STOP
0295..02AF	#	L&	[27]	(ʕ..ʯ)	LATIN LETTER PHARYNGEAL VOICED FRICATIVE..LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL
02B0..02C1	#	Lm	[18]	(ʰ..ˁ)	MODIFIER LETTER SMALL H..MODIFIER LETTER REVERSED GLOTTAL STOP
02C6..02D1	#	Lm	[12]		MODIFIER LETTER CIRCUMFLEX ACCENT..MODIFIER LETTER

				(˄..˞)	HALF TRIANGULAR COLON
02E0..02E4	#	Lm	[5]	(ˠ..ˤ)	MODIFIER LETTER SMALL GAMMA..MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
02EC	#	Lm		(□)	MODIFIER LETTER VOICING
02EE	#	Lm		(□)	MODIFIER LETTER DOUBLE APOSTROPHE
0300..036F	#	Mn	[112]	(◌̀..◌ͯ)	COMBINING GRAVE ACCENT..COMBINING LATIN SMALL LETTER X
0370..0373	#	L&	[4]	(□..□)	GREEK CAPITAL LETTER HETA..GREEK SMALL LETTER ARCHAIC SAMPI
0374	#	Lm		(ʹ)	GREEK NUMERAL SIGN
0376..0377	#	L&	[2]	(□..□)	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA..GREEK SMALL LETTER PAMPHYLIAN DIGAMMA
037A	#	Lm		(◌ͺ)	GREEK YPOGEGRAMMENI
037B..037D	#	L&	[3]	(□..□)	GREEK SMALL REVERSED LUNATE SIGMA SYMBOL..GREEK SMALL REVERSED DOTTED LUNATE SIGMA SYMBOL
0386	#	L&		(Ά)	GREEK CAPITAL LETTER ALPHA WITH TONOS
0388..038A	#	L&	[3]	(Έ..Ί)	GREEK CAPITAL LETTER EPSILON WITH TONOS..GREEK CAPITAL LETTER IOTA WITH TONOS
038C	#	L&		(Ό)	GREEK CAPITAL LETTER OMICRON WITH TONOS

Total: 725 ...(omitting 100957 from listing)...

============================

POSIX locale definition file constraints

$upper ‖ [$cntrl $digit $punct $space]
$upper ⊇ [A-Z]

$lower ‖ [$cntrl $digit $punct $space]
$lower ⊇ [a-z]

$alpha ‖ [$cntrl $digit $punct $space]
$alpha ⊇ [$lower $upper]

$digit ⊇ [0-9]

$alnum = [$alpha $digit]

$space ‖ [$upper $lower $alpha $digit $graph $xdigit]
$space ⊇ [$SP $FF $LF $CR] # $TAB $VTAB $NEL]
$space ⊇ $blank

$cntrl ‖ [$upper $lower $alpha $digit $punct $graph $print $xdigit]

$punct ‖ [$upper $lower $alpha $digit $cntrl $xdigit $SP]

$graph ⊇ [$upper $lower $alpha $digit $xdigit $punct]

$graph ‖ [$SP $cntrl]

$print ⊇ [$upper $lower $alpha $digit $xdigit $punct $graph $SP]
$print ‖ $cntrl

$xdigit ⊇ [$digit [a-f A-F]]

$blank ⊇ [$SP $TAB]

Extra POSIX 'POSIX locale' constraints

Let $C0Controls = [\u0000-\u001F]

0000..001F	#	Cc	[32]	(□..□)	..

Total: 32

$cntrl ⊇ $C0Controls

$punct ⊇ [[\u0021-\u007E] - [0-9 A-Z a-z]]

[$alpha $lower $upper $punct $digit $xdigit $alnum $space $blank $cntrl $graph $print $word] = [^\p{gc=unassigned} \p{gc=surrogate}]

**** SUMMARY ****

ParseErrorCount=0
TestFailureCount=0