

```
=====
== PXbase バンドル v0.5 <2010/06/15> ==
== by 「ZR」 (八登 崇之/Takayuki YATO) ==
== <zrbabbler@yahoo.co.jp> ==
=====

(This file is encoded in UTF-8,)
```

```
-----
pxcjkcat パッケージ (v0.4a) -- 和文文字カテゴリ操作
-----
```

■ 対応環境

upLaTeX2e (v0.11 以降)

■ 読込

```
\usepackage[<オプション>]{pxcjkcat}
```

オプションとして以下のものが指定できる。

ccv1

「モード CCV」を 1 (upTeX v0.11～0.28 と互換) とする。

ccv2

「モード CCV」を 2 (upTeX v0.29 以降と互換) とする。

※「モード CCV」については後で詳述する。

\cjkcategorymode で有効なモード値

kcatcode がモードに従って設定される。これ以外の場合は、パッケージ

読込時の kcatcode の変更はない。

■ 機能

```
\cjkcategor{<ブロック>,...}{<カテゴリ>}
```

<ブロック> で表される文字ブロック (複数指定が可能) の kcatcode を

<カテゴリ> に変更する。<ブロック> は「ブロック ID (後述)」または

「ASCII 以外の文字」で指定され、後者の場合はその文字の属するブロック

を指す。<カテゴリ> は以下に示す様に「カテゴリ ID」または数値で指定

する。この変更は局所的 (グルーピングに従う) である。

- noncjk (15): 欧文扱い
- kanji [または han] (16): 漢字扱い
- kana (17): 仮名扱い
- cjk (18): 「その他の和文」扱い
- hangul (19): ハングル扱い

ブロック ID の一覧および各々のカテゴリの意味については後掲。

`\cjkcategorymode{<モード>}`

全てのブロックの `kcatcode` の一括設定を行う。モードには以下のものが

ある。後に掲げるものほど `noncjk` であるブロックが増える。各モードで

の具体的な設定値については「各モードでの `kcatcode` の値」の節を参照。

- `forcecjk` : `upTeX` の既定の設定と同じ。(ただし「文字分類バージョン」

による小さい差異がある。詳細については「文字分類バージョン」の節を

参照。) ASCII 文字のみ `noncjk` で他のブロックは全て和文扱い (`noncjk`

以外; 具体的な値は後述)。和文フォントの中の Unicode 値の割当がある

全ての文字を和文文字として直接用いることができる。

- `prefercjk` : 和文扱いのブロックとして、`forcecjk` のものに加えて、

Adobe の定める CJK 文字集合 (Adobe-Japan1, Adobe-GB1, Adobe-CNS1,

Adobe-Korea1) の何れかと共通部分をもつ文字ブロックのみを加えて

(具体的な値は `forcecjk` と同じ)、残りを `noncjk` に設定する。

- `prefercjkvar` : `prefercjk` とほぼ同じで、違いは `grek`, `grek1`, `cyrl` の 3 つ (結果的にギリシャ・キリル文字の全て) が `noncjk` であること。

- `prefernoncjk` : 以下に掲げる「必要最低限」のブロックを除き、全てを

`noncjk` にする。

* `kanji`: 漢字・部首・注音字母: `hani`, `haniA`, `haniB`, `haniC`, `hani1`, `hani2`, `cjk01`, `cjk02`, `cjk03`, `cjk05`, `cjk06`, `bopo`, `bopo1`.

* `kana`: ひらがな・カタカナ: `hira`, `kana`, `kana1`.

* `cjk`: CJK 記号の一部・全角/半角互換形・彝文字: `cjk04`, `cjk08`, `cjk07`, `cjk09`, `cjk10`, `cjk11`, `cjk12`, `cjk13`, `sym15`, `yiii`, `yiii1`.

* `hangul`: ハングル完成形・ハングル字母: `hang`, `hang1`, `hang2`, `hangA`, `hangB`.

※モード `CCV` が 2 の場合、`cjk12` の再分割の中の `cjk1b`, `cjk1c` は

`kana` に変更される。

■ `kcatcode` の値の意味

`upTeX` (v0.11~) では Unicode 文字を複数の文字ブロックに分類し、各文字

ブロック毎に `kcatcode` と呼ばれるパラメタを持たせている。(文字ブロック

の分類は Unicode 文字ブロックとほぼ一致する。) これはそのブロックに

属する文字が入力ソース中に現れた時の `upTeX` の字句解析の動作を規定する。

なお以下では、入力文字コードが UTF-8 であると仮定する。また文字が属する

ブロックの `kcatcode` のことを単にその文字の `kcatcode` と呼ぶ。

(1) ある文字の `kcatcode` が 15 (`noncjk`) である場合、その文字は和文文字

として扱われず、UTF-8 表現のバイト列とみなされる。字句解析の結果は

それらのバイトの `catcode` の値に依存する。例えば、文字 α (`U+03B1`) の

`kcatcode` が 15 の状態で α がソースに現れたら、欧文 TeX において

`^^8e^^b1` というバイト列が現れた時と同じ結果になる。ここで `inputenc` で `utf8` (または `utf8x`) を指定してあれば、`inputenc` における「文字 α 」

の処理に回るはずである。

(2) ある文字の `kcatcode` が 15 以外である場合、その文字は 1 つの文字と

して扱われ、次の `catcode` をもつ文字と同様に字句解析される。

- `kcatcode` が 18 (`cjk`) \rightarrow `catcode` 12 と同様

- `kcatcode` が 16, 17, 19 \rightarrow `catcode` 11 と同様

そして、コントロールシーケンスの一部でないと判断された場合は、それ

は指定された `kcatcode` を保持する和文文字トークンとなる。

(3) `kcatcode` が 16, 17, 18 の和文文字トークンの直後の改行文字は無視

される (pTeX の和文と同じ)。 `kcatcode` が 19 の和文文字トークンの直後

の改行文字は空白トークンとなる (欧文と同じ)。これはハングルかななる

文書の組版に好都合だからである。

補足:

- ASCII 文字は常に欧文 TeX と同じ扱いになる。従って、ASCII ブロック

(latn) の kcatcode 指定は意味をもたない。

- kcatcode を表す upTeX のプリミティブは `\kcatcode<数値>` で、これはその数値を符号位置とする文字が属するブロックの kcatcode を指し示すレジスタを表す。

■ 文字分類バージョン

upTeX での文字ブロックの分割 (および各ブロックの kcatcode 値の既定値) は

改版時に変更され、これが互換性の問題を起こす可能性がある。

- a. Unicode の改版での文字ブロックの追加に追従する為の変更。
- b. 複数の文字種が混在する 1 つの Unicode 文字ブロックに対し、その中で文字種により処理を変えたいとする要望に応えるための Unicode 文字ブロックの再分割。例えば v0.29 の改版では Halfwidth and Fullwidth Forms (ブロック ID cjk12) が「再分割」されている。

本パッケージでは、ブロックの分割の互換性について以下の方針を採る。

- a のタイプの変更は「ある版の upTeX で未対応の文字ブロックの文字はその版では決して使われない」ことを仮定すれば互換性を損なうことがない。従って、これに対しては特に対策を行わない。
- b のタイプの変更 (「再分割」と呼ぶ) は、既存のソースの動作を変更することになるので、これに対しては互換性の為の対策を行う。
- ブロック ID を用いたブロック指定については、その対象文字集合が upTeX の版により変わらないようにする。
- モード指定 (`\cjkcategorymode`) を行う場合は、それによる kcatcode の

設定値が upTeX の版により変わらないようにする。(ただし、その版に

対応する Unicode の版で追加された文字を除く。)

- パッケージを読み込むだけでは kcatcode への変更は一切行われない。

従って、この場合の設定値は当然 upTeX の版に依存する。

具体的な対応をこれから述べる。まず、文字ブロックの「再分割」の違いを

「文字分類バージョン (CCV; Character Category Version)」と呼ぶことに

する。現状では次のものが存在する。

CCV 1 : upTeX v0.11 での定義

CCV 2 : upTeX v0.29 での定義

その上で、分割の変更について、以下のように対応する。ここでは CCV 2 に

おける cjk12 の再分割を例にする。

- cjk12 の再分割に関しては、再分割後のブロックに新たに ID を与える

(cjk1a, cjk1b, cjk1c)。cjk12 も引き続き使用可能である。

CCV が 2

以降の upTeX で cjk12 の kcatcode を変更する場合には、内部では

cjk1a~cjk1c を同時に変更する動作を行う。当然ながら CCV 1 の upTeX では cjk1a~cjk1c の指定は使えない。

CCV 2 では分割が変更されただけでなく、cjk1b と cjk1c の kcatcode の

既定値が 18(cjk) から 17(kana) に変更される。本パッケージでのモード

設定における kcatcode の設定値は upTeX の既定値を基礎として

いて、特に

`forcecjk` は既定値と全く同じ設定としている。既定値の変更については以下
のように対応する。

- パッケージオプションにおいて、「モード設定の際の設定値の基礎とする

CCV」(これを「モード CCV」と呼ぶ)を指定できるようにする。

すなわち

'ccvN' (N=1~2) でモード CCV が N になる。

- `forcecjk` の設定値は CCV が「モード CCV」である `upTeX` の既定値と

一致する。それ以外のモードの設定値もそれに応じて変わる。

- モード CCV の既定値は 1 とする。従って、*モード設定を使用した場合*

は、`kcatcode` の設定値は用いる `upTeX` の CCV に依存しない。

- モード設定オプション無しでパッケージを読み込んだだけの場合は、

`kcatcode` の設定は何も変更されない。従ってこの場合の設定値は用いる

`upTeX` の CCV に依存する。

例として次の場合を考える。

[`upTeX` の CCV が 1]

(1a) `\usepackage{pxcjkcat}`

(1b) `\usepackage[forcecjk]{pxcjkcat}`

(1c) `\usepackage[ccv1,forcecjk]{pxcjkcat}`

[`upTeX` の CCV が 2]

(2a) `\usepackage{pxcjkcat}`

(2b) `\usepackage[forcecjk]{pxcjkcat}`

(2c) `\usepackage[ccv1,forcecjk]{pxcjkcat}`

(2d) `\usepackage[ccv2,forcecjk]{pxcjkcat}`

(1a)(1b)(1c)(2b)(2c) は CCV 1 の既定値、(2a)(2d) は CCV 2 の既定値と

同じ値に設定される。

最後の注意点として、`\cjkcategory` でブロック指定に「文字」を使った場合

は単純にその文字の属する文字ブロックとみなされ、その動作は用いる `upTeX` の `CCV` に依存し、モード `CCV` とは無関係である。

■ Unicode ブロック ID 一覧

文字分類バージョン (`CCV`) 1 でのブロック。ただし [2] の注釈を付したもの

は `CCV 2` の `upTeX` で追加されたブロックである (先述の通り、Unicode 文字

ブロック追加による変更に関しては `CCV` は関知しない)。`CCV 1` の `upTeX` で

は、例えば `nkoo` の範囲は実際には `U+07C0~08FF` となる。

| ID | Unicode 範囲 | 名称 |
|--------------------|--------------|-----------------------------|
| <code>latn</code> | 0000 .. 007F | Basic Latin |
| <code>latn1</code> | 0080 .. 00FF | Latin-1 Supplement |
| <code>latnA</code> | 0100 .. 017F | Latin Extended-A |
| <code>latnB</code> | 0180 .. 024F | Latin Extended-B |
| <code>latn2</code> | 0250 .. 02AF | IPA Extensions |
| <code>sym01</code> | 02B0 .. 02FF | Spacing Modifier Letters |
| <code>sym02</code> | 0300 .. 036F | Combining Diacritical Marks |
| <code>grek</code> | 0370 .. 03FF | Greek and Coptic |
| <code>cyr1</code> | 0400 .. 04FF | Cyrillic |
| <code>cyr11</code> | 0500 .. 052F | Cyrillic Supplement |
| <code>armn</code> | 0530 .. 058F | Armenian |
| <code>hebr</code> | 0590 .. 05FF | Hebrew |
| <code>arab</code> | 0600 .. 06FF | Arabic |
| <code>sycr</code> | 0700 .. 074F | Syriac |
| <code>arab1</code> | 0750 .. 077F | Arabic Supplement |
| <code>thaa</code> | 0780 .. 07BF | Thaana |

| | | |
|-----------|---------|--|
| nkoo | 07C0 .. | 07FF Nko |
| samr [2] | 0800 .. | 08FF Samaritan |
| deva | 0900 .. | 097F Devanagari |
| beng | 0980 .. | 09FF Bengali |
| guru | 0A00 .. | 0A7F Gurmukhi |
| gujr | 0A80 .. | 0AFF Gujarati |
| orya | 0B00 .. | 0B7F Oriya |
| taml | 0B80 .. | 0BFF Tamil |
| telu | 0C00 .. | 0C7F Telugu |
| knda | 0C80 .. | 0CFF Kannada |
| mlym | 0D00 .. | 0D7F Malayalam |
| sinh | 0D80 .. | 0DFF Sinhala |
| thai | 0E00 .. | 0E7F Thai |
| laoo | 0E80 .. | 0EFF Lao |
| tibt | 0F00 .. | 0FFF Tibetan |
| mymr | 1000 .. | 109F Myanmar |
| geor | 10A0 .. | 10FF Georgian |
| hang1 | 1100 .. | 11FF Hangul Jamo |
| ethi | 1200 .. | 137F Ethiopic |
| ethi1 | 1380 .. | 139F Ethiopic Supplement |
| cher | 13A0 .. | 13FF Cherokee |
| cans | 1400 .. | 167F Unified Canadian Aboriginal Syllabics |
| ogam | 1680 .. | 169F Ogham |
| runr | 16A0 .. | 16FF Runic |
| tglg | 1700 .. | 171F Tagalog |
| hano | 1720 .. | 173F Hanunoo |
| buhd | 1740 .. | 175F Buhid |
| tagb | 1760 .. | 177F Tagbanwa |
| khmr | 1780 .. | 17FF Khmer |
| mong | 1800 .. | 187F Mongolian |
| cans1 [2] | 1880 .. | 18FF Unified Canadian Aboriginal Syllabics Extended |
| limb | 1900 .. | 194F Limbu |
| tale | 1950 .. | 197F Tai Le |
| talv | 1980 .. | 19DF New Tai Lue |

| | | |
|-----------|---------|--|
| khmr1 | 19E0 .. | 19FF Khmer Symbols |
| bugi | 1A00 .. | 1A1F Buginese |
| lana [2] | 1A20 .. | 1AFF Tai Tham |
| bali | 1B00 .. | 1B7F Balinese |
| sund [2] | 1B80 .. | 1BFF Sundanese |
| lepc [2] | 1C00 .. | 1C4F Lepcha |
| olck [2] | 1C50 .. | 1CCF Ol Chiki |
| sym38 [2] | 1CD0 .. | 1CFF Vedic Extensions |
| latn4 | 1D00 .. | 1D7F Phonetic Extensions |
| latn5 | 1D80 .. | 1DBF Phonetic Extensions Supplement |
| sym03 | 1DC0 .. | 1DFF Combining Diacritical Marks Supplement |
| latn3 | 1E00 .. | 1EFF Latin Extended Additional |
| grek1 | 1F00 .. | 1FFF Greek Extended |
| sym04 | 2000 .. | 206F General Punctuation |
| sym05 | 2070 .. | 209F Superscripts and Subscripts |
| sym06 | 20A0 .. | 20CF Currency Symbols |
| sym07 | 20D0 .. | 20FF Combining Diacritical Marks for Symbols |
| sym08 | 2100 .. | 214F Letterlike Symbols |
| sym09 | 2150 .. | 218F Number Forms |
| sym10 | 2190 .. | 21FF Arrows |
| sym11 | 2200 .. | 22FF Mathematical Operators |
| sym12 | 2300 .. | 23FF Miscellaneous Technical |
| sym13 | 2400 .. | 243F Control Pictures |
| sym14 | 2440 .. | 245F Optical Character Recognition |
| sym15 | 2460 .. | 24FF Enclosed Alphanumerics |
| sym16 | 2500 .. | 257F Box Drawing |
| sym17 | 2580 .. | 259F Block Elements |
| sym18 | 25A0 .. | 25FF Geometric Shapes |
| sym19 | 2600 .. | 26FF Miscellaneous Symbols |
| sym20 | 2700 .. | 27BF Dingbats |
| sym21 | 27C0 .. | 27EF Miscellaneous Mathematical Symbols-A |
| sym22 | 27F0 .. | 27FF Supplemental Arrows-A |
| brai | 2800 .. | 28FF Braille Patterns |
| sym23 | 2900 .. | 297F Supplemental Arrows-B |
| sym24 | 2980 .. | 29FF Miscellaneous Mathematical Symbols-B |

| | | |
|-----------|---------|--|
| sym25 | 2A00 .. | 2AFF Supplemental Mathematical Operators |
| sym26 | 2B00 .. | 2BFF Miscellaneous Symbols and Arrows |
| glag | 2C00 .. | 2C5F Glagolitic |
| latnC | 2C60 .. | 2C7F Latin Extended-C |
| copt | 2C80 .. | 2CFF Coptic |
| geor1 | 2D00 .. | 2D2F Georgian Supplement |
| tfng | 2D30 .. | 2D7F Tifinagh |
| ethi2 | 2D80 .. | 2DDF Ethiopic Extended |
| cyr1A [2] | 2DE0 .. | 2DFF Cyrillic Extended-A |
| sym27 | 2E00 .. | 2E7F Supplemental Punctuation |
| cjk01 | 2E80 .. | 2EFF CJK Radicals Supplement |
| cjk02 | 2F00 .. | 2FEF Kangxi Radicals |
| cjk03 | 2FF0 .. | 2FFF Ideographic Description Characters |
| cjk04 | 3000 .. | 303F CJK Symbols and Punctuation |
| hira | 3040 .. | 309F Hiragana |
| kana | 30A0 .. | 30FF Katakana |
| bopo | 3100 .. | 312F Bopomofo |
| hang2 | 3130 .. | 318F Hangul Compatibility Jamo |
| cjk05 | 3190 .. | 319F Kanbun |
| bopo1 | 31A0 .. | 31BF Bopomofo Extended |
| cjk06 | 31C0 .. | 31EF CJK Strokes |
| kana1 | 31F0 .. | 31FF Katakana Phonetic Extensions |
| cjk07 | 3200 .. | 32FF Enclosed CJK Letters and Months |
| cjk08 | 3300 .. | 33FF CJK Compatibility |
| haniA | 3400 .. | 4DBF CJK Unified Ideographs Extension A |
| sym28 | 4DC0 .. | 4DFF Yijing Hexagram Symbols |
| hani | 4E00 .. | 9FFF CJK Unified Ideographs |
| yiii | A000 .. | A48F Yi Syllables |
| yiii1 | A490 .. | A4CF Yi Radicals |
| lisu [2] | A4D0 .. | A4FF Lisu |
| vaii [2] | A500 .. | A63F Vai |
| cyr1B [2] | A640 .. | A69F Cyrillic Extended-B |
| bamu [2] | A6A0 .. | A6FF Bamum |
| sym29 | A700 .. | A71F Modifier Tone Letters |
| latnD | A720 .. | A7FF Latin Extended-D |

| | | |
|-----------|----------|------------------------------------|
| sylo | A800 .. | A82F Syloti Nagri |
| sym39 [2] | A830 .. | A83F Common Indic Number Forms |
| phag | A840 .. | AB7F Phags-pa |
| saur [2] | A880 .. | A8DF Saurashtra |
| deva1 [2] | A8E0 .. | A8FF Devanagari Extended |
| kali [2] | A900 .. | A92F Kayah Li |
| rjng [2] | A930 .. | A95F Rejang |
| hangA [2] | A960 .. | A97F Hangul Jamo Extended-A |
| java [2] | A980 .. | A9FF Javanese |
| cham [2] | AA00 .. | AA5F Cham |
| mymrA [2] | AA60 .. | AA7F Myanmar Extended-A |
| tavt [2] | AA80 .. | ABBF Tai Viet |
| mtei [2] | AB00 .. | ABFF Meetei Mayek |
| hang | AC00 .. | D7AF Hangul Syllables |
| hangB [2] | D7B0 .. | D7FF Hangul Jamo Extended-B |
| spc01 | D800 .. | DB7F High Surrogates |
| spc02 | DB80 .. | DBFF High Private Use Surrogates |
| spc03 | DC00 .. | DFFF Low Surrogates |
| spc04 | E000 .. | F8FF Private Use Area |
| hani1 | F900 .. | FAFF CJK Compatibility Ideographs |
| latn6 | FB00 .. | FB4F Alphabetic Presentation Forms |
| arab2 | FB50 .. | FDFD Arabic Presentation Forms-A |
| spc05 | FE00 .. | FE0F Variation Selectors |
| ckj09 | FE10 .. | FE1F Vertical Forms |
| sym30 | FE20 .. | FE2F Combining Half Marks |
| ckj10 | FE30 .. | FE4F CJK Compatibility Forms |
| ckj11 | FE50 .. | FE6F Small Form Variants |
| arab3 | FE70 .. | FEFF Arabic Presentation Forms-B |
| ckj12 | FF00 .. | FFEF Halfwidth and Fullwidth Forms |
| spc06 | FFF0 .. | FFFF Specials |
| linb | 10000 .. | 1007F Linear B Syllabary |
| linb1 | 10080 .. | 100FF Linear B Ideograms |
| sym31 | 10100 .. | 1013F Aegean Numbers |
| grek2 | 10140 .. | 1018F Ancient Greek Numbers |
| sym40 [2] | 10190 .. | 101CF Ancient Symbols |

| | | | |
|-------|-----|----------------|-----------------------------------|
| sym41 | [2] | 101D0 .. 1027F | Phaistos Disc |
| lyci | [2] | 10280 .. 1029F | Lycian |
| cari | [2] | 102A0 .. 102FF | Carian |
| ital | | 10300 .. 1032F | Old Italic |
| goth | | 10330 .. 1037F | Gothic |
| ugar | | 10380 .. 1039F | Ugaritic |
| xpeo | | 103A0 .. 103FF | Old Persian |
| dsrt | | 10400 .. 1044F | Deseret |
| shaw | | 10450 .. 1047F | Shavian |
| osma | | 10480 .. 107FF | Osmanya |
| cpri | | 10800 .. 1083F | Cypriot Syllabary |
| armi | [2] | 10840 .. 108FF | Imperial Aramaic |
| phnx | | 10900 .. 1091F | Phoenician |
| lydi | [2] | 10920 .. 109FF | Lydian |
| khar | | 10A00 .. 10A5F | Kharoshthi |
| sarb | [2] | 10A60 .. 10AFF | Old South Arabian |
| avst | [2] | 10B00 .. 10B3F | Avestan |
| prti | [2] | 10B40 .. 10B5F | Inscriptional Parthian |
| phli | [2] | 10B60 .. 10BFF | Inscriptional Pahlavi |
| orkh | [2] | 10C00 .. 10E5F | Old Turkic |
| sym42 | [2] | 10E60 .. 1107F | Rumi Numeral Symbols |
| kthi | [2] | 11080 .. 11FFF | Kaithi |
| xsux | | 12000 .. 123FF | Cuneiform |
| xsux1 | | 12400 .. 12FFF | Cuneiform Numbers and Punctuation |
| egyp | [2] | 13000 .. 1CFFF | Egyptian Hieroglyphs |
| sym32 | | 1D000 .. 1D0FF | Byzantine Musical Symbols |
| sym33 | | 1D100 .. 1D1FF | Musical Symbols |
| sym34 | | 1D200 .. 1D2FF | Ancient Greek Musical Notation |
| sym35 | | 1D300 .. 1D35F | Tai Xuan Jing Symbols |
| sym36 | | 1D360 .. 1D3FF | Counting Rod Numerals |
| sym37 | | 1D400 .. 1EFFF | Mathematical Alphanumeric Symbols |
| sym43 | [2] | 1F000 .. 1F02F | Mahjong Tiles |
| sym44 | [2] | 1F030 .. 1F0FF | Domino Tiles |
| sym45 | [2] | 1F100 .. 1F1FF | Enclosed Alphanumeric Supplement |
| cjk13 | [2] | 1F200 .. 1FFFF | Enclosed Ideographic Supplement |

| | | |
|-----------|------------------|---|
| haniB | 20000 .. 2A6FF | CJK Unified Ideographs Extension B |
| haniC [2] | 2A700 .. 2F7FF | CJK Unified Ideographs Extension C |
| hani2 | 2F800 .. 2FFFF | CJK Compatibility Ideographs Supplement |
| spc07 | E0000 .. E00FF | Tags |
| spc08 | E0100 .. EFFFF | Variation Selectors Supplement |
| spc09 | F0000 .. FFFFF | Supplementary Private Use Area-A |
| spc10 | 100000 .. 10FFFF | Supplementary Private Use Area-B |

※ ID の命名規則

- Unicode ブロック名にスクリプト (用字系) の名前が含まれるものは、それ
 に対する ISO 15924 のコードを用いた。単一のスクリプトのブロックが
 複数ある場合は、名前が "Extended-A, B, ..." のものは A, B, ... を、

それ以外のものは 1, 2, ... (一部符号値順でない) を末尾に付加した。

"Arabic" → arab ; "Latin Extended-C" → latnC

なお、Hiragana は hira だが Katakana は kana であることに注意。

- それ以外は、CJK 関係 (cjk)、特殊用途 (spc)、それ以外 (sym) の 3 つに

恣意的に分類して、2 桁の番号を付けた。この番号は基本的に符号値の

順とするが、後から追加されたブロックはそうならない。

CCV 2 では以下の変更がある。

- cjk12 (Halfwidth and Fullwidth Forms) が次の 3 つに分割される。

* cjk1b : U+FF10..U+FF19, U+FF21..U+FF3A, U+FF41..U+FF5A
 すなわち ASCII 英数字の全角互換形。

* cjk1c : U+FF66..U+FF6F, U+FF71..U+FF9D
 すなわちカタカナの半角互換形。句読点等の記号は含まない。

* cjk1a : cjk12 から cjk1b, cjk1c を除いた残り。

cjk12 の指定も使用可能である。

■ 各モードでの kcatcode の値

モード CCV が 1 の場合。

| 記号 | fc | pc | pcv | pnc | JIS | CID | 名称 |
|-------|----|----|-----|-----|-----|-----|---|
| latn | X | - | - | - | | * | Basic Latin |
| latn1 | 0 | - | - | X | * | * | Latin-1 Supplement |
| latnA | 0 | - | - | X | | * | Latin Extended-A |
| latnB | 0 | - | - | X | | * | Latin Extended-B |
| latn2 | 0 | - | - | X | | * | IPA Extensions |
| sym01 | 0 | - | - | X | | * | Spacing Modifier Letters |
| sym02 | 0 | - | - | X | | * | Combining Diacritical Marks |
| grek | 0 | - | X | - | * | * | Greek and Coptic |
| cyrl | 0 | - | X | - | * | * | Cyrillic |
| hang1 | H | - | - | - | | + | Hangul Jamo |
| latn3 | 0 | - | - | X | | * | Latin Extended Additional |
| grek1 | 0 | - | X | - | | * | Greek Extended |
| sym04 | 0 | - | - | X | * | * | General Punctuation |
| sym05 | 0 | - | - | X | | * | Superscripts and Subscripts |
| sym06 | 0 | - | - | X | | * | Currency Symbols |
| sym07 | 0 | - | - | X | | * | Combining Diacritical Marks for Symbols |
| sym08 | 0 | - | - | X | * | * | Letterlike Symbols |
| sym09 | 0 | - | - | X | | * | Number Forms |
| sym10 | 0 | - | - | X | * | * | Arrows |
| sym11 | 0 | - | - | X | * | * | Mathematical Operators |
| sym12 | 0 | - | - | X | * | * | Miscellaneous Technical |
| sym13 | 0 | - | - | X | | * | Control Pictures |
| sym15 | 0 | - | - | - | | * | Enclosed Alphanumerics |
| sym16 | 0 | - | - | X | * | * | Box Drawing |
| sym17 | 0 | - | - | X | | * | Block Elements |
| sym18 | 0 | - | - | X | * | * | Geometric Shapes |
| sym19 | 0 | - | - | X | * | * | Miscellaneous Symbols |
| sym20 | 0 | - | - | X | | * | Dingbats |

| | | | | | | |
|-------|---|---|---|---|---|---|
| sym23 | 0 | - | - | X | * | Supplemental Arrows-B |
| sym24 | 0 | - | - | X | * | Miscellaneous Mathematical Symbols-B |
| sym26 | 0 | - | - | X | * | Miscellaneous Symbols and Arrows |
| cjk01 | I | - | - | - | * | CJK Radicals Supplement |
| cjk02 | I | - | - | - | * | Kangxi Radicals |
| cjk03 | I | - | - | - | + | Ideographic Description Characters |
| cjk04 | 0 | - | - | - | * | CJK Symbols and Punctuation |
| hira | K | - | - | - | * | Hiragana |
| kana | K | - | - | - | * | Katakana |
| bopo | I | - | - | - | + | Bopomofo |
| hang2 | H | - | - | - | + | Hangul Compatibility Jamo |
| cjk05 | I | - | - | - | * | Kanbun |
| bopo1 | I | - | - | - | + | Bopomofo Extended |
| cjk06 | I | - | - | - | + | CJK Strokes |
| kana1 | K | - | - | - | * | Katakana Phonetic Extensions |
| cjk07 | 0 | - | - | - | * | Enclosed CJK Letters and Months |
| cjk08 | 0 | - | - | - | * | CJK Compatibility |
| haniA | I | - | - | - | * | CJK Unified Ideographs Extension A |
| hani | I | - | - | - | * | CJK Unified Ideographs |
| yiii | 0 | - | - | - | + | Yi Syllables |
| yiii1 | 0 | - | - | - | + | Yi Radicals |
| hangA | H | - | - | - | | Hangul Jamo Extended-B |
| hang | H | - | - | - | + | Hangul Syllables |
| hangB | H | - | - | - | | Hangul Jamo Extended-B |
| spc04 | 0 | - | - | X | * | Private Use Area |
| hani1 | I | - | - | - | * | CJK Compatibility Ideographs |
| latn6 | 0 | - | - | X | * | Alphabetic Presentation Forms |
| cjk09 | 0 | - | - | - | * | Vertical Forms |
| cjk10 | 0 | - | - | - | * | CJK Compatibility Forms |
| cjk11 | 0 | - | - | - | + | Small Form Variants |
| cjk12 | 0 | - | - | - | * | Halfwidth and Fullwidth Forms |
| cjk13 | 0 | - | - | - | | Enclosed Ideographic Supplement |
| haniB | I | - | - | - | * | CJK Unified Ideographs Extension B |
| haniC | I | - | - | - | | CJK Unified Ideographs Extension C |
| hani2 | I | - | - | - | * | CJK Compatibility Ideographs Supplement |

0 X - - (その他のブロック)

all, prc, pc1, pnc の欄はそれぞれ forcecjk, prefercjk, prefercjkvar, prefernoncjk モードでの kcatcode の値。値の意味: X=noncjk(15), I=kanji(16), K=kana(17), O=cjk(18), H=hanguk(19)。JIS 欄の * はそのブロック内の文字に JIS X 0208 に含まれるものを示す。

また CID 欄の * は Adobe-Japan1 に、+ は Adobe-GB1, Adobe-CNS1, Adobe-Korea1 に含まれるものを示す。

モード CCV が 2 の場合、cjk12 の再分割について以下になる。

| 記号 | fc | pc | pcv | pnc | JIS | CID | |
|-------|----|----|-----|-----|-----|-----|-------------------------|
| cjk1a | 0 | - | - | - | * | * | cjk12 で cjk1b, cjk1c 以外 |
| cjk1b | K | - | - | - | * | * | ASCII 英数字の全角互換形 |
| cjk1c | K | - | - | - | * | * | カタカナの半角互換形 (記号は含まず) |

※ upTeX の CCV が 2 でモード CCV 指定が 1 の場合は cjk1a～cjk1c の設定値は cjk12 のものに従う。

■ 更新履歴

README ファイルを参照。