Character Description Language (CDL):
The Set of Basic CJK Unified Stroke Types

Source: Tom Bishop <tbishop@wenlin.com> and Richard Cook <rscook@unicode.org>
Status: Expert Contribution
Date: May 23, 2004, 4:28 pm [revision of 20040523:16:27:07]
Action: For UTC and IRG consideration

This document is part of the *Specification of CDL* outlined in L2/03-404. See also L2/03-387 for additional discussion and examples of CDL usage. For information on the CDL specification and its implementations, see <http://www.wenlin.com/cdl/>.

1) The Set of Types

Table 1 below lists the set of 39 Basic Stroke Types currently implemented in the CDL descriptions of more than 40,000 ISO/IEC 10646 "CJK Unified Ideographs" (including all BMP, and 12,000 SIP forms).

The eleven headers *A*..*K* in Table 1 are as follows:

- *A* Sequential numbering [1..39] of all current types;

- *B* Numeric index for the 5 札 *zhá* types [1..5], with alphabetic sub-types [a..z];

- *C* Total number of 折 *zhé* 'transitional bends' (+1 = number of segments) in the type;

- *D* Total number of control points currently implemented for the type;

- *E* Frequency (non-recursive) of this type in current descriptions, as a percentage of total;

- *F* Glyph exemplifying the type in isolation (outside of compounds);

- *G* Provisional assignment of an ISO/IEC 10646 *UCS Scalar Value* for each exemplar in F, or *PUA* (Private Use Area) for unencoded forms;

- *H* Name of the type in Han characters;

- *I* Romanization in *pīnyīn* of H;

- *J* Abbreviation for the *pīnyīn* name of the type in I (acronymic, except for 39);

- *K* Notes on the type, including structural analysis (not necessarily tied to the actual implementation), unified variants of the type, examples of usage in compounds, and cross-references to similar types.

Table 1: Set of Basic CJK Unified Stroke Types

#	札	折	點	分	體	碼	名	拼	縮	注
A	B	C	D	E	F	G	H	I	J	K
1	1a	0	2	26.87	⼀	U+4e00	橫	*héng*	**h**	horizontal; as in 三十卅; or as ⼀ in 七毛戈; *cp.* ╱ **t**
2	1b	0	2	03.45	╱	PUA	提	*tí*	**t**	⼀ **h** + taper; 3rd stroke of 扌 as in 地; stroke 5 of 虫
3	2a	0	2	15.77	│	U+4e28	竪	*shù*	**s**	vertical; as in 中卜上; or as ∕ in 五 and 丑
4	2b	1	3	01.13	⌋	U+4e85	竪钩	*shù-gōu*	**sg**	│ **s** + left hook; as in 小丁宁才; *cp.* ⌊ **st**
5	3a	0	2	12.54	∕	U+4e3f	撇	*piě*	**p**	falling to left, not very curved; as in 八彳行; *cp.* ∕ **wp** and ∫ **sp**
6	3b	0	2	03.95	∫	PUA	弯撇	*wān-piě*	**wp**	curve + ∕ **p**; as in 大; *cp.* ∕ **p** and ∫ **sp**
7	3c	1	3	03.22	∫	PUA	竪撇	*shù-piě*	**sp**	│ **s** + ∕ **wp**; as in 厂; *cp.* ∫ **wp** and ∕ **p**
8	4a	0	2	09.59	╲	U+4e36	点	*diǎn*	**d**	taper + clockwise curve; as in 主; sometimes to left, as ╱ 1st in 火
9	4b	0	2	03.52	╲	PUA	捺	*nà*	**n**	falling right counter-clockwise curve; as in 人; *cp.* ∕ **p** and ╲ **pn**
10	4c	0	3	00.03	╲	PUA	点捺	*diǎn-nà*	**dn**	╲ **d** + ╲ **n**; 2nd stroke in 入, only in 入 (=入) and its compounds, *e.g.* 仚

## Table 1: Set of Basic CJK Unified Stroke Types

#	札	折	點	分	體	碼	名	拼	縮	注
A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
11	4d	1	3	00.43	⟍	*PUA*	平捺	*píng-nà*	**pn**	㇏ **n** + 一 **h**; last stroke in 是走辶; *cp.* ㇏ **n** and ㇄ **sw**
12	4e	1	3	00.11	㇏	U+4e40	提捺	*tí-nà*	**tn**	㇀ **t** + ㇏ **n**; last stroke in 夂 (=夊); or as ㇀ in 八入吏; *cp.* ㇈ **tpn**
13	4f	1	4	00.08	㇏	*PUA*	提平捺	*tí-píng-nà*	**tpn**	㇀ **t** + ⟍ **pn**; last stroke in 辶之; *cp.* ㇏ **tn**
14	5a	1	3	03.28	㇕	U+200cd	橫折	*héng-zhé*	**hz**	一 **h** + ㇑ **s**; 2nd stroke in 口; *cp.* ㇆ **hzg**
15	5b	1	3	00.90	㇇	*PUA*	橫撇	*héng-piě*	**hp**	一 **h** + ノ **p**; 1 in 又; 4 in 今; *cp.* ㇖ **hg**
16	5c	1	3	01.36	㇖	U+4e5b	橫钩	*héng-gōu*	**hg**	一 **h** + left hook; 2 in 冖写; *cp.* ㇇ **hp**
17	5d	1	3	02.54	㇄	U+200ca	竖折	*shù-zhé*	**sz**	㇑ **s** + 一 **h**; as in 山; or as ㇄ (ノ **s** + 一 **h**) in 乐牛东互; *cp.* ㇗ **pz**
18	5e	1	4	00.17	㇄	*PUA*	竖弯	*shù-wān*	**sw**	㇑ **s** + ⟍ **pn**; stroke 4 in 四
19	5f	1	3	01.36	㇙	U+2010c	竖提	*shù-tí*	**st**	㇑ **s** + ㇀ **t** (right hook); as in 民辰; *cp.* �亅 **sg**
20	5g	1	3	00.51	㇗	*PUA*	撇折	*piě-zhé*	**pz**	ノ **p** + ㇀ **t**; 3 in 公; stroke 1 in 厶; *cp.* ㇄ **sz**
21	5h	1	3	00.11	㇛	U+21fe8	撇点	*piě-diǎn*	**pd**	ノ **p** + 丶 **d**; stroke 1 in 女巜巛粵
22	5i	1	3	00.00	ノ	*PUA*	撇钩	*piě-gōu*	**pg**	ノ **p** + left-rising hook; as in 乂; *cp.* ノ **p**

Table 1: Set of Basic CJK Unified Stroke Types

| #	札	折	點	分	體	碼	名	拼	縮	注
A	B	C	D	E	F	G	H	I	J	K
23	5j	1	4	00.24	⟩	PUA	弯钩	*wān-gōu*	**wg**	curving ⟩ **sg**; 3 in 豕, 6 in 家
24	5k	1	3	01.81	⟍	PUA	斜钩	*xié-gōu*	**xg**	⟍ **n** + up hook; 5 in 我; 2 in 弋; *cp.* ⌞ **swg**
25	5l	2	4	00.14	⅂	PUA	横折折	*héng-zhé-zhé*	**hzz**	一 **h** +⌞ **sz** or ⅂ **hz** + 一 **h**; 2nd stroke in 凹; *cp.* ⅂ **hzw**, ⅂ **hzwg**
26	5m	2	5	00.03	⅂	PUA	横折弯	*héng-zhé-wān*	**hzw**	一 **h** +⌞ **sw**; 2 in 朵殳; *cp.* ⅂ **hzwg**
27	5n	2	4	00.18	⅃	PUA	横折提	*héng-zhé-tí*	**hzt**	一 **h** + ⌡ **st**; 2 in 讠, as in 记鸠; *cp.* ⅂ **hzz**, ⅂ **hzw**
28	5o	2	4	02.22	⅂	U+200cc	横折钩	*héng-zhé-gōu*	**hzg**	一 **h** + ⟩ **sg**; 2 in 月丹; or as ⅂ in 勺万; or as ⟋ in 也乜; *cp.* ⅂ **hz**
29	5p	2	4	00.28	⅂	U+2e84	横斜钩	*héng-xié-gōu*	**hxg**	一 **h** + ⟍ **xg**; 1st stroke in 飞卂; also in 凬气; *cp.* ⅂ **hzwg**
30	5q	2	4	00.44	⌐⅂	U+200d1	竖折折	*shù-zhé-zhé*	**szz**	⏐ **s** + ⅂ **hz** or ⌞ + ⏐; as 4 in 亞, 6 in 鼎, 11 in 龍; *cp.* ⅃ **szp**, ⌐ **szzg**
31	5r	2	4	00.11	⅄	PUA	竖折撇	*shù-zhé-piě*	**szp**	⟋ **s** + 一 **hg** / ⁊ **hp**; as in 专咼; or as ⅄ in 旻設叚夆; *cp.* ⌐ **szzg**
32	5s	2	5	01.84	⌞	U+4e5a	竖弯钩	*shù-wān-gōu*	**swg**	⌞ **sw** + up hook; as in 儿礼心; *cp.* ⌞ **sw**

Table 1: Set of Basic CJK Unified Stroke Types

#	札	折	點	分	體	碼	名	拼	縮	注
A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
33	5t	3	5	00.06	㇅	*PUA*	横折折折	*héng-zhé-zhé-zhé*	**hzzz**	㇕ **hz** + ㇕ **hz**; 4th stroke in 凸㒳; *cp.* ㇡ **hzzp**
34	5u	3	5	00.09	㇋	*PUA*	横折折撇	*héng-zhé-zhé-piě*	**hzzp**	⇁ **hg** + ㇇ **hp**; 1 in ㇃建; 2 in 及; *cp.* ㇅ **hzzz**, ㇆ **hzzzg**
35	5v	3	6	00.60	乙	U+4e59	横折弯钩	*héng-zhé-wān-gōu*	**hzwg**	一 **h** + ㇄ **swg**; stroke 19 in 黻; or as ㇈ stroke 2 in 九几風; *cp.* ㇉ **hzw**, ㇅ **hzz**, ㇂ **hxg**
36	5w	3	6	00.03	㇌	*PUA*	横撇弯钩	*héng-piě-wān-gōu*	**hpwg**	⇁ **hg** + ㇀ **wg**; 1 in 阝队; *cp.* ㇆ **hzzzg**
37	5x	3	5	00.92	㇗	*PUA*	竖折折钩	*shù-zhé-zhé-gōu*	**szzg**	㇓ **s** + ㇖ **hzg**; 2 in 马丂; *cp.* ㇄ **szz**, ㇙ **szp**
38	5y	4	6	00.11	㇆	U+2010e	横折折折钩	*héng-zhé-zhé-zhé-gōu*	**hzzzg**	⇁ **hg** + ㇖ **hzg**; 1 in 乃仍; *cp.* ㇌ **hpwg**, ㇅ **hzzz**, ㇋ **hzzp**
39	5z	1	2	00.06	◯	U+3007	圈	*quān*	**o**	circle; bottom of 智쑹쓯; points are for bounding rectangle

2)   Analysis of the set of Unified Basic Stroke Types

Table 2 below presents multi-dimensional feature analysis of the set of basic types. This analysis is given in terms of basic *segments* and transitional *junctures* between segments, and in terms of *vertical* (X), *horizontal* (Y), and *curvature* (Z) dimensions. For each of the *X, Y, Z* dimensions, the *directionality* of the stroke is indicated as follows:

$X =>$ **lr** 'left-to-right', **rl** 'right-to-left, **0** 'zero lateral movement';

$Y =>$ **tb** 'top-to-bottom', **bt** 'bottom-to-top', **0** 'zero longitudinal movement';

$Z =>$ **cw** 'clockwise', **ccw** 'counter-clockwise', **0** 'zero curvature'.

The total number of segments (T) for a given type may be written as T = C + 1, where *C* is equal to the number of junctures (column *C*). Each type is described with T elements in each of the *X, Y* and *Z* columns, where "+" indicates the juncture. Junctures are of two types, *curved* (gradual) and *sharp* (corner), and all curved junctures are associated with curvature of at least one of the conjoined segments. The relation between the number of transitions (column *C*) and the number of points (column *D*) is D = *sharp* + (*curved* * 2) + 2; when C = 0, D = 2. Elements separated by "|" indicate unified variants (parenthesized for T > 1), and column *F* includes several examples of such unifications (*i.e.* 2a, 4a, 4e, 5d, 5o, 5r, also given with examples in the notes in column K of Table 1). A trailing "+" in column *Z* indicates additional curvature, differentiating two pairs of types (3a,b and 4e,f).

See column K of Table 1 for analysis of the complex types into basic segments. The set of 7 basic segmental elements (those with zero transitions) is as follows:

<div align="center">一 h, ╱ t, ｜ s, ╱ p, ╰ wp, ╲ d, ╲ n</div>

This set may be reduced by 2 to a set of 5, by applying transformations to the **t** (+taper) and **wp** (+curve) types, relative to base types **h** and **p**, respectively.

Note that combination of segments, basic or not, always results in a number of transitions equal to one less than the number of combined segments. So, for example, the ╰ **pn** stroke has 1 transition (it is composed of ╲ **n** + 一 **h**), while ╰ **tpn** also has 1 transition (from ╱ **t** to ╰ **pn**) rather than 2 (╱ **t** + ╲ **n** + 一 **h**), which would ignore the higher level grouping for ╰ **pn** (= ╲ **n** + 一 **h**). Similarly, it should perhaps be emphasized that stroke count for the each of the 39 types is always 1, no matter how many junctures.

Future treatment of the 3 撇 *piě* types ╱ **p**, ╯ **wp**, and ╯ **sp** might involve unification, using a variable number of control points, though these are currently distinct in the implementation (note that there is at present only one encoded *piě* type, U+4e3f). Other unifications within the overall set might be possible as well, *e.g.* ╲ **n** with ╰ **pn**, and ╲ **tn** with ╰ **tpn**. The set of 39 given here does however seem to have general validity and wide acceptance, in terms of modern orthographic practices, especially as evident in the representative forms appearing in modern typography, and in the ISO/IEC 10646 codecharts. Refinements to the set of types will likely involve additions needed to accommodate very rare forms.

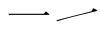# Table 2: Analysis of the set of Unified Basic Stroke Types
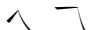
札	折	點	縮	體	橫	竖	弯
B	C	D	J	F	X	Y	Z
1a	0	2	**h**	一 ⁻	lr	0|bt	0
1b	0	2	**t**	⁄	lr	bt	0
2a	0	2	**s**	| ╱	0|rl	tb	0
2b	1	3	**sg**	⌡	0+lr	tb+bt	0+(0|cw)
3a	0	2	**p**	╱	rl	tb	cw
3b	0	2	**wp**	╱	rl	tb	cw+
3c	1	3	**sp**	╰	0+rl	tb+tb	0+cw
4a	0	2	**d**	丶 ╱	lr|rl	tb	cw
4b	0	2	**n**	╲	lr	tb	ccw
4c	1	3	**dn**	╲	lr+lr	tb+tb	cw+ccw
4d	1	3	**pn**	╰	lr+lr	tb+0	ccw+0
4e	1	3	**tn**	⟍ ⟋	lr+lr	(bt|0)+tb	0+ccw
4f	1	4	**tpn**	╰	lr+lr	bt+tb	0+ccw+
5a	1	3	**hz**	⌐	lr+0	0+tb	0+0
5b	1	3	**hp**	⌐	lr+rl	0+tb	0+cw
5c	1	3	**hg**	→	lr+rl	0+tb	0+0
5d	1	3	**sz**	∟ ╱	(0|rl)+lr	tb+0	0+0
5e	1	4	**sw**	∪	0+lr	tb+0	0+ccw
5f	1	3	**st**	↓	0+lr	tb+bt	0+0
5g	1	3	**pz**	╱	rl+lr	tb+bt	cw+0

Table 2: Analysis of the set of Unified Basic Stroke Types

札	折	點	縮	體	横	竖	弯
B	*C*	*D*	*J*	*F*	*X*	*Y*	*Z*
5h	1	3	**pd**	〈	rl+lr	tb+tb	cw+cw
5i	1	3	**pg**	╱	rl+rl	tb+bt	cw+0
5j	1	4	**wg**	⌐	lr+rl	tb+bt	cw+cw
5k	1	3	**xg**	⌐	lr+0	tb+bt	ccw+0
5*l*	2	4	**hzz**	⌐	lr+0+lr	0+tb+0	0+0+0
5m	2	5	**hzw**	⌐	lr+0+lr	0+tb+0	0+0+ccw
5n	2	4	**hzt**	⌐	lr+0+lr	0+tb+bt	0+0+0
5o	2	4	**hzg**	⌐⌐⌐	lr+(0|rl)+rl	(0|bt)+tb+bt	0+(0|cw)+0
5p	2	4	**hxg**	⌐	lr+lr+(0|lr)	0+tb+bt	0+ccw+(0|ccw)
5q	2	4	**szz**	⌐	0+lr+0	tb+0+tb	0+0+0
5r	2	4	**szp**	⌐⌐	(0|rl)+lr+rl	tb+0+tb	0+0+(0|cw)
5s	2	5	**swg**	⌐	0+lr+0	tb+0+bt	0+ccw+0
5t	3	5	**hzzz**	⌐	lr+0+lr+0	0+tb+0+tb	0+0+0+0
5u	3	5	**hzzp**	⌐	lr+rl+lr+rl	0+tb+0+tb	0+(0|cw)+0+cw
5v	3	6	**hzwg**	乙⌐	lr+(0|rl)+lr+0	0+tb+0+bt	0+(0|ccw)+ccw+0
5w	3	6	**hpwg**	⌐	lr+rl+lr+(0|rl)	0+tb+tb+bt	0+(0|cw)+cw+(0|cw)
5x	3	5	**szzg**	⌐	rl+lr+rl+(0|lr)	tb+0+tb+bt	0+0+cw+(0|cw)
5y	4	5	**hzzzg**	⌐	lr+rl+lr+rl+(0|rl)	0+tb+0+tb+bt	0+(0|cw)+0+cw+(0|cw)
5z	4	2	**o**	◯	lr+rl+rl+lr	tb+tb+bt+bt	cw+cw+cw+cw ?