

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: November 4, 2003, 3:23 pm
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	—	U+4e00	橫	<i>héng</i>	h	horizontal; as in 三十卅; or as 一 in 七毛戈; <i>cp.</i> 丿 t
2	1b	0	2	03.45	㇏	PUA	提	<i>tí</i>	t	一 h + taper; 3rd stroke of 𠂔 as in 地; stroke 5 of 虫
3	2a	0	2	15.77	丨	U+4e28	豎	<i>shù</i>	s	vertical; as in 中卜上; or as / in 五 and 丑
4	2b	1	3	01.13	㇏	U+4e85	豎钩	<i>shù-gōu</i>	sg	丨 s + left hook; as in 小丁子才; <i>cp.</i> ㇏ st
5	3a	0	2	12.54	㇏	U+4e3f	撇	<i>piě</i>	p	falling to left, not very curved; as in 八彳行; <i>cp.</i> ㇏ wp and ㇏ sp
6	3b	0	2	03.95	㇏	PUA	弯撇	<i>wān-piě</i>	wp	curve + ㇏ p ; as in 大; <i>cp.</i> ㇏ p and ㇏ sp
7	3c	1	3	03.22	㇏	PUA	豎撇	<i>shù-piě</i>	sp	丨 s + ㇏ wp ; as in 厂; <i>cp.</i> ㇏ wp and ㇏ p
8	4a	0	2	09.59	㇏	U+4e36	点	<i>diǎn</i>	d	taper + clockwise curve; as in 主; sometimes to left, as ㇏ 1st in 火
9	4b	0	2	03.52	㇏	PUA	捺	<i>nà</i>	n	falling right counter-clockwise curve; as in 人; <i>cp.</i> ㇏ p and ㇏ pn
10	4c	0	3	00.03	㇏	PUA	点捺	<i>diǎn-nà</i>	dn	㇏ d + ㇏ n ; 2nd stroke in 入, only in 入 (=入) and its compounds, <i>e.g.</i> 叕

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	<i>píng-nà</i>	pn	ㄥ n + 一 h ; last stroke in 是走乚; cp. ㄥ n and 乚 sw
12	4e	1	3	00.11		U+4e40	提捺	<i>tí-nà</i>	tn	ㄥ t + ㄥ n ; last stroke in 夂 (=攴); or as ㄥ in 八入吏; cp. ㄥ tpn
13	4f	1	4	00.08		PUA	提平捺	<i>tí-píng-nà</i>	tpn	ㄥ t + ㄥ pn ; last stroke in 乚之; cp. ㄥ tn
14	5a	1	3	03.28		U+200cd	橫折	<i>héng-zhé</i>	hz	一 h + s ; 2nd stroke in 口; cp. 冂 hgz
15	5b	1	3	00.90		PUA	橫撇	<i>héng-piě</i>	hp	一 h + 丿 p ; 1 in 又; 4 in 今; cp. 一 hg
16	5c	1	3	01.36		U+4e5b	橫鈎	<i>héng-gōu</i>	hg	一 h + left hook; 2 in 冂写; cp. 冂 hp
17	5d	1	3	02.54		U+200ca	豎折	<i>shù-zhé</i>	sz	s + 一 h ; as in 山; or as ㄥ (/ s + 一 h) in 乐牛东互; cp. ㄥ pz
18	5e	1	4	00.17		PUA	豎弯	<i>shù-wān</i>	sw	s + ㄥ pn ; stroke 4 in 四
19	5f	1	3	01.36		U+2010c	豎提	<i>shù-tí</i>	st	s + ㄥ t (right hook); as in 民辰; cp. 冂 sg
20	5g	1	3	00.51		PUA	撇折	<i>piě-zhé</i>	pz	丿 p + ㄥ t ; 3 in 公; stroke 1 in ㄥ; cp. ㄥ sz
21	5h	1	3	00.11		U+21fe8	撇點	<i>piě-diǎn</i>	pd	丿 p + ㄥ d ; stroke 1 in 女ㄥㄥㄥㄥ
22	5i	1	3	00.00		PUA	撇鈎	<i>piě-gōu</i>	pg	丿 p + left-rising hook; as in ㄥ; cp. 丿 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 丿 sg ; 3 in 豕, 6 in 家
24	5k	1	3	01.81		PUA	斜钩	xié-gōu	xg	ㄣ n + up hook; 5 in 我; 2 in 弋; <i>cp.</i> 乚 swg
25	5l	2	4	00.14		PUA	横折折	héng-zhé-zhé	hzz	一 h + 乚 sz or 乚 hz + 一 h ; 2nd stroke in 凹; <i>cp.</i> 乚 hzw , 乚 hzwg
26	5m	2	5	00.03		PUA	横折弯	héng-zhé-wān	hzw	一 h + 乚 sw ; 2 in 朵, 豕; <i>cp.</i> 乚 hzwg
27	5n	2	4	00.18		PUA	横折提	héng-zhé-tí	hzt	一 h + 乚 st ; 2 in 讠, as in 记, 鳩; <i>cp.</i> 乚 hzz , 乚 hzw
28	5o	2	4	02.22		PUA	横折钩	héng-zhé-gōu	hzg	一 h + 乚 sg ; 2 in 月, 丹; or as 乚 in 勺, 万; or as 乚 in 也; <i>cp.</i> 乚 hz
29	5p	2	4	00.28		U+2e84	横斜钩	héng-xié-gōu	hxg	一 h + 乚 xg ; 1st stroke in 飞, 𠂇; also in 风, 气; <i>cp.</i> 乚 hzwg
30	5q	2	4	00.44		U+200d1	竖折折	shù-zhé-zhé	szz	丨 s + 乚 hz or 乚 + 丨; as 4 in 亞, 6 in 鼎, 11 in 龍; <i>cp.</i> ㄣ szp , ㄣ szzg
31	5r	2	4	00.11		PUA	竖折撇	shù-zhé-piě	szp	/ s + ㄣ hg / フ hp ; as in 专, 吳; or as ㄣ in 𠂇, 設, 𠂇, 𠂇; <i>cp.</i> ㄣ szzg
32	5s	2	5	01.84		U+4e5a	竖弯钩	shù-wān-gōu	swg	乚 sw + up hook; as in 儿, 礼, 心; <i>cp.</i> 乚 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	橫折折折	<i>héng-zhé-zhé-zhé</i>	hzzz	┐ hz + ┐ hz ; 4th stroke in 凸𡵗; <i>cp.</i> ㄣ hzzp
34	5u	3	5	00.09		PUA	橫折折撇	<i>héng-zhé-zhé-piě</i>	hzzp	→ hg + ㄣ hp ; 1 in 𠂇建; 2 in 及; <i>cp.</i> ㄣ hzzz , ㄣ hzzzg
35	5v	3	6	00.60		U+4e59	橫折彎鉤	<i>héng-zhé-wān-gōu</i>	hzwg	一 h + 乚 swg ; stroke 19 in 𠂇; or as 乚 stroke 2 in 九几風; <i>cp.</i> 乚 hzw , 乚 hzz , 乚 hxg
36	5w	3	6	00.03		PUA	橫撇彎鉤	<i>héng-piě-wān-gōu</i>	hpwg	→ hg + ㄣ wg ; 1 in 𠂇队; <i>cp.</i> ㄣ hzzzg
37	5x	3	5	00.92		PUA	豎折折鉤	<i>shù-zhé-zhé-gōu</i>	szzg	/ s + ┐ hzg ; 2 in 马𠂇; <i>cp.</i> ㄣ szz , ㄣ szp
38	5y	4	6	00.11		U+2010e	橫折折折鉤	<i>héng-zhé-zhé-zhé-gōu</i>	hzzzg	→ hg + ┐ hzg ; 1 in 乃仍; <i>cp.</i> ㄣ hpwg , ㄣ hzzz , ㄣ hzzp
39	5z	1	2	00.06		U+3007	圈	<i>quān</i>	o	circle; bottom of 𠂇𠂇𠂇; 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 $D = \text{sharp} + (\text{curved} * 2) + 2$; 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:

— **h**, / **t**, | **s**, \ **p**, J **wp**, \ **d**, \ **n**

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 $\cup$ **pn** stroke has 1 transition (it is composed of $\backslash$ **n** + — **h**), while $\wedge$ **tpn** also has 1 transition (from / **t** to $\cup$ **pn**) rather than 2 ($\wedge$ **t** + $\backslash$ **n** + — **h**), which would ignore the higher level grouping for $\cup$ **pn** (= $\backslash$ **n** + — **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**, J **wp**, and J **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.* $\backslash$ **n** with $\cup$ **pn**, and $\wedge$ **tn** with $\wedge$ **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	豎	弯
<i>B</i>	<i>C</i>	<i>D</i>	<i>J</i>	<i>F</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
1a	0	2	h	— ↗	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	↘ ↗	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	橫	豎	彎
<i>B</i>	<i>C</i>	<i>D</i>	<i>J</i>	<i>F</i>	<i>X</i>	<i>Y</i>	<i>Z</i>
5h	1	3	pd	𠂇	rl+lr	tb+tb	cw+cw
5i	1	3	pg	𠂇	rl+rl	tb+bt	cw+0
5j	1	4	wg	𠂇	lr+rl	tb+bt	cw+cw
5k	1	3	xg	𠂇	lr+0	tb+bt	ccw+0
5l	2	4	hzz	𠂇	lr+0+lr	0+tb+0	0+0+0
5m	2	5	hzw	𠂇	lr+0+lr	0+tb+0	0+0+ccw
5n	2	4	hzt	𠂇	lr+0+lr	0+tb+bt	0+0+0
5o	2	4	hzg	𠂇 𠂇 𠂇	lr+(0 rl)+rl	(0 bt)+tb+bt	0+(0 cw)+0
5p	2	4	hxg	𠂇	lr+lr+(0 lr)	0+tb+bt	0+ccw+(0 ccw)
5q	2	4	szz	𠂇	0+lr+0	tb+0+tb	0+0+0
5r	2	4	szp	𠂇 𠂇	(0 rl)+lr+rl	tb+0+tb	0+0+(0 cw)
5s	2	5	swg	𠂇	0+lr+0	tb+0+bt	0+ccw+0
5t	3	5	hzzz	𠂇	lr+0+lr+0	0+tb+0+tb	0+0+0+0
5u	3	5	hzzp	𠂇	lr+rl+lr+rl	0+tb+0+tb	0+(0 cw)+0+cw
5v	3	6	hzwg	𠂇 𠂇	lr+(0 rl)+lr+0	0+tb+0+bt	0+(0 ccw)+ccw+0
5w	3	6	hpwg	𠂇	lr+rl+lr+(0 rl)	0+tb+tb+bt	0+(0 cw)+cw+(0 cw)
5x	3	5	szzg	𠂇	rl+lr+rl+(0 lr)	tb+0+tb+bt	0+0+cw+(0 cw)
5y	4	5	hzzzg	𠂇	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 ?