Subject: [unihan] Re: Fwd: resent: 12 additional characters for Unicode CJK Stroke group (ref: U31C0)

From: Richard COOK <rscook@unicode.org>

L2/14-007

Date: Mon, 13 Jan 2014 12:00:06 -0800

To: Rick McGowan <rick@unicode.org>

CC: Richard COOK <rscook@unicode.org>, "UniHan unihan@unicode.org" <unihan@unicode.org>, Ian Low <ianlow23@gmail.com>

On Jan 13, 2014, at 9:32 AM, Rick McGowan wrote:

> Hi Everyone,
>
> This also came in. I see that Tom replied to the previous message about this issue. Is this proposal something to consider? Or is it not in the realm of possiblity, or desirability?
>
> If needed, I could just forward this into the UTC document register.

Rick,

The short answer is: These are all unified variants, not suited to UCS standardization (except possibly in IVD).

If you'd like to fold Ian's two documents (and my reply here) into a single UTC doc, it would at least document my opinion, and might make for some interesting discussion. UTC can decide whether to submit it "FYI" to IRG, or what to do.

As Ian notes ("variant of") in the 2nd document, these are all variants of encoded CJK Strokes. But more than that, they are all minor glyph variants, unified (explicitly or implicitly) in the block and already supported (and distinguished) by CDL.

As Tom mentioned, IRG did decide to unify certain things in creating that block (and some rare types evident in the charts were intentionally excluded, pending further study). Many of the unifications agreed upon in the original IRG work on the CJK Strokes block are documented in TUS Appendix F, but some are only implicit.

Appendix F (built with Wenlin CDL):
<http://www.unicode.org/versions/Unicode6.2.0/appF.pdf>

CDL Spec (built by Wenlin):
<http://www.wenlin.com/cdl/>

CDL spec (and supporting documents) that inspired work on the CJK Strokes block in IRG are at the Wenlin link above. Ian's 2nd document reminds me of some of the early "distinctive feature" analysis of CJK Strokes that we did when the CDL spec was first published (more than a decade ago).

Besides that information, we (at Wenlin) maintain other data documenting the relation of the original CDL spec to eventual the CJK Strokes block. Some of this went into Appendix F, and some is as yet unpublished (but might be made available, see below). These data files reflect the current state of the CDL database and engine; the CDL database covers all UCS CJK, and so precisely documents the relation of the CJK Strokes to the CJK character set and to the glyphs evident in UCS code charts, SuperCJK14, and primary dictionary sources.

As I said above, my impression is that ***if*** the minor variants Ian lists require

additional formalization in UCS, they are glyph-level variants that would be better suited to formalization in IVD. (But IVD is immature without a formal CDL spec and public CDL implementations, tools needed for proper for IVD management; see below.) Certainly, these minor stroke variants are already handled outside of UCS by CDL, using two methods: (1) the CDL spec itself defines certain attributes further defining the stroke types (for example, the "points" attribute; the CDL engine defines others); (2) CDL uses private-use variation selectors to exhibit certain CJK Stroke variant glyphs, and ranges of variation.

BTW, besides these unified "minor variant" types that Ian mentions, we've collected a number of rare strokes as candidates for addition to the block (some of these were known in the original IRG work, some were not). I'm surprised that Ian doesn't mention any of these, since they can be found in the CJK charts themselves. But they are problematic for standardization because they are so rare, the forms are poorly attested, idiosyncratic, ill-defined.

At any rate, any proposal to augment the CJK Strokes block would need to come through IRG. Likewise, it would seem to me that any proposal to standardize CJK Stroke variants in IVD would need to come from IRG. But certain work remains to be done before further CJK Stroke work can even be attempted in IRG.

All of that having been said, I think it would be best to formalize CDL in W3C as an XML application, so that all of the ***implicit*** things that went into definition of the CJK Strokes block are made ***explicit*** (even more so than in Appendix F) and accessible in public standards (for use in IVD and UCS CJK work). Last year we (at Unicode) took some steps toward this, and 2014 will hopefully see more.

-Richard

> Rick
>
>
> -------- Original Message --------
> Subject: resent: 12 additional characters for Unicode CJK Stroke group (ref: U31C0)
> Date: Mon, 13 Jan 2014 13:06:50 +0000
> From: Ian Low <ianlow23@gmail.com>
> To: human3@unicode.org
>
> Hi Human3,
>
> This email has been resent due to a problem with the attachment in the first attempt.
>
> In addition to my previous email requesting 2 characters CJK Stroke SZP (竖折撇)
> and   CJK Stroke HXG (横斜钩) to be included in the Unicode CJK Strokes group
> (http://www.unicode.org/charts/PDF/U31C0.pdf), this message is a request for 12
> additional characters to be included within the same group.
>
> Please find attached RTF file containing details and relevant images.
>
> Yours Sincerely,
> Ian Low
>
> <cjk_strokes2.rtf>

On Jan 13, 2014, at 8:13 AM, Tom Bishop, Wenlin Institute wrote:

Dear Ian Low,

Concerning your recommendation:

> The following 2 strokes need to be included in the Unicode CJK Strokes group:
> 1)CJK Stroke SZP (竖折撇"down then kink left")
> 2) CJK Stroke HXG (横斜钩-"across then oblique hook")

The Wenlin CDL spec includes types SZP and HXG:

http://www.wenlin.com/cdl/cdl_strokes_2004_05_23.pdf

I was in the IRG discussion where it was decided to merge SZP and HXG with other types. SZP was merged into SZZ. HXG was merged into HZWG. HXWG was split off from HZWG. I would have been happy to encode all the CDL types as-is. Still, I was glad that nearly all the rest of the CDL types did get encoded. There was bound to be some compromise given the variety of preferences of the IRG participants.

Best wishes,

Tom

文林 Wenlin Institute, Inc.   Software for Learning Chinese
E-mail: wenlin@wenlin.com   Web: http://www.wenlin.com
Telephone: 1-877-4-WENLIN (1-877-493-6546)
☯

On Jan 12, 2014, at 11:46 AM, Rick McGowan wrote:

FYI. This came in, and I think this group might have an interest in reviewing it. I have not looked at it myself.
    Rick

-------- Original Message --------
Subject: Feedback and recommendation regarding CJK Strokes group re: U31C0.pdf
Date: Sun, 12 Jan 2014 18:27:12 +0000
From: Ian Low <ianlow23@gmail.com>
To: human3@unicode.org

Hi Human3,

Please find attached rtf file containing my feedback and recommendations in response to the CJK strokes grouping as defined in http://www.unicode.org/charts/PDF/U31C0.pdf 11 January 2014. I hope this message will be of use to the Unicode review team. I am sending this via gmail attachment instead of the webpage form as it contains images.

Regarding my background, I am an amateur researcher, author and publisher of Chinese dictionaries.

Yours Sincerely,
Ian Low
ianlow23@gmail.com

To: human3@unicode.org

Hi Human3,

This message contains my feedback and recommendations in response to the CJK strokes grouping as defined in http://www.unicode.org/charts/PDF/U31C0.pdf 11 January 2014. I hope this message will be of use to the Unicode review team. I am sending this via gmail instead of the webpage form as it contains images.

Regarding my background, I am an amateur researcher, author and publisher of Chinese dictionaries.

My sources for this feedback are:
A）http://baike.baidu.com/view/168278.htm
B）http://baike.baidu.com/view/988376.htm
C）http://wenku.baidu.com/view/1b71451b59eef8c75fbfb31f.html - does not define the oblique stroke
D）http://wenku.baidu.com/view/b4bcc1165f0e7cd1842536ab.html
E）http://www.360doc.com/content/12/0223/09/4295303_188816135.shtml - defines the oblique stroke
F) http://www.zdic.net

The above pages contain information regarding the definitions of the "offical" strokes as taught in primary schools in China.

**Commentary on Individual Characters**

1) UC31C0　　　CJK Stroke T - 提ti3-"lift" – attested in sources C) and E)

2) UC31C1　　CJK Stroke WG - 弯钩wan1gou1-"curved hook" – attested in sources C) and E)

3) UC31C2　　CJK Stroke XG - 斜钩xie2gou1-"oblique hook"– attested in sources C) and E)

4) UC31C3　　CJK Stroke BXG - Source C) defines this as 卧钩wo4gou1-"lying hook", Source E) does not define this stroke

5) UC31C4　　CJK Stroke SW - 竖弯shu4wan1-"down then curve" e.g. 4th stroke in 四 – attested in sources C) and E)

6) UC31C5　　CJK Stroke HZZ - 横折折heng2zhe2zhe2-"across then double kink" – e.g. 2nd stroke in 凹 – attested in source E) but not in C)

7) UC31C6　　CJK Stroke HZG - 横折钩heng2zhe2gou1-"across then kink hook" – e.g. 2nd stroke in 月 – attested in sources C) and E)

8) UC31C7　　CJK Stroke HP - 横撇heng2pie3-"across then right" – e.g. 2nd stroke in 水 – attested in sources C) and E)

9) UC31C8　　CJK Stroke HZWG - 横折弯钩heng2zhe4wan1gou1-"across then kink curve hook" – note that source E) distinguishes this stroke from the following stroke　　, which it defines as 横斜钩heng2xie2gou4-"across then oblique hook" e.g. 2nd stroke in 凤. Note that source C) does not
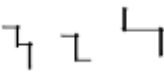
make this distinction.

10) UC31C9 　CJK Stroke SZWG - 竖折折钩shu4zhe4zhe2gou1-"downward then doublekink hook" – attested in sources C) and E)

11) UC31CA 　CJK Stroke HZT - 横折提heng2zhe2ti3-"across then kink lift" – attested in sources C) and E)

12) UC31CB 　CJK Stroke HZZP - 横折折撇heng2zhe2zhe2pie3-"across then kink kink right" – attested in sources C) and E)

13) UC31CC 　CJK Stroke HPWG - 横撇弯钩-"across then left curve hook" – attested in sources C) and E)

14) UC31CD 　CJK Stroke HZW - 横折弯-"across then kink curve" – attested in sources C) and E)

15) UC31CE 　CJK Stroke HZZZ - 横折折折-"across then triple kink" – attested in source E) only

16) UC31CF 　CJK Stroke N - 捺-"right" – attested in sources C) and E)

17) UC31D0 　CJK Stroke H - 横-"across" – attested in sources C) and E)

18) UC31D1 　CJK Stroke S - 竖-"down" – attested in sources C) and E)

19) UC31D2 　CJK Stroke P - 撇-"left" – attested in sources C) and E)

20) UC31D3 　CJK Stroke SP - defined as 　CJK Stroke P in source C)

21) UC31D4 　CJK Stroke D - 点dian3-"dot" – attested in sources C) and E)

22) UC31D5 　CJK Stroke HZ - 横折-"across then kink" – attested in sources C) and E)

23) UC31D6 　CJK Stroke HG - 横钩-"across then hook" – attested in sources C) and E)

24) UC31D7 　CJK Stroke SZ - 竖折-"down then kink" – attested in sources C) and E)

25) UC31D8 　CJK Stroke SWZ - not defined in simplified hanzi sources C) and E). The example given in the pdf file 肅 is a traditional hanzi character.

26) UC31D9 　CJK Stroke ST - 竖提-"down then lift" – attested in sources C) and E)

27) UC31DA 　CJK Stroke SG - 竖钩-"down then hook" – attested in sources C) and E)

28) UC31DB 　CJK Stroke PD - 撇点-"left then dot" – attested in sources C) and E)

29) UC31DC ╱ CJK Stroke PZ - 撇折-"left then kink" - attested in sources C) and E)

30) UC31DD ⌒ CJK Stroke TN - variant of stroke ╲ CJK Stroke N

31) UC31DE ⌐┘ CJK Stroke SZZ - 竖折折 - attested in source E) but not C)

32) UC31DF └ CJK Stroke SWG - 竖弯钩 - attested in sources C) and E)

33) UC31E0 乙 CJK Stroke HXWG - not defined in C) and E)

34) UC31E1 ㇆ CJK Stroke HZZZG - 横折折折钩 - attested in sources C) and E)

35) UC31E2 ∨ CJK Stroke PG - not defined in C) and E)

36) UC31E3 ◯ CJK Stroke Q - not defined in C) and E)

In addition to the above strokes, the stroke ↳ which is the 3rd stroke in the simplified character 专 is not defined in the unicode.org proposal

Summary

It appears that the difference between source C) and source E) is that even though both are used for simplified hanzi characters, the latter also accommodates traditional hanzi characters, as it contains the

strokes ㇆ ㇗ ⌐┘ which are normally found in traditional hanzi.

The following table showing strokes from source C) in blue and source E) in red with black tick marks

illustrating strokes present in the Unicode CJK stroke group, illustrate that the stroke ↳ CJK Stroke SZP (竖折撇"down then kink left") is missing from the Unicode group.

笔画	名称	例字	笔画	名称	例字
丶	丶 点 ✓	广	一	一 横钩 ✓	写
一	一 横 ✓	王		乛 横折钩 ✓	月
丨	丨 竖 ✓	巾	乁	乚 横折弯钩 ✓	九
丿	丿 撇 ✓	白	乛	乛 横撇弯钩 ✓	那
乀	乀 捺 ✓	八	乛	乛 横折折折钩 ✓	奶
✓	乀 提 ✓	打	乛	乛 竖折折钩 ✓	与
乁	乀 撇点 ✓	巡	乚	乚 竖弯 ✓	四
乚	乚 竖提 ✓	农	乚	乚 横折弯 ✓	沿
乛	乛 横折提 ✓	论	乛	乛 横折 ✓	口
亅	亅 弯钩 ✓	承	乚	乚 竖折 ✓	山
亅	亅 竖钩 ✓	小	∠	∠ 撇折 ✓	云
乚	乚 竖弯钩 ✓	屯	乛	乛 横撇 ✓	水
乚	乚 斜钩 ✓	浅	乛	乛 横折折撇 ✓	建
乚 乀	卧 钩 ✓	心	乚	乚 竖折撇 ✗	专

In addition, 乚 CJK Stroke HXG (横斜钩-"across then oblique hook") is also missing from the Unicode group.

Recommendation

The following 2 strokes need to be included in the Unicode CJK Strokes group:

1) 乚 CJK Stroke SZP (竖折撇"down then kink left")

2) 乚 CJK Stroke HXG (横斜钩-"across then oblique hook")

I hope this information has been of use. If I can be of any further assistance, please let me know.

Yours Sincerely,
Ian Low
ianlow23@gmail.com

Hi Human3,

In addition to my previous email requesting 2 characters CJK Stroke SZP (竖折撇) and

 CJK Stroke HXG (横斜钩) to be included in the Unicode CJK Strokes group (http://www.unicode.org/charts/PDF/U31C0.pdf), this message is a request for 12 additional characters to be included within the same group.

These 12 additional new strokes consist of 11 variants and 1 partial of the main strokes. These strokes are a result of my research in trying to reduce the hanzi characters to their fundamental components. They are listed below in order of importance.

11 Variants

1) CJK Stroke DT 丶(点提) - 2nd stroke of 安 – variant of 丶(D点)
For: This stroke flows in the opposition direction to the stroke from which it forms a variant, i.e. it is written from bottom left to top right, whereas Stroke D点 is written from top left to bottom right. There is no difficulty distinguishing between DT (点提) and (D点).

2) CJK Stroke TZG ⁊(提折钩) - 2nd stroke of 也 – variant of 丁 (HP横撇)
For: This stroke is very distinct from (HP横撇) from which it forms a variant. It is always invariably written with an upward angle and a vertical descend after the kink. There is no case in which there exists any difficulty in distinguishing between these two strokes.

3) CJK Stroke WP ——(卧撇) - 1st stroke of 妥 – variant of 丿 (P撇)
For: This stroke is very distinct from (P撇) from which it forms a variant. It is always invariably written at a much flatter angle than the latter. I am not aware of any case in which there exists any difficulty distinguishing between these two strokes.

4) CJK Stroke HXP ⁊ (横小撇) - 1st stroke of 子 – variant of ⁊ (HG横钩)
For: This stroke is very distinct from (HG横钩) from which it forms a variant. The latter always ends with a hook ending at a point right of centre of the horizontal line, whereas this stroke has a tail which is clearly much larger than a hook ending at a point directly under or near the middle of the horizontal line. I am not aware of any case in which there exists any difficulty distinguishing between these two strokes.

5) CJK Stroke X 丨 (斜) - 2nd stroke of 丑 – variant of 丨 (S竖)
For: This stroke is very distinct from S竖 from which it forms a variant. The latter is always right-angle vertical, whereas this stroke needs to be written with a slant. I am not aware of any case in which there exists any difficulty in distinguishing between these two strokes.

6) CJK Stroke DP 丿(点撇) - 2nd stroke of 前 – variant of 丿 (P撇)
For: In many characters, this is clearly written as a "dot" (e.g. 1st stokes in 鸟 and 舅, 2nd stroke in 迷) instead of a "stroke"(e.g. as in 木 or 人).
Against: In many cases, it is hard to distinguish whether it should be a "dot" or a "stroke", e.g. 1st stroke in 臼 and 2nd stroke in 米.

7) CJK Stroke TZ ⁊(提折) - 2nd stroke of 也 – variant of 丁 (HP横撇)

For: same as in TZG(提折钩)

Against: TZ(提折) is invariably an alternate form of TZG(提折钩) in that it only occurs as an alternate form of writing the latter, i.e. the hook is included in some fonts but excluded in others.

8) CJK Stroke QN ﹨ (起捺) - 2nd stroke of 卜 - variant of ﹨(D点)

For: In several characters, this is clearly written as a "stroke" (e.g. 卜 and 补) instead of a "dot"(e.g. 2nd stroke in 惧 and 慎).

Against: In many cases, it is hard to distinguish whether it should be a "dot" or a "stroke", e.g. as in 补 and 鳸.

9) CJK Stroke XHZ ﹁(斜横折) - 1st stroke of 丑 - variant of ﹁(HZ横折)

For: In many cases (e.g. as in 丑etc) this is clearly distinct from its classic main form (e.g. as in 日).

Against: Depending on font and writing style this stroke is often written exactly the same as ﹁(LHZ), i.e. without the vertical slant.

10) CJK Stroke WHZ ﹁ (卧横折) - 2nd stroke of 皿 - variant of ﹁(HZ横折)

For: In many cases (e.g. as in 盃孟etc) this is clearly distinct from its classic main form (e.g. as in 日).

Against: In many cases (e.g. 皿) it is difficult to distinguish from its main form.

11) CJK Stroke LHZ ﹁(立横折) - 2nd stroke of 且 - variant of ﹁(HZ横折)

For: In many cases (e.g. as in 且瞄etc) this is clearly distinct from its classic main form (e.g. as in 日).

Against: In many cases (e.g. 目) it is difficult to distinguish from its main form.

1 Partial

12) CJK Stroke XWG ∠ (撇弯钩) - 2nd stroke of 乙 - partial of 乙 (HXWG横斜弯钩)

Against: This stroke does not occur in any character except as a partial within 乙 (HXWG横斜弯钩).

For: Despite the fact that this stroke does not except in any character except as a partial within (HXWG横斜弯钩), following my research into trying to establish the fundamental components of hanzi characters, I feel that this stroke is important because it constitutes one of the 39 or thereabouts elemental* i.e. irreducible non-duplicating and non-overlapping "strokes" within hanzi characters. e.g. 乙 (HXWG横斜弯钩)can be reduced to ﹣(H横) and ∠(撇弯钩), but the latter cannot be reduced any further. Similarly ﹃ (HZZZG) can be reduced to ↗ (HXP横小撇), ﹣(H横) and ノ (PG), but the latter 3 cannot be reduced any further. (* note: these approx. 39 "elemental" strokes can be defined as ﹅ ﹃ ﹣ ↗ ｜﹨ ﾍ ﹃ ﹁ ﹃ ∠ ノ ﹂ ∠ ‹ ↓ ↳ ﹂ �L ↬ ﹁ ﹨ ノ ノ ｜ ﹣ ﹨ ﹃ ﹃ L ﹨ ﹨) ✓)
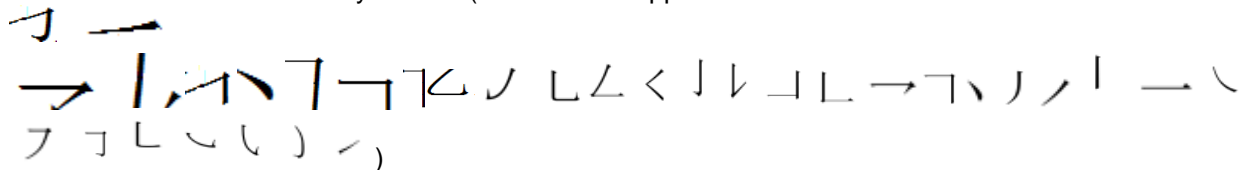
Recommendation

I recommend that 1, 2, 3, 4 and 5 from the above should be included in Unicode CJK Strokes group as these in all cases are very distinct and easily distinguishable from the main strokes from which they form a variant. They require a different application of the brush/pen, and therefore even though defined as "variants" they clearly are very distinct strokes.

I recommend that 6, 7, 8, 9, 10, 11, and 12 from the above should be included in the group for experimental and calligraphic reasons, so as to enable IT users to more easily describe instances where these strokes are clearly distinct from their main forms.

Yours Sincerely,
Ian Low
ianlow23@gmail.com