

Title: Proposal to encode (disunify) two stroke characters in CJK Strokes block

Author: Night Koo

Date: 2023-10-01

This is a proposal to encode (or disunify) two stroke characters in CJK Strokes block, namely CJK STROKE HZXG ㇇ (橫折斜鈎) and CJK STROKE SZP ㇇ (豎折撇).

The current form of CJK Strokes block is based on the [L2/06-212 \(WG2 3063\)](#), where the IRG Rapporteur proposed the current encoding for U+31D0..31E3. However, when comparing the document with its predecessors, [L2/04-221](#) and [L2/04-367](#), there are some strokes that were unified together. The unification discussion seems to be based on [IRGN 1137](#), which is reflected in the [Appendix F: Documentation of CJK Strokes](#) of the Unicode Standard.

However, two of the strokes are unified when their use is actually very prevalent, which is CJK STROKE HZXG ㇇ and CJK STROKE SZP ㇇ (or its variant form ㇇). These 2 characters are listed separately in [GF 2001-2001 《GB 13000.1 字符集汉字折笔规范》](#) with example usage. Thus, I propose to encode these two characters in the CJK Strokes block.

CJK STROKE HZXG ㇇ (橫折斜鈎)

This stroke is unified to U+31C8 CJK STROKE HZWG ㇇ based on information in Appendix F. However, based on GF 2001-2001, HZWG ㇇ should actually be unified to U+31E0 CJK STROKE HXWG ㇇ based on the number of zags (折) of the stroke. HZWG ㇇ and HXWG ㇇ both have 3 zags, while HZXG ㇇ only have 2 zags. Thus, HZWG ㇇ and HZXG ㇇ are listed separately in this document, while HXWG ㇇ is listed as a variant of (or unified to) HZWG ㇇.

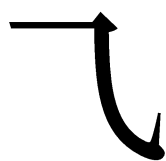
Although not as evident as this, the original document as submitted by TCA in [IRGN 1086A](#) also suggested the difference between HZXG ㇇ and HXWG ㇇, where HZXG ㇇ is listed under 2 turning points and HXWG ㇇ is listed under 3 turning points. It is thus possible to state that the unencoded stroke HZXG ㇇ is not equivalent to U+31C8 CJK STROKE HZWG ㇇ as suggested by Appendix F, and should be separately encoded to correctly represent the information as shown in GF 2001-2001 (since U+31C8 CJK STROKE HZWG ㇇ and U+31E0 CJK STROKE HXWG ㇇ are unified in GF 2001-2001).

CJK STROKE SZP ㇇ (豎折撇)

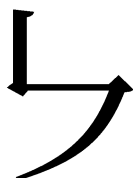
This stroke is unified to U+31DE CJK STROKE SZZ ㇇ based on information in Appendix F. However, based on GF 2001-2001, these two strokes are listed separately. The difference between these two strokes is the final stroke, where SZP ㇇ has a throw stroke (撇) and SZZ ㇇ has a vertical stroke (豎). These two strokes are not exchangeable unlike other unified variant strokes in many cases, thus requiring a separate character to be encoded.

There are two popular forms of this stroke currently in used: the first section being slanted 𠃉 and being vertical 𠃊. These exists in U+4E13 专 (3rd stroke) and U+594A 隹 (8th stroke) respectively. However, this vertical/slanted difference is not as significant as SZP 𠃊 vs SZZ 𠃉, and a similar unification is made to U+31D1 CJK STROKE S 丨 when in the character U+4E94 五 (2nd stroke) as 𠃊, and to U+31D7 CJK STROKE SZ 𠃊 when in the character U+7259 牙 (2nd stroke) as 𠃊. Referencing both stroke S and SZ, the vertical form for stroke SZP (𠃊) should be encoded, and the slanted form may be registered as a SVS in the future.

Proposed Characters



U+31E4 CJK STROKE HZXG



U+31E5 CJK STROKE SZP

Unicode Character Properties

The proposed characters should have the following Unicode properties matching those in CJK Strokes block.

31E4;CJK STROKE HZXG;So;0;ON;;;;;N;;;;;

31E5;CJK STROKE SZP;So;0;ON;;;;;N;;;;;

Additional Changes

The proposed update to Appendix F in [L2/23-116](#) will need to be modified to reflect the above changes. The diagrams are attached after the figures.

A font file containing both characters will be provided to update the *CJK Symbols* font used for the block as provided in [L2/20-058](#).

Additionally, the annotation under U+31C8 CJK STROKE HZWG 𠃊 for “1st stroke of 98DE 𠃊” will need to be moved to stroke HZXG 𠃊. A new annotation for stroke SZP 𠃊 can be “8th stroke of 594A 隹”.

That is all.

Figures (black is encoded, red is unencoded)

2 折	5.12	横折竖折横	横折折	ㄇ	凹 卮
	5.13	横折竖弯横	横折弯	ㄣ	朵
	5.14	横折竖折提	横折提	ㄥ	计 颓 鳩
	5.15	横折竖钩	横折钩	ㄎ (フ)	同 门 却 永 耍 万 母 仓 / 也
	5.16	横折捺钩	横斜钩 (俗称)	ㄨ (フ)	飞 风 执
	5.17	竖折横折竖	竖折折	ㄣ	鼎 卮 亞 吳
	5.18	竖折横折撇	竖折撇	ㄣ (フ、フ)	专 / 隼 / 矢
3 折	5.19	竖弯横钩	竖弯钩	ㄥ	己 匕 电 心
	5.20	横折竖折横折竖	横折折折	ㄣ	凸
	5.21	横折竖折横折撇	横折折撇	ㄣ	及 延
	5.22	横折竖弯横钩	横折弯钩	ㄣ (乙)	几 丸 / 艺 亿
	5.23	横折撇折弯竖钩	横撇弯钩 (俗称)	ㄣ	阳 部
	5.24	竖折横折竖钩	竖折折钩	ㄣ (フ)	马 与 钙 / 号 弓

GF 2001-2001

			Turning point 1 折數 1	Turning point 2 折數 2	Turning point 3 折數 3	Turning point 4 折數 4
5	Corner	ㄎ 折 Zhe	ㄎ 横折	ㄣ 横折横	ㄣ 横折横折	ㄣ 横折横折钩
			フ 横钩	ㄣ 横折弯	乙 横折曲钩	
			フ 横撇	ㄣ 横折钩	ㄣ	
			ㄣ 豎折	ㄨ 横斜钩		
			ㄣ 豎彎折右	ㄣ 豎橫折 1		
			ㄣ 豎彎折左	ㄣ 豎橫折 2		
			ㄣ 豎挑	ㄣ 豎曲钩		
			ㄣ 豎钩			

IRGN 1086A

汉字笔画名称表

- 一、基本笔画：一(横) 丨(竖) 丿(撇) 丶(点) ㇇(折)
 二、变形笔画：

笔形	名称	例字	笔形	名称	例字
㇇	提	刁 江	㇇	捺	大 进
㇇	横折	尺 口	㇇	横斜钩	飞 汽
㇇	横撇	又 水	㇇	横折弯钩	亿 九
㇇	横钩	予 卖	㇇	横撇弯钩	队 都
㇇	横折折	凹	㇇	横折折折	凸
㇇	横折弯	朵 船	㇇	横折折撇	延 及
㇇	横折提	说	㇇	横折折折钩	乃 场
㇇	横折钩	幻 月	㇇	竖提	长 以
㇇	竖钩	丁 小	㇇	竖折折钩	与 弓
㇇	竖折	山 母	㇇	撇折	台 红
㇇	竖弯	西 四	㇇	撇点	女 巡
㇇	竖弯钩	礼 巳	㇇	斜钩	戈 式
㇇	竖折撇	专	㇇	弯钩	狗 家
㇇	竖折折	鼎			

笔画	名称	例字	笔画	名称	例字	笔画	名称	例字
1	横	大	11	横钩	你	21	弯钩	了
2	竖	十	12	竖弯钩	元	22	横折弯钩	九
3	撇	八	13	撇折	去	23	竖弯	四
4	点	主	14	竖提	良	24	横折弯	没
5	横折	口	15	竖折	山	25	横折折折钩	仍
6	捺	人	16	撇点	女	26	横斜钩	凰
7	提	地	17	竖折折钩	弟	27	横折折撇	及
8	横折钩	月	18	斜钩	我	28	竖折撇	专
9	竖钩	小	19	横撇弯钩	那	29	竖折折	鼎
10	横撇	水	20	横折提	课	30	横折折	凹
						31	横折折折	凸

Education site showing the different strokes (gcjxkhd.com/nd.jsp?id=1138)

Appendix F Update

㇇ U+31C8	HZWG	héng-zhé-wān-gōu	横折弯钩	
-------------	-------------	------------------	------	--

九 U+4E5D, 几 U+51E0, 丸 U+4E38, 旭 U+65ED, 杂 U+6742

㇈ U+31E4	HZXG	héng-zhé-xié-gōu	横折斜钩	
-------------	-------------	------------------	------	--

飞 U+98DE, 风 U+98CE, 气 U+6C14, 虱 U+8671, 瘋 U+760B, 訊 U+8A0A

㇉ U+31DE	SZZ	shù-zhé-zhé	竖折折	
-------------	------------	-------------	-----	--

㇉ U+200D1, 𠂇 U+5350, 亞 U+4E9E, 鼎 U+9F0E, 吳 U+5433

㇊ U+31E5	SZP	shù-zhé-piě	竖折撇	㇊
-------------	------------	-------------	-----	---

专 U+4E13, 隹 U+594A, 𠂇 U+5928, 𠂇 U+279AE, 𠂇 U+244F7, 𠂇 U+249A1

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.
See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

- | | |
|--|---|
| 1. Title: | Proposal to encode (disunify) two stroke characters in CJK Strokes block |
| 2. Requester's name: | Night Koo |
| 3. Requester type (Member body/Liaison/Individual contribution): | Individual |
| 4. Submission date: | 2023-10-01 |
| 5. Requester's reference (if applicable): | |
| 6. Choose one of the following: | |
| This is a complete proposal: | Yes |
| (or) More information will be provided later: | |

B. Technical – General

- | | |
|---|---|
| 1. Choose one of the following: | |
| a. This proposal is for a new script (set of characters): | |
| Proposed name of script: | |
| b. The proposal is for addition of character(s) to an existing block: | Yes |
| Name of the existing block: | CJK Strokes |
| 2. Number of characters in proposal: | 2 |
| 3. Proposed category (select one from below - see section 2.2 of P&P document): | |
| A-Contemporary <input type="checkbox"/> B.1-Specialized (small collection) <input checked="" type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/> | |
| C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/> | |
| F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/> | |
| 4. Is a repertoire including character names provided? | Yes |
| a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? | |
| b. Are the character shapes attached in a legible form suitable for review? | Yes |
| 5. Fonts related: | |
| a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? | Submitter (Night Koo) |
| b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): | Modified based on Source Han Serif (https://github.com/adobe-fonts/source-han-serif) by Adobe under OFL |
| 6. References: | |
| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | Yes |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | Yes |
| 7. Special encoding issues: | |
| Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? | No |

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>Unified to other stroke characters in IRGN 1137</i>	Yes
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>IRG experts</i>	Yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:		<i>Users of Han character</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:		<i>Common</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>See the representative glyphs in the published ISO/IEC 10646 Standard</i>	Yes
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>Belongs to the CJK Strokes block</i>	Yes Yes
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>See proposal text</i>	Yes Yes
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:		No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		No

Title: Feedback to “IRGN2647 Proposal to disunify two strokes in CJK”
Source: Henry Chan
Date: 2023-10-16
Status: Individual Contribution to IRG #61
Action: For consideration by IRG
Pages: 1

CJK STROKE HZXG (橫折斜鉤):

I believe there is no need to add a new character CJK STROKE HZXG. Instead, it should be added as an alias of U+31C8 乚 CJK STROKE HZWG, and the representative glyph of U+31C8 be updated to 𠄎.

The annotation of “2nd stroke of 4E5D 九” should then be moved from U+31C8 乚 CJK STROKE HZWG to U+31E0 𠄎 CJK STROKE HXWG.

These changes will align ISO 10646 with the GF 2001-2001 《GB 13000.1 字符集汉字折笔规范》，where the stroke of 乙 is considered the same stroke as the second stroke of 九.

Of all the evidences provided in IRGN2647, 九 is grouped with 乙 in the evidences where both characters are present as examples.

Note in some typefaces the second part of the second stroke of 九 does not go straight down but bends slightly to the left. This is even more noticeable in the Kai script compared to the Song script.

No new character needs to be encoded.

CJK STROKE SZP (豎折撇):

Agree.

Title: Feedback to Henry Chan for feedback to IRGN2647

Author: Night Koo

Date: 2023-10-29

𠃉 31C8	HZWG	héng-zhé-wān-gōu	橫折彎鉤	𠃉
九4E5D, 几51E0, 飞98DE, 风98CE, 气6C14, 虱8671, 瘋760B				
乙 31E0	HXWG	héng-xié-wān-gōu	橫斜彎鉤	
乙4E59, 𠃉6C39, 乞4E5E				

Current information on unification in [Appendix F: Documentation of CJK Strokes](#) related to stroke **HZXG** and relatives (HZWG/HXWG)

This is a response to Henry Chan’s “Feedback to ‘IRGN2647 Proposal to disunify two strokes in CJK’” submitted on 2023-10-16. I will suggest UTC to continue to encode CJK STROKE HZXG 𠃉 (橫折斜鉤). The reasons are as follow.

The CJK Strokes in range U+31C0..U+31CF is first proposed through [L2/04-263](#) as part of the HKSCS extension. The current U+31C8 CJK STROKE HZWG 𠃉 is encoded in HKSCS as 0x884D with the current glyph as shown in the Code Charts. Below shows the glyphs of U+31C8 in past and current HKSCS documents.

		884C 𠃉 F30F F30F 21FE8			31C7 𠃉 H-884A
C	𠃉 F30F F30F	884D 𠃉 F310 F310 F310	31C8 𠃉 F310 F310 (884D)	31C8 𠃉 F310 [F310] (884D)	31C8 𠃉 H-884D
D	𠃉 F310 F310	884E 𠃉 F311 F311 200CA	31C9 𠃉 F312 F312 (884F)	31C9 𠃉 F312 [F312] (884F)	31C9 𠃉 H-884F
E	𠃉 F311 F311				

0x884D in versions of HKSCS, from left to right: 1999, 2001, 2004, 2008, latest 2016.

This glyph shape is stable throughout the encoding process. Thus, the suggestion to change the glyph of U+31C8 from HZWG 𠃉 to HZXG 𠃉 will break backwards-compatibility with HKSCS and previous versions of Unicode.

It is also noted that the proposed stroke HZXG 𠃉 has a similar shape at U+2E84 CJK RADICAL SECOND THREE 𠃉. However, the proposed stroke HZXG 𠃉 is not unifiable with U+2E84 due to following considerations:

1. U+2E84 is considered a variant of Kangxi Radical 5 乙 (which can be considered as stroke HXWG 乙) in the chart annotation, which is opposite of the mapping as shown in GF 2001-2001.

2E84 𠃉 CJK RADICAL SECOND THREE
 • variant of Kangxi Radical 5
 → 4E59 乙

Annotation in [Code Charts of CJK Radicals Supplement](#)

折	2	5.12	横折竖折横	横折折	𠃉	凹 𠃉
		5.13	横折竖弯横	横折弯	𠃉	朵
		5.14	横折竖折提	横折提	𠃉	计 颞 鸠
		5.15	横折竖钩	横折钩	𠃉 (フ)	同 门 却 永 耍 万 母 仓 / 也
		5.16	横折捺钩	横斜钩 (俗称)	𠃉	飞 风 执
		5.17	竖折横折竖	竖折折	𠃉	鼎 𠃉 亞 吳
		5.18	竖折横折撇	竖折撇	𠃉 (フ、フ)	专 / 隼 / 𠃉
		5.19	竖弯横钩	竖弯钩	𠃉	己 匕 电 心
	折	3	5.20	横折竖折横折竖	横折折折	𠃉
		5.21	横折竖折横折撇	横折折撇	𠃉	及 延
		5.22	横折竖弯横钩	横折弯钩	𠃉 (乙)	几 丸 / 艺 亿
		5.23	横折撇折弯竖钩	横撇弯钩 (俗称)	𠃉	阳 部
		5.24	竖折横折竖钩	竖折折钩	𠃉 (フ)	马 与 钙 / 号 弓
4 折	5.25	横折竖折横折竖钩	横折折折钩	𠃉 (フ)	乃 / 杨	

GF 2001-2001

2. U+2E84 has the property *Radicals* in the Unicode Character Database, while the proposed CJK STROKE HZXG 𠃉 should not have the property *Radicals*; instead it should have the same properties as other CJK strokes in U+31C0..31E3.
3. U+2E84 is separately defined in HKSCS at 0xC8D7, which shows that the HKSCS treats HZWG 𠃉 and HZXG 𠃉 as different characters (or strokes), and changing the glyph of U+31C8/0x884D from HZWG 𠃉 to HZXG 𠃉 will make 0x884D and 0xC8D7 have equivalent shape in HKSCS, breaking compatibility.

It is possible to change the mapping of 0xC8D7 in HKSCS to the newly proposed CJK STROKE HZXG 𠄎, however as the range of 0xC8D6-0xC8F1 in HKSCS are clearly related to the CJK radicals, it is unadvisable to change the mapping of 0xC8D7 in HKSCS, which will also break backward-compatibility. There is also no apparent benefit of changing the mapping in HKSCS.

					2E80 𠄎 2E80 F820 (C8D6)	2E80 𠄎 2E80 [F820] (C8D6)	2E80 𠄎 H-C8D6
𠄎 F820 2E80	𠄎 F830 2EB6	C8D7 𠄎 F821 2E84 2E84	C8E1 𠄎 F82B 2EA5 2EA5	C8EB 𠄎 F835 2ECC 2ECC	2E84 𠄎 2E84 F821 (C8D7)	2E84 𠄎 2E84 [F821] (C8D7)	2E84 𠄎 H-C8D7
𠄎 F821 2E84	𠄎 F831 2EBC	C8D8 𠄎 F822 2E86 2E86	C8E2 𠄎 F82C 2EA7 2EA7	C8EC 𠄎 F836 2ECD 2ECD	2E86 𠄎 2E86 F822 (C8D8)	2E86 𠄎 2E86 [F822] (C8D8)	2E86 𠄎 H-C8D8
𠄎 F822 2E86	𠄎 F832 2EBE						

U+2E84 in versions of HKSCS, from left to right: 1999, 2001, 2004, 2008, latest 2016. All have clearly shown 0xC8D7 is grouped with other radical variants.

There are also two other stroke characters separately encoded in the CJK Radicals Supplement, which are U+2E82 CJK RADICAL SECOND ONE 𠄎 and U+2E83 CJK RADICAL SECOND TWO 𠄎, corresponding to U+31D6 CJK STROKE HG 𠄎 and U+31DF CJK STROKE SWG 𠄎 respectively. Thus, the encoding of U+2E84 should not affect the proposed encoding of CJK STROKE HZXG 𠄎.

That is all.