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 ㇈.
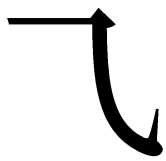
Although not as evident as this, the original document as submitted by TCA in [IRGN 1086A](#) also suggested the difference between HZXG ㇈ and HXWG ㇠, where HZXG ㇈ is listed under 2 turning points and HXWG ㇠ is listed under 3 turning points. It is thus possible to state that the unencoded stroke HZXG ㇈ 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).
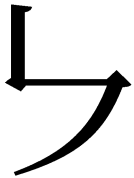
CJK STROKE SZP ㇏ (豎折撇)

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 ㇈ and being vertical ㇈. These exists in U+4E13 专 (3rd stroke) and U+594A 奊 (8th stroke) respectively. However, this vertical/slanted difference is not as significant as SZP ㇈ vs SZZ ㇈, and a similar unification is made to U+31D1 CJK STROKE S ｜ when in the character U+4E94 五 (2nd stroke) as ／, and to U+31D7 CJK STROKE SZ ㇗ when in the character U+7259 牙 (2nd stroke) as ∠. Referencing both stroke S and SZ, the vertical form for stroke SZP (㇈) 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 ㇈ for "1st stroke of 98DE 飞" will need to be moved to stroke HZXG ㇈. A new annotation for stroke SZP ㇈ can be "8th stroke of 594A 奊".

That is all.

Figures *(black is encoded, red is unencoded)*

	5.12	横折竖折横	横折折	乚	凹 卍
	5.13	横折竖弯横	横折弯	乁	朵
2	5.14	横折竖折提	横折提	ㄥ	计 颃 鸠
折	5.15	横折竖钩	横折钩	𠃌 (乛)	同 门 却 永 耍 万 母 仓 / 也
	5.16	横折捺钩	横斜钩（俗称）	乁	飞 风 执
	5.17	竖折横折竖	竖折折	𠃑	鼎 𠄧 亞 吳
	5.18	竖折横折撇	竖折撇	ㄥ (𠃋、ㄑ)	专 / 臾 / 矢
	5.19	竖弯横钩	竖弯钩	乚	己 匕 电 心
3	5.20	横折竖折横折竖	横折折折	𠄎	凸
折	5.21	横折竖折横折撇	横折折撇	㇋	及 延
	5.22	横折竖弯横钩	横折弯钩	乁 (乙)	几 丸 / 艺 亿
	5.23	横折撇折弯竖钩	横撇弯钩（俗称）	㇌	阳 部
	5.24	竖折横折竖钩	竖折折钩	𠃑 (㇉)	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
			𠃍 橫折	乙 橫折橫	㇋ 橫折橫折	㇋ 橫折橫折鉤
			⌐ 橫鉤	乁 橫折彎	乙 橫折曲鉤	
			𠃋 橫撇	𠃌 橫折鉤	ㄅ	
			ㄴ 豎折	乁 橫斜鉤		
			ㄴ 豎彎折右	ㄅ 豎橫折 1		
			ㄴ 豎彎折左	ㄣ 豎橫折 2		
			丨 豎挑	乚 豎曲鉤		
5	Corner	𠃍 折 Zhe	亅 豎鉤			

IRGN 1086A

汉字笔画名称表

一、基本笔画：一(横) 丨(竖) 丿(撇) 丶(点) 乛(折)

二、变形笔画：

笔形	名称	例字		笔形	名称	例字	
⺀	提	刁	江	乀	捺	大	进
⺜	横折	尺	口	乁	横斜钩	飞	汽
⺹	横撇	又	水	乙	横折弯钩	亿	九
⼀	横钩	予	卖	孑	横撇弯钩	队	都
⺄	横折折	凹		㇅	横折折折	凸	
⺄	横折弯	朵	船	㇋	横折折撇	延	及
⺄	横折提	说		㇋	横折折折钩	乃	场
⺄	横折钩	幻	月	㇄	竖提	长	以
亅	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	巴	㇂	斜钩	戈	式
ㄣ	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	一	横	大	11	㇐	横钩	你	21	㇕	弯钩	了
2	丨	竖	十	12	乚	竖弯钩	元	22	乁	横折弯钩	九
3	丿	撇	八	13	㇛	撇折	去	23	㇄	竖弯	四
4	丶	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	ㄣ	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	橫折彎鉤	
九 U+4E5D, 几 U+51E0, 丸 U+4E38, 旭 U+65ED, 杂 U+6742				

㇟ U+31E4	**HZXG**	héng-zhé-xié-gōu	橫折斜鉤	
飞 U+98DE, 风 U+98CE, 气 U+6C14, 虱 U+8671, 瘋 U+760B, 訊 U+8A0A				

㇗ U+31DE	**SZZ**	shù-zhé-zhé	豎折折	
㇗ U+200D1, 卐 U+5350, 亞 U+4E9E, 鼎 U+9F0E, 吳 U+5433				

ㄣ U+31E5	**SZP**	shù-zhé-piě	豎折撇	ㄣ
专 U+4E13, 奊 U+594A, 头 U+5928, 設 U+279AE, 顺 U+244F7, 萛 U+249A1				

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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 ☐ B.1-Specialized (small collection) *X* B.2-Specialized (large collection) ☐
   C-Major extinct ☐ D-Attested extinct ☐ E-Minor extinct ☐
   F-Archaic Hieroglyphic or Ideographic ☐ G-Obscure or questionable usage symbols
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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Yes*

    If YES explain *Unified to other stroke characters in IRGN 1137*

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.)? *Yes*

        If YES, with whom? *IRG experts*

        If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included? *Users of Han character*

    Reference:

4. The context of use for the proposed characters (type of use; common or rare) *Common*

    Reference:

5. Are the proposed characters in current use by the user community? *Yes*

    If YES, where? Reference: *See the representative glyphs in the published ISO/IEC 10646 Standard*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
    in the BMP? *Yes*

        If YES, is a rationale provided? *Yes*

          If YES, reference: *Belongs to the CJK Strokes block*

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? *No*

        If YES, is a rationale for its inclusion provided?

        If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either
    existing characters or other proposed characters? *No*

        If YES, is a rationale for its inclusion provided?

        If YES, reference:

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? *Yes*

        If YES, is a rationale for its inclusion provided? *Yes*

        If YES, reference: *See proposal text*

11. Does the proposal include use of combining characters and/or use of composite sequences? *No*

    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:

12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? *No*

        If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? *No*

    If YES, are the equivalent corresponding unified ideographic characters identified?

        If YES, reference: