

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Ideographic Research Group Document
Title: Comments on how to handle the strokes in the digitized running text
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Arthur Ng Hou Man (吳浩文, University of Macau)
Status: Individual Contribution to IRG #60, online meeting
Action: For consideration by IRG
Date: 2023-02-28

At IRG #59, we had the serious disagreements on [WS2021-00001:SAT-05947](#). So many experts pointed out their comments, but we didn't reach an agreement at that time. IRG decided to keep it in M-Set temporarily.

00001	1.0 一			SAT-05947				
	二							
	SC=0, FS=0 TS=1							

Fig. 0.1 WS2021-00001

The most common uses of the ideographic strokes are listed in tables to teach the students recognize the writing elements of the ideographs and how to write well. The followings are two tables listed in the East Asian textbooks.

常用笔画名称表

笔画	名称	例字	笔画	名称	例字
一	横	开 土	乚	竖弯	四 西
丨	竖	中 上	丿	竖钩	小 可
ノ	撇	天 禾	㇏	弯钩	了 手
㇏	捺	人 尺	㇏	斜钩	我
丶	点	头 下	㇏	撇折	去 东
㇏	提	虫 把	㇏	卧钩	心
㇏	横折	口 五	㇏	撇点	女
㇏	横撇	子 水	㇏	横折钩	力 月
㇏	横钩	你	㇏	竖弯钩	儿 巴
㇏	竖折	山 牙	㇏	横折弯钩	儿
㇏	竖提	长 比	㇏	竖折折钩	马 鸟

Fig. 0.2 中华人民共和国教育部, 温儒敏: 《义务教育教科书 语文 一年级 上册》, 北京: 人民教育出版社, 2016.7, ISBN 978-7-107-31240-3, p. 119

한자의 획				
①	가	로	획	一 生 五
②	가	로	꺾어	ㄱ 子 牙 守
③	내	리	획	丨 丨 千 木
④	내	리	꺾음	ㄱ ㄱ 日 口
⑤	내	리	꺾어	ㄱ ㄱ 月 力
⑥	가	로	꺾음	ㄴ 亡 直
⑦	엇	꺾음		ㄴ ㄴ 女 母
⑧	치	삐	침	ㄷ 海 指
⑨	내	삐	침	ㄷ 手 左
⑩	꺾은	내	삐	ㄷ 水 又
⑪	갈	고	리	ㄷ 民 事 子
⑫	훤		획	ㄷ 九 川
⑬	휘어	돋은	획	ㄷ 凡 見
⑭	과		임	ㄷ 人 建 之

Fig. 0.3 박종원: 《한문 (중학교 제 1 학년용) 3 판》, 평양: 교육도서출판사, 2009.10, 교-09-583, p. 7

However, the uses of the ideographic strokes in the running tests are unfamiliar for us. We try to show some relevant knowledge in this document, and suggest to clarify the encoding principles of component/stroke.

1. Definition

From different perspectives, we can summarize different definitions on the ideographic strokes.

1) We can define the ideographic stroke as the minimum writing unit of the ideograph in the official script (隸書) and regular script (楷書). Note: the definition is not suitable for the ideographs in the seal scripts.

2) we can also define it as the dot or bar written from the movement to touch the paper to the movement to leave the paper.

1. 笔画

笔画是构成现代汉字字形的最小单位。从落笔到起笔所写的点或线叫一笔或一画。

汉字的基本笔画有五类,即“横(一)、竖(丨)、撇(丿)、点(丶)、折(乚)”,又称为“札”字法。传统上还有八类说,包括点(丶)、横(一)、竖(丨)、撇(丿)、捺(㇏)、提(㇀)、折(乚)、钩(亅),即通常所说的“永”字八法。五类说是由八类说归并而成的,比较简明,所以有些辞书的笔画检字法就参考笔画数和“横、竖、撇、点、折”的笔画顺序来编排。

Fig. 1 张文忠, 夏军: 《语言文字规范十五讲》, 上海: 上海大学出版社, 2019.8, ISBN 978-7-5671-3646-5/H·372, p. 26

Based on these two definitions, we can abstract the strokes in the ideographic boxes easily.

2. Handling in regional character set

As we know, the glyphs for some ideographic strokes are the same as some ideographs, such as the horizontal stroke (一) and U+4E00 (一), the vertical stroke (丨) and U+4E28 (丨). Take Big Five as an example here. The original version published by III didn't include any appropriative stroke and component out of Hanzi. The extended version used in Eten system included some strokes which could be shared with the Kangxi radicals firstly.

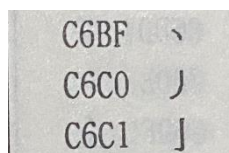


Fig. 2.1 倚天資訊股份有限公司技術資料編輯組:《倚天中文系統使用手冊》,台北:倚天資訊股份有限公司,1990.6,ISBN 957-504-008-2, p. 附錄 6-3

Secondly, they include others with some components which could not be shared with the Kangxi radicals.

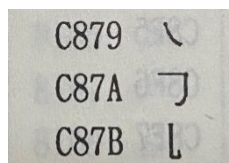


Fig. 2.2 倚天資訊股份有限公司技術資料編輯組:《倚天中文系統使用手冊》,台北:倚天資訊股份有限公司,1990.6,ISBN 957-504-008-2, p. 附錄 6-5

According to the specification of Eten system, the above strokes were treated as the special symbols not Hanzi, and they didn't include some strokes which could be shared with common Hanzi twice. Note: the code points of these strokes in ET1 are different, and Eten extended the TCA code and developed the Eten code based on the same character set of Eten version of Big Five code.

▲ 特殊符號 (SPCFONT、SPCFSUPP)

1. SPCFONT

	使用範圍	字數	保留範圍	字數
標準字	A140~A3BF	408	-----	---
控制碼	A3C0~A3E0	33	A3E1~A3FE	30
合計		441		30

2. SPCFSUPP

	使用範圍	字數	保留範圍	字數
標準字	C6A1~C8FE	408	-----	---
合計		408		0

Fig. 2.3 倚天資訊股份有限公司技術資料編輯組:《倚天中文系統使用手冊》,台北:倚天資訊股份有限公司,1990.6,ISBN 957-504-008-2, p. 附錄 7-2

3. Use in the running text

Mr. Fu Yonghe (傅永和) pointed out the ideographic strokes were abstracted by Madam Wei (卫夫人) firstly in Chinese Jin Dynasty (晋朝), but she didn't name the ideographic strokes at that time. It looks Madam Wei treated them not as the common Hanzi without the readings. Note: another possibility is that the article was written by Wang Xizhi (王羲之) who was the eminent apprentice of Madam Wei.

汉字笔画是汉字从篆书变到隶书之后才产生的。汉字在甲骨文、金文、篆书阶段，可以说是由线条构成的，直到隶书阶段才从线条变成笔画。

晋朝卫夫人在其《笔阵图》一文中，把汉字的笔画分成“一”“丶”“丿”“丨”“㇏”“乚”“丁”七种，但没有给七种笔画定名。

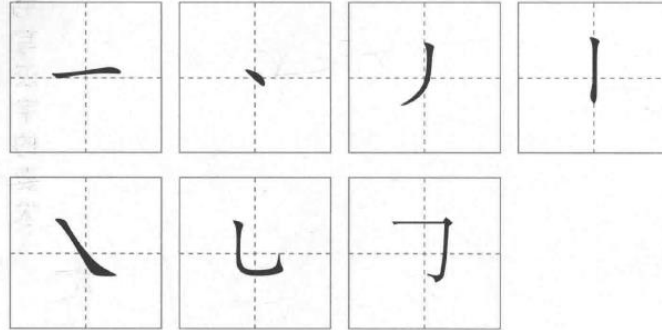


Fig. 3.1 傅永和, 刘栋慧: 《书写汉字的要求》, 杭州: 西泠印社出版社, 2019.1, ISBN 978-7-5508-2470-6, p. 2

When we check one version of the original article of Madam Wei, we found the typographical features of the ideographic strokes are the same as the common Hanzi. That means we can treat them as the common Hanzi totally for typography.

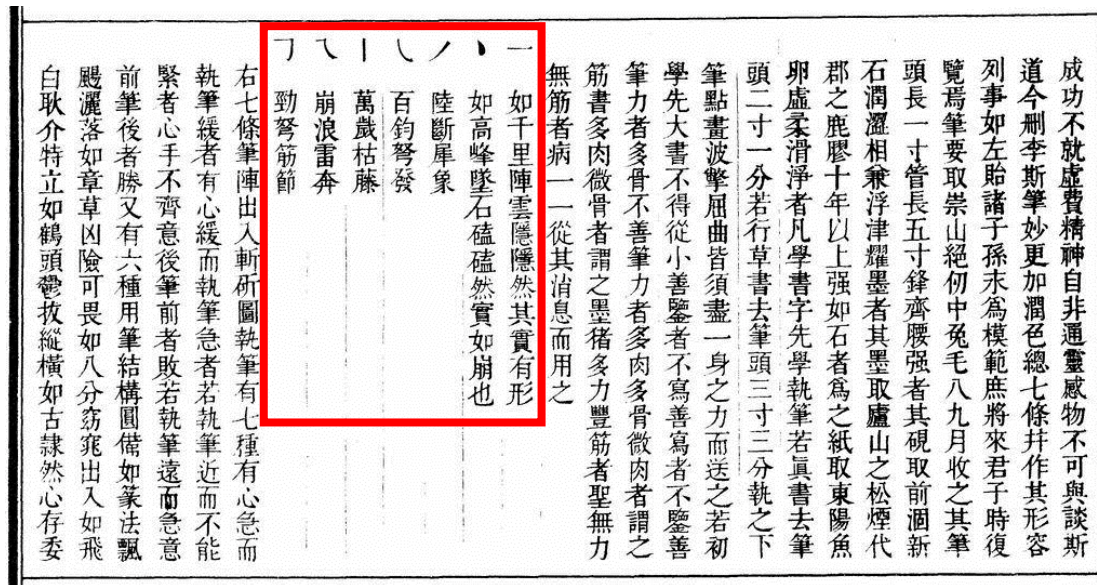


Fig. 3.2 《古今圖書集成字學典》, 民國影本, 第八十三卷

The following two books show the typographic features clearer as we mentioned above. These two books both quoted some sentences of Madam Wei's article, and the ideographic strokes were typeset as common Hanzi without any difference as well.

一分執之下筆點畫波屈曲皆須盡一身之力而送之若初學先大書不得從小善鑒者不寫善寫者不鑒善筆力者多骨不善筆力者多肉多骨微肉者謂之筋骨多肉微骨者謂之墨猪多力豐筋者聖無力無筋者病一一從其消息而用之

一如千里陣雲隱隱然其實有形、如高峯墜石磕磕然實如崩也、陸斷犀象、百鈞弩發、萬歲枯藤、崩浪雷奔、勁弩筋節

Fig. 3.3 汪砢玉：《珊瑚網》，文淵閣四庫全書本，卷二十三上

強如石者紙取東陽魚卵虛柔滑淨者然後靜據意思
揮襟作之先學執筆若真書去筆頭二寸一分若行草
去筆頭三寸一分執之下筆點畫波掣屈曲皆須盡一
身之力而送之初學先須大書不得從小善鑒者不寫
善寫者不鑒善筆力者多骨不善筆力者多肉多骨微
肉為筋書多肉微骨為墨猪多力豐筋者聖無力豐肉
者病一一從其消息而用之一如千里陣雲隱隱然其
寔有形、如高峯墜石磕磕然寔如崩也、陸斷犀象
、百鈞弩發、萬歲枯藤、崩浪雷奔、勁弩筋節用
筆有七種有心急而執法緩者有心緩而執法急者若
執筆近而不能緊者心手不齊意後筆前者敗若執筆
遠而急意前筆後者勝又有六種用筆結構圓備如篆
法飄颻灑落如章草凶險可畏如八分窈窕出入如飛
白耿介特立如鶴頭鬱拔縱橫如古隸然心存委曲每
為一字各象其形斯造妙矣義之題其後曰夫紙者陣
也筆者刀稍也墨者鍔甲也研水者城池也心意者將

Fig. 3.4 鄭杓, 劉有定: 《衍極》, 文淵閣四庫全書本, 卷下

We also found the ideographic strokes were typeset in the Han running text like common Hanzi in the books on the traditional Chinese medicine and Chinese calligraphy.

心说

心，火脏，身之主，神明之舍也。《小篆》尝言，心字篆文只是一倒火字耳。盖心，火也，不欲炎上，故颠倒之，以见调燮之妙也。祝无功曰：庖氏一画，直竖之则为丨、左右倚之则为丿为㇏，缩之则为丶，曲之则㇏，㇏、圆而神，一、丨、丿、八方以直，世间字变化浩繁，未有能外一、丨、丿、结构之者。独心字欲动欲流，圆妙不居，出之乎一、丨、丿、八之外，更索一字与作对不得。正以心者，新也。神明之官，变化而日新也。心主血脉，血脉日新，新新不停，则为平人，否则疾矣。其合脉也，其荣色也，开窍于舌。

Fig. 3.5 陈修园; 刘从明, 王明惠: 《陈修园医学歌诀五种》, 郑州: 河南科学技术出版社, 2021.3, ISBN 978-7-5725-0280-4, p. 53

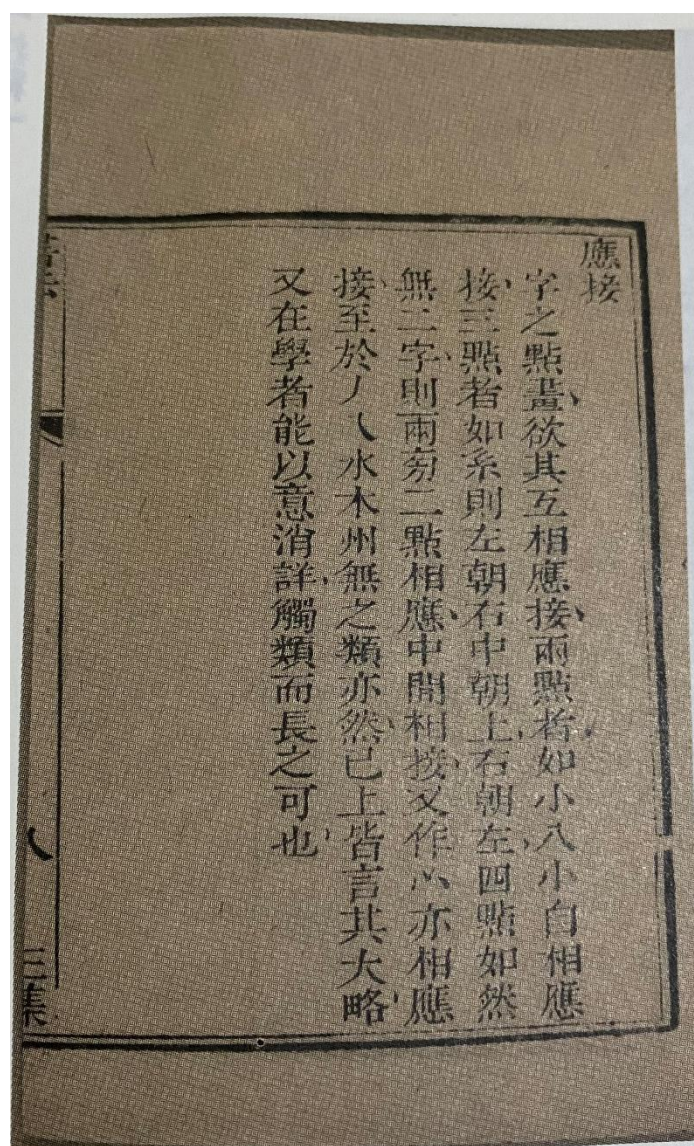


Fig. 3.6 欧阳询, 王道焜, 房弘毅, 卜希旸: 《欧阳询的书论与书法》, 北京: 文化艺术出版社, 2010.8, ISBN 978-7-5039-4579-3, p. 26

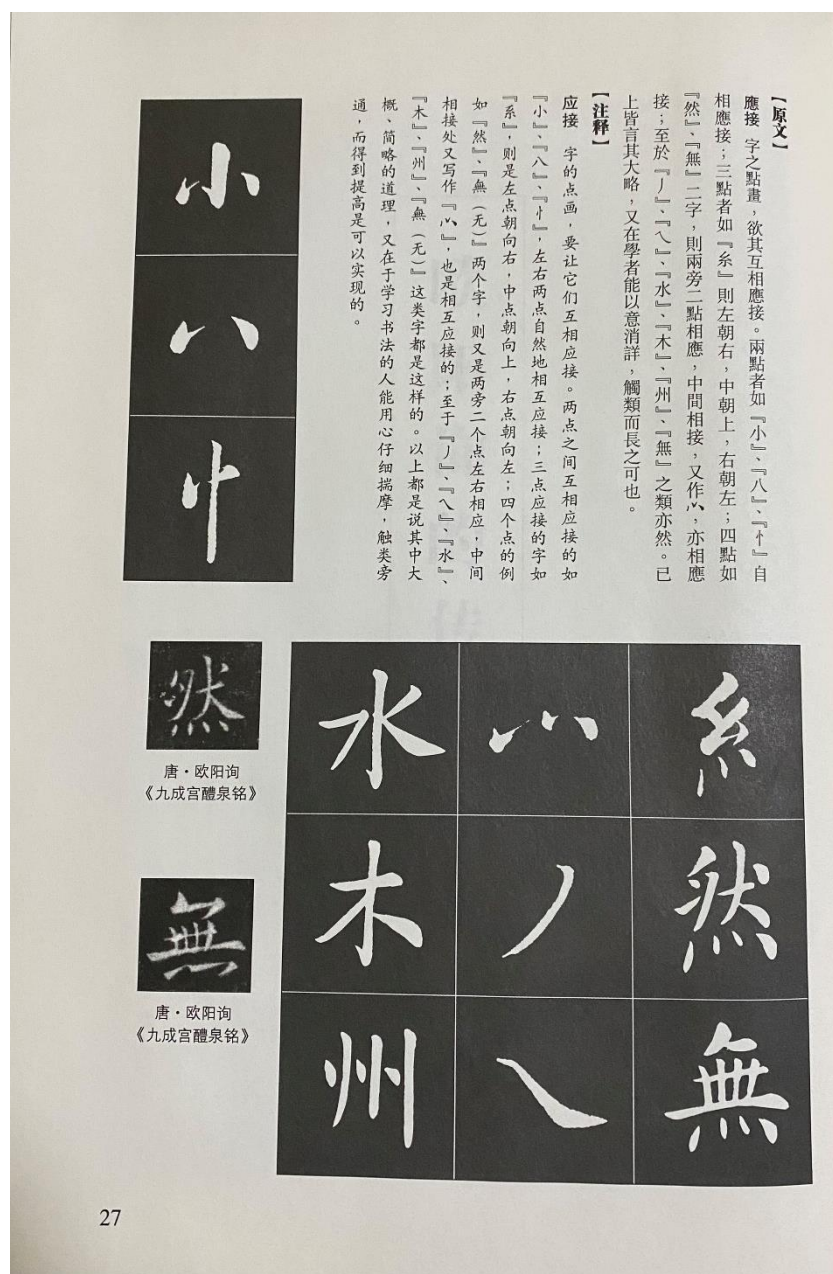


Fig. 3.7 欧阳询, 王道焜, 房弘毅, 卜希旸: 《欧阳询的书论与书法》, 北京: 文化艺术出版社, 2010.8, ISBN 978-7-5039-4579-3, p. 27

4. Previous experience in UCS

As we know, HKSCS was developed on the basis of Eten version of Big Five. HK SAR inherited the definition mentioned in Section 2, and they also defined more ideographic strokes later. HK SAR once submitted the same ideographic stroke as WS2021-00001:SAT-05947 to CJK Ext. C1 project with other common Hanzi together in IRGN893, which the SN is also 00001 in that project.


0075.011	0	
—	0	
HK-8840		

Fig. 4.1 HK-8840 in IRGN893

00001	
0075.011	✓
00	
HongKong	HK-8840

Fig. 4.2 C1-00001 in IRGN907

However, HK-8840 had been removed since Version 3.0. This one is mapped to U+31C0 (ㄟ) in the later version of HKSCS.

31C0
✓
H-8840

Fig. 4.3 H-8840 in HKSCS-2016

Other ideographic strokes all have been mapped to the CJK strokes if there is no more Hanzi uses.

5. Typographic rendering effects

Let's see how UCD shows on the CJK strokes and CJKUI first before we observe the typographic rendering effects.

The UCD information is shown as below.

```
nnnn;<CJK Stroke>;So;0;ON;;;;N;;;;;
nnnn;<CJK Radical>;So;0;ON;;;;N;;;;;
nnnn;<Kangxi Radical>;So;0;ON;;;;N;;;;;
nnnn;<CJK Ideograph>;Lo;0;L;;;;N;;;;;
```

It is not hard to know that one CJK stroke is assigned as So (Other_Symbol: *a symbol of other type*) like CJK radical and Kangxi radical, but one CJKUI is assigned as Lo (Other_Letter: *other letters, including syllables and ideographs*).

The script information is shown as below.

```
31C0..31E3 ; Common # So [36] CJK STROKE T..CJK STROKE Q
2E80..2E99 ; Han # So [26] CJK RADICAL REPEAT..CJK RADICAL RAP
2F00..2FD5 ; Han # So [214] KANGXI RADICAL ONE..KANGXI RADICAL
FLUTE
4E00..9FFF ; Han # Lo [20992] CJK UNIFIED IDEOGRAPH-4E00..CJK
UNIFIED IDEOGRAPH-9FFF
```

It is also not hard to know that one CJK stroke is related to the common script (Zyyy), but CJKUI, CJK radical and Kangxi radical are all related to the Han script (Hani, etc). However, the Script_Extensions value for the CJK stroke is also Hani, that means it could be used as CJKUI when it is used in the running Han texts.

```
# Script_Extensions=Hani
31C0..31E3 ; Hani # So [36] CJK STROKE T..CJK STROKE Q
```

The line breaking information is shown as below.

```
31C0..31E3;ID # So [36] CJK STROKE T..CJK STROKE Q
2E80..2E99;ID # So [26] CJK RADICAL REPEAT..CJK RADICAL RAP
2F00..2FD5;ID # So [214] KANGXI RADICAL ONE..KANGXI RADICAL
FLUTE
4E00..9FFF;ID # Lo [20992] CJK UNIFIED IDEOGRAPH-4E00..CJK
UNIFIED IDEOGRAPH-9FFF
```

The line breaking classes for CJK stroke, CJK radical, Kangxi radical and CJKUI are all ID (Ideographic: *break before or after, except in some numeric context*). We found the typographic rendering effect on the line breaking is not good in Microsoft Office, but well in the browsers.

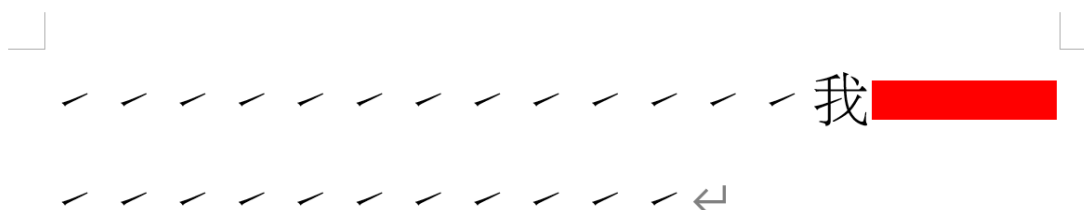


Fig. 5.1 Multiple U+31C0 (-) with one CJKUI on Word (Microsoft 365)

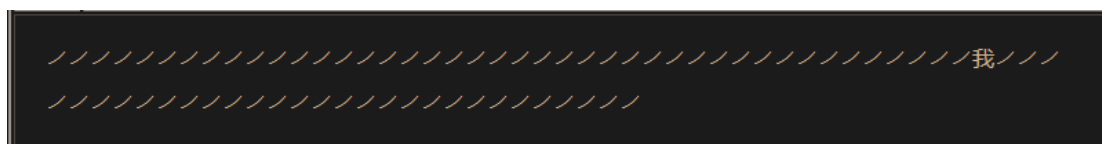


Fig. 5.2 Multiple U+31C0 (-) with one CJKUI on Microsoft Edge with html

We try to change the recording way in document.xml under word folder in the corresponding docx file to make the CJK strokes and CJKUI included in the same element, but the typographic rendering effect is still the same as Fig. 5.1 shows in Word. Another expert helps us confirm this behavior is not related to [Uniscribe](#) and [DirectWrite](#).

To sum up, it is OK to use the CJK stroke as common Hanzi for UCS and Unicode, if there is not more sufficient evidence to support as an individual Hanzi, but we suggest Microsoft update the line breaking behaviors on CJK stroke, which is not a reason to encode it and other ideographic strokes as CJKUI.

6. Suggestions for encoding

Based on the discussion at the previous IRG meeting and the above clarifications, we feel the encoding principles of component/stroke seem to be needed to clarify, which will help define the way of accepting character in our future works. The purpose to delimit the definition of component/stroke is to clarify whether it belongs to component/stroke, especially when it comes from the explanation notes of characters/vocabulary in ancient documents.

The followings are some preliminary ideas.

- 1) Reject more abstract CJK stroke in future;
- 2) For the abstract component, there must be more than one Hanzi which is composed by the component;
- 3) If any expert can provide the running text as the evidence for one component and it is commonly used in some areas, we can consider treating it as a special CJKUI, but it is not encouraged;
- 4) It should be used rarely in the daily life.

7. Acknowledgement

Clerk Ma provides some helps on the rendering issues.

(End of Document)