

ISO/IEC JTC1/SC2/WG2/IRGN2154b

Doc. # : Korea JTC1/SC2 k2326b ROK Normalization Rule(Draft2)

Korea JTC1/SC2, Committee on Character Codes

ISO/IEC JTC1/SC2/WG2/ Ideographic Rapporteur Group

Authors: Shin Sanghyun, Cho Sungduk, Choi Jinswook, Pyo Seungju,
Heo Chul, Choi Jungdo

Title: ROK Normalization Rule(Draft2)

Date: 2016. 10. 13.

Status: National Body Position, Rep. of KOREA

Action: For consideration by IRG

This document refers to the glyph normalization rule(Draft2) of ROK.

1. Introduction

This document describes the rule for normalizing the shape of glyph(Hanja) that can be found on the historical books. There is some characters(Hanja) that can be distinguished from the general variant form in the historical books. We cannot handle this kind of characters by unification rule because the unification mainly deals with the issue of variant form in the printed script, not in the written script. Written script is sometimes represented as cursive script, sometimes as semi-cursive script and so on. So we need to normalize these written script into regular or standard script. The aim of this document is to make a normalization rule for making a regular script and make a consensus about the normalization standard for IRG work.

1.1. Scope of this document

The aim of this document is to clarify the normalized character from the various glyphs in the form of written script that can be found in woodblock script or manuscript without any change in their meaning. That is only different glyphs, same pronunciation, same meaning by writer, this case is very temporary, and personal case, cannot generalized. so, don't need set new character code point. Contrariwise, if we found new glyphs look like very similar with some variant glyphs, but have different meaning, this is not included normalization rule, in this case, we have to think unification with other characters or set new character code point.

1.2. Scope of Normalization

We can apply the normalization rule when we meet the below

conditions.

a. when we need to normalize the cursive script found in woodblock script or manuscript

b. there seems to have a kind of typographical error considering the context of the full text in the non-printed script


c. when we meet the glyphs that have the same component(element or module), but their shape are different because of the writing custom fixed by tradition in the non-printed script

1.3. Differences in 'Normalization' and 'Unification'

a. A case of cursive script


① 麗


i  → 麗 → 麗

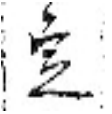
ii  → 麗

iii  → 麗 → 麗

② 定

i  → 定 → 定

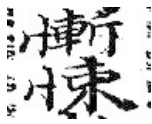
ii  → 定 → 定

iii  → 定 → 定

In case of ①-i and ①-iii, mostly normalization to “麗”, but some case normalization to “麗”. If the certain peoples name, place name, book title and others, writers deliberately write a certain variant form “麗”, normalization to “麗”. In case of ②-i, ②-ii, ②-iii, not Proper noun, not normalization to “芝”, normalization to “定”.

b. customary written form to normalization

① 忄

 → 慚悚

② 日

特 → 時

The ① and ② is the other component shape which was used widely in the 18th and 19th centuries in Korea. This case is not writers deliberately write, is case was used as a customary written form. Among this customary written form, such component shape case occurs very commonly like a custom rule, this is typical example of normalization, see below case.

2. Source reference (please refer to the IRGN2083c (KR's supplementary evidences))

The source references used in this document are originated from the IRG document N2083c.

- source reference book

A: 韓國文集叢刊 本輯

B: 韓國文集叢刊 續輯

- the source reference consists of 5 components.

-> book number (e.g.: 92 means 92輯)

-> page number (e.g.: 15 means 15 page)

-> the position within a page

a: upper right

b: upper left

c: lower right

d: lower left

-> a vertical line number

-> publication date

甲: 17C

乙: 18C

丙: 19C

丁: 20C


As an example, consider "步 [B078;316d;6(丙)]"

[B078;316d;6(丙)] can be understood as: 韓國文集叢刊 續輯, 78輯 (book 78), page 316, lower left, in vertical line 6, published at 19C.

3. Normalization Table (ROK, 2016. 7. 10.)

This table shows 9 types of normalization component shape like 步, 革, 轟, 尢, 必, 竹, 亥, 攴, 辛. Besides this 9 types, Normalization type of entry can be cleaned up with about 900 species.

Id	Variant component shape	Normalization component shape	Evidence example & source	comment
1-1	步 _(少)	步 _(少)	步-  [B078;316d;6(丙)] 涉-  [A317;526c;8(丁)] 賓-  [A272;27b;8(丙)], 賓-  [A12;372c;10(鮮)] 頻-  [A50;14d;5(甲)] 輦-  [A218;391c;1(乙)]	※ refer to the UCV_N009. This type is the Variant component shape with 少 or a component look like 少.
1-2	尢	尢	砂-  [A272;27c;4(丙)], 紗-  [A272;32c;2(丙)], 沙-  [A1;293c;7(乙)], 陟-  [A104;175d;6(丙)] 殍-   KC-05220	Same above.
1-3	尢	尢	尢-  [A50;30b;7(甲)],	Same above.
1-4	尢	尢	尢-  [A40;445b;6(甲)],	Same above.
1-5	尢	尢	尢-  [B061;322d;8(乙)]	Same above.
1-6	尢	尢	尢-  [B094;516b;1(丙)]	Same above.

1-7	少	少	隹 [B081;501a;4]	Same above.
1-8	尒 _(小)	尒 _(少)	歲 [B069;097c;1(乙)] 穢 [A59;180d;7(甲)]	※ refer to the UCV_N015.
1-9	歲	歲	歲 [B081;500d;9] 穢 [B081;059;a;8(18後)],	same above.
2-1	革		鞞 [B088;100d;4(丁)], 鞋 [B068;396c;2(乙)],	This type is the Variant component shape of 革.
2-2	鞞		鞞 [A327;266d;6(丁)],	same above.
2-3	鞞	革	鞞 [A50;23a;6(甲)], 鞞 [A015;296d;6(甲)], 鞋 [B023;036b;1(丙)], 鞞 [B084;202d;9(丙)], 鞞 [B113;541b;1(丙)], 鞞 [A316;123c;2(丙)] 鞞 KC-04469 	same above.
2-4	鞞		鞞 [B071;471c;7(丙)], 鞞 [B064;237c;9(丙)], 鞞 [B071;477a;4(丙)], 鞞 [B064;241a;7(丙)],	same above.
2-5	鞞		鞞	same above.

2-6	𦍋		鞮-鞮 [A217;22a;3(乙)], 勒-勒 [B080;325a;5(乙)], 鞅-鞅 [B091;044c;3], 鞋-鞋 [B080;335a;9], 鞭-鞭 [A291;512a;1],	same above.
2-7	𦍌		革-革 [A316;142a;3(丙)],	same above.
2-8	𦍍		鞞-鞞 [B061;322b;10(乙)],	same above.
3-1	𦍎	聶	攝-攝 [B023;027b;2(丙)], 囁-囁 [A217;13b;9(乙)], 囁-囁 [A1;298d;8(乙)], 躡-躡 [A2;346a;5(丙)], 鑷-鑷 [A50;18c;2(甲)],	This type is Variant component shape when 耳 is repeated some times. 聶 is 3 times repeated 耳 component, when write on purpose drop a one stroke in the middle.
3-2	𦍇	聶	攝-攝 [A69;216d;4(丙)], 躡-躡 [B104;087c;3(丙)], 囁-囁 [B104;383d;6(丙)], 攝-攝 [B113;442c;8(19後)], 𦍇-𦍇 [A047;406a;4]	same above.
3-3	𦍈	聶	躡-躡 [B064;283d;10(丙)],	same above.

3-4	聶	聶	聶-聶 [B091;056b;2],	same above.
3-5	聶	聶	聶-聶 [B080;371d;3(18後)],	This type is repeated 聶聶 change write to 𠂔.
3-6	聶	聶	聶-聶 [B082;392a;10(乙)],	same above.
4-1	流	流	疏-疏 [b104;126c;3(丙)] 疏-疏 [b104;133d;7(丙)] 流-流 [B064;031b;10(丙)] 梳-梳 [B084;028c;5(丁)]	This type is Variant component shape of 流 Top component 𠂔 change to 二. show often on character which include 流 or 疏.
4-2	流	流	流-流 [A2;242b;1(乙)],	same above.
4-3	流	流	流-流 [A92;14a;8(甲)], 流-流 [A128;237c;2(丁)], 流-流 [A15;117c;6(鮮)], 流-流 [B070;314a;8(丙)], 流-流 [A216;131b;9(丙)], 流-流 [B064;313d;5(丙)], 流-流 [B066;267d;8(丙)], 流-流 [A319;512a;7(丁)], 𠂔-𠂔 KC-04749 - 𠂔	same above.

4-4	𣶒	充	流-流 [A58;357d;6(乙)], 梳-梳 [A17;331a;6(鮮)], 毓-毓 [A7;492b;7(麗1)], 疏-疏 [A1;415b;10(乙)], 蔬-蔬 [A1;300a;1(乙)], 踰-踰 [A3;88b;6(麗)], 琉-琉 [a12;296c;4(鮮2)],	same above.
4-5	𣶒	亢	荒-荒 [A7;495c;9(麗1)],	same above.
4-6	𣶒	充	流-流 [安東金氏己未譜序, 1b;5]	same above. 儿 changed to 𣶒.
5-1	宀(又)	宀(必)	密-密 [A316;277c;6(丙)], 蜜-蜜 [b104;087a;1(丙)], 密-密 [A289;424d;9]	This type is various form of 必. 必 is from 宀→宀→宀→宀 change to 又 or 又.
5-2	又	必	密-密 [A316;140a;1(丙)],	same above.
6-1	𣶒	竹	筍-筍 [A218;417c;4(乙)],	This type is various form of 竹. In hand writing change to 竹, 𣶒, 𣶒, 𣶒, 𣶒, 𣶒.
6-2	𣶒	竹	筍-筍 [A53;115a;7(甲)], 鑄-鑄	same above.
6-3	𣶒	竹	嗒-嗒 [B086;323c;4],	same above.

			塔 [B062;529c;4(乙)], 簞 [a12;238b;6(鮮2)],	
6-4	𦵏	竹	簞 [kc_04516]	same above.
6-5	𦵏	竹	籠 [B091;060c;4],	same above.
6-6	𦵏	竹	籬 [B081;505d;5], 簞 [B091;045b;7], 筑 [B091;049b;1], 簞 [B091;046b;8], 簞 [B091;045b;10],	same above.
6-7	𦵏	竹	筍 [B082;082b;8(乙)],	same above.
7-1	𦵏	亥	𦵏 [KC-04768 - 𦵏] 亥 [B071;477c;4(丙)], 該 [A333;099b;6(丁)], 骸 [A327;252c;2(丁)], 賅 [A328;013d;6(丁)], 駭 [b113;445a;6(19後)], 菱 [B094;513d;2(丙)],	This type is various form of 亥.
7-2	亥	亥	骸 [A272;18d;5(丙)],	same above.

7-3	亥	亥	亥-亥 [A59;140c;3(甲)], 該-該 [A59;136d;1(甲)], 駭-駭 [A59;136d;5(甲)], 骸-骸 [A42;145c;3(甲)], 菱-菱 [a19;117c;8(鮮2)], 核-核 [a19;114b;3(鮮2)], 劾-劾 [A1;289c;4(乙)],	same above.
7-4	亥	亥	咳-咳 [B068;520d;8(丙)],	same above.
7-5	亥	亥	亥-亥 [B079;231c;9(丙)],	same above.
7-6	亥	亥	亥-亥 [B094;295a;9(丙)],	same above.
7-7	亥	亥	亥-亥 [a3;23c;9(麗)], 駭-駭 [a3;35a;8(麗)],	same above.
7-8	亥	亥	亥-亥 [a3;16d;6(麗)],	same above.
7-9	亥	亥	骸-骸 [B091;045b;8(18後)],	same above.
8-1	友	友	髮-髮 [B080;104c;1(乙)], 髮-髮 [B064;342a;2(丙)], 黻-黻 [A330;013a;6(丁)],	※refer to UCV_N256. This type is various form of 友.

8-2	𦍋	𦍋	拔-拔 [B094;694c;2(丁)], 跋-跋 [a7;472c;5(鮮1)], 𦍋-𦍋 [a19;123d;4(鮮2)], 被-被 [A272;34b;10(丙)], 芟-芟 [A010;394b;10(乙)], 絨-絨 [A5;319d;2(乙)],	same above.
8-3	𦍋	𦍋	拔-拔 [B094;350c;3(丙)], 跋-跋 [B061;278a;3(乙)], 𦍋-𦍋 [B104;218a;4(丙)], 被-被 [B112;133d;4(丙)], 𦍋-𦍋 [B102;271b;6(丙)], 絨-絨 [A273;16a;3(丙)],	same above.
8-4	𦍋𦍋	𦍋	拔-拔 [A1;275a;3(乙)], 𦍋-𦍋 [b101;485b;4(丙)], 跋-跋 [A59;160b;1(甲)],	same above.
8-5	夭	𦍋	拔-扶 [A032;014d;1(丁)] 跋-跃 [b113;629c;8(19後)],	same above.
8-6	犬	𦍋	髮-𦍋 [a3;34b;6(麗)],	same above.
8-7	𦍋	𦍋	髮-𦍋 [a3;78a;1(麗)],	same above.

8-8	火	𤇀	髮-髮 [B071;477a;5(丙)],	same above.
8-9	犬	𤇀	跋-跋 [a3;49b;2(麗)], 拔-拔 [a018;335d;5(鮮2)],	same above.
8-10	友	𤇀	拔-拔 [B104;082d;5(丙)], 跋-跋 [B104;182a;8(丙)], 紱-紱 [B104;276c;1(丙)],	same above.
8-11	𤇀	𤇀	跋-跋 [a3;48b;8(麗)], 拔-拔 [a25;360c;8(鮮2)] 紱-紱 [b104;276c;1(丙)],	same above.
8-12	𤇀	𤇀	拔-拔 [a3;55d;10(麗)], 跋-跋 [A42;100b;1(甲)],	same above.
8-13	𤇀	𤇀	𤇀-𤇀 [a12;324d;8(鮮2)], 被-被 [A079;162b;1(甲)], 芟-芟 [B091;053b;8], 紱-紱 [B091;036d;2],	same above.
8-14	𤇀	𤇀	跋-跋 [A1;294a;1(乙)], 拔-拔 [A22;242b;6(甲)],	same above.
9-1	辛	辛	辛-辛 [A3;21d;3(麗)], 辭-辭 [A17;359b;4(鮮)],	This type is various form of 辛. under 立 add 1 stroke on 十.

			辭- 辭 [A3;531d;6(甲)], 辭- 辭 [A40;469a;3(甲)], 宰- 宰 [A216;136d;9(丙)], 滓- 滓 [A42;40a;8(甲)], 梓- 梓 [B075;260a;2(甲)], 壁- 壁 [B073;189d;5(乙)], 躡- 躡 [B085;095a;10(丙)], 僻- 僻 [A19;120c;8(鮮)], 辨- 辨 [A316;147a;4(丙)],	
--	--	--	--	--