INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 2/WG 2

Universal Multiple-Octet Coded Character Set (UCS)

ISO/IEC JTC 1/SC 2/WG 2 N4572
2014-04-01

Title:	**Preliminary Proposal to Standardize Variation Selectors for U+3013**
Source:	**suzuki toshiya**
Status:	**Individual Contribution**
Action:	**For expert consideration**
Distribution:	**ISO/IEC JTC 1/SC 2/WG 2, ISO/IEC JTC 1/SC 2 and Liaison Organizations**

Abstract

Since the standardization of Emoticons (so-called Emoji) and Variation Selectors for their visual shapes, some typographic requirements to control the visual shapes are being introduced to the plain text encoding by using VS technology. The GETA MARK ("=" U+3013) could be considered as one of the characters which have so many visual shapes to be distinguished. In this document, the merits to standardize Geta VS (GVS) are discussed.

1. Background of Separately Coded GETA MARK

The geta mark was originally coded by Japanese coded character set JIS C 6226-1978, and Chinese standard GB 2312-1980 included it too (Figure 1). The origin of geta is recognized as the upside-down metal sort while the typesetters could not find the appropriate glyph from their collection. If the upside-down metal sort is included, the inked image by 2 feet and 1 groove (see Figure 2) is printed.

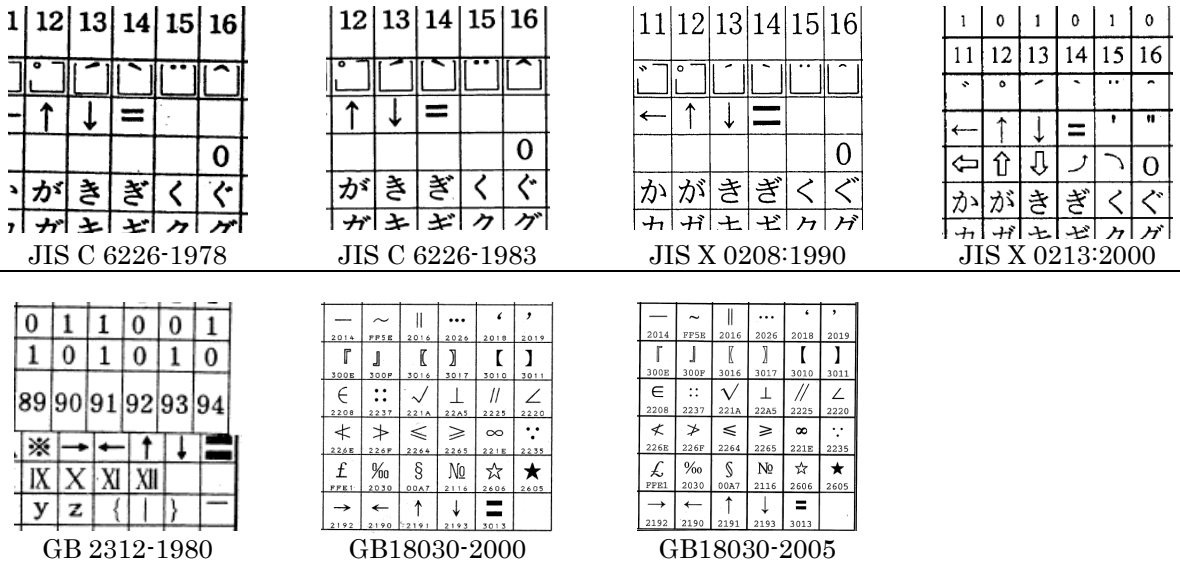
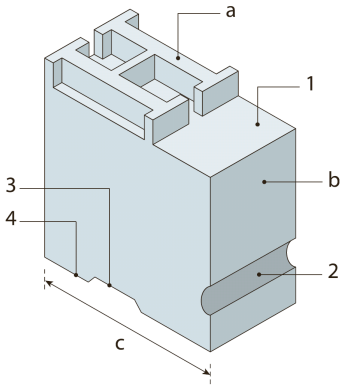


Figure 1: GETA glyphs in Japanese and Chinese Standard Charset Specs.

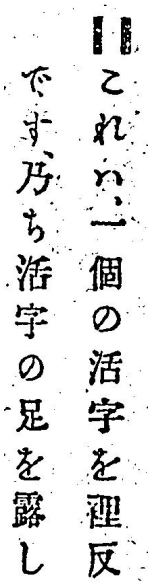
2. Minimum Requirement of the Variation Selectors for GETA MARK

As the inked image of the original geta mark was determined by the feet and groove of the metal sort, the it was looking like a heavy 2 lines at the left and right (in the vertical writing mode, Figure 2) ends of an em square, or at the top and bottom ends of an em square (in the horizontal writing mode). As a result, the glyph looking like a heavy equal is not truthful image for the original geta mark shown in Figure 2.

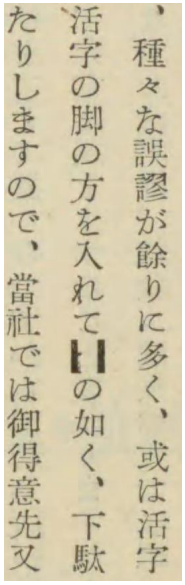


a: Face, b: Body (Shank),
c: Point size,
1: Shoulder, 2: Nick,
3: Groove, 4: Foot

Diagram of a cast metal sort (Wikipedia)



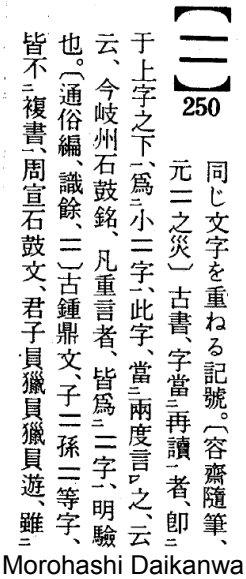
活版術楷梯(Kappanjutsu Kaitei), p.10



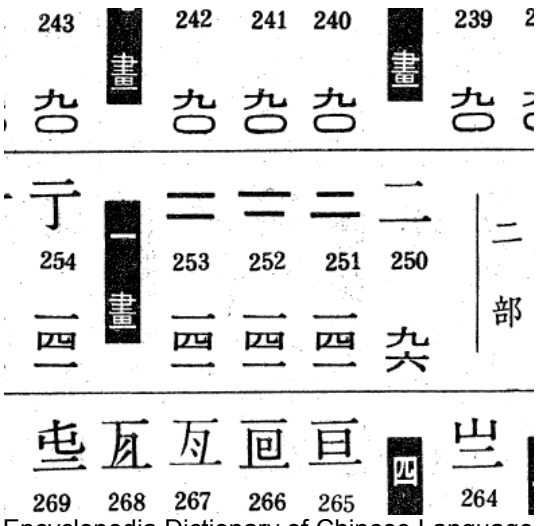
印刷の栞(Insatsu No Shiori), p. 36

Figure 2: Original GETA appearance in Japanese Typesetted Materials

JIS C 6226-1978 (and its successors like JIS X 0208, JIS X 0213) used light weight glyph looking like a heavy equal. In Chinese standards from GB 2312-1980 to GB 18030-2000, the glyph for geta mark was more similar to the original one, but GB 18030-2005 changed it to new one. Considering such historical changes, it would not be reasonable to revert the representative glyph in UCS code chart from current one to the original style. Also an idea to allocate new codepoint to interchange the original geta shape is questionable, because there is no semantic difference. So the variation sequence to distinguish old style (Figure 2, GB 2312-1980, GB 18030-2000) and new style (JIS and GB 18030-2005) would be a considerable option.
It should be noted that it is difficult to distinguish a geta mark with too light weight from the repetition mark (U+20120) in Morohashi Daikanwa, because Morohashi Daikanwa and its derivertives (e.g. Encyclopedia Dictionary of Chinese Language, Mojikyo, etc) uses sans serif typeface for this character exceptionally (Figure 3).



Morohashi Daikanwa Vol.1, p. 453



Encyclopedia Dictionary of Chinese Language

Figure 3: Repetation Marks looking like light-weight GETA glyphs

3. Requirement of the Registry for GETA MARK Variation Selectors

As described in above, the visual representation of the geta mark was not standardized in the metal typesetting era, because the feet and groove were not specified in the industrial standards (the groove was a result to remove the extra metal). Some tutorial texts for Japanese metal typesetters list several options when the typesetters could not find the appropriate glyph; the upside-down metal sort, the 90-degree rotated ideographic glyph (Figure 4).

> "活字がないため、あり合わせの活字を裏向けたり、或いは表向けても横にしたり、○や□を入れたりして組み込んで置くこと"
>
> *Because the typesetter could not find the (right) metal sort, the different metal sort is put in the upside-down pose, or over sideways, or substituted by ○ or □"*

It is expected that the visual shape of the geta mark could not be limited to 2 or 3 appearances. Therefore standardization with the registry like IVS is expected. The significant advantage of the GVS registry is that it could be useful to interchange the possibly-unencoded characters with the ideographic metrics.

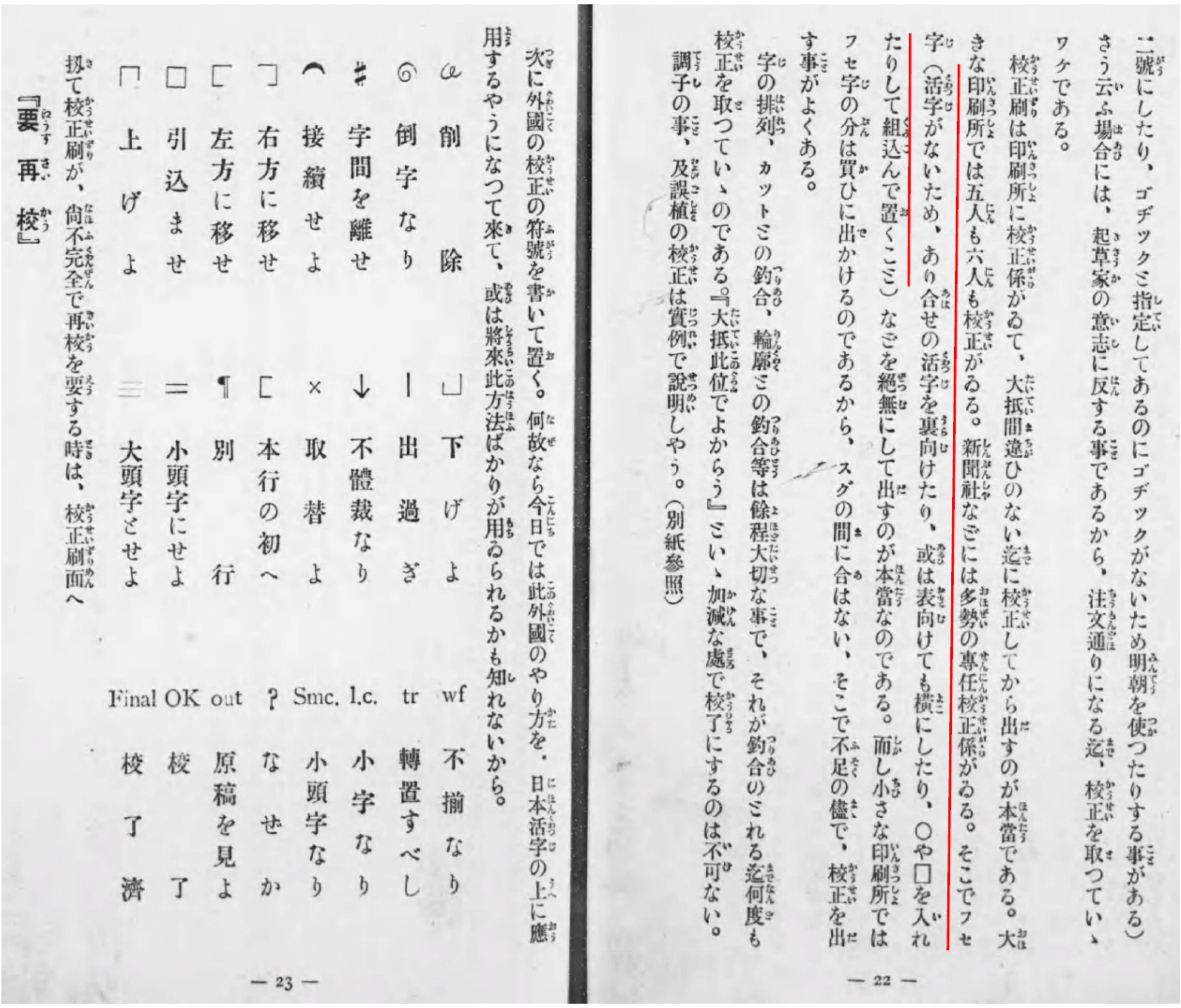
- On 2007, WG2/IRG N1327 proposed the idea to have new block for the speedy standardization of the ideographs for the proper names, without the applying the unification rules, by the consideration of the long working time to standardize CJK Unified Ideographs Extension C, D and E. Unfortunately, the consensus to propose it from IRG to WG2 was not formed in IRG. There was no official conclusion, but the remarkable objection to the idea is that it is not easy for users to select appropriate block to interchange a glyph, therefore the shortcut for the speedy standardization of the possibly-unencoded characters may induce the future works for IRG to check the duplicated encoding of the character. After the standardization of CJK Unified Ideograph Extension D, the process of urgently needed character (UNC) was documented, but the definition of the urgency is controversial especially for the requests by non governmental organization.
- The registration of IVS was started to help the visual distinction of the unifiable glyphic variants since 2007. Also the required information for the IVS registration is fewer than that for the unified ideograph submission (e.g. printed matter evidence, typographic image of the character, IDS, radical-stroke, first stroke, etc). But IVS is not the complete solution for the speedy standardization of the possibly-unencoded ideographs either, because the base character (with unifiable glyphic difference) is required to allocate an IVS; as far as the registrant (or IRG) could not be sure the unification of the submitted glyph with an existing coded character, IVS could not be allocated. As a result, the time to standardize the possibly-unencoded characters is not reduced by IVS.
- Recent submissions to CJK Unified Ideographs may have the difficulty to identify the character by the printing style glyph. Many evidences are handwritten material or faintly printed (and it is difficult to collect multiple evidences for the rarely used character), therefore the discussion to identify the stable glyph shapes is becoming the time consuming process. Many edge cases are found and the consensus in IRG is not easy to form (IRG N1907, N1914, N1921, N1978, N1983).

Recent discussion in Unicode mailing list (`http://unicode.org/pipermail/unicode/2014-March/000275.html`) shows the expect of the shortcut for the speedy standardization of the method to interchange possibly-unencoded ideograph-like glyph via the plain text. Considering that the speedy standardization could not request the consensus and the responsibility of the ideographic experts, the information interchange via the geta mark would be a conventional and considerable option with appropriate trade-off of the stability. Of course, the glyph specified by GVS should not be prohibited to submit to the future versions of the ISO/IEC 10646.

4. Items to be Discussed by the Experts

The important item for further discussion is who manages the variation sequences and the representative glyph. Possible candidates would be following organizations;

- Unicode Consortium, as IVS is managed by.
- ConScript Unicode Registry, as they work for PUA assignment coordination.
- Registration Authority of ISO/IEC 10036, as they assign a unique sequence number to the glyphic image as a part of the JTC1 standard.
- New working group under WG2, *Geta Experts Tentative Association*, to maintain the registry before the official relegate to the appropriate organization.


二號にしたり、ゴヂックと指定してあるのにゴヂックがないため明朝を使つたりする事がある）さう云ふ場合には、起草家の意志に反する事であるから、注文通りになる迄、校正を取つていゝワケである。

校正刷は印刷所に校正係がゐて、大抵間違ひのない迄に校正してから出すのが本當である。大きな印刷所では五人も六人も校正がゐる。新聞社などには多勢の專任校正係がゐる。そこでフセ字（活字がないため、あり合せの活字を裏向けたり、或は表向けても横にしたり、〇や□を入れたりして組込んで置くこと）などを絶無にして出すのが本當なのである。而し小さな印刷所ではフセ字の分は買ひに出かけるのであるから、スグの間に合はない、そこで不足の儘で、校正を出す事がよくある。

字の排列、カットとの釣合、輪廓との釣合等は餘程大切な事で、それが釣合のとれる迄何度も校正を取つていゝのである。『大抵此位でよからう』といゝ加減な處で校了にするのは不可ない。

調子の事、及誤植の校正は實例で説明しやう。（別紙參照）

— 22 —

次に外國の校正の符號を書いて置く。何故なら今日では此外國のやり方を、日本活字の上に應用するやうになつて來て、或は將來此方法ばかりが用ゐられるかも知れないから。

ℓ	削除	⊔	下げよ	wf	不揃なり
⑥	倒字なり	|	出過ぎ	tr	轉置すべし
♯	字間を離せ	↓	不體裁なり	l.c.	小字なり
⁀	接續せよ	×	取替よ	Smc.	小頭字なり
⌉	右方に移せ	[	本行の初へ	?	なせか
[	左方に移せ	¶	別行	out	原稿を見よ
□	引込ませ	=	小頭字にせよ	OK	校了
⌈	上げよ	≡	大頭字とせよ	Final	校了濟

扱て校正刷が、尚不完全で再校を要する時は、校正刷面へ『要再校』

— 23 —

Figure 4: Tutorial Text describing about the possible geta variants
(広告印刷物の拵え方(Koukoku Insatubutsu No Kosiraekata), p.22-23)

References

[1] “活版術楷梯” (Kappanjutsu Kaitei), 梅渓生 (Baikeisei), 清文堂活版所(Seibundou Kappanjo), Tokyo, Japan, 1902.

[2] “広告印刷物の拵らへ方” (Koukoku Insatubutsu No Kosiraekata), 清水正巳(Simizu Masami), 十合書籍部(Sogou Shosekibu), Kyoto, Japan, 1924.

[3] “印刷の栞” (Insatsu No Shiori), 交進社印刷所 (Koushinsha Insatujo), Osaka, Japan, 1936.

[4] “Proposal to Code Proper Names in a Separate Area”, Lu Qin, JTC1/SC2/WG2/IRG N1327, 2007.

[5] “two Hanzi”, Andrew West, `http://unicode.org/pipermail/unicode/2014-March/000275.html`, 20/Mar/2014.

[6] “エイプリルフール” (Eipuriru Fûru), `http://en.wikipedia.org/wiki/April_Fools%27_Day`