Title: On the Encoding Model of Characters Known as Shinobi-Iroha

Source: Wang Yifan

Status: Individual Contribution

Action: For consideration by IRG

Date: 2018-10-19

1.   Background

The UK National Body submitted 12 characters from a book *Mansenshūkai*[1] (萬川集海) for IRG Working Set 2017 to be included in CJK Unified Ideographs (UK 2017; Fig. 1), of which 7 characters are solely supported by that book as of WS2017 v2.0. Those characters, according to the book, was being used among a school of ninja to encrypt confidential messages, and have become relatively widely known among Japanese culture via ninja-themed fictional works throughout the latter half of the 20th century, with the name of Shinobi-iroha (忍びいろは). This paper intends to provide some background

Figure 1: Characters in the UK submission (UK 2017)

---
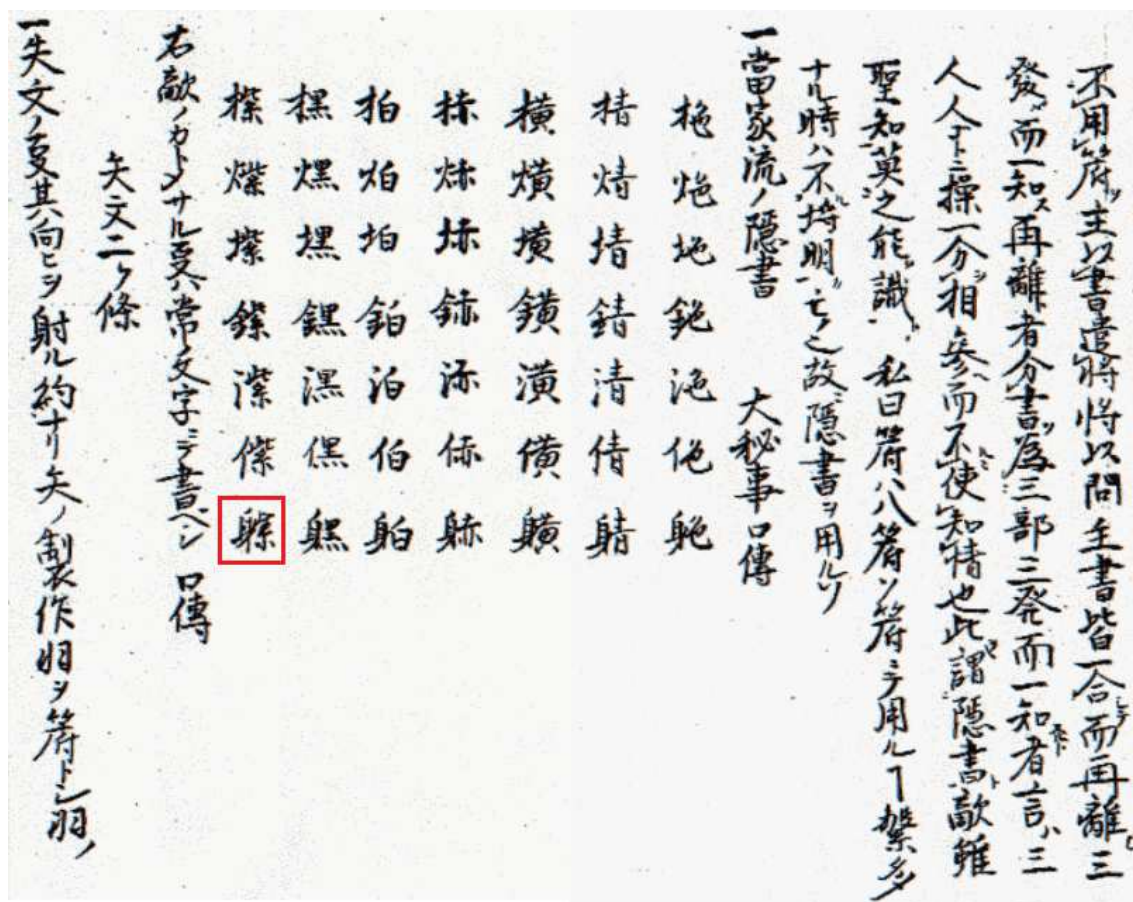
[1] Or, *Bansenshūkai*. Thought to be compiled by Fujibayashi Yasutake (藤林保武).

information regarding the characters, raise concerns about encoding Shinobi-iroha as a part of CJK Ideographs, and propose alternative encoding possibilities.

I observed another possible evidence capacity problem in original UK proposal, but it is out of this paper's scope.

2. The nature of Shinobi-iroha

2.1. Origin and history

Figure 2: Characters in Nakajima (2015)
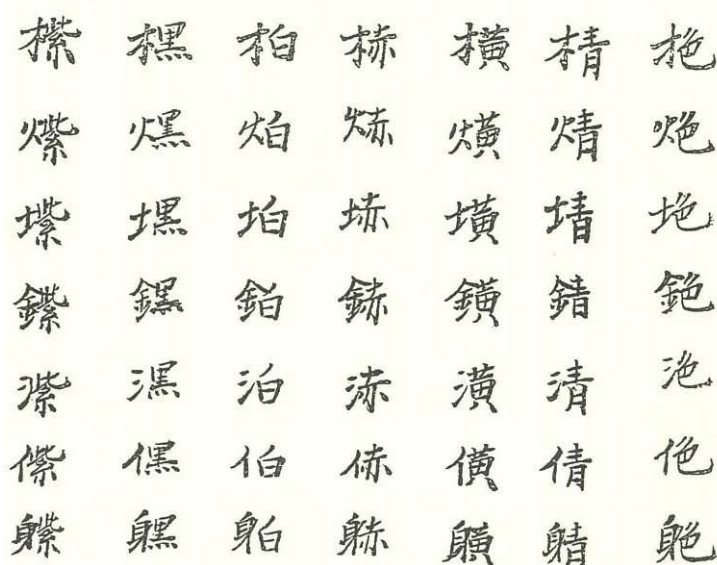
Those characters appear in vol. 5 of *Mansenshūkai*, with a caption that reads 當家流ノ隠書 "secret writing in our family's style." The 49 symbols are aligned in 7×7 grid, sorted by the component on the right, then the left side.

The reading of each character is unspecified in the original book, only noted that: 大秘事口傳 "top secret matters (will be instructed) by word of mouth." The characters are usually called Shinobi-iroha nowadays, though Nakajima (2015) states that:

> 「忍びいろは」という言葉は各種忍術伝書の中に見当たらない。後世、おそらく現代の研究者がそう呼び、検証されないまま固有名詞として使用されていると思われる。
>
> "The term 'Shinobi-iroha' is not attested in *ninjutsu* instruction books. It seems being named by newer, presumably modern researchers, and has been in currency as if a proper name without examination." [Translation by me, same hereinafter]

(Nakajima, 2015: 75)

Their historical usage outside *Mansenshūkai* is not yet known. The signs are thought to have gained wider present-day recognition after Sanpei Shirato's (白土三平) influential comic works utilizing

them.

2.2. System

Although the exact reading was not provided by the original book, Yamaguchi (1981) pointed out that those characters in the first chart should be arranged (from right to left, then top to bottom) in *iroha* ordering (Fig. 3), deriving from the traditional pangram verse of the same name, so that the arrangement of signs in the succeeding chart—which is to explain how to encode each sign into conch-blowing notes (Fig. 4)—can be understood as *gojūon* ordering, the systematic consonant-vowel sorting also serving as the standard pedagogical order in today's Japan. Given that traditional Japanese syllabary only contains 48 units, Yamaguchi assumes that the last symbol should represent some delimiting punctuation such as comma or period.

This interpretation is currently the most accepted, established view. Sanpei Shirato also employs those characters in his works as per Yamaguchi's hypothesis.
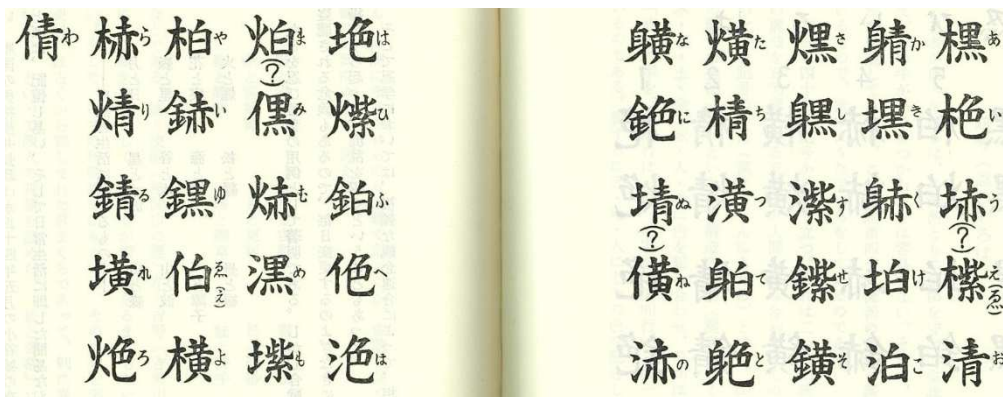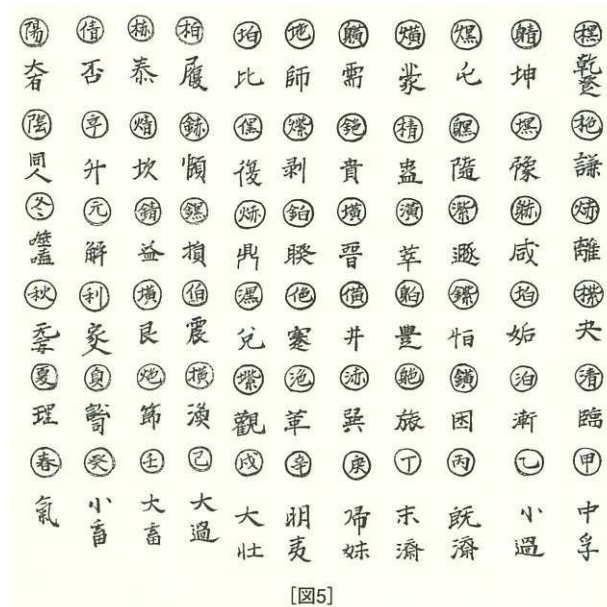
Figure 4: Decoding by Yamaguchi (1981)

Figure 3: Conch signs chart (Nakajima 2015)

Consequently, Shinobi-iroha can be validly deemed to be a syllabary (or moraic writing) as a writing system. The character inventory size which nearly comparable to that of kana also supports it. Note that, as Nakajima (2015) states, the assignment of syllables is likely to be swapped every time to keep messages secure, hence they virtually do not have fixed sound values.

2.3. Rights

The characters at least date back to 1676, thus the copyright has already been exhausted, even supposing them as Fujibayashi Yasutake's own creation. No person or entity is known so far holding trademark rights regarding those graphical shapes.

3. Encoding problems

Due to striking similarity between Shinobi-iroha and typical Han characters, it is now an important question whether they should be identified as a subset of Han, therefore, a part of CJK Ideographs in UCS implementation, if we aim systematic encoding of Shinobi-iroha. In light of this, the proposal to IRG by UK NB is a rightful decision, only that, to my knowledge, there exist concern to accept Shinobi-iroha as CJK Ideographs, because of various issues mentioned below.

3.1. Typological difference

As stated above, Shinobi-iroha is, or at least currently accepted as, a syllabary that is analogous to Kana. This characteristic is particularly problematic considering Unicode conventions, where the term "ideograph" as in CJK Ideographs actually indicates "logograph."

> *Ideograph* (or *ideogram*). […] (2) A generic term for the unit of writing of a logosyllabic writing system. In this sense, ideograph (or ideogram) is not systematically distinguished from logograph (or logogram). […]
>
> *Logograph* (or *logogram*). (1) Any symbol that primarily represents a word (or morpheme) in contrast to a sound or pronunciation. […]
>
> ("Glossary of Unicode Terms", 2017)

And this is, of course, not an abstract definition, but a practical criterion or rationale that has been applied to actual encoding process, for example, during discussion on merge of Han and Khitan scripts (cf. WG2 N3820, N3925, N3934, N3942).

> Khitan Small Script is more analogous to Katakana or Hangul than to CJK Ideographs: its shapes reflect phonetic values. As a result, the group agreed that Khitan Small Script be separated from CJK Unified Ideographs as the functions are different.
>
> (Anderson et al., 2010 [WG2 N3942])

Incorporation of Shinobi-iroha into CJK Ideographs would lead to confusion to the Unicode and UCS character encoding model, compared to that of Katakana into Han, or Cherokee into Latin.

Admittedly, as of Unicode 11.0, two genuinely Shinobi-iroha-sourced characters 𠎁 (U+20381) and

𨊂 (U+28282) are contained in the CJK Unified Ideographs Extension B block, which was proposed by Japan NB with JIS X 0213 as basis. However, the reason these characters exist is to keep the bibliographic record of Sanpei Shirato's comic *Ishimitsu* (柧𨊂偀潢), whose title is represented in Shinobi-iroha. Considering Japan has been proposed several glottographic characters of questionable identity, and actually succeeded to add some thereof[2], being expedient "kanji" for administrative use, we should not take the existence of the two characters as the immediate ground that the Shinobi-iroha system is a subset of Han.

3.2. Unification problem

Following Yamaguchi's decryption, it is now most commonly understood that the left-hand components of Shinobi-iroha are 木, 火, 土, 金, 水, 人 and 身 (in this order), roughly based on the Chinese Five Elements (五行) scheme. This, however, as also pointed out by Nakajima (2015), is not so consistent across versions of *Mansenshūkai*, as some glyphs apparently have 扌 instead of 木 (Fig. 5). Since the overall arrangement is undisputed, and only the distinction among 49 characters is

Figure 5: The dubious component in multiple versions

Figure 6: Characters in Yamaguchi (1981)

---

[2] Including 乄 (U+4E44), 𪜈 (U+2A708) and 〳 (U+2CF00).

significant in this system, the components 扌 and 木 could be interchangeable in Shinobi-iroha. Obviously, it is incompatible with normal CJK Ideographs framework, where unifying 木 ("tree") with 扌, which is a variant of 手 ("hand"), is unacceptable.

Also, if we rigidly comply to Yamaguchi's interpretation, we should also accept full-form 人 as a left-hand component instead of 亻 shape (Fig. 6; though I could not find out what version Yamaguchi had referred). This would add to confusion in CJK Ideographs' unification policy, if not unacceptable, when they are merged into the block.

3.3.   Recognition mismatch

It should not be ignored that most literature which introduce Shinobi-iroha in Japanese, printed or online as far as I have found, implicitly or explicitly hint their assumption that Shinobi-iroha and Han are distinct systems. Though the user base size of the system is unknown, it should be taken into account that many users do not conceive them as Han simply because of shape similarity.

左図の四十九の記号から成り立つ。

"It consists of 49 **symbols** shown in figure on left." [emphasis by me, same hereinafter]

……記号の一部（漢字のへんにあたる）とする。……

"... used as components of symbols (**equivalent to *pian* in Chinese characters**) ..."

……構造の一部（漢字のつくりにあたる）に配置している。……

"... put them in a part of structure (**equivalent to *pang* in Chinese characters**) ..."

……漢字の字母を造りかえて組み立てた系統的暗号である。

"... a systematic cryptograms by transforming and combining Chinese character elements."

(Yamaguchi, 1981: 118–120)

CJK IDEOGRAPHS **SIMILAR TO THE SHINOBI IROHA LETTERS**, FOUND IN THE UCS [text in a chart; see Fig. 7]

かようなわけで、この漢字づらをした「軅」「偶」の 2 字は純粋に忍びいろはを出自として UCS に忍び込んでいる文字だったりします。

"That's how the two characters, 軅 and 偶, **set up themselves as Chinese characters** and sneaked into UCS, while having pure Shinobi-iroha background."

(Umihotaru, 2011)

当然、漢字の様に見えるわけだが、漢字でないものも含まれている。

"Naturally they **look like** Chinese characters, but some are not."

("Ninja no Chie", n.d.)

「や」を表す「柏」など一部は漢字として成立しますが、「て」などは「身」＋「白」となり独自の文字となります。

"Some, such as 柏 that represents や, are valid as Chinese characters, while others such as て which is 身+白, are original characters."

("Angōhōshiki", n.d.)

Figure 7: Chart from Umihotaru (2011)

UCSに存在する、忍びいろはと同形の漢字
CJK IDEOGRAPHS SIMILAR TO THE SHINOBI IROHA LETTERS, FOUND IN THE UCS

い 桗 682C	ち 楮 68C8	よ 横 6A2A	ら 㫒 2AC92	や 柏 67CF	あ 標 23638	ゑ 橴 6A74
ろ 炮	り 㷊	た 熿 71BF	む 烺 7103	ま 炟 241E2	さ 燺 3DF5	ひ 燤
は 地	ぬ 堷 57E5	れ 墴 58B4	う 㻽 212FD	け 坰 2129A	き 墂	も 墟
に 鈀 92AB	る 錆 9306	そ 鑕 9404	ゐ 鉌 4932	ふ 鉑 9251	ゆ 鑆 28B46	せ 鑢
ほ 㳊 23D0A	を 清 6E05	つ 潢 6F62	の 㵾 6D7E	こ 泊 6CCA	め 潶 6F76	す 潒
へ 㠜 2021C	わ 倩 5029	ね 儛 50D9	お 㳻 2A73B	え 伯 4F2F	み 儁 20381	ん 僣
と 皰 28246	か 鯖	な 鱝 28287	く 鮇	て 鮊	し 鱲 28282	

❖異体字 / VARIANTS: 清 6DF8   横 6A6B

4. Suggestion/solution

From the above, I would like to propose possible ways to encode Shinobi-iroha, if they are otherwise qualified to be included in UCS.

4.1. As syllabic characters

As argued in section 3.1, Shinobi-iroha is typologically syllabic (or moraic) writing, highly compatible with Kana. Consequently, the most straightforward implementation will be encoding them as a new script with its own block in somewhere, probably SMP. The set will contain 49 characters, but the existing two in Extension B could be omitted in order to save a hexadecimal column, as there is little possibility new members join them.

4.2. As letterlike symbols

As its use for substitution cipher that results in variable linguistic correspondence, encoding them as glottographic characters may not be appropriate. Or, its visual collision with Han characters may become problem. Thus, they could also be encoded as pseudo-script symbols that somewhat analogous to those in Letterlike Symbols and Mathematical Alphanumeric Symbols blocks.

4.3. As composite Ideographs

Even with what is stated in section 3.3, the fact that they are made of Ideographic elements is usually accepted. Given this situation, they may be encoded as composite Ideographs analogous to some already in Enclosed CJK Letters and Months, CJK Compatibility, and Enclosed Ideographic Supplement blocks. Auxiliary compatibility equivalence to sequence of Ideographs may be defined too.

Alternatively, they could be defined as named sequences according to each structure (headed by IDS, if allowed). Since named sequences do not take up code points, one could add needed units on demand.

References

Anderson, Deborah, SEI and UC Barkeley (2010). "Ad hoc report on Khitan Small Script". ISO/IEC JTC1/SC2/WG2 N3942. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3942.pdf [Accessed 2018-10-17].

"Angōhōshiki" [暗号方式] (n.d.). *CypherTec Inc*. https://www.cyphertec.co.jp/techno/t_encryption.html [Accessed 2018-10-17].

"Glossary of Unicode Terms" (2017). https://www.unicode.org/glossary/ [Accessed 2018-10-17].

Nakajima, Atsumi [中島篤巳] (ed.) (2015). *Kanpon Mansenshūkai* [完本　萬川集海]. Tokyo: Kokushokankokai.

"Ninja no Chie" [忍者の知恵] (n.d.). *Kōkaryū Ninjutsu Yashiki* [甲賀流忍術屋敷]. http://www. kouka-ninjya.com/info.html [Accessed 2018-10-17].

UK (2017). "UK Submission for IRG Working Set 2017". ISO/IEC JTC1/SC2/WG2/IRG N2232. http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg49/IRGN2232.summary.pdf [Accessed 2018-10-17].

Umihotaru (2011). Untitled article. *Mugen Kajo* [無限花序] (blog). http://d.hatena.ne.jp/Umihotaru/20111216/1324033352 [Accessed 2018-10-17].

Yamaguchi, Masayuki [山口正之] (1981[1963]). *Ninja no Seikatsu* [忍者の生活]. Seikatsu-shi Sōsho [生活史叢書], 2. Tokyo: Yuzankaku.

(End of Document)