

PROPOSAL TO DISUNIFY U+5F50

Individual Contribution

For consideration by UTC

Alexander Zapryagaev

08 April 2020

rudetection@gmail.com

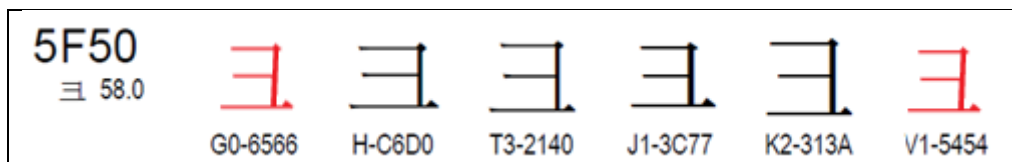
TABLE OF CONTENTS

The Proposal.....	2
The Characters	3
Local Data.....	5
Can't They Stay Unified?.....	9
Possible Solution	10
Appendix: IDS.....	12

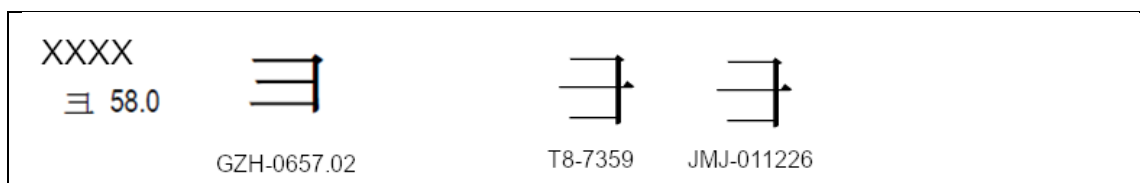
THE PROPOSAL

This document proposes a series of changes that could serve as a solution to the current situation when three highly graphically similar characters, mostly used as components of other characters were, I believe, mistakenly unified in Unicode, which now leads to unwanted consequences. The following changes are introduced.

1. Alter the glyphic representation of the character G0-6566 in the GB 2312-1980 standard so it conforms to the 𠄎 style; do the same with V1-5454 (unless the Vietnamese standard already shows 𠄎, in which case only the representative glyph in the code chart requires replacement).



2. Encode, appending to the URO, a new character with the representative glyphs following the 𠄎 or 𠄏 standards.



(The references offered here are: Zhōnghuá Zìhǎi for simplified “Snow” and the Taiwan and Moji-Joho references mentioned later in the document.)

3. Alter the presentation of the relevant radicals in the following way:

2F39		KANGXI RADICAL SNOUT ≈ 5F50 𠄎
2E94		CJK RADICAL SNOUT ONE → 5F51
2E95		CJK RADICAL SNOUT TWO <input type="checkbox"/> Despite the name, is actually a version of 2F1C KANGXI RADICAL AGAIN → XXXX 𠄎

4. Introduce the following new IVS in the Adobe-Japan1 collection:

𠄎 U+XXXX U+E0100 Adobe-Japan1 CID+15391

and establish that U+5F50 U+E0101 is deprecated and the one above is recommended instead.

5. Introduce the following new IVS in the Moji-Joho collection:

𠄎 U+XXXX U+E0101 Moji_Joho MJ011226

𠄏 U+XXXX U+E0102 Moji_Joho MJ011227

and establish that U+5F50 U+E0103 and U+5F50 U+E0104 are deprecated and the ones above are recommended instead.

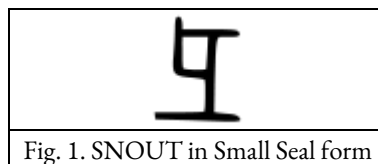
Repeat the steps above for the Hanyo-Denshi collection.

6. (Optional) Redefine the IDS according to new mappings, as delineated in Appendix.

This document is the response to the feedback ([Feedback 1](#), [Feedback 2](#)) to [IRGN 2414 Proposal to De-Unify One Obsolete Simplified Chinese Character](#), where a different solution was offered which, upon discussion, cannot be maintained anymore.

THE CHARACTERS

U+5F50 𠄎 (J-Source form shown) opens the segment of the URO dedicated to the Kāngxī radical 58 (U+2F39 𠄎 KANGXI RADICAL SNOUT according to the naming system of the code chart). It is followed by U+5F51 𠄏, providing the alternate form of the radical-character as given by U+2E94 𠄏 CJK RADICAL SNOUT ONE in the “CJK Radicals Supplement” block. Both represent the graphical variations of the character with the Early Middle Chinese (EMC) encoding k-jeH (Baxter), defined by *Shuōwén* as 豕之頭 “head of a boar”, which is probably what the original form of the character depicts (Fig. 1):



U+5F50 is thus a well-known and universally spread alternate graphic representation of the same character. In fact, in modern simplified and shinjitai forms the occurrence of 𠄎 is much more probable than that of the unchanged elements; see, for example, U+9332 録, which was simplified in Japan from U+9304 錄, and its Chinese version 录 (in 录, which originally denoted “to trim/carve wood,” the whole grapheme probably depicts a decapitated tree, but the top was early reinterpreted as the “decapitated” head of a “pig”¹).

The characteristic lower line of 𠄏, frequently extended in length, is retained in 𠄎 as well, according to its usage. Currently, the word 𠄏 is not used to describe a pig’s snout anymore, with general words like 口鼻部 *kǒubíbù* taking

¹ <https://wenlin.co/wow/Zi:%E5%BD%95>.

the job, but there is at least one situation where the character is still relevant: that is, speaking about the radical itself (𠂇部 *jìbù* “Radical 58”). We will refer to 𠂇 as SNOUT.

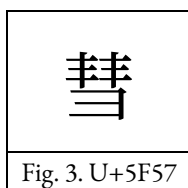
However, this character possesses the extreme graphical similarity to another one, also popular as a component: I refer to 𠂇. As the graphical form shows, the difference lies in the manner of the contact between the right vertical and lower horizontal strokes (here the vertical slightly protrudes² and, more importantly, the horizontal stops at the vertical). This character is one, the most moderate, of the descendants of the glyphs for “right hand” (Fig. 2), and we will apply the name HAND to it:



This character directly represents a right hand, now normatively 又, which is the descendant of the same glyph.

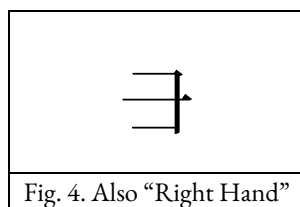
Later, the basic glyph suffered a phonetic loan, now standing for the identically read adverb: 又 *yòu* “again;” this required to modify the glyph meaning “right,” now 右 *yòu*, with a 口 “mouth” appended.

Still, many characters containing “right hand” as a component show an explicitly 𠂇-like form, such as (Fig. 3):



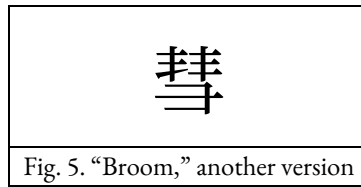
This character denotes a broom, with the top side being a drawing of the twigs of the broom, while the bottom is the hand carrying it.

Perhaps, the more recognizable form of this HAND is, however, one with the horizontal line in the middle protruding beyond the vertical, like in the *Kāngxī Dictionary* (Fig. 4):



We will denote it as LONG HAND and, once again, demonstrate its usage on the broom (Fig. 5):

² Or exactly touches the end of the horizontal, as might be rendered in several fonts and is, at least according to the representative glyphs, the normal Vietnamese rendering.



Thus, we consider three elements:

SNOUT	𠃉
HAND	𠃊
LONG HAND	𠃋
Table 1. The characters	

The situation becomes complex as no character beside one is encoded in Unicode, and which one is actually depends on the locale.

LOCAL DATA

PRC

The source of the complexity is probably the decision of the PRC to unify the graphical form of these components and choose 𠃊 as the joint representation of them. To quote the PRC feedback of 2019.10.23 to IRGN2414:

‘The only reason why 𠃊𠃊𠃊 is are always mix used mistakenly in ancient times, and Ministry of Culture and Reform and arrangement Committee of Chinese characters of China released a document called 《印刷通用汉字字形表》 in 1965 to solve this chaotic situation in general use. This document use 𠃊 instead of 𠃊𠃊𠃊<...>, so that the dictionaries and GB2312-1980 followed that decision.’

Thus, the 1980 establishment of GB 2312-1980 produced the following aberrant representative glyph G0-6566:

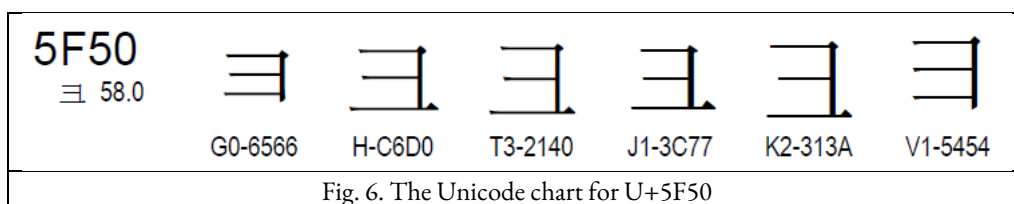
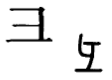




Fig. 6. The Unicode chart for U+5F50

As we observe, even the Vietnamese form was influenced by the PRC decision. Note, however, that the dictionaries of the PRC are not consistent with implementing the decision (acknowledged by the PRC response as well), and the most authoritative editions in the matter of graphic correctness separate them (see Fig. 7–8).

 <p>彑 說文·彑部</p> <p>《說文》：“彑，豕之頭，象其銳而上見也。讀若屬。”</p> <p>jì 《廣韻》居例切，去祭見。月部。</p> <p>①猪頭。《廣韻·祭韻》：“彑，《說文》作彑，云：‘豕之頭。’”</p> <p>②獾屬。《玉篇·彑部》：“彑，彑類也。”按：《爾雅·釋獸》“彑”宋邢昺疏：“彑即獾也。”</p>	 <p>录 粹一三七六 𣎵 宰崗麓 彑 厂叔多父盤</p> <p>彑 麓伯麓 彑 說文·录部 彑 吾作鏡</p> <p>《說文》：“录（旧字形作‘𣎵’），刻木录录也。象形。”按：甲骨文、金文像井上轆轤打水之形，当为轆轤之“轆”的初文。</p> <p>lù 《廣韻》盧谷切，入屋來。屋部。</p> <p>①[录录]一一可数貌。《說文·录部》：“录，刻木录录也。”徐鍇繫傳：“录录，猶歷歷也，一一可數之兒。”</p> <p>②本。《廣韻·屋韻》：“录，本也。”</p> <p>③刻木。《廣韻·屋韻》：“录，刻木也。”</p> <p>④姓。《集韻·屋韻》：“录，姓。”</p> <p>⑤今为“録”的简化字。</p>
<p>Fig. 7. 漢語大字典 (2 ed.) on SNOOT, separate and component</p>	

 <p>彗 說文古文</p> <p>彗 說文·又部</p> <p>彗 說文或體</p> <p>彗 天文雜占二·六</p>	<p>《說文》：“彗，掃竹也。从又持彗。彗，彗或从竹。箒，古文彗，从竹从習。”徐灝注箋：“彗蓋象竹彗之形，非彗字。猶鳥足从匕而非匕，魚尾似火而非火也。”</p> <p>huì (旧读 suì) 《廣韻》于歲切，去祭云。又徐醉切，祥歲切。月部。</p> <p>①扫帚。《說文·又部》：“彗，掃竹也。”《史記·孟子荀卿列傳》：“如燕，昭王擁彗先驅，請列弟子之座而受業。”</p>
<p>Fig. 8. 漢語大字典 (2 ed.) on HAND as component</p>	

The LONG HAND is not used in the PRC standard, as HAND 彑 is treated as the graphic norm; thus, any character containing a HAND will be rendered with 彑; for example, ‘snow’ 雪, possessing a simplified 彗 in the bottom part, is not rendered with 彑.

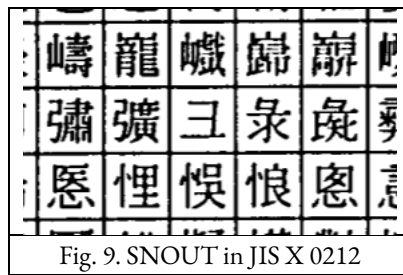
Additionally, I should mention that not even all fonts aimed at PRC usage follow the “U+5F50 = HAND” principle; while this seems to be generally true for Song (sans-serif) fonts as well as heiti (sans-serif), fonts derived from the regular script consistently show SNOOT, though perfectly able to produce HAND as component (see FangSong in Table 2):

Font name	5F50	'Snow'
FangSong	𠄎	雪
KaiTi	𠄎	雪

Table 2. Kaiti fonts shipped with OS Windows

Japan

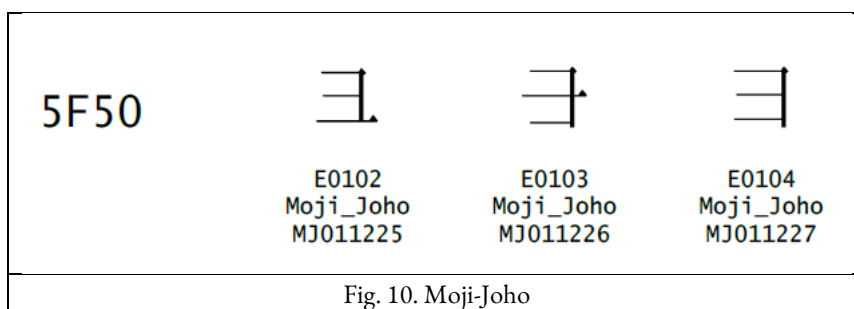
Japanese standard JIS X 0212-1990 (J1), to which the code chart refers, includes a version of the characters under consideration. Under 28-87, it incorporates the character U+5F50 (Fig. 9):



The characters surrounding it make it clear that it is SNOUT that is meant. The combination U+5F50 U+E0100 as well as just U+5F50 in both the 1990 and 2004 versions of the mappings (as well as U+2F39) are mapped to this glyph.

But this is far from the end of the story, as both of the remaining forms occur in Japanese names, and thus an *ad hoc* solution was required. Adobe-Japan1 incorporated U+5F50, U+E0101 as the denotation of U+2E95, depicted there as 𠄎, LONG HAND. But this is a conservative solution.

Moji-Joho, on the other hand, uses all three (Fig. 10):



Obviously, as the characters are non-cognate, such a solution is suboptimal, but it allows to encoded all the necessary forms.

Finally, we must observe that, unlike PRC, in Japan the difference between the forms of HAND was considered not a question of style but a method of separation between the frequent and rare characters (approximately Jōyō vs. non- Jōyō), and thus both occur: for comparison, see the 'broom' 𠄎 [Jinmeiyō] vs. 'snow' 雪 [Jōyō].

Hong Kong, Taiwan, Macau

All the sources from these locations provided in the code chart do refer to SNOOUT.

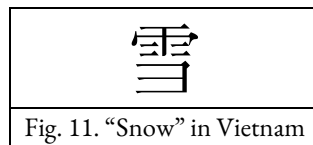
Taiwan, however, maintains HAND as well, which is stored at Plane 8, 7359 of CNS-11643 (<https://www.cns11643.gov.tw/wordView.jsp?ID=553817>); the Unicode mapping it is forced to use is U+2E95 in the “CJK Radicals Supplement” block.

Korea

Korea maintains only the 冫 glyph, as K2 [KS X 1027-1:2011 (formerly PKS C 5700-1 1994)], character 313A.

Vietnam

As alluded before, HAND as component takes in Vietnamese glyphs a unique form with the horizontal and the vertical joining at the same location; the Nom Na Tong form of “Snow” will be sufficient as an example (Fig. 11):



Yet, their representative glyph of U+5F50 is a copy of the Chinese one; as the definitions and readings in Nom Foundation dictionary show, this is wrong: the intended use of U+5F50 in Vietnamese refers to SNOOUT, not HAND (Fig. 12):

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
kệ	冫	U+5f50	mặc kệ, thầy kệ	vhn	not to pay attention to
kẹ	冫	U+5f50	lúa kẹ	btcn	snout; KangXi radical 58
kí	冫	U+5f50	bộ kí	gdhn	snout; KangXi radical 58

Fig. 12. What is U+5F50 in Vietnamese

The source given in the code chart is V1 [TCVN 6056:1995], character 5454. The actual glyph should be checked to establish the intention, but the fact that it appears in the substring

彈疆冫形彩

is very much suggestive of SNOUT.

CAN'T THEY STAY UNIFIED?

No, they cannot, I argue, and this is why. Though it worked out before, consistency required the disunifications, for at least two reasons (beside the fact that they are obviously non-cognate): one conceptual and one practical.

Para-Source Separation

Though there is no direct opposition in a single source here, one may argue there is: it is in Unicode itself. U+5F50 is considered the Unified version of the radical 2F39 (Fig. 13):

2F39	𠄎	KANGXI RADICAL SNOUT
		≈ 5F50 𠄎
2E94	𠄎	CJK RADICAL SNOUT ONE
		→ 5F51 𠄎
2E95	𠄎	CJK RADICAL SNOUT TWO
		→ 5F50 𠄎

Fig. 13. Code charts for radicals

However, the unfortunately named SNOUT TWO (which should have better been CJK RADICAL AGAIN, referring to U+2F1C 𠄎, the main allomorph of HAND in modern Sinitic writing) is also matched to U+5F50, without a better place to be sent (not to U+53C8 𠄎, surely?). Furthermore, the Taiwan evidence that the encoded T3-2140 SNOUT and T8-7359 HAND/LONG HAND are separate also points to the fact that the Source Separation Rule *should* have been invoked during the initial composition of the URO.

Second Stage

A more eminent and practical reason, however, lies in the overload of characters. There is no much harm as long as these characters refer only to full forms of components. But the addition of Extension G changed that.

The problem comes from the Second Stage Simplifications. These Second Stage Simplifications (第二次 汉字简化方案—草案, SSS) were an abortive project of PRC government in late 1970s. They were supposed to become a continuation of the highly successful campaign of simplifications implemented throughout the 1950s-60s. Part One was released on 20th December 1977 and was consistently used in all the publications in 人民日報 *Rénmín Ribào* until July 1978. It was widespread during that period and gained mass currency, but, announced a failure, soon practically dropped out of usage, though an official withdrawal (though not declaration of abandoning any simplification plans) was postponed until 24th June 1986. Many of the forms proposed in the reform were in fact popular forms of characters, that could be found in informal use ever before. Some of the characters introduced in this part still remain in everyday use, such as in private letters and in signage, despite the cancellation. Part Two was

published simultaneously with Part One but declared not for immediate use but rather for implementation in case the first one succeeds. Currently, unlike the first part, these characters are not recognized by the majority of the literate Chinese.

The characters of Part One (except those produced by analogy, that have been separately offered to WS-2020) are now encoded in Unicode. However, they include the following (Fig. 14):

1.1.136	𠂇 ¹¹	𠂇	𠂇	96EA [TGH 1]
Fig. 14. The new “Snow”				

Basically, for “Snow” the top is cut, retaining just HAND. And, just like that, HAND gets a new semantic meaning. Now U+5F50 is overloaded, playing the roles of two different morphemes of Mandarin: *jì* ‘pig’s head, name of a radical’ – and *xuě* ‘snow.’ And while their graphical identity was probably desirable by the compilers of the actual simplification scheme in the 70s – after all, they tried to make 𠂇 and 𠂇 graphically identical all this time – but not to the current use, when any font maker deciding to support Second Stage Simplifications but render SNOUT and any kind of HAND differently is forced to the complex decision of keeping them the same, though they might co-occur, even contrastively, in the same sentence.

Just imagine the sentence “the character for ‘snow’ is not identical to the radical SNOUT,” but written in the Second Stage.

The arguments proposed above, I hope, make it sufficiently clear that U+5F50 should be split in two, with the forms of HAND being separate (though there is no sense to separate 𠂇 and 𠂇, not even distinguished by Taiwan).

POSSIBLE SOLUTION

However, the solution offered by the PRC response is utterly untenable. It offers the following:

‘...China decides to move the G glyph and source reference out of the code point and put it in the next working set, and submit a new one for U+5F50 to match the common view.’

No change should ever alter the correspondences of the standard as basic as GB 2312-1980, frequently used on simpler devices that will never be able to update to the newer versions of Unicode; furthermore, if a new one would be submitted, it will likely be encoded outside the BMP, which is highly undesirable.

In fact, the following is clear: the location in the URO, the matching with the standards (especially Japanese JIS 0212) already implied in the table, even the graphical forms of the radicals as now presented in Unicode 13.0 lead to the conclusion that the codepoint U+5F50 must be the Unified correspondence of the radical SNOUT. And this means only one solution.

I propose to alter the graphical form of G0-6566 itself. If the representative glyph in the actual standard GB 2312-1980 is made to represent SNOUT and not HAND, the relationships between various standards and intentions become clearer. It is HAND (preferable in the LONG HAND form which is orthodox in the Kāngxī Dictionary) that needs to be re-encoded, preferably appended to the URO as this would (if the corresponding, though optional, rewriting of the IDS is undertaken) fix the contradiction that a TIP character can be used in the IDS of BMP characters. The Second Stage simplified “Snow” can be just unified to this new character.

APPENDIX: IDS

Unicode #	Glyph	Contains	Comments
U+520D	𠂇	HAND	G-only form; thus can be considered containing HAND by default
U+591B	𠂈	SNOUT	Top
U+5987	𠂉	HAND	As in 5E1A
U+5BFB	𠂊	HAND	As in 5C0B
U+5C0B	尋	HAND	
U+5E1A	帚	HAND	
U+5F50	𠂋	SNOUT	
U+5F52	𠂌	SNOUT	Under radical SNOUT
U+5F53	当	SNOUT	
U+5F55	录	SNOUT	
U+5F57	𠂍	HAND	Broom
U+5F5A	彙	SNOUT	
U+5F5B	彙	SNOUT	Also 彙 where this is obvious
U+5F5E	彙	SNOUT	
U+626B	扫	HAND	As in 5E1A
U+698B	𠂎	HAND	𠂎 without left sides
U+7075	灵	HAND	
U+7E01	縁	SNOUT	
U+96EA	雪	HAND	As in 5F57

U+3777	尋	HAND	As in 5C0B
U+37AA	𪗇	SNOUT	
U+FA7A	喙	SNOUT	Variant of 5599 喙
U+2019F	𪗈	SNOUT	
U+20295	𪗉	HAND	See HDZ
U+202B8	𪗊	HAND	As in 20295
U+20767	𪗋	HAND	
U+20A14	𪗌	HAND	
U+20B4A	𪗍	HAND	
U+20C1F	𪗎	HAND	
U+20F7D	𪗏	HAND	
U+213F7	𪗐	HAND	
U+2163C	𪗑	SNOUT	
U+21758	𪗒	HAND	
U+21A5A	𪗓	SNOUT	
U+21A5D	𪗔	HAND	
U+21AD4	𪗕	HAND	
U+21AE7	𪗖	HAND	
U+21B36	尋	HAND	
U+21C29	𪗗	SNOUT	
U+22125	𪗘	SNOUT	
U+222EA	𪗙	HAND	
U+2239D	𪗚	SNOUT	
U+2244F	𪗛	HAND	
U+22451	𪗜	HAND	
U+22452	𪗝	HAND	

U+22454	𠄎	SNOUT	
U+22455	𠄏	SNOUT	
U+22456	𠄐	HAND	
U+22458	𠄒	SNOUT	
U+22459	𠄓	SNOUT	
U+2245C	𠄔	SNOUT	
U+2245F	𠄗	HAND	
U+22462	𠄙	SNOUT	
U+22464	𠄚	HAND	
U+22468	𠄜	SNOUT	
U+2246A	𠄞	SNOUT	
U+2246F	𠄟	HAND	
U+22472	𠄢	HAND	
U+22473	𠄣	SNOUT	
U+22474	𠄤	HAND	
U+22478	𠄨	SNOUT	
U+2247F	𠄩	HAND	
U+22493	𠄭	SNOUT	
U+2268C	𠄰	HAND	
U+229B3	𠄳	HAND	
U+22B48	𠄷	HAND	
U+22BAE	𠄹	HAND	
U+22ECD	𠄻	HAND	
U+22F3F	𠄿	HAND	
U+22F64	𠄾	HAND	
U+231B8	𠄿	HAND	
U+231D7	𠄿	HAND	

U+231F6	習	HAND	
U+23414	椴	HAND	
U+23A8F	設	HAND	
U+23AA3	毅	SNOUT	
U+23D4C	淶	HAND	
U+23F19	冪	HAND	
U+2450C	晉	HAND	
U+24C5A	畱	HAND	
U+256C0	稜	HAND	
U+257BE	稞	HAND	
U+25B7C	策	HAND	
U+25BD0	策	HAND	
U+26133	繇	SNOUT	
U+261DA	繇	SNOUT	
U+2642E	習	HAND	
U+2642F	閔	HAND	
U+26431	翯	HAND	
U+26518	耨	HAND	
U+2673A	腴	HAND	
U+2678B	翯	HAND	
U+26951	臼	HAND	
U+26B91	莒	SNOUT	
U+26C97	菝	HAND	
U+26C99	菬	HAND	
U+27122	藻	HAND	
U+284F6	遯	SNOUT	
U+28521	遂	SNOUT	

U+28604	遶	HAND	
U+287CC	鄴	HAND	
U+28D0C	閭	HAND	
U+28D9F	闔	HAND	
U+291A5	霰	HAND	
U+29E8E	鱒	HAND	
U+29FEE	鴉	HAND	
U+2A4E5	鼃	HAND	
U+2A82C	叟	HAND	
U+2AAC6	汨	HAND	
U+2B29E	虐	HAND	
U+2B846	昌	HAND	
U+2B930	與	HAND	
U+2BE44	婦	HAND	
U+2CBCD	隙	HAND	
U+2CCE8	駟	HAND	
U+2CEB9	冂	HAND	
U+2CF99	帛	HAND	
U+2D055	與	HAND	
U+2D184	虎	HAND	
U+2D1A2	錘	HAND	
U+2D4D5	學	HAND	
U+2D4D7	學	HAND	
U+2D4D8	學	HAND	
U+2D612	帛	HAND	
U+2D6CD	昂	HAND	
U+2D6CE	泥	HAND	

U+2D6D0	羣	HAND	
U+2D6D2	羣	HAND	
U+2D723	急	HAND	
U+2D733	急	HAND	
U+2D8EB	攷	HAND	
U+2DB1F	藁	HAND	
U+2DB9A	𪗇	HAND	
U+2DD48	魚	HAND	
U+2E2B8	習	HAND	
U+2E2C0	𪗇	HAND	
U+2E638	課	HAND	
U+2F838	𪗇	SNOUT	
U+2F847	喙	SNOUT	
U+2F858	墜	SNOUT	
U+2F866	婦	HAND	
U+2F874	当	HAND	
U+2F8BC	掃	HAND	
U+2F8DD	東	HAND	
U+2F904	浸	HAND	
U+2F95B	穩	HAND	
U+2F962	篆	HAND	