

Source:	CheonHyeong Sim (沈天珩, aka TianHeng Shen)
Title:	Feedback on IRGN2717, N2731, N2738 and N2742
Status:	Individual Contribution on IRG #63
Action:	To be considered by IRG

## Table of contents

### [0. Background](#)

#### [1. Personal Overall Opinion](#)

#### [2. Some Tricky Issues](#)

#### [3. Unification Issues](#)

##### [3.1. Han–Hangul Ideographs](#)

##### [3.2. Han–Katakana Ideographs](#)

##### [3.3. Han–Latin Ideographs](#)

##### [3.4. Han–Hiragana Ideographs](#)

##### [3.5. Other Possible Cases](#)

#### [4. Conditions for Determining Script–hybrid Ideographs](#)

#### [5. Ideographs without Han Components](#)

#### [6. Conclusion](#)

## Background

For how to encode the script–hybrid ideographs, there were already lots of discussions during the CJK/Unihan Working Group meeting and the Script Encoding Working Group meeting:

对于应该如何编码混合文种表意文字，在CJK/Unihan工作小组和文种编码工作小组的会议中已有许多讨论：

スクリプトハイブリッド表意文字をいかに符号化するについて、すでにCJK/Unihanワーキンググループとスクリプトエンコーディングワーキンググループの会議でたくさんの議論がある：

Doc No.	Author	Title
<a href="#">L2/24-125</a>	Gen Kojitani	Proposal to add a new CJK Abbreviations block
<a href="#">L2/24-126</a>	C. H. Sim	Comments on L2/24-125
<a href="#">L2/24-135</a>	Gen Kojitani	Response to L2/24-126
<a href="#">L2/24-201</a>	Gen Kojitani	Proposal to add a new Script–Hybrid CJK Ideographs block

Also some discussions at <https://github.com/unicode-org/sah/issues/516> (but the SEWG repo is private, only few people in the IRG can get access to it).

It was brought to the IRG by Dr. Ken Lunde (IRGN2717), with three more inputs later respectively by China NB (IRGN2731), Ming Fan (IRGN2738) and Wang Xieyang (IRGN2742). And after some private communications with Kushim Jiang on this topic, I decided to create a written document to elaborate my opinions detailedly.

小林剑博士把这个议题带到了IRG (IRGN2717), 随后中国、范铭和王谢杨又各提交了一份儿反馈意见 (分别为IRGN2731、IRGN2738、IRGN2742)。在与姜兆勤就这个话题进行了一些私下交流后, 我决定写一份书面文件来详细阐述我的观点。

小林劍博士はこの議題をIRGに持ち込み (IRGN2717)、その後、中国・范銘・王謝楊は一つずつのフィードバックを提出した (それぞれIRGN2731・IRGN2738・IRGN2742)。姜兆勤とこの話題について私的な交流を行った後、書面で自分の意見を詳しく述べることにした。

What is more, in WS2024, UTC submits some script-hybrid ideographs (all of them are Han-Katakana ones), originally from [L2/23-139](#) authored by Gen Kojitani. They are:

此外, 在WS2024中, UTC提交了一些混合文种表意文字 (均为“汉字-片假名”字符), 源自小路谷源的提案[L2/23-139](#)。它们是:

さらに、WS2024では、UTCが小路谷源の提案[L2/23-139](#)からの数個のスク립トハイブリッド表意文字 (すべて「漢字-片假名」文字) を提出した。次のようになる:

[WS2024-01144](#); UTC-03390; 𠄎尸ソ

[WS2024-01351](#); UTC-03385; 𠄎广マ

[WS2024-01356](#); UTC-03391; 𠄎广𠄎タカ

[WS2024-03257](#); UTC-03392; 𠄎\*\*𠄎月ト

## Personal Overall Opinion

In general, I neither fully support nor completely oppose to encode such characters as CJKUIs. My opinion is that, the Han-Katakana ones and the Han-Hangul ones are suitable to be treated as CJKUIs, meanwhile, the Han-Latin ones and the Han-Hiragana ones are not.

总体而言, 我既不完全支持也不完全反对将此类文字编码为中日韩统一表意文字。我的观点是, “汉字-片假名”字符与“汉字-谚文”字符适合被视为中日韩统一表意文字, 而“汉字-拉丁字母”字符与“汉字-平假名”字符不适合。

全体的に言えば、このような文字を日中韓統合表意文字に符号化することを完全に支持するわけではないし、完全に反対するわけでもない。私の考えでは、「漢字-片假名」文字と「漢字-ハングル」文字は日中韓統合表意文字と見なされるのに適しながら、「漢字-アルファベット」文字と「漢字-平假名」文字は適しない。

The strokes of Katakana and Hangul are very similar to Han, they could blend together without any acosmia feelings. What is more, in Unicode, we have already encoded so

many [Han-Katakana ideographs](#) and [Han-Hangul ideographs](#) as CJKUIs (links point to UTN#43, kStrange).

片假名和谚文的笔画与汉字非常相似，它们混合在一起没有违和感。此外，在Unicode中，我们已经以中日韩统一表意文字编码了许多 [“汉字-片假名”字符](#)与 [“汉字-谚文”字符](#)（链接指向UTN#43, kStrange）。

片假名とハングルの筆画は漢字によく似ており、それらが混ざっていても違和感がない。また、Unicodeには、すでに多くの [「漢字-片假名」文字](#)と [「漢字-ハングル」文字](#)は日中韓統合表意文字として符号化されている（リンクはUTN#43、kStrangeを指している）。

## Some Tricky Issues

Basically I do not support to encode the Han-Latin ideographs and Han-Hiragana ideographs as CJKUIs because Latin and Hiragana look too different to Han ideographs. However, we still have such examples already encoded or about to be encoded as CJKUIs in Unicode.

大体上我不支持将“汉字-拉丁字母”字符和“汉字-平假名”字符编码为中日韩统一表意文字是因为拉丁字母和平假名与汉字看起来差别太大。但是，我们仍然存在已经编码或即将编码为中日韩统一表意文字的这类字符。

基本的には「漢字-アルファベット」文字と「漢字-平假名」文字を日中韓統合表意文字として符号化することを支持しないのは、アルファベットと平假名は漢字と見た目の違いが大きすぎるからである。しかし、そのような文字を日中韓統合表意文字としてもう符号化されているの、またはすぐに符号化されるのも存在する。

The 1<sup>st</sup> example is 坊 (U+2D92A, CJK-ExtF). Its left part is a Hentaigana component ゑ (U+1B0A6) derived from 者, while its right part is a Han component 方. Currently this is the only Han-Hiragana ideograph in Unicode.

第一个例子是“坊”（U+2D92A，中日韩统一表意文字扩展区F）。它的左侧是一个由“者”字变化而来的变体假名“ゑ”（U+1B0A6），而它的右侧是一个汉字部件“方”。就目前而言这是Unicode中唯一的一个“汉字-平假名”字符。

一番目の例は「坊」（U+2D92A、日中韓統合表意文字拡張F）である。その左側は「者」からなった変体仮名の「ゑ」（U+1B0A6）でありながら、右側は漢字の部品の「方」である。今のところでは、これはUnicodeでの唯一の「漢字-平假名」文字である。

2D92A  
方 70.2  
坊  
JMJ-057632

The 2<sup>nd</sup> example will be 炮 (tentatively U+323BF, CJK-ExtJ). Its left part is a Latin letter “X”, while its right part is a Han component 也. Currently this is about to be the

only Han–Latin ideograph in Unicode. No matter how the glyph looks like, semantically the left part should be a Latin letter.

第二个例子将是“地”（暂为U+323BF，中日韩统一表意文字扩展区J）。它的左侧是一个拉丁字母X，而它的右侧是一个汉字部件“也”。就目前而言这即将成为Unicode中唯一的一个“汉字–拉丁字母”字符。无论字形长什么模样，从字理上讲它的左侧就应是一个拉丁字母。

二番目の例は「地」（一時的にU+323BF、日中韓統合表意文字拡張J）である。その左側はアルファベットのXでありながら、右側は漢字の部品の「也」である。今のところでは、これはすぐにUnicodeでの唯一の「漢字–アルファベット」文字になる。グリフがどんな形であれ、意味的には左側がアルファベットであるべきである。



<https://hc.jsecs.org/irg/ws2021/app/?id=00029>

However this leads to another question that, what on earth is the X–like component in the Chinese simplified characters like 赵 and 风? Since they are all just symbols without any meanings, could we reconstruct the cognition, and force to treat the letter “X” in 地 equivalent to the X–like component in 赵 or 风? Where is the boundary between Han symbols and a letter from a different script if they look similar?

然而这又引申出了另一个问题，简化字“赵”和“风”等字中的“X”部件到底是什么？既然它们都只是没有意义的符号，我们是否可以理据重构，强行认为“地”中的X等同于“赵”或“风”中的“X”？当外形相似时，“汉字中的符号”和“另一个文种中的某个字母”的界限在哪里？

しかし、これはまた別の問題に導き、中国語簡体字の「赵」や「风」などの字の「X」という部品はいったい何であろうか。いずれも意味のない記号である以上、「地」のXの語源を強いて「赵」または「风」の「X」に等しくように再構築することができるであろうか。形が似ている場合、「漢字の記号」と「別のスクリプトの文字」の境界線はどこであろうか。

## Unification Issues

If we treat any one kind of script–hybrid ideographs as CJKUIs, then we have to face the unification issues. Here I would like to discuss each kind of script–hybrid ideographs separately.

当我们把任何一类混合文种表意文字视为中日韩统一表意文字，我们就不得不面对认同问题。这里将分别讨论各类混合文种表意文字。

スクリプトハイブリッド表意文字のいずれかの種類を日中韓統合表意文字と見なすと、統合の問題に直面しなければならない。ここでは各種類のスクリプトハイブリッド表意文字を別々に議論する。

## Han–Hangul Ideographs

Let us first see an example in CJK–ExtA.

我们先来看一组中日韩统一表意文字扩展区A中的例子。  
 まず日中韓統合表意文字拡張Aでの例を見てみよう。



These two characters looks almost the same, however, actually they are completely different.

这两个字符看起来几乎一样，但实际上它们完全不同。  
 この二つの文字はほとんど同じように見えるが、実際には全く違う。

<a href="#">U+3588</a>	𪚩	GDM-00510	信息技术信息交换用字符集 第八辅助集 (SJ/T 11239-2001)
------------------------	---	-----------	--------------------------------------

Fig.1 [IRGN2729\\_G–Source horizontal extensions for 2,145 characters](#)

39区	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
01-19	鸚	鵑	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻	鶻
20-39	鶻	鶻	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲	痲
40-59	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞	窞
60-79	窞	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤	衤
80-94	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆	耆

Fig.2 SJ/T 11239–2001, aka 八辅

𪚩	wo	A950
𪚩	wo	A951
𪚩	guo	A952

Fig.3 《地名库外字代码对照表》，thanks to 王谢杨 for providing the evidence

The one from 八辅 is read as “guo”, which seems to be an error form derived from 耆 (gǒu/こう), so semantically its lower part should be a Han component 口.

八輔中的这一个读作guo，似乎是一个从“耇”（gǒu）变来的讹字，因此字理上它的下半部分应当是汉字部件“口”。

八輔からのはguoと読み、「耇」（こう）からなった誤字のようであるので、意味的にはその下半分が漢字の部品の「口」であるべきである。

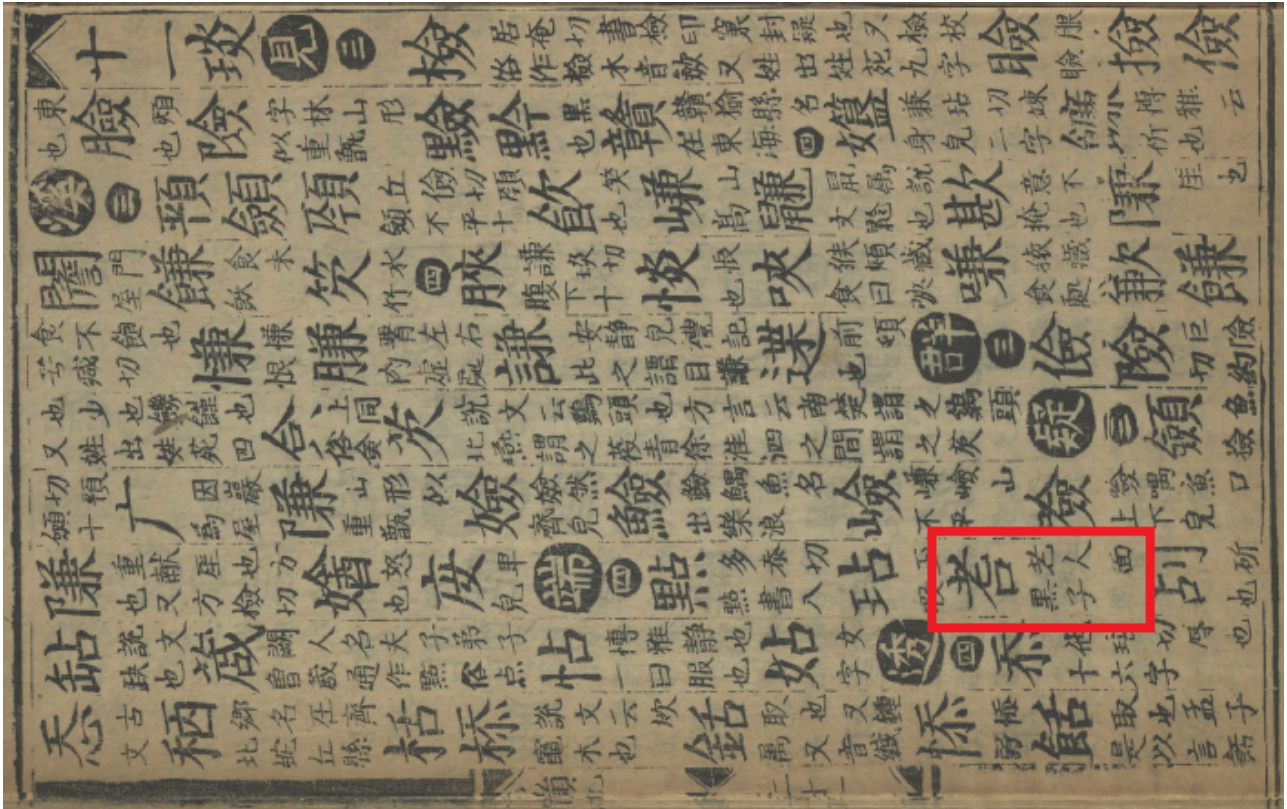


Fig.4 《五音集韻》

The one from 五音集韻 is read as 端咸鹽三開 (diǎn/てん), which is no doubt an error form derived from 耇, so semantically its lower part should also be a Han component 口.

五音集韻中の这一个读作“端咸盐三开”（diǎn），无疑是从“耇”变来的讹字，因此字理上它的下半部分也应当是汉字部件“口”。

五音集韻からのは「端咸塩三開」（てん）と読み、「耇」からなった誤字に違いないので、意味的にはその下半分も漢字の部品の「口」であるべきである。

However, [it is also a character used for Korean transliteration with the reading “nom”](#), which combines the reading of 老 (“no” in Korean) and a coda “m”, so semantically its lower part should be a Hangeul component “m”.

然而，[它还是个韩国的音译用字，读作nom](#)，结合了“老”字的韩语读音no和韵尾m，因此字理上它的下半部分应当是谚文部件m。

しかし、[これはまだnomと読む韓国の音訳用漢字であり](#)、「老」の韓国語読みnoと韻尾mを結合しているので、意味的にはその下半分がハングルの部品のmであるべきである。

I do not know whether the disunification for 耇 (U+3588) and 耇 (U+439B) were deliberately done, most likely not, but what if we face a similar case in the later working sets?

Shall we add a rule to unify or to not unify a Hangul component to a similar Han component?

我不知道耆 (U+3588) 和耆 (U+439B) 未被认同是否是刻意为之，大概率不是，但如果我们在后续的工作集中遇到类似的情况如何处理？我们是否应该加一条规则来认同或阻止认同一个谚文部件与与之相似的汉字部件？

耆 (U+3588) と耆 (U+439B) が統合されていないのは意図的なことなのかどうかは分からず、たぶんそうではないが、もし後続のワーキングセットで同じような状況に遭遇したらどうするであろうか。ハングルの部品とそれに似た漢字の部品を統合したり、統合しなかったりするルールを加えるべきであろうか。

## Han-Katakana Ideographs

We can see that there are some Hangul components very similar to Han components from the previous example, which Katakana components also do. However, it will be more complicated when analyzing the cognition since the Katakana script is originally derived from the Han script.

从前面的例子中我们可以看出，有一些谚文部件与汉字部件非常相似，而片假名部件也是如此。然而，在分析同源性时会更复杂，因为片假名最初来源于汉字。

前の例から分かるように、漢字の部品とよく似ているハングルの部品があるが、片仮名の部品もそうである。しかし、片仮名はもともと漢字に由来するため、同源性进行分析する時にはより複雑になる。

Let us first see the example 舳 (U+2C711). It is a Zhuang character with the reading “liz” or a Nôm character with the reading “lác”, which is no doubt a phonogram, whose phonetic component is obviously a Han component 力.

我们先来看一个例子“舳” (U+2C711)。这是一个读作liz的古壮字或读作lác的喃字，无疑是形声字，且其声旁显然为汉字部件“力”。

まず「舳」という例を見てみよう。これはlizと読む古壮字あるいはlácと読むチュノムであり、形声字に違いなく、その音符は明らかに漢字の部品の「力」である。

2C711  
舟 137.2  
舳  
V4-5261

<a href="#">U+2C711</a>	舳	GZ-2902201	《古壮字字典》 page 290 column 2 entry 2 character 1
-------------------------	---	------------	---

Fig.5 [IRGN2729\\_G-Source horizontal extensions for 2,145 characters](#)

Meanwhile, UTN#43 (kStrange) mentioned that it is also an abbreviated form that means “cutter” (カッター), which means that its right part is a Katakana component 力. However, Katakana 力 is originally derived from the left part of 加, which also points to 力.

同时，UTN#43 (kStrange) 提到它也可以是一个意为“卡特帆船”（日语：カッター）的略字，这说明它的右半部分是一个片假名部件“カ”。然而，片假名“カ”最初来源

于“加”字的左半部分，这也指向了“力”。

同時に、UTN#43 (kStrange) は、この字は「カッター」を意味する略字でもあり、右半分が片仮名の部品の「カ」であることを示している。しかし、片仮名の「カ」はもともと漢字の「加」の左半分に由来するので、これも「力」を指している。

Here, we may say that we should unify the Katakana component カ and Han component 力 because they not only look similar but also cognate; we may say that we should not unify them because they are from different scripts.

这里，我们可以说片假名部件“カ”和汉字部件“力”需要被认同，因为它们不但外观相似，而且同源；我们也可以说它们不能被认同，因为它们来自不同的文种。

ここでは、形が似ているだけでなく、同源的であるので、片仮名の部品の「カ」と漢字の部品の「力」を統合すべきだと言える；一方で、異なるスクリプトからなったので、それらを統合すべきではないとも言える。

Do not rush, let us see another example 𪗇 (U+282A3). It is a Zhuang character with the reading “boek” or a character used for a person name in Taiwan with the reading “pū”, which is no doubt a phonogram, whose phonetic component is obviously a Han component 卜.

别急，我们再来看一个例子“𪗇”（U+282A3）。这是一个读作boek的古壮字或读作pū的台湾人名用字，无疑是形声字，且其声旁显然为汉字部件“卜”。

焦らずに、もう一つの例「𪗇」（U+282A3）を見てみよう。これはboekと読む古壮字あるいはpūと読む台湾人名用漢字であり、形声字に違いなく、その音符は明らかに漢字の部品の「卜」である。

282A3  
車 159.2  
𪗇  
TF-2D52

<a href="#">U+282A3</a>	𪗇	GZ-0281303	《古壮字字典》page 28 column 1 entry 3 character 3
-------------------------	---	------------	---

Fig.6 [IRGN2729\\_G—Source horizontal extensions for 2,145 characters](#)

Meanwhile, UTN#43 (kStrange) mentioned that it is also an abbreviated form that means “track” (トラック), which means that its right part is a Katakana component ト. However, Katakana ト is originally derived from a part of 止, which does not point to ト.

同时，UTN#43 (kStrange) 提到它也可以是一个意为“卡车”（日语：トラック）的略字，这说明它的右半部分是一个片假名部件“ト”。然而，片假名“ト”最初来源于“止”字的一部分，这并不指向“ト”。

同時に、UTN#43 (kStrange) は、この字は「トラック」を意味する略字でもあり、右半分が片仮名の部品の「ト」であることを示している。しかし、片仮名の「ト」はもともと漢字の「止」の一部に由来するので、これは「ト」を指していない。



Here, we may say that we should unify the Katakana component ト and Han component 卜 because they look similar; we may say that we should not unify them because they are not only from different scripts but also non-cognate.

这里，我们可以说片假名部件“ト”和汉字部件“卜”需要被认同，因为它们外观相似；我们也可以说它们不能被认同，因为它们不但来自不同的文种，而且非同源。

ここでは、形が似ているので、片仮名の部品の「ト」と漢字の部品の「卜」を統合すべきだと言える；一方で、異なるスクリプトからなっただけでなく、同源的でもないので、それらを統合すべきではないとも言える。

To sum up, we may organize the main info into a table as following:

综上所述，我们可以将主要信息整理成如下表格：

以上のように、主要な情報を次の表にまとめることができる：

character	similar shape	component cognition	same script
机 (jī & つくえ)	○	○	○
舫 (liz/lǎc & カッター)	○	○	×
軌 (boek/pū & トラック)	○	×	×

We may call these three different case Lv.1, Lv.2 and Lv.3 respectively. Then we will say that 迁 (qiān, the Chinese simplified form of 遷) & 𠄎チ (see L2/24–125) are Lv.2, while 耆 (gǒu/こう or diǎn/てん) and 耆 (nom) are Lv.3. We need to decide from which level we should unify the characters and add the corresponding rules into the PnP if we decide to encode Han–Katakana ideographs as CJKUIs.

我们可以把这三类情况分别称为Lv.1、Lv.2和Lv.3。那么我们就可以说，“迁”（qiān，“遷”的简化字）和“𠄎チ”（见L2/24–125）属于Lv.2，而“耆”（gǒu/diǎn）和“耆”（nom）属于Lv.3。如果我们决定将“汉字–片假名”字符作为中日韩统一表意文字收录，则我们需要决定要从哪个级别起进行认同，并将对应的规则加入PnP中。

この三つの状況をそれぞれLv.1、Lv.2、Lv.3と呼ぶことができる。そうしたら、「迁」（qiān、「遷」の中国語簡体字）と「𠄎チ」（L2/24–125を参照）はLv.2であり、「耆」（こう/てん）と「耆」（nom）はLv.3である。「漢字–片假名」文字を日中韓統合表意文字として符号化することにすれば、どのレベルから統合するかを決定し、また対応する規則をPnPに加える必要がある。

Also, I would like to have one more question for Gen Kojitani. Why did you exclude the four Han–Katakana ideographs mentioned at the top of this document in your newest L2/24–201 document? What is the essential difference between 𠄎マ and 𠄎言コ?

另外，我还有一个问题想要问小路谷源。为什么本文最开始处提到的四个“汉字–片假名”字符在您的最新文档L2/24–201中被排除了？“𠄎マ”和“𠄎言コ”的本质差异是什么？

また、小路谷源に聞きたい質問がもう一つある。なぜ本文の最初に言及した四つの「漢字–片假名」文字があなたの最新の文書L2/24–201で除外されたのであろうか。「𠄎マ」

と「言コ」の本質的な違いは何であろうか。

## Han–Latin Ideographs

With the above preparation, let us see the Han–Latin cases. Since there is no connection between the Latin script and the Han script, there will be no Lv.2 cases.

有了上面的铺垫，我们来看“汉字–拉丁字母”这一类。由于拉丁字母和汉字之间没有关联，所以不会有Lv.2的情况出现。

上記で持ち出した下地により、「漢字–アルファベット」の種類を見てみよう。アルファベットと漢字の間には繋がりがないため、Lv.2の場合は全然出て来ない。

In IRGN2742, Wang Xieyang gave out a good example, 𠄎 & 𠄎T, which will be a Lv.3 case according to this document. The former one is [a variant form of 问](#), while the latter one is an abbreviated form that means “question” (问题, i.e. 𠄎口ti). They are phonetically and semantically different, with the similar shape.

在IRGN2742中，王谢杨给出了一个很好的例子，“𠄎”和“𠄎T”，按此文档这属于Lv.3。前者为[“问”的异体字](#)，而后者为“问题”（即𠄎口ti）的略字。它们音义皆不同，但拥有相似的字形。

IRGN2742の中で、王謝楊は「𠄎」と「𠄎T」といういい例を提供し、この文書によるとこれはLv.3である。前者は[「问」の異体字](#)であり、後者は「問題」（中国語：问题；すなわち𠄎口ti）の略字である。発音も意味も異なるが、字形だけが似ている。

We could see that Lv.3 cases really exist in Han–Hangul ideographs, Han–Katakana ideographs and Han–Latin ideographs. This issue is more complicated than we might think, just as China NB says in IRGN2731, “there is a general lack of understanding and study among linguists and coding experts on script–hybrid ideographs”. Personally I think that we may unify a Lv.2 case but we should not unify a Lv.3 case due to the non–cognition, which we have many not unified examples in normal Han characters such as 𠄎/𠄎/𠄎/𠄎 & 𠄎/𠄎/𠄎/𠄎.

我们可以发现，Lv.3的情况在“汉字–谚文”字符、“汉字–片假名”字符和“汉字–拉丁字母”字符中均有实例出现。这个问题比想象中的复杂，正如中国在IRGN2731中所说的：“文字学和编码专家对汉字–表音文字混合体文字尚普遍缺乏认知和研究”。个人认为我们可以认同Lv.2的情况，但我们不能认同Lv.3的情况，因为它们非同源，而我们在普通的汉字中也有许多未认同的例子，如“𠄎/𠄎/𠄎/𠄎”和“𠄎/𠄎/𠄎/𠄎”。

Lv.3の場合は、「漢字–ハングル」文字、「漢字–片假名」文字、「漢字–アルファベット」文字のいずれにも実例があることがわかる。この問題は想像以上に複雑で、中国がIRGN2731で述べたように、「文字学と符号化の専門家はスクリプトハイブリッド表意文字に対してまだ普遍的に認知と研究を欠いている」。個人的にはLv.2の場合を統合することができるが、Lv.3の場合を統合することはできないと考えている。なぜなら、それらは同源的不是からである。普通の漢字でも「𠄎/𠄎/𠄎/𠄎」と「𠄎/𠄎/𠄎/𠄎」のように、統合していない例がたくさんある。

Compared with Han–Katakana ideographs, there comes out one more issue for Han–Latin ideographs that, how to calculate the strokes. For example, is “I” a 1–stroke letter or a 3–stroke letter? How many strokes does “G” have? Indeed we have the similar question in normal Han characters such as whether “卩” is a 3–stroke component or a 4–stroke component. We can directly define how many strokes each letter has; or Kushim Jiang has an opinion that, all such ideographs might have an SC (residual stroke count) value “max\_value” to make sure they will be sorted at the last of a radical.

与“汉字–片假名”字符相比，对于“汉字–拉丁字母”而言还有一个问题，即怎么数笔画数。例如，I是1画还是3画？G有多少个笔画？在普通的汉字中确实也有类似的问题，例如“卩”是3画还是4画。我们可以直接定义每个字母有多少个笔画；或者姜兆勤有一个想法，所有这类字符的部外笔画数均为“max\_value”（最大值）来确保它们被排在相同部首的字的最后。

「漢字–片假名」文字と比べ、「漢字–アルファベット」には画数をどのように数えるかという問題がある。例えば、Iは1画であろうか、3画であろうか。Gはいくつの筆画があるであろうか。普通の漢字でも確かに似たような問題がある。例えば、「卩」は3画なのか4画なのか。文字ごとにいくつの筆画があるかを直接定義することができ、あるいは、姜兆勤には、これらすべての文字の部外画数が「max\_value」（最大値）であり、同じ部首の字の最後に配列されることを保証する考えがある。

There is not any technical issues for the stroke count, but we must have a clear rule to prevent the possible ambiguities and unnecessary debates. Meanwhile, the characters with cursive strokes already encoded in Unicode can also be used for reference, such as 夂.

在笔画数计算方面并没有什么技术问题，但是必须有一个明确的规则，避免可能出现的含糊不清与不必要的争论。同时，已编码的带有弧形笔画的字也能作为参考，例如“夂”。

画数計算には技術的な問題はないが、起こりうる曖昧さや不必要な論争を避けるためには、明確なルールが必要である。また、もう符号化された円弧の筆画を持つ字も参考にすることができ、例えば、「夂」。

Besides, Wang Xieyang asked a question that how to determine the glyphs of script–hybrid ideographs in regular script and whether we can normalize the glyph of 𠂇K to 𠂇K in IRGN2742, and Ming Fan thought that it was not necessary to localize these non–Han components to convert them to “normal” Han characters in IRGN2738.

另外，王谢杨在IRGN2742中问道，混合文种表意文字的楷体字形应该如何确定，以及我们是否能将“𠂇K”规范化为“𠂇K”，而范铭在IRGN2738中则认为我们没有必要将这些非汉字的部件本地化为“普通”汉字。

また、王謝楊はIRGN2742の中で、スクリプトハイブリッド表意文字の楷書体の字形をどのように決定すべきか、および「𠂇K」を「𠂇K」に標準化できるかどうかを尋ねたが、范銘はIRGN2738の中で、これらの非漢字の部品を「普通」の漢字にローカライズする必要はないと考えている。

Here I support Ming Fan. We may want to normalize “K” to 𠂇 since there already exists a similar Han ideograph, but what if we meet something like “G” or “Q”? And there

will be some advantages if we do not normalize, for example, we have 𠄎<sup>ㄨ</sup>R in L2/24–201, without the normalization, we can easily distinguish it from 宀.

这里我支持范铭。我们可能会想到要把K规范化为“ㄨ”，因为已经存在相似的汉字了，然而如果我们遇到的是G或Q呢？并且，不进行规范化也有一些好处，例如在L2/24–201中有“𠄎<sup>ㄨ</sup>R”字，而若我们不进行规范化，则可以很容易将其与“宀”区别开来。

ここでは范銘を支持する。似たような漢字がすでに存在するので、Kを「ㄨ」に標準化しようとするかもしれないが、もし出会ったのはGやQであったら？また、標準化しないことにもいくつかのメリットがあります。例えば、L2/24–201には「𠄎<sup>ㄨ</sup>R」の字があり、標準化しなかったら「宀」と区別しやすくなる。

## Han–Hiragana Ideographs

In the previous text, I have already raised the issues about the unification rules for the Lv.3 cases and the stroke count rules. And about the cases like 坊 & 𠄎者方 (if exists), we will never unify them even if they are cognate, since they are too different, just like 馬 & 马, 鳥 & 鸟, etc. So there will be nothing new to discuss here.

前文中我已经提过有关Lv.3的认同规则和笔画数计算规则的问题。关于类似于“坊”和“𠄎者方”（若存在）的情况，即使其同源，我们也从来不会认同，因为差异太大，就像“馬”和“马”、“鳥”和“鸟”等。因此这里没有什么新的问题待讨论。

上記では、もうLv.3の場合の統合ルールと画数計算のルールについて質問した。「坊」と「𠄎者方」（存在するなら）のような状況については、「馬」と「马」や「鳥」と「鸟」などのように、違いが大きすぎるため、同源的であってもけっして統合しない。したがって、ここには議論すべき新しい問題がない。

## Other Possible Cases

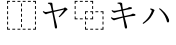
In IRGN2742, Wang Xieyang also mentioned that there were some other cases of script–hybrid ideographs such as Han–Bopomofo ideographs. Since the Bopomofo script is derived from the Han script, and the strokes of Bopomofo are even more similar to Han than the strokes of Katakana are, we can just apply the rules for Han–Katakana ideographs to Han–Bopomofo ideographs.

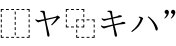
在IRGN2742中，王谢杨也提到了存在一些其它种类的混合文种表意文字，例如“汉字–注音符”字符。由于注音符也源于汉字，且比起片假名，注音符的笔画更接近汉字，我们可以直接将“汉字–片假名”字符的规则套用至“汉字–注音符”。


IRGN2742の中で、王謝楊は「漢字–注音符」文字など、他の種類のスク립トハイブリッド表意文字が存在することも言及している。注音符も漢字に由来し、片仮名よりも注音符の筆画が漢字により近いため、「漢字–片仮名」文字のルールを「漢字–注音符」にそのまま当てはめることができる。

## Conditions for Determining Script–hybrid Ideographs

In IRGN2742, Wang Xieyang asked if it is necessary to have Han components when determining script–hybrid ideographs. The answer should be “no”. Assume that you create a

Katakana–Hangul ideograph, it completely meets the definition of “script–hybrid”. This may be a more serious problem, however, actually it seems that such cases could hardly exist. For something like  from IRGN2637, it is no doubt not a script–hybrid ideograph. However, for something like that, we cannot assert that “it is not a script–hybrid ideograph” equals “it should not be encoded as a CJKUI”. This turns to another question, how to encode such “ideographs without Han components”.

在IRGN2742中，王谢杨问道，我们在确定一个字符是否为混合文种表意文字时，拥有汉字部件是否必需。答案应该是否定的。假设你创造了一个“片假名–谚文”字符，它完全符合“混合文种”的定义。这似乎是一个更严重的问题，然而这类情况事实上很难存在。对于IRGN2637中的“”，它毫无疑问不是混合文种表意文字。然而，对于这类情况，我们不能断言“它不是混合文种表意文字”等同于“它不应该编码为中日韩统一表意文字”。这转变成了另一个问题，即如何编码这些“没有汉字部件的表意文字”。

IRGN2742の中で、王謝楊は、文字がスクリプトハイブリッド表意文字であるかどうかを判断する時、漢字の部品を持つ必要があるかどうかを尋ねた。答えは否定であるべきである。仮に「片仮名–ハングル」文字を作るとしたら、完全に「スクリプトハイブリッド」の定義に合っている。これはもっと深刻な問題のようであるが、このような状況は事実上存在しにくい。IRGN2637に現れる「」については、間違いなくスクリプトハイブリッド表意文字ではない。しかし、このような状況に対して、「スクリプトハイブリッド表意文字ではない」ということは「日中韓統合表意文字として符号化すべきではない」ということに等しいとは断言できない。「漢字の部品のない表意文字」をどのように符号化するかという別の問題に変わっている。

## Ideographs without Han Components

Let us see an example ㇗ (U+2BCCD).

我们来看一个例子“㇗”（U+2BCCD）。

「㇗」（U+2BCCD）という例を見てみよう。

2BCCD  
㇗ 40.7  
㇗  
JK-65130

MJ057333



コピーフィールド  
玊

窠

MJ文字図形名 MJ057333

MJ文字図形バージョン 1.0

総画数 10

音読み

訓読み うつほ・うつぼ


部首内画数 宀部7画


Fig.7 [文字情報基盤検索システム for JMJ-057333](#)


It is no doubt combined with three Katakana components ウ, ツ and ホ, but not three Han components 宀, ツ and 木, according to its reading.

根据它的读音,毫无疑问它由三个片假名部件“ウ”、“ツ”和“ホ”组成,而非三个汉字部件“宀”、“ツ”和“木”。

読みによると、三つの漢字の部品の「宀」・「ツ」・「木」ではなく、三つの片仮名の部品の「ウ」・「ツ」・「ホ」で構成されていることは間違いない。

It is reasonable to identify it as an ideograph since it has a stable reading うつぼ, a stable glyph ウツホ and a stable meaning “feather pot”.

将其认定为表意文字是合理的,因为它有稳定的读音“うつぼ”、稳定的字形“ウツホ”和稳定的字义“羽壺”。

安定した読み「うつぼ」があり、安定したグリフ「ウツホ」もあり、安定した意味「羽壺」もあり、表意文字として認識するのは合理的である。

What is the core difference between a kana and a Han ideograph? A kana has only a stable reading and a stable glyph, but it does not have a stable meaning. Kana itself does not represent for anything.



假名和汉字(表意文字)的核心区别是什么?假名只有稳定的读音和稳定的字形,却没有稳定的意义。假名本身不代表任何东西。





仮名と漢字(表意文字)の核心的な違いは何であろうか。仮名は安定した読みと安定したグリフだけがあり、安定した意味はない。仮名そのものは何も意味しない。

窠 is an ideograph, and it is used in Japan, so it should be encoded as a CJKUI.

“窠”是表意文字,并且用于日本,因此它应该被编码为中日韩统一表意文字。

「窠」は表意文字でありながら、日本で使用されているので、日中韓統合表意文字として符号化されるべきである。

For the same reason, ヤキハ has also to be encoded as a CJKUI.

同理，“ヤキハ”也应被编码为中日韩统一表意文字。  
同様に、「ヤキハ」も日中韓統合表意文字として符号化されるべきである。

There is no lack of characters that wholly derived from another script encoded as CJKUIs, for example:

被编码为中日韩统一表意文字的字符中并不乏整个儿地从另一文种衍生而来的字符，例如：

日中韓統合表意文字として符号化された文字の中には、全体的に別のスクリプトから派生した文字も少なくない。例えば、

宋洪遵《泉志・外國品中・屋馱國梵書錢》載“梵子錢”有“𠵽、𠵾、𠵿、𠶀”四字，云“文不可辨”。——《漢語大字典》

These four characters from CJK-ExtB are originally derived from the Tangut script, the corresponding Tangut characters are 𠵽𠵾𠵿𠶀.

这四个中日韩统一表意文字扩展区B中的字来源于西夏文，对应的西夏文字符为“𠵽𠵾𠵿𠶀”。

この四つの日中韓統合表意文字拡張Bの字は西夏文字に由来し、対応する西夏文字の文字は「𠵽𠵾𠵿𠶀」である。

(Ref: <https://hc.jsecs.org/irg/ws2021/app/?id=02305#c14447>)

Besides Tangut, there are also many other characters from Siddham, Khitan Large Script, etc. No reason only to reject the ideographs from Katakana – even 卫 (the Chinese simplified form of 衛) is derived from the Katakana letter エ!

除西夏文之外，也有很多来源于悉曇文、契丹大字等文种的字符。我们没有理由只拒绝来源于片假名的表意文字——甚至于，连“卫”（“衛”的简化字）都源自片假名“エ”！

西夏文を除けば、悉曇文字や契丹大字などのスクリプトに由来する文字も多い。片假名由来の表意文字だけを拒否する理由はない——さらに、「卫」（「衛」の中国語簡体字）も片假名の「エ」に由来している！

The next tricky thing is the “E” from IRGN2731. If a hospital is very big and has many branches, we may say A院, B院, C院, D院, E院 in order to refer a specific branch; however, if “E” is a regional informal simplified form of 医, then E院 only refers to “hospital” itself. These two “E”s are very different, the former one is just a Latin letter used as a serial number, meanwhile, the latter one is an ideograph with its stable reading, stable glyph and stable meaning.

下一个棘手的東西是IRGN2731中的E。如果一个医院很大并且有很多分院，我们可能会用“A院”、“B院”、“C院”、“D院”、“E院”来指代某个特定的分院；然而，若E作为“医”的地域性俗字，那么“E院”就只代表“医院”本身。这两个E非常不同，前者只是一个用作序列号的拉丁字母，而同时后者却是一个拥有稳定音形义的表意文字。

次の厄介なものはIRGN2731でのEである。もし病院が大きくて多くの分院があれば、「A院」・「B院」・「C院」・「D院」・「E院」を使って特定の分院を指すかもしれない。しかし、Eが「医」の地域的な俗字であれば、「E院」は病院（中国語：医院）そのもの

のを意味するだけである。この二つのEは非常に異なり、前者はシリアル番号として使われるアルファベットにすぎず、後者は安定した音・形・義を持つ表意文字になっている。

As China NB says in their document, “Encoding such characters together with CJKUIs may confuse the boundaries between letters and Chinese characters, seriously interfering with the existing theories of Chinese characters. For example, the nature of letters which are used to replace Chinese characters may become controversial if we do so.”

正如中国在他们的文档中所述的那样：“将此类字符与中日韩统一表意文字放在一起编码，可能会混淆字母与汉字的界限，与现有汉字理论产生严重冲突。比如，用于代替汉字的字母的性质可能会产生争议。”

中国が彼らの文書で述べたように、「このような文字を日中韓統合表意文字と一緒に符号化すると、アルファベットと漢字の境界線が混同され、既存の漢字理論と深刻な衝突が生じる可能性がある。例えば、漢字の代わりに使用されるアルファベットの性質が論争になるかもしれない。」

Indeed. However, it is neither a combination of multiple characters from another script, nor a glyph corruption like 𠃉𠃊𠃋𠃌. Based on this, I suggest not to encode a new character but only to use the **fullwidth** Latin letter E (U+FF25), since we do not want to see a rotated ideograph in the vertical texts. We may consider to add a `script_extension` property for E (U+FF25) as “Hani” to indicate that it could also be used as an ideograph in Chinese texts – [during the July, 2024 SEWG meeting, people decided to encode some phonetic characters of Greek origin with the script property “Latn” since their use is in phonetic notation systems that are extensions of the Latin script.](#) The suggestion for E (U+FF25) here is similar.

确实。然而它既不是来源于另一个文种的多个字母的结合体，也不是一个类似于“𠃉𠃊𠃋𠃌”的损坏的字形。基于此，我建议不重新编码该字符，而是使用**全角**拉丁字母“E”（U+FF25），因为我们不希望在竖排文本中看到一个被旋转了的表意文字。我们也许可以考虑给“E”（U+FF25）增加一个`script_extension`属性“Hani”以指示出它也能够在中国文本中作为表意文字使用——[在2024年7月的文种编码工作小组会议中，人们决定编码一些来源于希腊字母的音标符号，其script属性为“Latn”，因为它们只在扩展拉丁字母的音标符号系统中使用。](#)此处对于“E”（U+FF25）的建议也类似。

確かに。しかし別のスクリプトからの複数の文字の結合ではなく、「𠃉𠃊𠃋𠃌」のような崩壊のグリフでもない。これに基づいて、この文字を新しく符号化せず、代わりに**全角**のアルファベットの「E」（U+FF25）を使うことを勧める。縦組みのテキストに回転された表意文字を見たくないからである。「E」（U+FF25）に`script_extension`の属性「Hani」を追加し、中国語のテキストでも表意文字として使えることを指し示すことも考えられるかもしれない——[2024年7月のスクリプトエンコーディングワーキンググループの会議では、ラテン文字拡張の音声記号のシステムでしか使用されていないため、scriptの属性が「Latn」であるギリシャ文字由来の音声記号を符号化することにした。](#)ここの「E」（U+FF25）についての提議も類似なことである。

Additionally, the same for Q (U+FF31) if we have more documentary evidences:



此外，对于“Q”（U+FF31）也同样，若有更多书面证据：  
 その他、もっと多くの書面の証拠があれば、「Q」（U+FF31）についても同様：

西文字母 [編輯]

全學聯漢字	原字形	繁體中文	簡體中文
Q <sup>[14]</sup>	級／ <b>kyū</b>	級	级
[[石 K]] <sup>[14]</sup>	確／ <b>kaku</b>	確	确
[[食 K]]	館／ <b>kan</b>	館	馆
[[言 S]]／[[讠 S]]	識／ <b>shiki</b>	識	识
[[J o]]	情／ <b>jō</b>	情	情
[[糸 N]]	繩／ <b>nawa</b>	繩	绳
[[言 N]]／[[讠 N]]	認／ <b>nin</b>	認	认
[[宀 R]] <sup>[19]</sup>	寮／ <b>ryō</b>	寮	寮
[[亻 宀 R]] <sup>[14]</sup>	僚／ <b>ryō</b>	僚	僚
[[广 K]]	慶／ <b>kei</b>	慶	庆
[[广 O]]	応／ <b>ō</b>	應	应
[[广 K O]]	慶応／ <b>keiō</b>	慶應	庆应
[[宀 K]]	警察／ <b>keisatsu</b> 憲法／ <b>kenpō</b>	警察 憲法	警察 宪法
[[M o]]	ムーブメント／ <b>movement</b>	不適用	不適用

Fig.8 [The article 全学联汉字 on Chinese Wikipedia](#)

## Conclusion

To sum up, technically nothing difficult, grammatologically a plenty of issues need to be solved with a clear rule and a consensus. Before all the issues are solved, IRG should not carelessly decide to handle the script–hybrid ideographs. Each kind of script–hybrid ideographs may be discussed separately, there are many choices other than accept all or reject all, e.g. my opinion as stated above.

总而言之，技术方面并无难点，而文字学方面有很多问题亟需通过明确的规则和共识来解决。在所有问题解决之前，IRG不应当草率地决定处理混合文种表意文字。每一类混合文种表意文字都可以被单独讨论，除了全部接受或全部拒绝之外，还有很多其它选项，例如上文中所陈述过的我的观点。

要するに、技術面に難点はないが、文字学面では明確なルールと共通認識を通じて解決する必要がある問題が多い。すべての問題が解決する前に、IRGはスクリプトハイブリッド表意文字を処理することを軽率に決定してはならない。各種類のスクリプトハイブリッド表意文字は別々に議論することができ、すべて受け入れるまたはすべて拒否する以外にも、上記で述べた私の考えなど、他の選択肢がたくさんある。

Any inputs of academic value are welcome.

欢迎任何具有学术价值的意见。

学术的価値のある意見を歓迎する。

(End of document)