

# FINAL PROPOSAL TO ENCODE TAIWANESE KANJI (臺-灣-語-假-名) IN THE UCS

Fredrick R. Brennan

孟 メイ  
福 フク  
黎 レイ  
Psihedelisto  
フレッド・ブレンナン  
*copypaste@kittens.ph*

LAST REVISED ON  
13 December 2020

*Changes since 9 November 2020:* Accepted all round two changes requested by Script Ad Hoc.

*Changes since 27 October 2020:* Accepted all changes requested by Script Ad Hoc.

*Changes since 23 September 2020:* Accepted all changes requested by Eiso Chan in L2/20-233.

古家時雄 ふるやときお を追悼して  
文字鏡研究会 もじきょう に心から感謝します

This document was typeset with S<sub>I</sub>L<sub>E</sub>.

# ACKNOWLEDGEMENTS

*in no particular order...*

Sachihiro Tanimoto (<sup>たにもと さちひろ</sup>谷本玲大), Waseda University, Kokugakuin University

For his patient explanation of the history of Mojikyō, and his priceless help in getting Mojikyō Character Map working on my computer.

Deborah Anderson, University of California @ Berkeley Script Encoding Initiative

For her tireless review of script proposals by n00bs like me.

Daisuke Miura (<sup>み うらだい すけ</sup>三浦大介), World Special-Characters Wiki (世界の特殊文字ウィキ)

For his recommendation that I name the tone letters like MODIFIER LETTER KATAKANA instead of just KATAKANA as I originally planned; I did not know modifier letters could be non-Latin, but he knew of the precedent of U+10FC —MODIFIER LETTER GEORGIAN NAR <sup>(6)</sup>.

Ryūsei Yamaguchi (<sup>やま ぐち りゅうせい</sup>山口隆成)

For his experience with Mojikyō, Unicode, and all around good advice.

Wil Lee (Lí Kho-Lūn, 李<sup>リ</sup>科<sup>コ</sup>潤<sup>潤</sup>), Patreon

For kindly giving me a Taiwanese Hokkien name (also usable for Mandarin Chinese), for use in this proposal.

Ken Lunde (<sup>こ ぼやしけん</sup>小林 劍), Unicode Consortium

For his font development advice, and helpful advice regarding Unihan.

Issei Yamazaki (<sup>やま ざき いっ せい</sup>山崎一生)

For helping me choose good shapes for the glyphs as a Japanese learner of Hokkien who writes in Taiwanese kana daily, and providing me with several difficult to find resources.

Simon Cozens (<sup>さいもん</sup>カズンズンオウ文) & Caleb Maclennan, S<sub>I</sub>L<sub>E</sub>

For their help with S<sub>I</sub>L<sub>E</sub>, an experimental typesetting engine and alternative to L<sup>A</sup>T<sub>E</sub>X.

# Table of Contents

<b>Quick summary . . . . .</b>	<b>5</b>
Tone marks . . . . .	5
Nasalized tone marks . . . . .	6
Combining characters . . . . .	7
<b>Overview . . . . .</b>	<b>8</b>
Linguistic . . . . .	8
Historic (Taiwanese...kana?) . . . . .	8
Legacy encodings . . . . .	11
Pseudo-Unicode encodings . . . . .	11
Mojikyō . . . . .	11
<b>Unicode considerations . . . . .</b>	<b>13</b>
UnicodeData.txt . . . . .	13
On tone six . . . . .	13
On the aspiration mark . . . . .	14
On glyph order . . . . .	14
Tone letters . . . . .	14
Combining characters . . . . .	14
Vertical typesetting . . . . .	15
On naming . . . . .	16
On shakuhachi notation . . . . .	17
<b>Evidence of all requested characters . . . . .</b>	<b>18</b>
Tone letters . . . . .	18
Combining characters . . . . .	18

Miscellaneous pages . . . . . 20

**Bibliography . . . . . 24**

# Chapter 1

## Quick summary

*For copyable names, see § Unicode considerations.*

I propose a new block, in the Supplementary Multilingual Plane, to be called Kana Extended-B. I propose that the new block be given range U+1AFF0 – U+1AFFF, and that it include:

### 1.1 Tone marks

/ KATAKANA LETTER MINNAN TONE-2  
 Mojikyō 69622; proposed representation U+1AFF0

- In Minnan, also used for tone six<sup>1</sup>

\ KATAKANA LETTER MINNAN TONE-3  
 Mojikyō 69623; proposed representation U+1AFF1

• KATAKANA LETTER MINNAN TONE-4  
 Mojikyō 69624; proposed representation U+1AFF2

< KATAKANA LETTER MINNAN TONE-5  
 Mojikyō 69625; proposed representation U+1AFF3

| KATAKANA LETTER MINNAN TONE-7  
 Mojikyō 69627; proposed representation U+1AFF5

• KATAKANA LETTER MINNAN TONE-8  
 Mojikyō 69628; proposed representation U+1AFF6

1. For more discussion of this issue, and why I'm proposing we leave U+1AFF4 and U+1AFFC open, please see § On tone six.

## 1.2 Nasalized tone marks

Ḷ KATAKANA LETTER MINNAN NASALIZED TONE-1  
Mojikyō 69629; proposed representation U+1AFF7

Ḹ KATAKANA LETTER MINNAN NASALIZED TONE-2  
Mojikyō 69630; proposed representation U+1AFF8

- In Minnan, also used for tone six<sup>1</sup>

Ḷ̇ KATAKANA LETTER MINNAN NASALIZED TONE-3  
Mojikyō 69631; proposed representation U+1AFF9

Ḷ̈ KATAKANA LETTER MINNAN NASALIZED TONE-4  
Mojikyō 69632; proposed representation U+1AFFA

Ḷ̉ KATAKANA LETTER MINNAN NASALIZED TONE-5  
Mojikyō 69633; proposed representation U+1AFFB

Ḷ̊ KATAKANA LETTER MINNAN NASALIZED TONE-7  
Mojikyō 69635; proposed representation U+1AFFD

Ḷ̋ KATAKANA LETTER MINNAN NASALIZED TONE-8  
Mojikyō 69636; proposed representation U+1AFFE

1. For more discussion of this issue, and why I'm proposing we leave U+1AFF4 and U+1AFFC open, please see § On tone six.

### 1.3 Combining characters

*Note: The two below are provided for informational purposes and would be unified with existing characters. Notes are about Taiwanese kana.*

#### ◌◌ COMBINING DOT BELOW

Mojikyō 69202; U+0323

- Indicates aspiration; native name *sàng-khì hú* [送氣符]. For example, while *タ* is /ta/, *タア* is /t<sup>h</sup>a/.
- At least one source renders it identically to the *nakaguro* (中黒);  
→ • U+30FB KATAKANA MIDDLE DOT

#### ◌◌ COMBINING OVERLINE

Mojikyō 69637–69642;<sup>2</sup> U+0305

- Makes the kana it combines with /t/ in the case of *チ* (/ti/) and *ツ* (/tu/); or adds a /t/ as in *チ* (/tʃa/ or /tʃa/) and *ツ* (/tʃe/ or /tʃe/); or changes vowel sound as in *チ* (/i/) and *チ* (/ə/)
- Sometimes rendered identically to the *chōonpu* (長音符; U+30FC; KATAKANA-HIRAGANA PROLONGED SOUND MARK), so *チ* in place of *チ*.

For the dotted and overline characters, no widely accepted names are in use. No Unicode Named Character Sequences are requested.

2. Mojikyō chose to encode the versions of the katakana with combining overline above separately, so this refers to the individual katakana they encoded. In Mojikyō order: *チチツモツウ*. They are missing *チ* and *ウ*.

## Chapter 2 Overview

### 2.1 Linguistic

Taiwanese Hokkien ( 臺語 ) is a Sinitic Southern Min language spoken throughout the world; as of 2013, it was spoken by 48 million people.<sup>1</sup> It is spoken as far afield as the Philippines, where it is known as *Fookien*. In ISO 639-3, it is known as Min Nan.

In Taiwan, Hokkien is a widely spoken language, and is even used in academic writing. The most popular way of writing Hokkien is known as *Hàn-lô* ( 漢羅 ), which to untrained eyes may appear to be *mojibake*...

“ M̄ 是講攏無台灣來 ê 學生來修。Tī 中級班，有一個讀電腦 ê 博士生；初級班有一個客家人，讀設計碩士班。<sup>2</sup> ”

However, it is not; what is happening is that 10–15% of Hokkien phonemes are not representable as a standardized Chinese character. So, reading Hokkien only in Chinese characters, called *Tng-lâng-jī* ( 唐儂字 ), is quite difficult. Therefore, other systems, among them the Taiwanese government’s official *Tâi-lô* ( 臺羅 ) and the older *Peh-ōe-jī* ( 白話字 ) upon which it is based, along with, more rarely, *chù-im hú-hō* ( 注音符號 ), known elsewhere as *bopomofo*, are mixed in with the Chinese characters. Indeed, there’s no reason Taiwanese kana can’t be used for this purpose:

“ ム卜是講攏無台灣來工<學生來修。子|中級班，有一個讀電腦工<博士生；初級班有一個客家人，讀設計碩士班。 ”

But perhaps we’re getting ahead of ourselves.

### 2.2 Historic (Taiwanese...kana?)

If I were given the task of naming Taiwanese kana, or, perhaps better said, translating their Hokkien/Japanese name, I would have named them “Hokkien kana”. Because,

1. Sun, Obed [孫暉皓] (2017). “台灣16年「母語教育」宣告失敗？我們需要「搶救台語」嗎？——我在新加坡的四個觀察”. Crossing [換日線]. Tianxia Magazine [天下雜誌]. Quote: “全球的閩南語（包括潮汕話、海南話等）使用者有超過 4,800 萬人”.

2. Khîn-hōa<sup>n</sup>, Lí [李勤岸] (2007). 台語飯桌. Sin Bōng Ài [信望愛].



indeed, that's what they are; historically speaking, there's nothing Taiwanese about Taiwanese kana except that they can be used to write Taiwanese Hokkien. Indeed, they are not even exclusively *Taiwanese*, as they are also suitable for writing the Quanzhou dialect.

Taiwanese kana were invented by Japanese linguists on behalf of the imperial Japanese government, which at the time of their introduction, was pursuing a policy of *dōka* (同化), meaning, “assimilation”. The period we most remember them for today was the period of the 1930's, when they were spread primarily to teach Taiwanese people Japanese.<sup>3</sup>

The primary person involved in their creation was the linguist Naoyoshi Ogawa [小川尚義].<sup>4</sup> The first of Ogawa's works that they make a notable appearance in is the *Japanese–Taiwanese Daijiten* (日台大辭典) of 1907.<sup>5</sup> Another early example is a textbook from 1902, the *Textbook for Taiwan* (台灣教科用書). Ogawa continued to refine them until he published his greatest work, the two volume *Comprehensive Taiwanese–Japanese Dictionary* [臺日大辭典] (1931). Indeed, it is entirely on the back of this work that modern interest in them continues, with the dictionary continuing to receive updates long after the collapse of the Japanese empire; even famed Taiwanese linguist Âng Ūi-jîn contributed substantially to an update of the dictionary in 1993, published as the *Minnan Classic Dictionary Collection* (閩南語經典辭書彙編).

Due to the high quality of this dictionary, which even informed the Taiwanese government's own *Dictionary of Frequently-Used Taiwan Minnan* [臺灣閩南語常用詞辭典] (1<sup>st</sup> ed. 2011),<sup>6</sup> Japanese learners of Hokkien continue to use Taiwanese kana even today.<sup>7</sup> But, of course, this is not the only reason to encode Taiwanese kana; they are interesting in and of themselves, as a historical writing system with a large body of work behind it.

Even in Ogawa's time, he was not the only author; other Japanese scholars recognized the quality of his work, and many other works exist which integrate Taiwanese kana, such as the *Taiwan Proverb Collection* [臺灣俚諺集覽] (1914) and the *State Monopoly Bureau's Taiwanese Dictionary* [專賣局台灣語典] (1923), which, unlike Ogawa's work, mixed *Peh-ōe-jī* with Taiwanese kana.

3. Chen, Chun-Hui [陳君慧] (June 2002); Lin, Ching-Hsiun [林慶勳], ed. “《訂正台灣十五音字母詳解》音系研究”. National Sun Yat-sen University Chinese Literature Department: Master's Thesis.

4. Ibid, p. 19.

5. Ibid, p. 19–20.

6. “Bibliography” [參考書目]. 臺灣閩南語常用詞辭典. Retrieved August 1, 2020.

7. For two examples of Japanese learners of Hokkien who use Taiwanese kana, see the Twitter profiles of Soa<sup>n</sup>-kiä (@suannkia) and Taipa (@taipalogy).

In our time, besides being used as a substantial basis for the *Dictionary of Frequently-Used Taiwan Minnan*, Âng Ūi-jîn's version of the dictionary lives on due to a project of Lîm Chùn-iok [林俊育] of the Academia Sinica: [taigi.fhl.net/dict](http://taigi.fhl.net/dict). This impressive work is a complete digitization of the 1993 dictionary, and when you search for terms, you are given links to individual pages of it. For example, searching 夢, meaning "dream", brings up [page 555 of volume 2](#), which brings up quite a detailed entry, which differentiates between frightening dreams/nightmares (恐ろしい夢) and merely bad dreams (悪い夢); between being told about a dream (夢の告を受ける) from dream interpretation (夢判断):

託<sup>トク</sup> | 出現<sup>ツツヒエス</sup> || 夢枕<sup>ゆめまくら</sup> に立つ。  
 い<sup>ゆめ</sup> 夢。完<sup>ウス</sup> | || 夢<sup>ゆめ</sup> の告<sup>つげ</sup> を受<sup>う</sup> ける。  
 圓<sup>チアス</sup> | || 夢<sup>ゆめ</sup> 判<sup>はん</sup> 断<sup>だん</sup>。  
 夢<sup>ゆめ</sup>。● (姓) 夢<sup>ぼう</sup>。● 夢<sup>ゆめ</sup>。悪<sup>オク</sup> | || 恐<sup>おそ</sup>ろしい夢<sup>ゆめ</sup>。歹<sup>パイ</sup> | || 惡<sup>わる</sup>。

## 2.3 Legacy encodings

### 2.3.1 Pseudo-Unicode encodings

A “pseudo-Unicode” encoding is currently in use on Wikipedia and elsewhere. This “encoding”, to the extent that it is one, mixes SVG glyphs, HTML markup, and a few Unicode characters intended for other purposes.

- U+1AFF0–1AFFE are represented with SVG’s;
- U+0323, the COMBINING DOT BELOW, is used as in this document; and
- U+0305, the COMBINING OVERLINE, is represented variably by a CSS border-top, or else by a COMBINING MACRON (U+0304) or other script-inappropriate combining mark. It seems this is done due to lack of font support for combining marks over katakana.

A second pseudo-Unicode encoding exists, which I call the “[Taipa encoding](#)”. It misuses characters like U+27E8, MATHEMATICAL LEFT ANGLE BRACKET, because of its similar appearance to tone 5, ㄥ. It does not seem to have widely caught on outside of Taipa’s tweets; certainly this kind of misuse is a good reason to introduce real characters for the task.

### 2.3.2 Mojikyō

Mojikyō almost fully supports Taiwanese kana. Its only deficiency is that it is missing ㄨ̄ and ㄛ̄, which means that the Quanzhou dialect is not reproducible in the Mojikyō encoding.

In Mojikyō, the Taiwanese kana are available in the font `Mojikm0D.TTF`; 20 code-points are devoted to them. Mojikyō chooses to encode the COMBINING OVERLINE as precombined with the kana it combines with, and it encodes the COMBINING DOT BELOW as a spacing mark.



As mentioned in § Acknowledgements, Mr. Tanimoto and Mr. Yamaguchi helped me a lot with getting this program working. I was quite touched by Mr. Tanimoto's story, so I chose to memorialize his friend, Mr. Tokio Furuya, who is now deceased, on the front cover of this paper. Mr. Furuya was the main developer of Mojikyō.

Below are the Taiwanese kana glyphs as they appear in Mojikm0D.TTF; the top row are the tone letters, middle row nasalized tone letters, and bottom row overlined katakana:

ㄨ ㄨˊ ㄨˊˊ ㄨˋ ㄨˋˊ ㄨˋˊˊ  
 ㄨˊ ㄨˊˊ ㄨˊˊˊ ㄨˋˊ ㄨˋˊˊ ㄨˋˊˊˊ  
 ㄨ ㄨ ㄨ ㄨ ㄨ ㄨ

## Chapter 3

# Unicode considerations

### 3.1 UnicodeData.txt

```

1AFF0;KATAKANA LETTER MINNAN TONE-2;Lm;0;L;;;;;N;;;;;
1AFF1;KATAKANA LETTER MINNAN TONE-3;Lm;0;L;;;;;N;;;;;
1AFF2;KATAKANA LETTER MINNAN TONE-4;Lm;0;L;;;;;N;;;;;
1AFF3;KATAKANA LETTER MINNAN TONE-5;Lm;0;L;;;;;N;;;;;
1AFF5;KATAKANA LETTER MINNAN TONE-7;Lm;0;L;;;;;N;;;;;
1AFF6;KATAKANA LETTER MINNAN TONE-8;Lm;0;L;;;;;N;;;;;
1AFF7;KATAKANA LETTER MINNAN NASALIZED TONE-1;Lm;0;L;;;;;N;;;;;
1AFF8;KATAKANA LETTER MINNAN NASALIZED TONE-2;Lm;0;L;;;;;N;;;;;
1AFF9;KATAKANA LETTER MINNAN NASALIZED TONE-3;Lm;0;L;;;;;N;;;;;
1AFFA;KATAKANA LETTER MINNAN NASALIZED TONE-4;Lm;0;L;;;;;N;;;;;
1AFFB;KATAKANA LETTER MINNAN NASALIZED TONE-5;Lm;0;L;;;;;N;;;;;
1AFFD;KATAKANA LETTER MINNAN NASALIZED TONE-7;Lm;0;L;;;;;N;;;;;
1AFFE;KATAKANA LETTER MINNAN NASALIZED TONE-8;Lm;0;L;;;;;N;;;;;

```

### 3.2 On tone six

In modern Hokkien, tone six is equal to tone two.<sup>1</sup> A symbol, therefore, for tone six, is not in frequent use. However, I found a single source with a symbol for tone six, Taigikho (2017), and recommend we leave a slot open for it should it become more popular in the future, or should more sources emerge, so we have a logical place to put it. Even in Taigikho (2017), however, the symbol is barely discernible, as shown below, and it seems to be that they are trying to discourage its use.

1. Chiung, Wi-vun Taiffalo (2003). “[Tone Change in Taiwanese: Age and Geographic Factors](#)”. University of Pennsylvania Working Papers in Linguistics. **8** (1).

表 0  
18  
◎ 聲調

調序	第1聲	第2聲	第3聲	第4聲	第5聲	第6聲	第7聲	第8聲
調類	陰平	陰上	陰去	陰入	陽平	陽上	陽去	陽入
臺羅調符	○	ó	ò	oh	ô	ǒ	ō	òh
臺語假名調符 <small>(臺北元音 臺北輔音)</small>	常音	/	\	●	<			●
	鼻音	ㄛ	ㄛ	ㄑ	ㄑ		ㄑ	ㄑ

✓ 後表標註「ㄛ」的音節，臺語假名使用「鼻音」調符。

### 3.3 On the aspiration mark

The aspiration mark is quite unusual in that in some sources it acts as a combining character,<sup>2</sup> and in other sources, it appears identically to the *nakaguro* (中黒).<sup>3</sup>

As they are semantically the same, this issue should be left to higher-level protocols, such as fonts, to solve, via e.g. stylistic alternates. It is likely that in the sources where it appears as a *nakaguro*, this is only due to subpar printing technology, as it's clear it's meant to have been combining by Ogawa.

### 3.4 On glyph order

It makes the most sense to use the glyph order from Âng and Ogawa (1992). It is a simple order: basic Japanese kana order (*gojūon*, 五十音), with tone letters in numerical order and nasalized versions after normal versions. Overlined versions of characters come after regular versions, and aspirated versions come after those.

So:

#### 3.4.1 Tone letters

/ < ㄛ < \ < ㄑ < ● < ㄑ < < < ㄑ < | < ㄑ < ● < ㄑ

#### 3.4.2 Combining characters

夕 < 夕<sup>4</sup>

- Ogawa (1932), Âng and Ogawa (1992), Taigikho (2017), among others.
- State Monopoly Bureau's Taiwanese Dictionary* [專賣局台灣語典] (1923), among others.
- Âng and Ogawa (1992), vol. 2, p. 2.

サ < 𑖑 < 𑖒 < ソ<sup>5</sup>

### 3.5 Vertical typesetting

Taiwanese kana are most often typeset vertically historically. All horizontally typeset examples I was able to find are from this century, e.g., on Wikipedia:

**Taiwanese kana** (タイ < ヲアヌ < ギイ / カア / ビエン < ) is a *katakana*-based *writing system* that was used to write *Taiwanese Hokkien* (commonly called "Taiwanese") when the *island of Taiwan* was *under Japanese rule*. It functioned as a phonetic guide to *hanzi*, much like *furigana* in *Japanese* or *Zhuyin fuhao* in *Chinese*. There were similar systems for other languages in Taiwan as well, including *Hakka* and *Formosan languages*.

As far as Unicode is concerned, all we need is to put this in `VerticalOrientation.txt`:

```
1AFF0..1AFFE ; U # Lc [13] KATAKANA LETTER MINNAN TONE-2..KATAKANA
LETTER MINNAN NASALIZED TONE-8
```

For this paper, I am using a SIL<sup>E</sup> package I wrote to do the vertical layout. Looking at the vertical layout, there are two ways to do it. In this paper, I consider it to be two lines, with the tone letter on the second line. It is also possible to do so in OpenType via contextual positioning, like this:

```
feature valt
  lookup valt2
    # Make glyph 0-width and move it back 1.5 glyphs;
    # takes advantage of fact kana are monospace.
    # x    y    xadv yadv
    pos @toneletters.valt <-1300 500 0    0>;
  valt2;
valt;
```

Vertical layout is really a problem to be handled by higher level protocols, and not a Unicode consideration, but is mentioned for completeness.

For a fully working OpenType implementation, see [FRB Taiwanese Kana](#), which is a font I made originally for this proposal but which can be used for Taiwanese kana in general. It works with HTML vertical layout, and has bolder glyphs for ruby:

5. Ång and Ogawa (1992), vol. 1, pp. 610–611.

# F R B 臺灣語假名

臺タイ<語ギイ/

タイ<ギイ/  
臺語

臺タイ<  
語ギイ/

No special actions need be taken, nor JavaScript be used, the HTML for the last example is just:

```
<p class=vert><ruby>臺<rt>タイ<</rt>語<rt>ギイ/</rt><</ruby></p>
```

## 3.6 On naming

Finally, I will explain my reasoning for why I am requesting the glyph names I am requesting. I have very deliberately chosen not to include a reference to Taiwan in either the names of the characters or the name of the requested block. In China, even the word Taiwan is politically sensitive, as is the period of Japanese rule. Further, Taiwanese kana write the Hokkien language, which is also spoken in mainland China. Indeed, the Quanzhou dialect, historically named the “Chinchew dialect”, gets its name from the city of Quanzhou (泉州市). Second of all, the Taiwanese kana were used occasionally in imperial Japan in the linguistic study of other Chinese languages, though the vast majority of surviving sources and modern use is in/for Hokkien.

The UTC should be aware of this issue when deciding on naming. My proposal originally asked the UTC to name the characters *Hokkien*. Eiso Chan, however, thought that *Min Nan* would be better, as this is the name used by ISO 639-3, and submitted the comment document L2/20-233. I strongly agree, and have revised the character names in my proposal accordingly.



### 3.7 On shakuhachi notation

The *shakuhachi* (尺八<sup>しゃくはち</sup>) is a traditional Japanese instrument, which looks similar to a flute. It has a traditional notation scheme known as *shakuhachi gakufu* (尺八楽譜<sup>しゃくはちがくふ</sup>).

As seen below,<sup>6</sup> in *shakuhachi*, the combining characters mentioned in this proposal do not combine the same way in vertical writing. They combine on the right, as is more common in CJK contexts.

情 熱 大 陸	の	尺 八	譜		
情熱大陸	の	尺八	譜		
情熱大陸	の	尺八	譜		
情熱大陸	の	尺八	譜		
情熱大陸	の	尺八	譜		

I agree with Eiso Chan's recommendations in L2/20-233. I am hopeless when it comes to musical subjects, so a more complete *shakuhachi* proposal will not be forthcoming from me, and it is out of the scope of this proposal. However, I immediately understand the wisdom of a registered OpenType feature such as `skhc`. As far as Unicode is concerned, Chan's recommended changes to `NameList.txt` and `UAX#50` are in my opinion sound.

6. Via Tenzan Yamada (山田典山): 情熱大陸の尺八譜.

## Chapter 4 Evidence of all requested characters

### 4.1 Tone letters

Âng and Ogawa (1992), vol. 1, p. 3.

			(2)	(1)	二、 符號。
鼻音	常音		八聲符號	出氣音符號	
ㄆ	ㄨ	上平			
ㄆ	ノ	上聲			
ㄑ	ノ	上去			
ㄆ	ノ	上入			
ㄨ	ノ	下平			
ㄆ	ノ	下去			
ㄨ	ノ	下入			

### 4.2 Combining characters

Âng and Ogawa (1992), vol. 1, p. 5.

ㄆ	ㄆ	ㄨ	ㄨ
ㄑ	ㄑ	ㄑ	ㄑ
ㄑ	ㄑ	ㄑ	ㄑ
ㄑ	ㄑ	ㄑ	ㄑ
ㄑ	ㄑ	ㄑ	ㄑ

Âng and Ogawa (1992), vol. 1, p. 3. Proof of ㄚ and ㄛ, missing from main table.

ㄛ「ツエ」の促りたる音、即ち(ㄛ)の音を表はす。

渣 這 祭 仄 齊 坐 絶

ㄚ「ツオ」の促りたる音、即ち(ㄚ)の音を表はす、此の母音に廣狹の二種

あり、「オ」列の假名の條参照。

租 祖 作 宗 槽 左 紙

ㄚ「テイ」の促りたる音、即ち(ㄚ)の音を表はす、

知 朝 稻 等 陣 中 直

ㄚ「トウ」の促りたる音、即ち(ㄚ)の音を表はす。

蛛 除 長 對 墩 唇 突

ㄚ唇「イ」舌「ウ」の位置にて發する音を表はす。

余 居 暑 鋸 除 慮 魚

ㄚ唇「エ」舌「ヲ」(狭き「オ」)の位置にて發する音を表はす。

銅 過 稅 火 飛 袋 尾

他の五十音假名及び符號假名の發音は、大體に於て、日本語と同一なりと雖も亦多少異なるものなきにあらず、今特に注意を要するものにつき左に説明すべし。

一、「ウ」列の假名。

(1) 「ウ」列の假名にして「クウ」「クウ」等の如き長音、又は「ウイ」

「クヌ」「クツ」の如く「イ」「ヌ」「ツ」の前に來るものは國語の「ウ」

より狭く唇を突出して發音する「ウ」の母音を有す。

有 龜 位 君 骨 恤 孫

### 4.3 Miscellaneous pages

Hirasawa (1914), p. 147.

<p>醜婦だとして舅姑に見えしなむ譯に行かない。  <b>將相本無種</b> <b>男兒當自強</b>      將相本種なし男兒當に自ら強むべし○史記の陳涉世家に壯士不死即已死即舉大名耳王侯將相寧有種乎。</p>	<p>男の残り者は有るが女の残り者は無い。  <b>存查</b> <b>無存查</b> <b>某</b>  <b>大目新娘</b> <b>尋不見灶</b>      目の大きい嫁が竈を見つけず○目の前にあるものを見附ることの出来ない時などに嘲る詞。</p>	<p>男の子が多ければ心配が多い。  <b>年冬好收</b> <b>查某人發嘴鬚</b>      豊年で忙しいと見えて女に鬚が生えてゐる○忙しくて女の頬に黒き物のつきたるなどにいふ。</p>
<p><b>多男</b> <b>則多懼</b>  <b>賢婦令夫貴</b> <b>惡婦令夫敗</b>      賢婦は夫をして貴からしめ惡婦は夫をして敗れ</p>	<p>悪い油は滓多く悪い女は涙多し。  <b>歹油厚滓</b> <b>歹查某厚目屎</b>  <b>歹查哺厚同年</b> <b>歹查某厚姑姨</b>      悪い男は同年多く悪い女に伯母さん多し○正しからぬ人は互に同類を作つて惡事をする意。</p>	<p>賢婦は夫をして貴からしめ惡婦は夫をして敗れ</p>

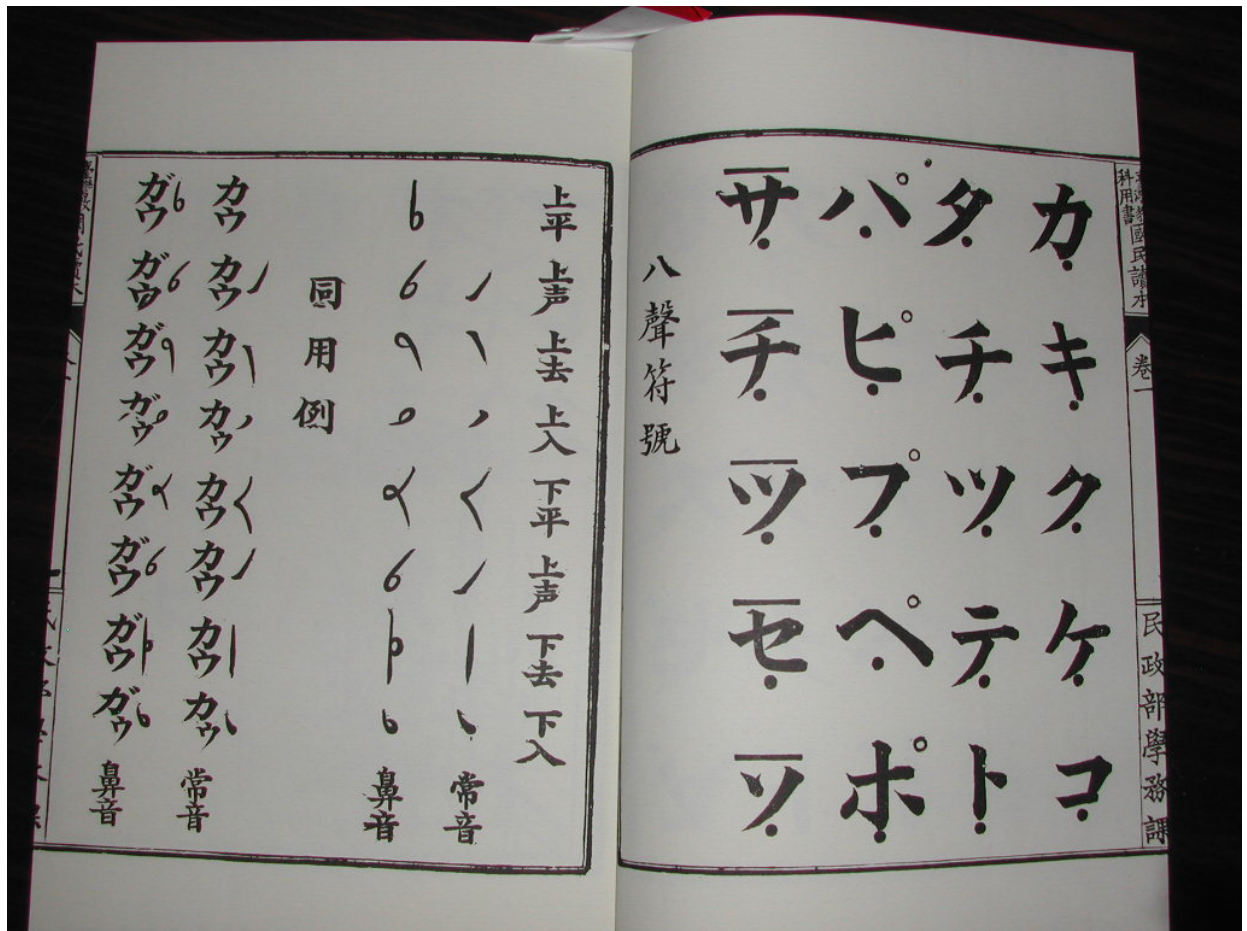
第五篇 人倫 三男女

一四七

Ogawa (1938). Unknown page, from Liong, et al (1999). “*The Japanese–Taiwanese Dictionary and The New Japanese–Taiwanese Dictionary : a Comparison*”. p. 16.

オチル(おつ)(自)落。 加落。加落落。葉ガ  
 落葉。日ガ山ニ。日落山。雷ガ  
 起雷。鷲カラ。跋落馬。試験ニ  
 脱名。落第。垢ガ。脱垢。肉ガ。肉  
 消去。力ガ。失望。餒志。瘡ガ。乞食  
 症脱去。色ガ。退色。鼻ガ。凹鼻  
 字ガ。字落去。脱字。評判ガ。敗名  
 聲。城カラ。對城逃走。計ニ  
 中計。城ガ。城陥落。品ガ。此  
 號較輪。較次。札ガ。標着。胸ニ  
 能曉。了解。罪ニ。受罪。此ノ  
 理ニ。歸此號道理。

*Textbook for Taiwan* [台湾教科用書] (1902).



The above book is another corroborating source for the equality of tone two (上声) and tone six (上声) in Hokkien, as well.



先生が本を読んでいます。生徒を、静かに、聞いています。

先生が講義し、學生が静かに静かに聴く。

## Chapter 5 Bibliography

- *Textbook for Taiwan* [台湾教科用書] (1902).<sup>1</sup> Volume 1 [国民読本卷一]. Governor-General of Taiwan, Imperial Government of Japan.
- Hirasawa, Teiakira (1914). *Taiwan Proverb Collection* [臺灣俚諺集覽]. Governor-General of Taiwan, Imperial Government of Japan.
- Ogawa, Naoyoshi (1932). *Comprehensive Taiwanese–Japanese Dictionary* [臺日大辭典]. Governor-General of Taiwan, Imperial Government of Japan.
  - Volume 1 (1931)
  - Volume 2 (1932)
- Ogawa, Naoyoshi (1938). *New Japanese–Taiwanese Dictionary* [新訂日台大辭典]. (One volume only, second never completed.)
- Âng, Ūi-jîn; Ogawa, Naoyoshi (1992). *Minnan Classic Dictionary Collection* (閩南語經典辭書彙編)
- *Tâi-lô–Taiwanese kana tables* (臺羅調符—臺語假名調符) (2017). Table 0/18. Taigikho (失控的台語課).

1. Meiji 35 (明治35年)



**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title: A proposal to encode Taiwanese kana in the UCS

2. Requester's name: Fredrick R. Brennan

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 23 September 2020 (rev. 13 December 2020)

5. Requester's reference (if applicable): N/A

6. Choose one of the following:

This is a complete proposal:

(or) More information will be provided later:

**B. Technical – General**

1. Choose one of the following:

a. This proposal is for a new script (set of characters):

Proposed name of script: Kana Extended-B

b. The proposal is for addition of character(s) to an existing block:

Name of the existing block: \_\_\_\_\_

2. Number of characters in proposal: 13

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary  B.1-Specialized (small collection)  B.2-Specialized (large collection)

C-Major extinct  D-Attested extinct  E-Minor extinct

F-Archaic Hieroglyphic or Ideographic  G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?

a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

b. Are the character shapes attached in a legible form suitable for review?

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Fredrick R. Brennan

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Fredrick R. Brennan <copypaste@kittens.ph>. Font name is FRB Taiwanese Kana. It's SIL OFL licensed.

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

I discuss the order characters should be in, and issues presented by vertical typesetting.

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database ( <http://www.unicode.org/reports/tr44/> ) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup>? Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	Yes <i>Mojikyo Institute, Academia Sinica, Eiso Chan</i> <i>Correspondence</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>Japanese learners of Hokkien, Taiwanese scholars of Japanese imperial era</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Twitter, Facebook, etc.</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	Yes <i>Existing combining characters are used. See page 7.</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No