

A PROPOSAL TO ENCODE TAIWANESE KANA (臺^{TĀI}-灣^{OĀN}-語^{GÍ}-假^{KHÁ}-名^{NAH}) IN THE UCS

Fredrick R. Brennan



18 August 2020

古家^{ふるや}時雄^{ときお}を追悼^{もじきょう}して
文字鏡研究会に心から感謝します

ACKNOWLEDGEMENTS

in no particular order...

たに もと さち ひろ
Sachihiro Tanimoto (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.

み うら だいすけ
Daisuke Miura (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⁽⁶⁾.

やま ぐち りゅう せい
Ryūsei Yamaguchi (山口 隆成)

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

Wil Lee (Lí Kho-Lūn, 李 /
科 /
潤 /), Patreon

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

こ ばやしけん
Ken Lunde (小林 劍), Unicode Consortium

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

やま ざき いっ せい
Issei Yamazaki (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.

さい もん
Simon Cozens (カズンズン才文) & Caleb Maclennan, SJLE

For their help with SJLE, an experimental typesetting engine and alternative to L^AT_EX.

Table of Contents

Quick summary	4
Tone marks	4
Nasalized tone marks	5
Combining characters	6
Overview	7
Linguistic	7
Historic (Taiwanese..kana?)	7
Legacy encodings	10
Pseudo-Unicode encodings	10
Mojikyō	10
Unicode considerations	12
UnicodeData.txt	12
On tone six	12
On the aspiration mark	13
On glyph order	13
Tone letters	13
Combining characters	13
Vertical typesetting	14
Evidence of all requested characters	16
Tone letters	16
Combining characters	16
Miscellaneous pages	18
Bibliography	22

Chapter 1

Quick summary

I propose a new block, in the Supplementary Multilingual Plane, to be called Katakana Phonetic Extensions Extended-A. I propose that the new block be given range U+1BA00 – U+1BA0F, and is to include:

1.1 Tone marks

/ MODIFIER LETTER KATAKANA TONE TWO

Mojikyō 69622; tentative encoding U+1BA01

- In Hokkien, also used for tone six¹

\ MODIFIER LETTER KATAKANA TONE THREE

Mojikyō 69623; tentative encoding U+1BA02

• MODIFIER LETTER KATAKANA TONE FOUR

Mojikyō 69624; tentative encoding U+1BA03

< MODIFIER LETTER KATAKANA TONE FIVE

Mojikyō 69625; tentative encoding U+1BA04

| MODIFIER LETTER KATAKANA TONE SEVEN

Mojikyō 69627; tentative encoding U+1BA06

• MODIFIER LETTER KATAKANA TONE EIGHT

Mojikyō 69628; tentative encoding U+1BA07

1. For more discussion of this issue, and why I'm proposing we leave U+1B306 and U+1B316 open, please see § On tone six.

1.2 Nasalized tone marks

⠁ MODIFIER LETTER KATAKANA NASALIZED TONE ONE
Mojikyō 69629; tentative encoding U+1BA08

⠃ MODIFIER LETTER KATAKANA NASALIZED TONE TWO
Mojikyō 69630; tentative encoding U+1BA09
• In Hokkien, also used for tone six¹

⠄ MODIFIER LETTER KATAKANA NASALIZED TONE THREE
Mojikyō 69631; tentative encoding U+1BA0A

⠅ MODIFIER LETTER KATAKANA NASALIZED TONE FOUR
Mojikyō 69632; tentative encoding U+1BA0B

⠆ MODIFIER LETTER KATAKANA NASALIZED TONE FIVE
Mojikyō 69633; tentative encoding U+1BA0C

⠈ MODIFIER LETTER KATAKANA NASALIZED TONE SEVEN
Mojikyō 69635; tentative encoding U+1BA0E

⠉ MODIFIER LETTER KATAKANA NASALIZED TONE EIGHT
Mojikyō 69636; tentative encoding U+1BA0F

1. For more discussion of this issue, and why I'm proposing we leave U+1B306 and U+1B316 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 (KATAKANA) DOT BELOW

Mojikyō 69202; tentative encoding U+0323

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

◎ COMBINING (KATAKANA) OVERLINE

Mojikyō 69637–69642;² tentative encoding U+0305

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

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: サチツセツヲ. They are missing ク and ツ.

Chapter 2

Overview

2.1 Linguistic

Taiwanese Hokkien (臺灣語) is a Sinitic Southern Min language spoken throughout the world; as of 2013, it was spoken by 48 million people.¹ It is spoken as far afield as the Philippines, where it is known as *Fookien*.

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。Tí 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。”

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 *Tāng-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 hû-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:

“ ム̄是講攏無台灣來工〈學生來修。矛|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]. Tiānxià 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-hōaⁿ, Lí [李勤岸] (2007). 台語飯桌. Sìn 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.³

The primary person involved in their creation was the linguist Naoyoshi Ogawa [小川尚義].⁴ The first of Ogawa's works that they make a notable appearance in is the *Japanese–Taiwanese Daijiten* (日台大辭典) of 1907.⁵ 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 Úi-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] (1st ed. 2011),⁶ Japanese learners of Hokkien continue to use Taiwanese kana even today.⁷ 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 Soaⁿ-kiā (@suannkia) and Taipa (@taipalogy).

In our time, besides being used as a substantial basis for the *Dictionary of Frequently-Used Taiwan Minnan*, Âng Úi-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. 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):

	バ
	ン
託す。	夢。
一	●
出現	完ヌ(姓)
二	夢。
夢枕	夢。
に立つ。	夢。
○	●
夢枕に立つ。	夢の告。
○	●
夢枕に立つ。	悪オ。
○	●
夢枕に立つ。	恐。
○	●
圓ヌ	夢。
一	夢。
二	夢。
夢判斷。	タバ。
一	●
夢判斷。	悪。

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+1BA01–1BA0F 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, ጀ. 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

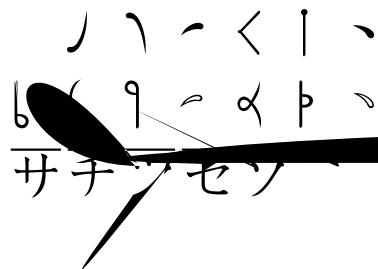
Mojikyō almost fully supports Taiwanese kana. Its only deficiency is that it is missing 𩁈 and 𩁉, which means that the Quanzhou dialect is not reproducible in the Mojikyō encoding.

In Mojikyō, the Taiwanese kana are available in the font `Mojikm0D.TTF` ; 20 codepoints are devoted to them. Mojikyō chooses to encode the COMBINING KATAKANA OVERLINE as precombined with the kana it combines with, and it encodes the COMBINING KATAKANA 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:



Chapter 3

Unicode considerations

3.1 UnicodeData.txt

```
1BA01;MODIFIER LETTER KATAKANA TONE TWO;Lm;0;L;;;;N;;;;;
1BA02;MODIFIER LETTER KATAKANA TONE THREE;Lm;0;L;;;;N;;;;;
1BA03;MODIFIER LETTER KATAKANA TONE FOUR;Lm;0;L;;;;N;;;;;
1BA04;MODIFIER LETTER KATAKANA TONE FIVE;Lm;0;L;;;;N;;;;;
1BA06;MODIFIER LETTER KATAKANA TONE SEVEN;Lm;0;L;;;;N;;;;;
1BA07;MODIFIER LETTER KATAKANA TONE EIGHT;Lm;0;L;;;;N;;;;;
1BA08;MODIFIER LETTER KATAKANA NASALIZED TONE ONE;Lm;0;L;;;;N;;;;;
1BA09;MODIFIER LETTER KATAKANA NASALIZED TONE TWO;Lm;0;L;;;;N;;;;;
1BA0A;MODIFIER LETTER KATAKANA NASALIZED TONE THREE;Lm;0;L;;;;N;;;;;
1BA0B;MODIFIER LETTER KATAKANA NASALIZED TONE FOUR;Lm;0;L;;;;N;;;;;
1BA0C;MODIFIER LETTER KATAKANA NASALIZED TONE FIVE;Lm;0;L;;;;N;;;;;
1BA0E;MODIFIER LETTER KATAKANA NASALIZED TONE SEVEN;Lm;0;L;;;;N;;;;;
1BA0F;MODIFIER LETTER KATAKANA NASALIZED TONE EIGHT;Lm;0;L;;;;N;;;;;
```

3.2 On tone six

In modern Hokkien, tone six is equal to tone two.¹ 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	ó	ō	óh
臺語假名調符	常音	/	\	•	<	/		•
	鼻音	ɓ	ɗ	ɳ	ɳ̥	ɿ	ɸ	ɳ̥

(第七元音·輔音轉音)

✓ 後表標註「o」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,² and in other sources, it appears identically to the *nakaguro* (中黒).³

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

/ < 6 < \ < ɳ < • < ɳ̥ < < ɳ̥̥ < | < ɸ < ɸ̥ < ɸ̥̥

3.4.2 Combining characters

夕 < 夕⁴

2. Ogawa (1932), Âng and Ogawa (1992), Taigikho (2017), among others.

3. *State Monopoly Bureau's Taiwanese Dictionary* [專賣局台灣語典] (1923), among others.

4. Âng and Ogawa (1992), vol. 2, p. 2.

サ < タ < ザ < ソ⁵

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.

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

```
1BA01..1BA0F ; U # Lc [13] MODIFIER LETTER KATAKANA TONE
TWO..MODIFIER LETTER KATAKANA NASALIZED TONE EIGHT
```

For this paper, I am using a S_IL^E 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. Âng and Ogawa (1992), vol. 1, pp. 610–611.

臺タイ〈語ギイ／

臺 語
タイ(ギイ)

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>
```

Chapter 4

Evidence of all requested characters

4.1 Tone letters

Âng and Ogawa (1992), vol. 1, p. 3.

		二	
		(2)	(1)
鼻音	常音	八聲	出氣音符號
l	०	上平	聲符號
l	/	上聲	號
ı	ı	上去	
º	º	上入	
d	⟨	下平	
þ		下去	
º	·	下入	

4.2 Combining characters

Âng and Ogawa (1992), vol. 1, p. 5.

ヰ	ປ	ດ	ກ
チ	ປ	ច	ຂ
ツ	ປ	ຜ	ຂ
セ	ປ	ຕ	ເ
ソ	ປ	ຕ	ໂ

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

モ「ヅエ」の促りたる音、即ち(モ)の音を表はす。

渣^モ/這^モ/祭^モ/仄^モ/齊^モ/坐^モ/絕^モ

ヲ「ヅオ」の促りたる音、即ち(モ)の音を表はす、此の母音に廣狹の一種あり、(「オ」列の假名の條参照)。

租^ヲ/祖^ヲ/作^ヲ/宗^ヲ/憎^ヲ/左^ヲ/紙^ヲ

チ「ティ」の促りたる音、即ち(チ)の音を表はす、

知^チ/朝^チ/稻^チ/等^チ/陣^チ/中^チ/直^チ

ツ「トウ」の促りたる音、即ち(ツ)の音を表はす。

蛛^ツ/除^ツ/長^ツ/對^ツ/墩^ツ/唇^ツ/突^ツ

ぢ唇「イ」舌「ウ」の位置にて發する音を表はす。

余^ヂ/居^ヂ/暑^ヂ/鋸^ヂ/除^ヂ/慮^ヂ/魚^ヂ

オ唇「エ」舌「ヲ」(狭き「オ」)の位置にて發する音を表はす。

鍋^ヲ/過^ヲ/稅^ヲ/火^ヲ/飛^ヲ/袋^ヲ/尾^ヲ

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。

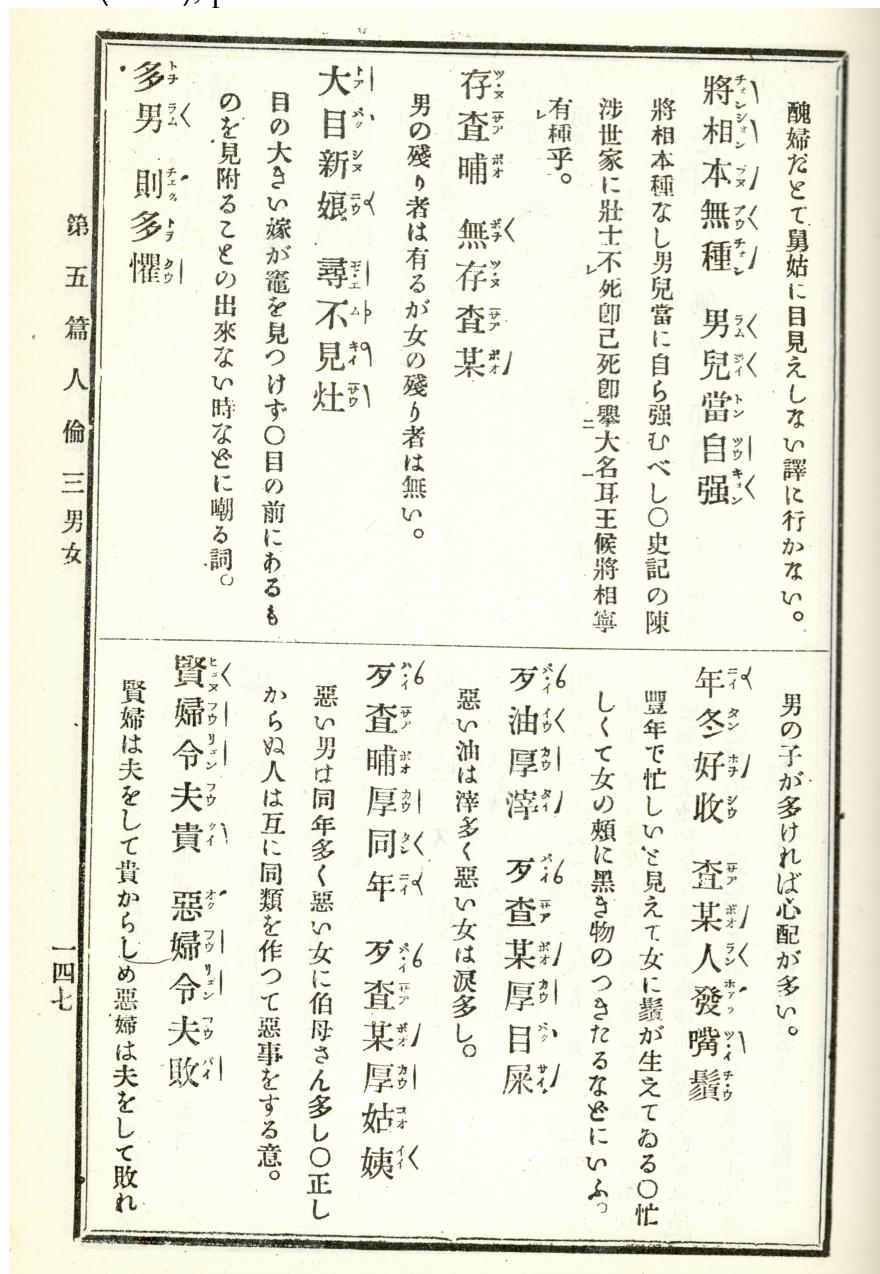
一、「ウ」列の假名。

(1) 「ウ」列の假名にして「ウウ」「クウ」等の如き長音、又は「ウイ」「クヌ」「クツ」の如く「イ」「ヌ」「ツ」の前に来るものは國語の「ウ」より強く唇を突出して發音する「ウ」の母音を有す。

有^ウ/多^ウ/位^ウ/君^ウ/骨^ウ/恤^ウ/孫^ウ

4.3 Miscellaneous pages

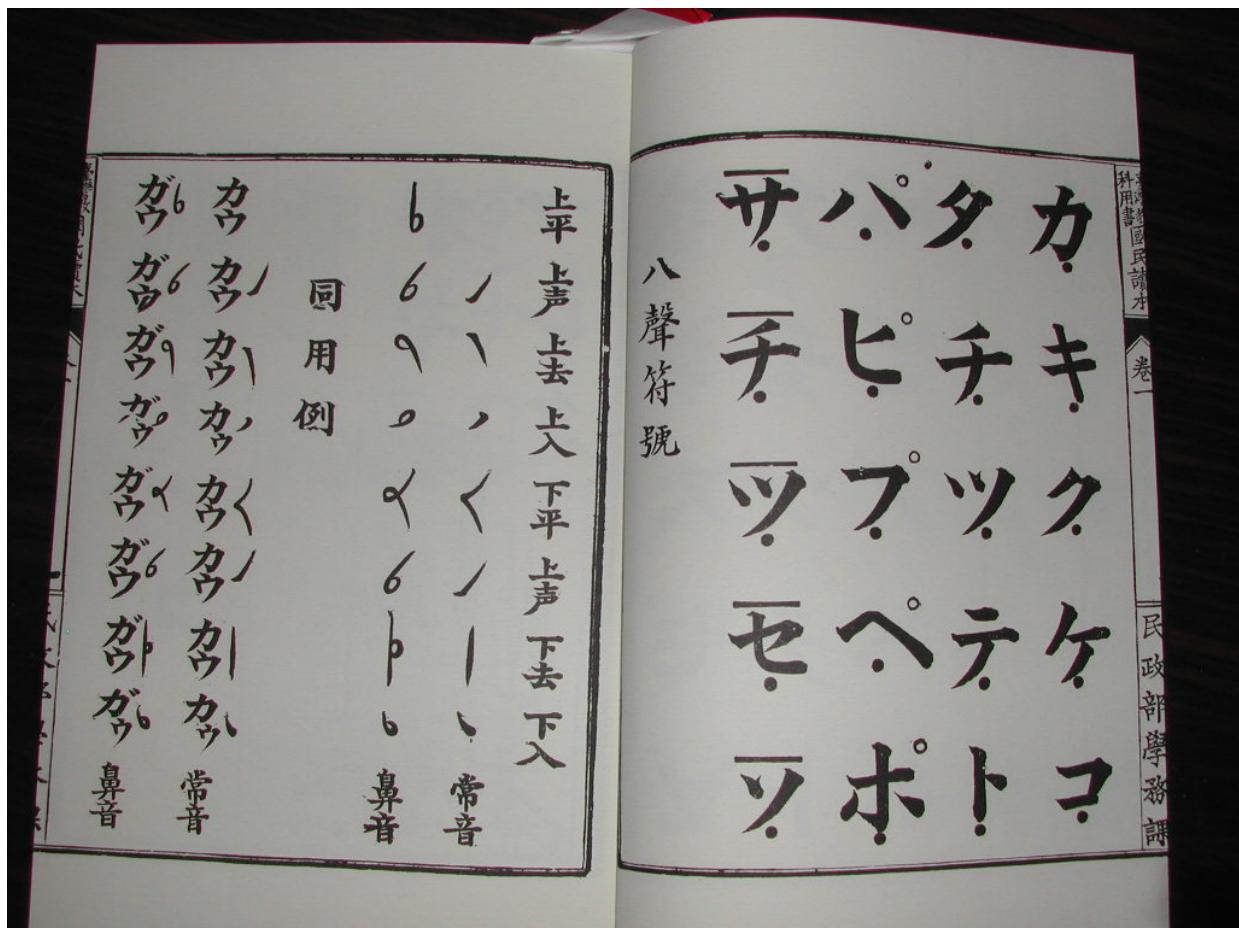
Hirasawa (1914), p. 147.



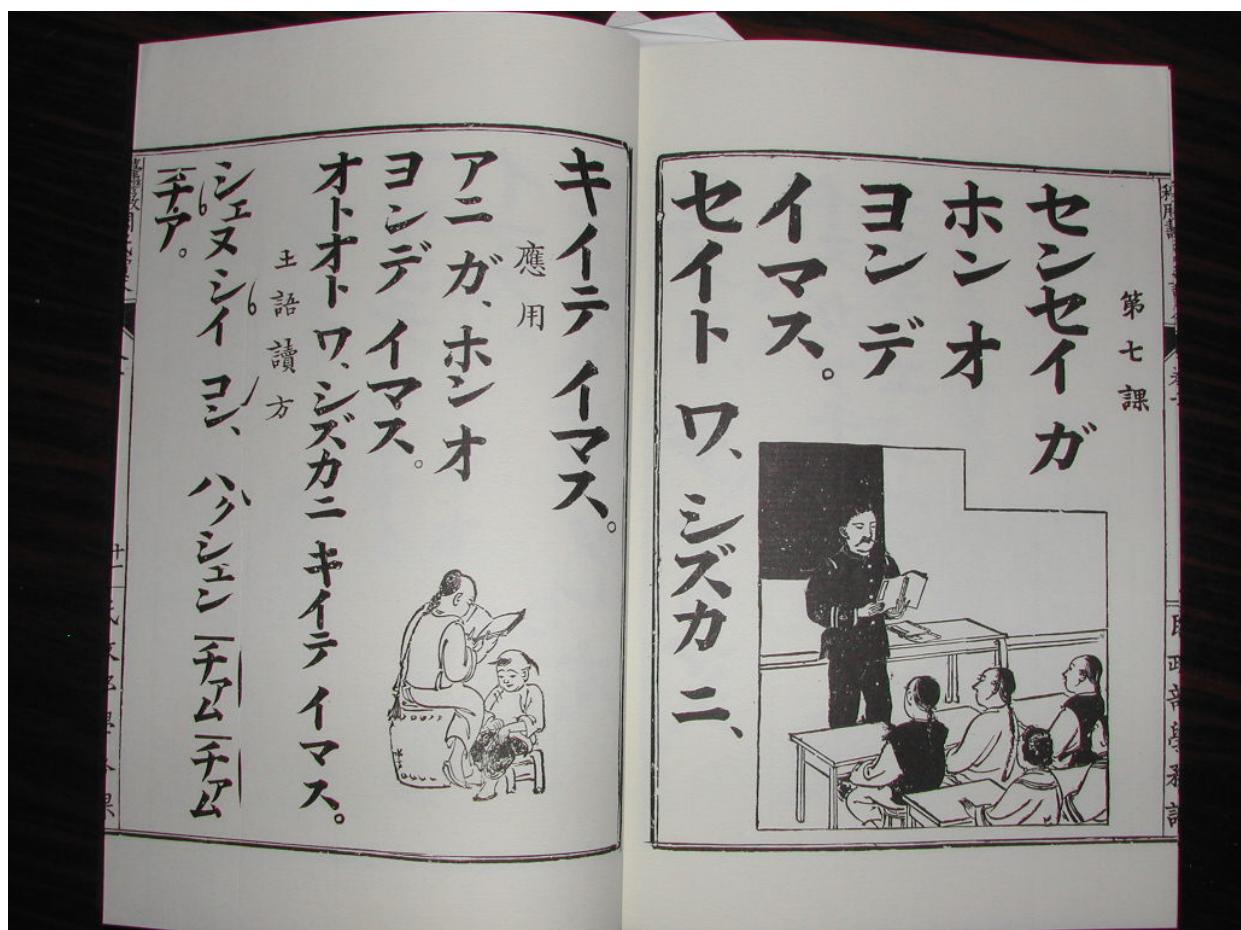
Ogawa (1938). Unknown page, from Lioung, et al (1999). “*The Japanese–Taiwanese Dictionary* and *The New Japanese–Taiwanese Dictionary*: a Comparison”. p. 16.

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。「馬カラ」跋落馬。 ○試験ニ
脱名落第。 ○「垢ガ」脱垢。「肉ガ」肉
消去。「力ガ」失望。 餅志。「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」城陷落。 ○「品ガ」此
號較輸。 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 [台灣教科用書] (1902).



The above book is another corroborating source for the equality of tone two (下声) and tone six (下声) 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。

先生シシテ生シジメテ講シヨウシ、 學シラフ生シジメテ恬チヤウ恬チヤウムチヤウム。

Chapter 5 Bibliography

- *Textbook for Taiwan* [台灣教科用書] (1902).¹ 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, Úi-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 10646.¹

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:	<i>A proposal to encode Taiwanese Kana in the UCS</i>	
2. Requester's name:	<i>Fredrick R. Brennan</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>17 August 2020</i>	
5. Requester's reference (if applicable):	<i>N/A</i>	
6. Choose one of the following:		
This is a complete proposal:	<input checked="" type="checkbox"/>	
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/>	
Proposed name of script:	<i>Katakana Phonetic Extensions Extension A</i>	
b. The proposal is for addition of character(s) to an existing block:	<input type="checkbox"/>	
Name of the existing block:	<i>15</i>	
2. Number of characters in proposal:	<i>15</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/>	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Fredrick R. Brennan</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Fredrick R. Brennan <copypaste@kittens.ph>. Font name is FRB Taiwanese Kana. It's SIL OFL licensed.</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/>	
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)?	<input checked="" type="checkbox"/>	
	<i>I discuss the order characters should be in, and issues presented by vertical typesetting.</i>	
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.

¹? 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?	<input type="checkbox"/> No
If YES explain	_____
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.)?	<input type="checkbox"/> Yes
If YES, with whom?	_____ <i>Mojikyo Institute</i>
If YES, available relevant documents:	_____ <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?	<input type="checkbox"/> Yes
Reference:	_____ <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)	<input type="checkbox"/> Rare
Reference:	_____
5. Are the proposed characters in current use by the user community?	<input type="checkbox"/> Yes
If YES, where? Reference:	_____ <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?	<input type="checkbox"/> No
If YES, is a rationale provided?	_____
If YES, reference:	_____
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<input type="checkbox"/> Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<input type="checkbox"/> No
If YES, is a rationale for its inclusion provided?	_____
If YES, reference:	_____
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<input type="checkbox"/> No
If YES, is a rationale for its inclusion provided?	_____
If YES, reference:	_____
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?	<input type="checkbox"/> Yes
If YES, is a rationale for its inclusion provided?	<input type="checkbox"/> Yes
If YES, reference:	_____ <i>The combiners are confusable with Latin ones, but kana is very different from Latin and new characters are needed.</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	<input type="checkbox"/> Yes
If YES, is a rationale for such use provided?	<input type="checkbox"/> Yes
If YES, reference:	_____ <i>See section "Evidence of all requested characters"</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	<input type="checkbox"/> Yes
If YES, reference:	_____ <i>See section "Evidence of all requested characters"</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<input type="checkbox"/> No
If YES, describe in detail (include attachment if necessary)	_____ _____
13. Does the proposal contain any Ideographic compatibility characters?	<input type="checkbox"/> No
If YES, are the equivalent corresponding unified ideographic characters identified?	_____
If YES, reference:	_____