Source:	CheonHyeong Sim (沈天珩)
Title:	Proposal to Encode One Small Kana Letter
Date:	2025−01−15
Action:	To be considered by CJK & Unihan WG and UTC

Proposed Character and Properties

One small Katakana letter is proposed:

Codepoint	Character Name	Representative Glyph
1B168	KATAKANA LETTER SMALL ARCHAIC YE	ㇱ

General Category and Other Properties

1B168;KATAKANA LETTER SMALL ARCHAIC YE;Lo;0;L;;;;;N;;;;;

Line Break Properties

1B168   ; CJ # Lo   KATAKANA LETTER SMALL ARCHAIC YE

Vertical Orientation

1B168   ; Tu[1] # Lo   KATAKANA LETTER SMALL ARCHAIC YE

Annotations for CodeCharts

1B168 ㇱ KATAKANA LETTER SMALL ARCHAIC YE
· used for Marshallese transliteration
→1B121 エ KATAKANA LETTER ARCHAIC YE

Proposed Codepoint

My early document in 2016 to propose to encode some small Kanas[2] attracted a proposal for more small Kanas from Ryusei Yamaguchi[3] and some considerations from Japan NB[4]. After the discussion, the conclusion is that, dozens of codepoints in the Small Kana Extension block (1B130..1B16F) are reserved for possible future small Kanas[5] as shown in the table below (grey cells indicate tentatively reserved codepoints):

Hiragana block: あいうえおつやゆよわかけ (codepoints omitted)																	
Katakana block: アイウエオツヤユヨワカケ (codepoints omitted)																	
U+	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
31Fx	ク	シ	ス	ト	ヌ	ハ	ヒ	フ	ヘ	ホ	ム	ラ	リ	ル	レ	ロ
1B13x	き	く	こ	さ	し	す	せ	そ	た	ち	て	と	な	に	ぬ	ね

1

1B14x	の	は	ひ	ふ	へ	ほ	ま	み	む	め	も	ら	り	る	れ	ろ
1B15x	ゐ	ゑ	を	ん	キ	コ	サ	セ	ソ	タ	チ	テ	ナ	ニ	ネ	ノ
1B16x	マ	ミ	メ	モ	ヰ	エ	ヲ	ン								

Thus, the next available codepoint would be U+1B168.

Attestations

The proposed small Kana letter could be found in at least the following two ancient books: ミクロネシア語の綜合研究 and マーシャル語の研究, both from the Shōwa（昭和）era. Let us see the former one first:
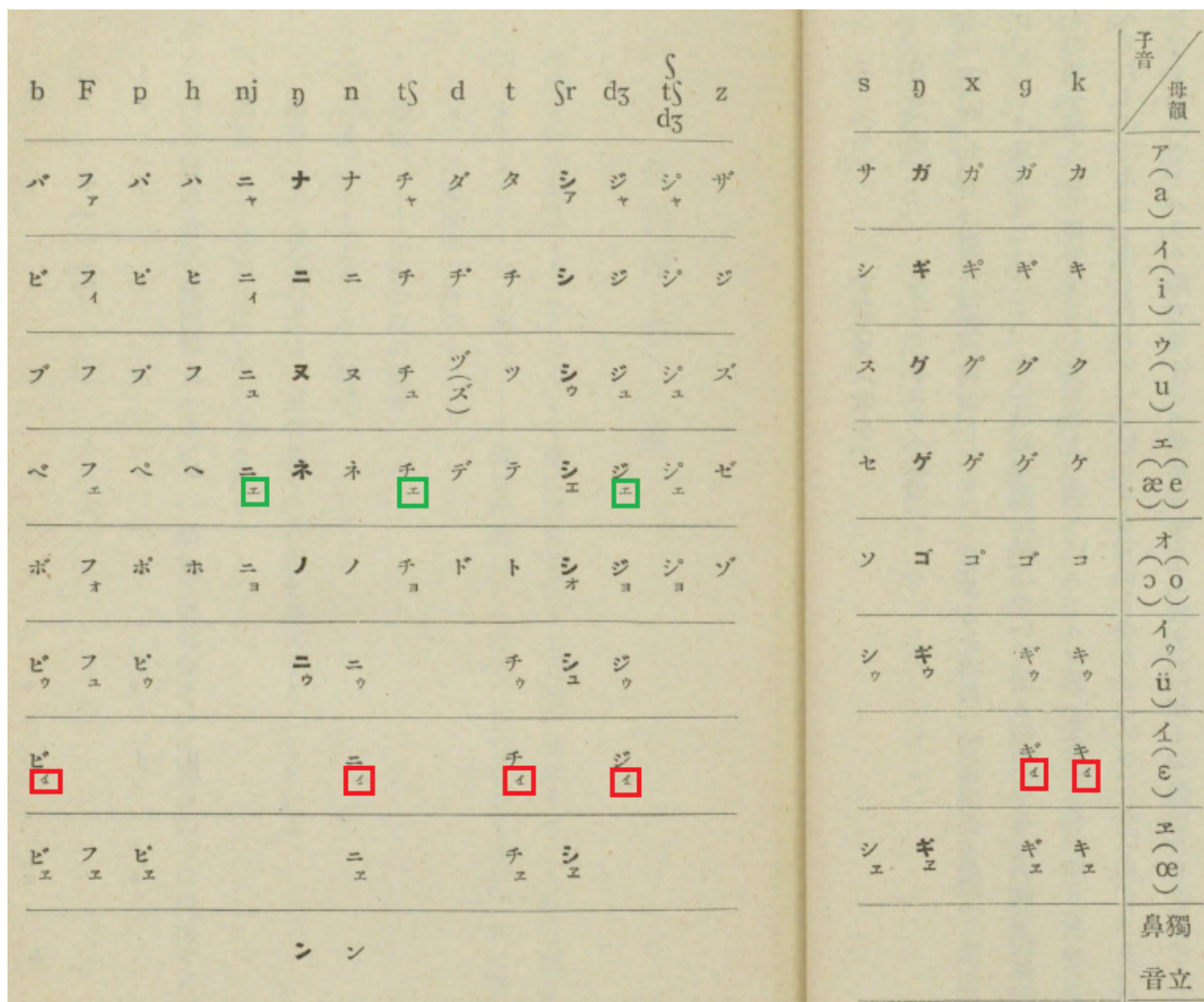
Fig.1 ミクロネシア語の綜合研究 P24−25

In Fig.1, we could see some small ㄈs marked in red; some small エs are marked in green for contrast − they represent for different pronunciations. However, in the main body of the same book, no small ㄈs could be found but only small エs − even if only キㄈ is listed in the table in Fig.1 but キエ is not, the main body still prefers キエ:

Fig.2−3　ミクロネシア語の綜合研究　P75, P140

What is more, this book only explains vocabulary and grammar, without providing paragraphs. So let us see the latter one. Also, a similar table is provided at the front of the book:
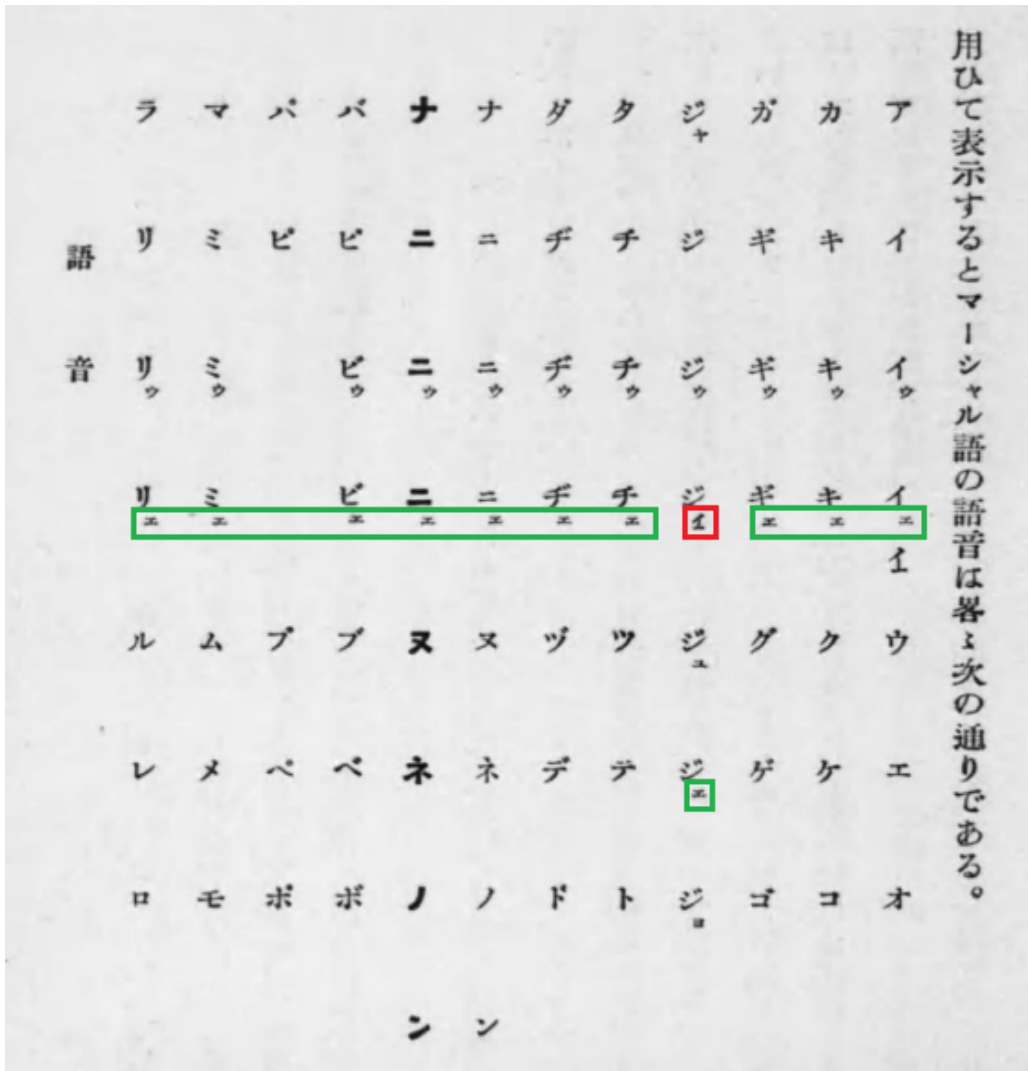
用ひて表示するとマーシャル語の語音は畧ゝ次の通りである。

Fig.4 マーシャル語の研究 P7

Here, the author uses the small エ as much as possible — but still, ジェ exists, to represent something different from ジェ in Marshallese.

After comparing the transliterated words in the main body of the book with the modern Marshallese orthography and phonetic system[6], we could get a rough correspondence for vowels as the following table:

Height	Unrounded		Rounded
	Front	Back	
Close	イ /i/	イゥ /ɯ/	ウ /u/
Mid	エ /e/~/ɛ/	イェ(エ) /ʌ/	オ /o/~/ɔ/
Open		ア /a/	

Note that, we could hardly get the exact correspondence since so many dialects in the Marshall Islands vary from each other. What is more, the book was published in Shōwa 4 (1929), which is almost a hundred years ago, the language itself may also change. Also, the transliteration is not that precise, e.g. it does not distinguish /l/ from /r/, so may be the vowels.
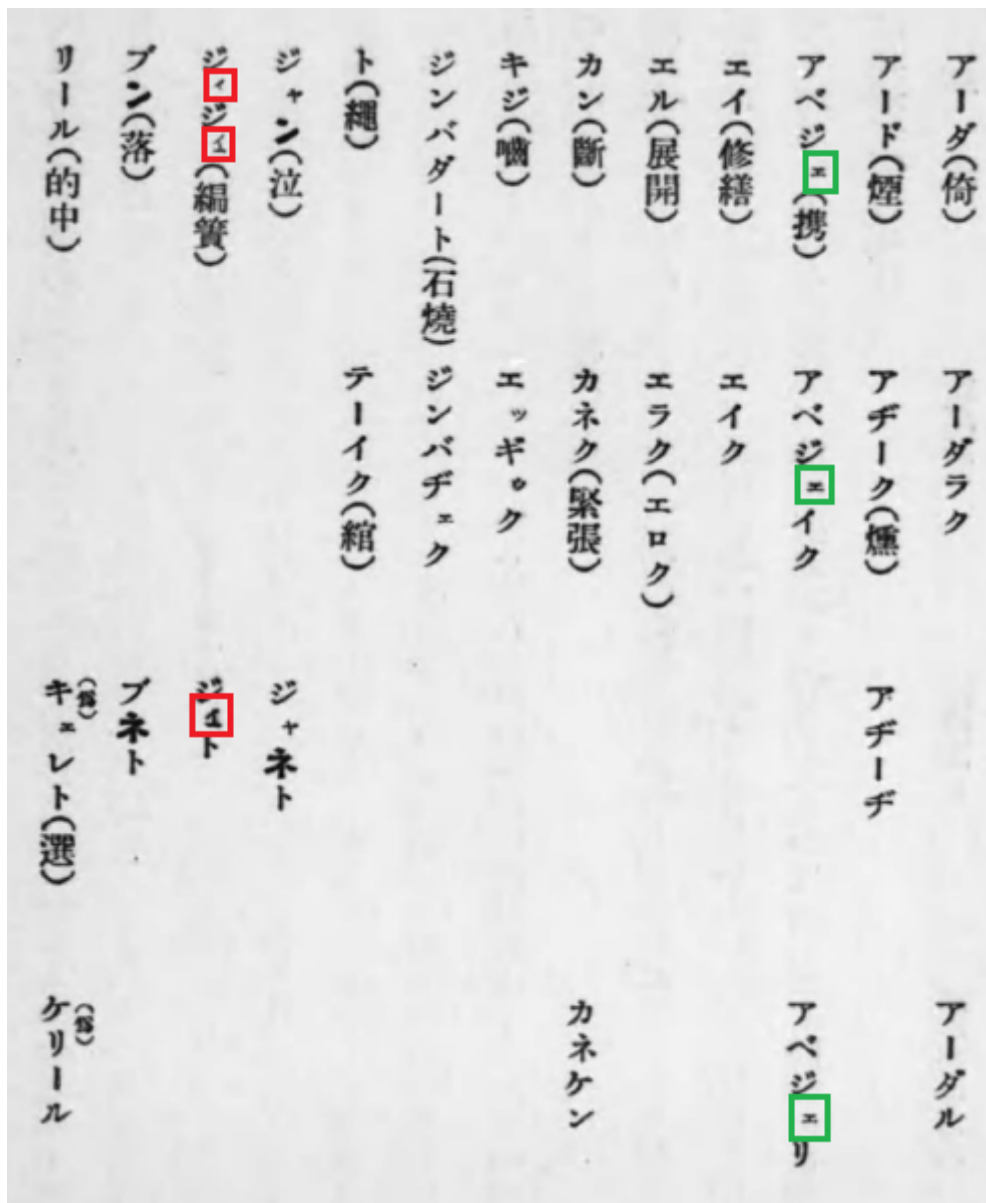
アーダ（倚）　アーダラク　　　　　　　　　　　アーダル
アード（煙）　アヂーク（燻）　　　　　　　　　アヂーヂ
アベジェ（携）　アベジェイク　　　　　　　　　アベジェリ
エイ（修繕）　エイク
エル（展開）　エラク（エロク）
カン（断）　カネク（緊張）　　　　　　　　　　カネケン
キジ（嚙）　エッギョク
ジンバダート（石燒）　ジンバヂェク
ト（繩）　テーイク（縮）　　テーイク（緒）
ジャン（泣）　ジャネト　　　　ジャネト
ジェ／ジェ（編簀）　　　　　　ジェト　　ブネト
プン（落）　　　　　　　　　　　　　　キュレト（選）
リール（的中）　　　　　　　　　　　　ケリール

Fig.5 マーシャル語の研究　P86

In Fig.5, a part of the main body explaining the grammar, both ジェ and ジェ appears, the former one is marked in red, while the latter one is marked in green.

Different from the previous book, this book provides so many meaningful paragraphs as valuable running texts. Let us see one piece which contains the proposed character:

エオ エ・リット。ジュオン ノー イム ルグチ。—ラク イ ペリジェト、エ バル アイ。ナジ テバレ エ・バル ブン ジュオン ノー、バル ルグチ。エダル エダル――ラク イジョ エ アル ワロク。エ ワロク、エジ ブイジ ミジ。エ カヌジ キジェロク。エ・ムイジ エ ワネネロク（ワン・エーネ・ロク）イム ジシェト、ブエ エ ナニン ミジ ウォト ジリク（今少しで死ぬ）。ケ エ・ムイジ アーン カッキゥジェ、エジ バ ナン エオ エ・リク「クオン アイ ブエ イ キジェロク」。インネム エジ アイ エラク イープ エーブ、エ ヅロク イム レビシ ラル。

だ。彼が將に達せんとする時、一つの波が起つて彼を拋り出した。押流されて岩礁に遲りついた時彼は再泳いだ。彼が將に達せんとするや、再び波が落ちかゝり再抛り出された。押流されて再彼處に現はれた其時には彼は甚疲れて殆ど死にさうであつた。乃で岩礁に上り、息も絶々に坐つた。休息した後彼は弟に對つて「私は疲れたから汝が泳がねばならぬ」といふた。乃で彼は泳いだ。波が打上げる時には彼は水に潜り、固く底を捉へ、其が去る

Fig.6 マーシャル語の研究 P168

エ・ムイジ エ ワネネロク イム ジジェト、ブエ エ ナニン ミジ ウォト ジリク。乃で岩礁に上り、息も絶々に坐つた。

If we write this sentence in the modern Marshallese orthography, it will be:

Eṃōj e wāānāneḷọk im jijet, bwe e nañin mij wōt jidik.

It seems that the vowel *e* in "jijet" is more likely to be /e/ rather than /ʌ/, so we would better clarify that the word ジジェト is really not a misprint. Fortunately, we could find the same word **in the main body** (grammar explanation) and **at the back** (glossary) of this book, as shown in Fig.7 and Fig.8 perspectively.

（ロ）第三形ナ、クェはとり別けて重くいふに用ひ、國語に於て「我ぞ」「汝

るぞ」の如く指定する場合の「我」「汝」に相當する。されば之を第一形の

こともある。例

ナ（我）ウ（嗟）ォトー　私のみ

ナイ（我）ナーナ（汝）ケ　私は悪いか

クェ（汝）コ（我）・アル　キ（汝）ムマネ　汝こそなしたれ（汝がしたのだ）

（八）第二段の形は音便の爲にジを添付したもので、主格としてのみ用ひられ

に用ひられることがあり、兩者の中いづれを用ひべきかは聽覺によつて決

が、次の點に注意することを要する。

一、ジャ行を以て始まる語の前には第二形を用ひざることを可とする。例

エジ（我）ジジェ下

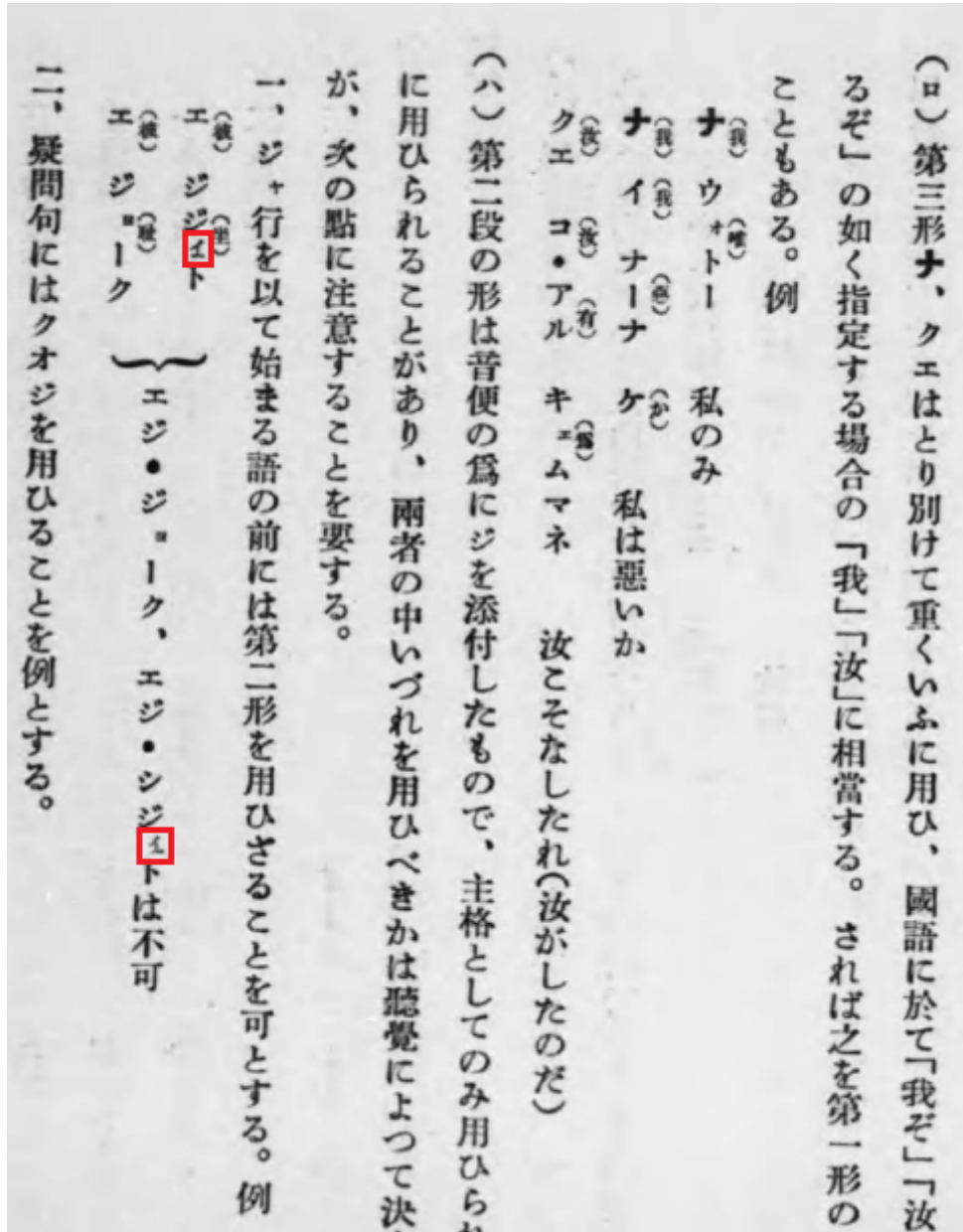エジ（被）ジ（疑）ーク

エジ・ジョーク、エジ・シジェ下は不可

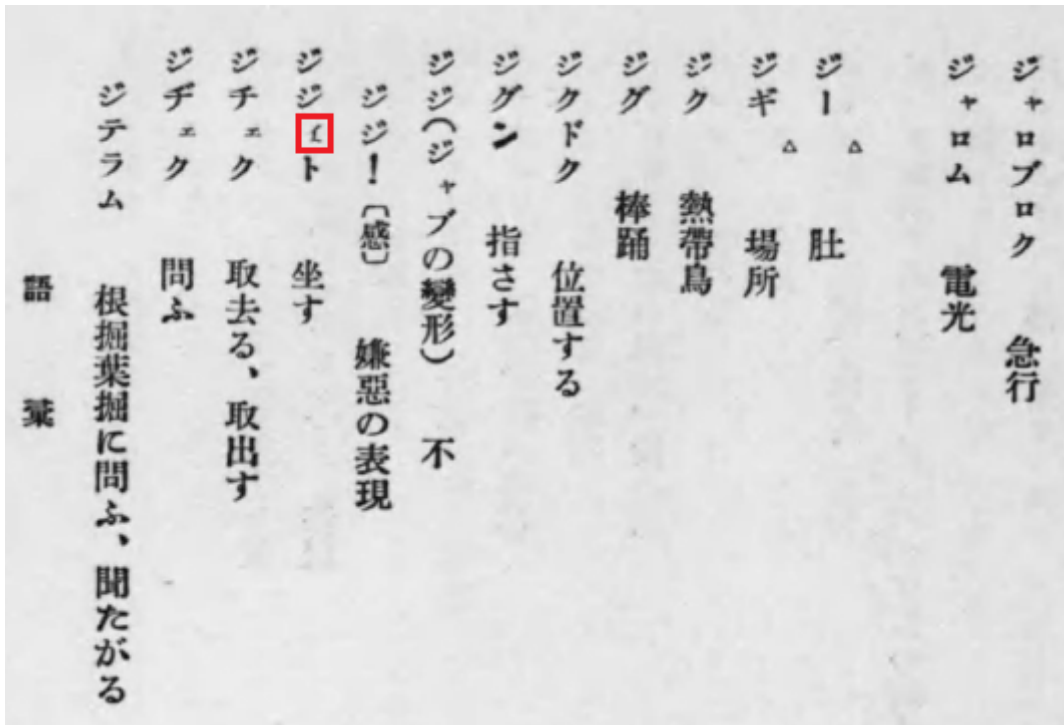二、疑問句にはクオジを用ひることを例とする。

Fig.7 マーシャル語の研究　P60

Fig.8 マーシャル語の研究 P223

Also, we may found some detailed pronunciation rules on Wikipedia[7]:

When a vowel phoneme appears between consonants with different secondary articulations, the vowel often surfaces as a smooth transition from one vowel allophone to the other.[23] For example, *jok* 'shy', phonemically /tʲɛkʷ/, is often realized phonetically as [tʲɛ�running kʷ].[23] It follows that there are 24 possible short diphthongs in Marshallese:[23]

Phoneme	oʲ_oˠ	oʲ_oʷ	oˠ_oʲ	oˠ_oʷ	oʷ_oʲ	oʷ_oˠ
/i/	[i͡ɯ]	[i͡u]	[ɯ͡i]	[u͡ɯ]	[u͡i]	[u͡ɯ]
/e/	[e͡ɤ]	[e͡o]	[ɤ͡e]	[ɤ͡o]	[o͡e]	[o͡ɤ]
/ɛ/	[ɛ͡ʌ]	[ɛ͡ɔ]	[ʌ͡ɛ]	[ʌ͡ɔ]	[ɔ͡ɛ]	[ɔ͡ʌ]
/æ/	[æ͡ɑ]	[æ͡ɒ]	[ɑ͡æ]	[ɑ͡ɒ]	[ɒ͡æ]	[ɒ͡ɑ]

Fig.9 Detailed Pronunciation Rules on Wikipedia

Since *j* is a palatalized consonant, while *t* is a velarized consonant, according to the Marshallese orthography, the word "jijet" may probably pronounced as [tʲizʲe͡ɤtˠ], which may explain why the author of this book chose to use /ʌ/ to represent the vowel *e* in "jijet" (even the vowel in the certain dialect a hundred years ago recorded by the author of the book might be closer to /ʌ/). Thus, ジジェト STANDS NO CHANCE of being a misprint.

References

[1]L2/25−035: Request to Change VO for some Small Kana Lettters and Bopomofo Letters, CheonHyeong Sim, 2025−01−13

[2]L2/16−334: Hiragana and Katakana (Small Letters), CheonHyeong Sim, 2016−11−03

[3]L2/16−354: Proposal to add Kana small letters, Ryusei Yamaguchi, 2016−11−07

[4]L2/17−091: Japanese National Body Contribution on Small Kana Characters (SC2 N4523), Japan NB, 2017−04−01

[5]WG2 N4825: Additional repertoire for ISO/IEC 10646:2017 (5th ed.) Amendment 2.0, Michel Suignard, 2017−06−05

[6]Marshallese−English dictionary (online)

[7]Marshallese language (Wikipedia)

A. Administrative

1. Title: *Proposal to Encode One Small Kana Letter*

2. Requester's name: *CheonHyeong Sim*

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

4. Submission date: *2025-01-15*

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

6. Choose one of the following:

    This is a complete proposal: *YES*

    (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): *NO*

        Proposed name of script:

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

        Name of the existing block: *Small Kana Extension*

2. Number of characters in proposal: *1*

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

A—Contemporary		B.1—Specialized (small collection)	X	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? *YES*

    a. If YES, are the names in accordance with the "character naming guidelines" *YES*

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

5. Fonts related:

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

    CheonHyeong Sim

    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

	CheonHyeong Sim

6. References:

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

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

    of proposed characters attached? **YES**

7. Special encoding issue

    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)? **NO**

8. 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 UAX#44: 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.

C. Technical − Justification

1. Has this proposal for addition of character(s) been submitted before? **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.)? **NO**

      If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:

    size, demographics, information technology use, or publishing use) is included? **NO**

      Reference:

4. The context of use for the proposed characters type of use; common or rare) **Rare**

    Reference:

5. Are the proposed characters in current use by the user community? **NO**

    If YES, where? Reference:

6. After giving due considerations to the principles in the P&P document must the proposed characters be

    entirely in the BMP? **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)? **YES**

8. Can any of the proposed characters be considered a presentation form of an existing

    character or character sequence? **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? **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? | *YES*

If YES, is a rationale for its inclusion provided? | *YES*

If Yes,
reference: | *U+1B121 is the regular form; proposed is the small form. We never unify the regular forms and the small forms for Kana since their functions are different.*

11. Does the proposal include use of combining characters and/or use of composite sequences? | *NO*

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? | *NO*

If Yes,
reference:

12. Does the proposal contain characters with any special properties such as

control function or similar semantics? | *NO*

If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? | *NO*

If YES, are the equivalent corresponding unified ideographic characters identified?

If Yes,
reference:

(End of document)