

Proposal for superscript katakana letters for Ryukyuan languages

Gen Kojitani(genkjtn@gmail.com)

2023 May 8

I am requesting the addition of several superscript katakana letters for Ryukyuan languages.

Background

This proposal requests the addition of Unicode for the kana letter extensions for Ryukyuan languages. Ryukyuan languages or "Shimakutuba" (しまくとぅば; the local name) is one of the minority languages spoken in Japan and is a member of the Japonic language family. They are considered linguistically distinct from the Japanese spoken in mainland Japan because of their different linguistic characteristics and the lack of continuity of dialects. The languages are classified into Amami, Kunigami, Okinawan, Miyako, Yaeyama, and Yonaguni, which are considered as separate languages, and all of them were designated as languages in danger of extinction by UNESCO in 2009.

In an official survey conducted by Okinawa Prefecture in 2008, 20.9% (approximately 251,156 people) of the 1,201,701 Okinawans (of whom 2,630 were surveyed) reported using Ryukyuan Languages (Shimakutuba) as often as they use Japanese as a common language, and 6.9% (approximately 82,917 people) indicated that they use Ryukyuan Languages (Shimakutuba) as their first language.[1]


Ryukyuan languages are a fundamental part of the cultural identity of the people of Okinawa, whose origins lie in the Ryukyu Kingdom, once a separate country from Japan, and it is important to be able to accurately record their languages in order to preserve and pass on their culture. Since some of the Ryukyuan languages have different phonological rules from the Japanese spoken on the mainland, various writing systems have been used to record the phonology of the Ryukyuan languages. Because Okinawans who speak Ryukyuan languages have their roots in the Ryukyu Kingdom and are also Japanese citizens as part of their identity, they have long used kana characters to write their language, just as they have done in Japanese language.

However, until recently, Ryukyuan languages were written using various writing systems, and it has been difficult to communicate what kind of pronunciation is used in the recorded documents. In order to reduce confusion among users of recorded documents, the Shimakutuba Orthography Review Committee (しまくとぅば正書法検討委員会) of the Okinawa Prefecture Department of Culture, Tourism and Sports established an orthography on March 30, 2022, and a new standardized orthography was formulated.[2] This orthography includes a notation for each language and phonetic symbols for all Ryukyuan languages. This writing system includes superscript katakana in addition to the kana used in normal Japanese writing. It is important to encode these characters so that they can be displayed in plain text so that Shimakutuba speakers can transmit information in their own language in the future. Therefore, I propose to add these superscript kana characters to Unicode.

Proposed Characters

First, a list of characters proposed for addition to Unicode is shown below. The list includes the letter form of each character, its name, and a description of its reading etc.

Table 1. list of proposed characters

letter form	name of character	description
	KATAKANA LETTER SUPERScript A	<p>U+1B168 <super> 30A2 ア used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a vowel /a/ without glottal stop as “アア”.</p> <p>Similar to U+30A1 ア KATAKANA LETTER SMALL A, but written above the baseline. The same applies to the following.</p>

letter form	name of character	description
イ	KATAKANA LETTER SUPERScript I	U+1B169 <super> 30A4 イ used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a vowel /i/ without glottal stop as “イイ”.
ウ	KATAKANA LETTER SUPERScript U	U+1B16A <super> 30A6 ウ used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a vowel /u/ without glottal stop as “ウウ”.
エ	KATAKANA LETTER SUPERScript E	U+1B16B <super> 30A8 エ used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a vowel /e/ without glottal stop as “エエ”.
オ	KATAKANA LETTER SUPERScript O	U+1B16C <super> 30AA オ used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a vowel /o/ without glottal stop as “オオ”.
ツ	KATAKANA LETTER SUPERScript TU	U+1B16D <super> 30C4 ツ used in Kunigami, Okinawan, Yonaguni and common Ryukyuan phonetic symbols. Indicates a syllable with a glottal stop and a cephalic consonant as “ツヤ”, “ツワ”.
フ	KATAKANA LETTER SUPERScript HU	U+1B16E <super> 30D5 フ used in common Ryukyuan phonetic symbols. Common Ryukyuan languages phonetic symbols to encompass the different Ryukyuan languages and describe their speech, used when it is necessary to distinguish between /ɸ/ and /f/. The /ɸa/ is written as “フア” and the /fa/ as “ファ”.
ン	KATAKANA LETTER SUPERScript N	U+1B16F <super> 30F3 ン used in Kunigami, Okinawan and common Ryukyuan phonetic symbols. Indicates a nasal consonant /n-/, /m-/, /ɲ-/, /ŋ-/ or /N/ with a syllable head that does not contain a voiced burst and an immediately following consonant as “ンン”.

Examples

国頭語：[?]ラー、^ツラー[[?]ra:] (お前) 〈伊江島〉
 沖縄語：[?]ワー、^ツワー[[?]wa:] (豚)、'ウトゥ、^ウウトゥ[utu] (夫)

Fig. 1: 『沖縄県における「しまくとぅば」の表記について』, p.3, 沖縄県文化観光スポーツ部, March 30, 2022, https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf, (last retrieved: April 29, 2023)

KATAKANA LETTER SUPERSCRIPT TU, KATAKANA LETTER SUPERSCRIPT U

国頭語 仮名表記一覧表

ア [[?] a] [a]	イ [[?] i] [i]	ウ [[?] u] [u]	エ [[?] e] [e]	オ [[?] o] [o]						
[?] ア ア [[?] a] [a]	[?] イ イ [[?] i] [i] [ji]	[?] ウ ウ [[?] u] [wu]	[?] エ エ [[?] e] [e]	[?] オ オ [[?] o] [o]						
カ [ka]	キ [ki]	ク [ku]	ケ [ke]	コ [ko]	キャ [kja]	キュ [kju]	キョ [kjo]	クワ [kwa]	クイ [kwi]	クエ [kwe]
[?] カ ツカ [[?] ka] [[?] ka]	[?] キ ツキ [[?] ki] [[?] ki]	[?] ク ツク [[?] ku] [[?] ku]	[?] ケ ツケ [[?] ke] [[?] ke]	[?] コ ツコ [[?] ko] [[?] ko]				[?] クワ ツクワ [[?] kwa] [[?] kwa]	[?] クイ ツクイ [[?] kwi] [[?] kwi]	[?] クエ ツクエ [[?] kwe] [[?] kwe]

Fig. 2: 『沖縄県における「しまくとぅば」の表記について』, p.8, 沖縄県文化観光スポーツ部, March 30, 2022, https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf, (last retrieved: April 29, 2023)

KATAKANA LETTER SUPERSCRIPT A, KATAKANA LETTER SUPERSCRIPT I, KATAKANA LETTER SUPERSCRIPT U, KATAKANA LETTER SUPERSCRIPT E, KATAKANA LETTER SUPERSCRIPT O, KATAKANA LETTER SUPERSCRIPT TU

<p>(長音) ー</p> <p>(促音) ッ</p> <p>(撥音) 【語頭】 'ン, ^ン 'n, 'm, 'j, 'η, 'N [n-,m-,j-,η-,N] (緩やかな声立て) [?]ン, ^ツ [[?]n,[?]m, [?]j,[?]η, [?]N] (声門閉鎖音) 【語中・語末】 ン [-n,-m, -j-, -η-, -N]</p> <p>ア行で始まる単語には「声門閉鎖音」と「緩やかな声立て」の区別があるが、「声門閉鎖音」で始まる語の方が多いため、「緩やかな声立て」の表記に記号を添え、「声門閉鎖音」で始まるア行の単語については何も記号を添えずに表記することで両者を書き分ける。</p> <p>例. ウトゥ[utu] (音) 'ウトゥ, ^ウウトゥ[utu] (夫)</p>
--

Fig. 3: 『沖縄県における「しまくとぅば」の表記について』, p.10, 沖縄県文化観光スポーツ部, March 30, 2022, https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf, (last retrieved: April 29, 2023)

KATAKANA LETTER SUPERSCRIPT N, KATAKANA LETTER SUPERSCRIPT TU, KATAKANA LETTER SUPERSCRIPT U

ファ [ɸa]	フィ [ɸi]	フ [ɸu]	フェ [ɸe]	フォ [ɸo]									
フ ^ァ [fa]	フ ^ィ [fi]	フ ^ウ [fu]	フ ^ェ [fe]	フ ^ォ [fo]				フ ^ャ [fja]	フ ^ュ [fju]	フ ^ョ [fjo]			
パ [pa]	ピ [pi]	プ [pu]	ペ [pe]	ポ [po]	パイ [pɰ] [pi]	プス [ps] [pʰi] [pi]	ペエ [pɛ] [pɛ]	ピャ [pja]	ピュ [pju]	ピョ [pjo]	プア [pwa]		

Fig. 4: 『沖縄県における「しまくとぅば」の表記について』, p.22, 沖縄県文化観光スポーツ部, March 30, 2022, https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf, (last retrieved: April 29, 2023)

KATAKANA LETTER SUPERSCRIPT HU

Properties

U+1B168; KATAKANA LETTER SUPERSCRIPT A;Lo;0;L;<super> 30A2;;;N;;;;;
U+1B169; KATAKANA LETTER SUPERSCRIPT I;Lo;0;L;<super> 30A4;;;N;;;;;
U+1B16A; KATAKANA LETTER SUPERSCRIPT U;Lo;0;L;<super> 30A6;;;N;;;;;
U+1B16B; KATAKANA LETTER SUPERSCRIPT E;Lo;0;L;<super> 30A8;;;N;;;;;
U+1B16C; KATAKANA LETTER SUPERSCRIPT O;Lo;0;L;<super> 30AA;;;N;;;;;
U+1B16D; KATAKANA LETTER SUPERSCRIPT TU;Lo;0;L;<super> 30C4;;;N;;;;;
U+1B16E; KATAKANA LETTER SUPERSCRIPT HU;Lo;0;L;<super> 30D5;;;N;;;;;
U+1B16F; KATAKANA LETTER SUPERSCRIPT N;Lo;0;L;<super> 30F3;;;N;;;;;

References

- 『平成28年度しまくとぅば県民意識調査報告書』, 沖縄県, March, 2009, <https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/simakutuba/documents/kenminishiki.pdf>, (last retrieved: May 8, 2023)
- 『沖縄県における「しまくとぅば」の表記について』, 沖縄県文化観光スポーツ部, March 30, 2022, https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf, (last retrieved: April 29, 2023)

Acknowledgements

I would like to express my gratitude for the permission from the Shimakutoba Promotion Office, Cultural Promotion Division, Department of Culture, Tourism and Sports, Okinawa Prefecture (沖縄県文化観光スポーツ部文化振興課しまくとぅば普及推進室), to submit a Unicode proposal on my individual behalf.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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>Proposal for superscript katakana letters for Ryukyuan languages</i>
2. Requester's name:	<i>Gen Kojitani</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual contribution</i>
4. Submission date:	<i>2023/05/08</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(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:	
b. The proposal is for addition of character(s) to an existing block:	<i>yes</i>
Name of the existing block:	<i>Small Kana Extension</i>
2. Number of characters in proposal:	<i>8</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/> B.1-Specialized (small collection) <input checked="" 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?	<i>yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Gen Kojitani (genkjtn@gmail.com)</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>IPA Font License</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>no</i>
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>no</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?	<i>no</i>
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.)?	<i>yes</i>
If YES, with whom? <i>沖縄県文化振興課しまくとぅば普及推進室 (Shimakutuba Promotion Office, Cultural Promotion Division, Okinawa Prefecture)</i>	
If YES, available relevant documents: <i>https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/documents/02_shimakutubahyouki.pdf</i>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>yes</i>
Reference: <i>https://www.pref.okinawa.lg.jp/site/bunka-sports/bunka/shinko/simakutuba/documents/kenminishiki.pdf</i>	
4. The context of use for the proposed characters (type of use; common or rare)	<i>rare</i>
Reference: <i>(Since the orthography is still being finalized, there are only a few examples of usage such as bifurcation.)</i>	
5. Are the proposed characters in current use by the user community?	<i>yes</i>
If YES, where? Reference: <i>Okinawa prefecture, Japan</i>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>no</i>
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)?	<i>yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>no</i>
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?	<i>no</i>
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?	<i>yes</i>
If YES, is a rationale for its inclusion provided? <i>yes</i>	
If YES, reference: <i>Table 1.</i>	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>no</i>
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: _____	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary) _____	

13. Does the proposal contain any Ideographic compatibility characters?	<i>no</i>
If YES, are the equivalent corresponding unified ideographic characters identified? _____	
If YES, reference: _____	

(Detach the following pages from the filled summary form before submitting).

Information accompanying submissions

The process of deciding which characters should be included in the repertoire of the standard by WG 2 depends on the availability of accurate and comprehensive information about any proposed additions. WG 2, at its San Francisco meeting 26, designed a form (template) that will assist the submitters in gathering and providing the relevant information, and will assist WG 2 in making more informed decisions.

This form is part of the [WG 2 Principles and Procedures document](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html) and is available on line at <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

A duly completed proposal summary form must accompany each new submission. The latest version of this form must be used in submissions. Such a form will assist WG 2 to better evaluate the proposal, and progress the proposal towards a speedier acceptance and inclusion in the standard. Submitters are also requested to ensure that a proposed character does not already exist in the standard.

Submitters are encouraged to visit the "[Where is my Character](#)" page on the Unicode web site for more information on checking if their proposed character or script is already encoded in the standard, or a similar proposal has already been made by someone else. There are also several electronic discussion lists maintained by the Unicode consortium that one could use to discuss with other experts internationally on various subjects related to the standard. Submitters are also encouraged to familiarize themselves with ISO/IEC TR15285 – Character Glyph Model. The latest version of ISO/IEC 10646 as well as ISO/IEC TR15285 are available on line under 'Freely Available Standards' at: <http://standards.iso.org/ittf/PubliclyAvailableStandards/>.

National bodies should take note of the guidelines in section *C.4 Some guidelines on proposing new material as ballot comments* in the [P&P document](#).

In addition to text extracted from the standard in *Annex L: Character-naming guidelines* in the [P&P document](#), the following definitions from the standard are also referenced in the proposal summary form:

Clause 4.12 Combining character:

A member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character (see also 4.14).

NOTE – ISO/IEC 10646 specifies several subset collections, which include combining characters.

Clause 4.14 Composite sequence:

A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters (see also 4.12).

NOTE 1 – A graphic symbol for a composite sequence generally consists of the combination of the graphic symbols of each character in the sequence.

NOTE 2 – A composite sequence is not a character and therefore is not a member of the repertoire of ISO/IEC 10646.

Submitter's responsibilities

The national body or liaison organization (or any other organization or an individual) proposing new character(s) or a new script shall provide:

1. Proposed category for the script or character(s), character name(s), and description of usage.
2. Justification for the category and name(s).
3. A representative glyph(s) image on paper:
If the proposed glyph image is similar to a glyph image of a previously encoded ISO/IEC 10646 character, then additional justification for encoding the new character shall be provided.
Note: Any proposal that suggests that one or more of such variant forms is actually a distinct character requiring separate encoding should provide detailed, printed evidence that there is actual, contrastive use of the variant form(s). It is insufficient for a proposal to claim a requirement to encode as characters in the Standard, glyphic forms which happen to occur in another character encoding that did not follow the Character-Glyph Model that guides the choice of appropriate characters for encoding in ISO/IEC 10646.
Note: WG 2 has resolved in Resolution [M38.12](#) not to add any more Arabic presentation forms to the standard and suggests users to employ appropriate input methods, rendering and font technologies to meet the user requirements.
4. Mappings to accepted sources, for example, other standards, dictionaries, accessible published materials.
5. Font resource:
A font must be provided to the Editor promptly after the characters have been approved by WG2 for use in printing the charts. If a font is not provided, the Editor cannot include the glyphs in the charts and, as a result, the repertoire of characters corresponding to these glyphs will not be included in draft amendments. Fonts must be in one of the following formats (in preferential order): OpenType, TrueType, Postscript Type 1.
Fonts submitted must have no license restrictions that prevent embedding into PDF documents. Because of synchronization between ISO/IEC 10646 and The Unicode Standard, any grant of license must cover use in publishing both standards and related documents (see resolution [M45.30](#)). For technical reasons, the editors must be able under that license to freely modify or replace glyph outlines in their copies of the fonts. The submitter should be prepared to provide fonts of suitable quality and license conditions, unless equivalent fonts are already available to the editors.
6. List of all the parties consulted:
Towards ensuring that all significant stakeholders have been consulted, WG 2 requests that submitters solicit feedback from other groups who use the proposed characters, and list these contacts. Submitters are encouraged to provide the email id-s of the submitters as well as other experts who have been consulted to facilitate any clarification queries.
7. Equivalent glyph images:
If the submission intends using composite sequences of proposed or existing combining and non-combining characters, a list consisting of each composite sequence and its corresponding glyph image shall be provided to better understand the intended use.
8. Compatibility equivalents:
If the submission includes compatibility ideographic characters, identify (per resolution [M45.29](#)):
 - the source, which contains two distinct code positions that correspond to a single unified CJK Ideograph character of ISO/IEC 10646
 - the ISO 10646 unified CJK Ideograph
 - the code position in the source for the unified CJK ideograph
 - the code position in the source for the proposed compatibility ideograph
9. Properties that may affect the BiDi processing
Any BiDirectional algorithm related properties associated with the characters should be spelled out (see UAX#9 - <http://www.unicode.org/reports/tr9/>).
10. Any additional information that will assist in correct understanding of the different characteristics and linguistic processing of the proposed character(s) or script.
11. If any of the proposed characters are suitable as *syntax* characters please take note of guideline D.2.5 (on page 24 and of Unicode Standard Annex 31 at <http://www.unicode.org/reports/tr31/>).
12. If the proposal is for a new script, identify the standardized characters that are commonly used directly in the script, or proposed to be unified with the characters of the script, in particular standardized characters allocated in different blocks. Examples include punctuation marks and combining marks. Such information will assist in assigning properties for characters shared across multiple scripts or in identifying character repertoires needed to support particular languages.
13. If you are aware of already standardized characters that are visually close to any of the proposed characters, you are invited to list them in the proposal. This will assist in the analysis of the script for 'visually confusables', towards providing additional guidance on use of the standard from a security perspective (see UTR#36 - Unicode Security Considerations – at <http://www.unicode.org/reports/tr36/>).