Proposal for missing three Kana-Ligatures

Gen Kojitani(genkjtn@gmail.com)

2024 June 21

I am requesting the addition of three Hiragana/Katakana ligatures. This proposed document is a revised version of the previous proposal L2/23-112, which was submitted in response to comment L2/23-163.

Background

This proposal requests the addition of Unicode for the ligatures of kana characters (合略仮名; gōryaku gana) used in old Japanese documents. Gōryaku gana are ligatures of hiragana or katakana used in Japanese writing, and were mainly used in literature until the Meiji era (until 1912). In the study of Japanese archives and the documentation of literary works, encoding these characters is important to accurately preserve the information of the originally written characters.

Each gōryaku gana is a single kana character that represents a word with two or more syllables, and is composed of a ligature of several different kana characters, as in the case of the Latin ligature ‘Æ’.

Currently, Unicode already includes several ligature kana, including U+309F 紋 (hiragana yori) in the Hiragana block, U+30FF 郑 (katakana koto) in the Katakana block, U+2A708 𪜈 (katakana tomo) in the CJK Ideographs Extended-C block and U+2CF02 𬼂 (hiragana nari), U+2CF00 𬼀 (katakana site), U+2CEFF خلاف (katakana nari) in the CJK Ideographs Extended-F block. In addition to these, there are several other gōryaku gana that occur frequently in the ancient literature, and these are proposed here.

Since some of the "multisyllabic" kana letters which have historically been referred to in the literature as gōryaku gana, are merely a sequence of kana characters (連綿; renmen) in vertically handwritten documents, the following criteria were considered in defining the gōryaku gana proposed for encoding in this document.

1. When found in handwritten documents (vertical writing) made before the development of movable type printing technology, multisyllabic kana letters have their kana-components arranged sideways, i.e. in a direction different from the normal writing direction, or have been significantly altered by omitting components of the original characters.

2. Multisyllabic kana characters that are clearly printed as a single character in books and other works produced after the spread of letterpress printing technology in the Meiji era.

3. Multisyllabic kana letters that meet criteria 1 or 2 are considered gōryaku gana.

Note that even if letters have historically been considered to be part of gōryaku gana, CJK letters and their abbreviations (such as ‘Measured’ for ‘iu’ and ‘寸/時’ for the abbreviation of ‘時 toki’) are not included in gōryaku gana in this proposal.

The following rules were used to determine the order of character names and letters.

1. Ligatures that combine multiple Hiragana characters are classified as "HIRAGANA", and ligatures that combine multiple Katakana characters are classified as "KATAKANA".

2. When the reading of a character is expressed in ordinary kana characters, a character composed of two kana characters is classified as DIGRAPH, as in " ./ HIRAGANA DIGRAPH YORI."

3. The general order of the characters follows the Unicode block order, with hiragana ligatures first and katakana after. Within each group, each gōryaku gana reading was sorted in "aiueo order," which is the current lexical order of kana in Japanese.
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 its descriptions.

Table 1. list of proposed characters

<table>
<thead>
<tr>
<th>letter form</th>
<th>name of character</th>
<th>description</th>
</tr>
</thead>
</table>
| こ と | HIRAGANA DIGRAPH KOTO | ≈ <vertical> 3053 こ 3068 と representing a Japanese word “koto”. This word means “thing”. The katakana-form is 30FF ト KATAKANA DIGRAPH KOTO.  
- It often appears in vertical writing contexts, but its use as a single-character ligature can also be found in many documents produced after the widespread use of letterpress printing technology.  
- Even in vertically written documents, it is not used for any other purpose than “thing”. For example, this character is never used in words such as “ことわる (to refuse)” or “ことわざ (proverb)”. |
| ト キ | KATAKANA DIGRAPH TOKI | ≈ 30C8 ト 30AD キ representing a Japanese word “toki”. This word means “when...,” or “time”.  
- The left part of the stroke has both left curved and straight variations (fig. 7).  
- Even in vertically written documents, it is not used for any other purpose than “time” or “when”. For example, this character is never used in words like “シフトキーエ (shift key)”. |
| ト テ | KATAKANA DIGRAPH TOTE | ≈ 30C8 ト 30C6 テ representing a Japanese word “tote”. This word means “but” or “although”.  
- Even in vertically written documents, it is not used for any other purpose than “but” or “although”. For example, this character is never used in words like “カットテープ (cut tape)”. |

Note:

In my previous proposal L2/23-112, I included ‘KATAKANA DIGRAPH YORI (ヨリ),’ but since dictionaries only state that “such a character exists” and any examples of its actual usage could not be found, I have removed it from this proposal.

L2/23-112 also included some Hiragana ligatures, but these have been removed from this proposal because they can be considered as simple character shape changes resulting from sequences of Hiragana characters in vertically hand-written documents.

In addition, since the component of "to-iu(とい)" is a CJK ideograph, it has been determined that it is appropriate to treat it as a CJK ideograph, and has been removed from this proposal.
Examples

The following images are examples for usage of gōryaku gana.

Fig. 1: 假字本末 下巻 p.26 伴信友 1850年, https://books.google.co.jp/books?id=OslZAAAAcAAJ&pg=frontcover#v=onepage&q=&f=false

KATAKANA DIGRAPH TOKI
Fig. 2: 手島隠蔵 編『国語漢文自習要覧 : 附・仮名遣詳解』p.20-21, 秀文社, 大正5. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/942681 (参照 2023-04-24)

KATAKANA DIGRAPH TOKI(right), HIRAGANA DIGRAPH KOTO(left)
Fig. 3: 落合直文著『日本大文典』第1編, p.42-43, 博文館, 明治27-30. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/864050 (参照 2023-04-24)

KATAKANA DIGRAPH TOKI
Fig. 4: 落合直文 著『日本大文典』第1編, p.44-45, 博文館, 明27-30, 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/864050 (参照 2023-04-24)
KATAKANA DIGRAPHTOTE
Fig. 5. 藤井乙男『操觚便覧』, p.16-17, 浜本明昇堂, 明32.12. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/863176 (参照2023-04-24)

KATAKANA DIGRAPH TOKI
Fig. 7: カッケンボス, ガノー 著ほか『改正物理全志』, p.414-415, 煙雨楼, 明12.3. 国立国会図書館デジタルコレクション
https://dl.ndl.go.jp/pid/829989 (参照 2023-04-24)

KATAKANA DIGRAPH TOKI
HIRAGANA DIGRAPH KOTO
Fig. 10: 糸左近 著『雅俗対照』, p.60-61, 金刺芳流堂, 明35.11. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/864212 （参照 2023-04-24）

HIRAGANA DIGRAPH KOTO(right), KATAKANA DIGRAPH TOKI(left)
假名の借用について

図13は落合直文の『ことばの泉: 日本大辞典』（大倉書店）、37.1頁に掲載されている図です。この図は、假名の借用を示すために使用されています。図中には、借用された假名の組み合わせが示されています。

図の図例は、借用された假名の組み合わせを示しています。例えば、「KATAKANA DIGRAPH TOKI」の借用は図の右下に示されています。


KATAKANA DIGRAPH TOKI
KATAKANA DIGRAPH TOKI
Fig. 13: 日比野綱雄 編『習字大全』上, p.5-6, 東壁堂, (明13). 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/852450 (参照 2023-04-24)
Fig. 14: AFPBB News: 『魯迅自筆の日本語の手紙が初公開 上海市』，2023年4月5日19:47 発信地：中国 [中国・中国・台湾]，https://www.afpbb.com/articles/-/3458675
KATAKANA DIGRAPH TOKI
末広鐵腸 著『落葉のはき寄せ: 政治小説』、p.26-27、青木嵩山堂、明33.6。国立国会図書館デジタルコレクション
https://dl.ndl.go.jp/pid/885631（参照 2023-04-25）

HIRAGANA DIGRAPH KOTO
Fig. 17: 朝日新聞縮刷版 昭和23年1~4月, 1948.4.27記事, 朝日新聞社, 1988.5.

HIRAGANA DIGRAPH KOTO
Fig. 18: 『平家物語：真字熱田本』巻第2, p.4, 育徳財団製和16. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/1142755 (参照 2023-04-25)

KATAKANA DIGRAPH TOTE
Fig. 19: 「平家物語：真字熱田本」巻第2, p.7, 育徳財団, 昭和16. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/1142755 (参照 2023-04-25)
2. 壱般之項  G.P. の初項  通比

第 8 項 = ar,
第 4 項 = ar^2,
以下同理

設 = a 異数へ此項数 = 1 少なか
之 = 由第 8 項アリストレヘ

(第壹例) G.P. の初項 5 ニュッ通比 8 ト求ム.
第 10 項 = ar^{10-1} = 5 \times 8^{9} = 5 \times 8^8.

(第貳例) G.P. の第 4 項 189 及び第 6 項 1701 ニュッ通比 8 項求ム．

除算 = 由ア a = 189 及び a^8 = 1701

又 = a \times (\pm 8)^4 = 189 及び a = \pm 7.

之= 由ア 第 8 項 = \pm 7 \times 189^7 = 1609.

3. 等比中項 トハ G.P. ナナス三数
ノ中数ナシフ.
Fig. 21: 近森出来治著『楽典摘要』，p.2-3，前川書店，明35.11。
国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/854815（参照2024-06-）
Fig. 22 音楽取調所編『音楽取調成績申報書』, p.8-9. 文部省, 明17.2. 国立国会図書館デジタルコレクション https://dl.ndl.go.jp/pid/854780 (参照 2024-06-21)

KATAKANA DIGRAPH TOKI
民衆芸術といふものが最近文壇の二三氏に依つて論議された。

私は今その人々の論議を一々委し読みだわけでないから、それの論議の中く何處にあるかを知るかが出来ないが、民衆芸術といふ言葉そのもの、概念が今以て可なりに曖昧であることがあるわかった。或る人は、これを以て普通芸術以外の芸術と解してゐるらしい。又或る人は之を以て所謂社会主義の芸術と解してゐるらしい。しかしこれらは何れも民衆芸術といふことの内容の真を因へた芸術に過ぎない。民衆芸術といふことには、単に通俗芸術、一般芸術、社会主義芸術といふ以外自ら別にその適確な意義がなければならぬ。
## ISO/IEC JTC 1/SC 2/WG 2

**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646**

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


Please ensure you are using the latest Form from [http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html) for latest Roadmaps.

---

### A. Administrative

1. **Title:** Proposal for missing Kana-Ligatures
2. **Requester’s name:** Gen Kojitani
3. **Requester type (Member body/Liaison/Individual contribution):** Individual contribution
4. **Submission date:** 2024/06/21
5. **Requester’s reference (if applicable):**
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):
   - Proposed name of script:
   - b. The proposal is for addition of character(s) to an existing block:
   - Name of the existing block: Kana Extended-A
2. **Number of characters in proposal:** 3
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?** yes
   - a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? 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? Gen Kojitani (the author of this proposal)
   - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): IPA font license, genkjtn@gmail.com
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 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)? no
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](http://www.unicode.org) for such information on other scripts. Also see Unicode Character Database ([http://www.unicode.org/reports/tr44/](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?**
   - If YES explain: *L2/23-112, my previous proposal*
   - Yes

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?
   - No

3. **Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?**
   - Reference: ___________
   - No

4. **The context of use for the proposed characters (type of use; common or rare) is included?**
   - Reference: ___________
   - For study

5. **Are the proposed characters in current use by the user community?**
   - Reference: ___________
   - No

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?
   - 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?
   - 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?
   - 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?
    - 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?
    - No

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?
    - No
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 online 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 online under 'Freely Available Standards' at: [http://standards.iso.org/ittf/PubliclyAvailableStandards/](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](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html).

In addition to text extracted from the standard in Annex L: Character-naming guidelines in the [P&P document](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html), 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.

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/](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/](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/](http://www.unicode.org/reports/tr36/)).