

PROPOSAL TO ENCODE A KANA CHARACTER FOR TRANSCRIPTION OF LATE
MIDDLE JAPANESE

Alexander Zapryagaev

22 November 2019

rudetection@gmail.com

TABLE OF CONTENTS

1. Introduction	2
2. Linguistic Use	2
3. The Evidence	3
4. Motivation for Encoding	4
5. Proposed Character	4
Bibliography.....	5
Font Usage	5

1. INTRODUCTION

This proposal offers to encode an additional character (small version of U+1B06A HENTAIGANA LETTER TU-2). The motivation lies in the correct and unambiguous representation of earlier forms of Japanese language, especially Late Middle Japanese, in academic contexts, scientific articles and dictionaries.

2. LINGUISTIC USE

Late Middle Japanese (LMJ) is a stage of the development of Japanese language associated, according to [Frellesvig 2010, p. 1] with the Kamakura and Muromachi periods of Japanese history (approximately AD 1'200–1'600), culminating with the Sengoku era; in other words, the intermediate stage between the classical Heian language and the speech of Edo. During most of the period, the current form of language was barely put to paper [ibid., pp. 297ff.], due to the prestige of the Heian language which formed a literary norm, except occasional appearance in dialogue parts of fictional works. However, from the second half of the fifteenth century, it is being increasingly written down [ibid., pp. 299–303]. It happened first in Japanese script, due to the development of *shōmono* (抄物) genre, representing the lecture notes of eminent scholars and thus the actual, albeit formalized, speech of the urban centres, and then, from mid-sixteenth century, in Latin alphabet, together with the proselytizing efforts of the Jesuit missionaries, mostly Portuguese. These and other sources leave us with a wealth of knowledge about the then-current state of the language; the grammars by João Rodrigues – *Arte da Lingoa de Iapam*/日本大文典 (1'604–8); *Arte Breve da Lingoa Iapoa*/日本小文典 (1'620) – and the grand Jesuit dictionary *Vocabulario da Lingoa de Iapam*/日葡辞書 (1'603–4) were probably not superseded by any description of comparable span until the twentieth century.

The data in Latin script provide us with priceless phonetic information, revealing contrasts and sound-changes neglected or reflected insufficiently in native Japanese script. One of these is the existence of the final moraic -t. While modern language possesses two moraic consonantal phonemes that can close the syllable and turn it heavy (the nasal /N/ and the geminating /Q/), the Portuguese data reveal the existence of another moraic element in the language of the sixteenth century: a final -t, which could close the syllable on itself. They reveal, for example, that the word 今日 was pronounced *connit* [kon:it], and not [kon:itɕi], as in modern Japanese. This could appear word-internally as well, such as 発熱する *fotnet suru* [fotnet-suuru] (now [hotsunetsui-suuru]).

The main source of the final -t were the 入声 (Entering Tone, that is, closed by a stop) syllables of Sino-Japanese borrowings, especially those not completely assimilated and/or belonging to a more

refined stratum (compare number ‘1,’ given explicitly as *ichi* in the *Vocabulario*, and not **it*). However, there is a limited number of native words that, by vocalic elision, came to possess variants with -t in LMJ; an example from *Vocabulario* would be *ximot* 楯 (also *ximoto*).

While for the Europeans the representation of such a feature was not problematic and required just employing ‘t,’ representing it was not a trivial task in Japanese script, especially still lacking standardization, both in kana use (complete under Motoori in the nineteenth century) and in choice between multiple possible hentaigana variants (fixed even later, in the start of the twentieth century). As demonstrated by [Sugahara 2000, pp. 21–82], some LMJ manuscripts demonstrate awareness of the distinction by systematically choosing from different sets of hentaigana to denote [tu] (including geminated) and [-t], but this is in no case spread to all documents; furthermore, nothing similar is known for the previous stage of Early Middle Japanese, when the final -t is supposed to also have existed in the Sino-Japanese borrowings.

3. THE EVIDENCE

We consider the *Hōyaku Nippo Jisho* (邦訳日葡辞書, ‘Japanese-translated Japanese-Portuguese Dictionary’), released by Iwanami Shoten 岩波書店 in 1980 under the chief editorship of Doi Tadao (土井忠生). It is a translated and annotated edition of the original *Vocabulario...* of 1’603, and thus a full-fledged dictionary of LMJ for Japanese users.

In Preface, p. 31, the editors admit: “入声の t には、促音‘ッ’と区別して、特に‘っ’を用いた。” (As for the Entering Tone -t, it is different compared to the geminate フ, thus we used ♪.) Hence, they acknowledge the need of a separate symbol to denote -t in the corpus of the dictionary, and choose ♪ as such (Fig. 1).

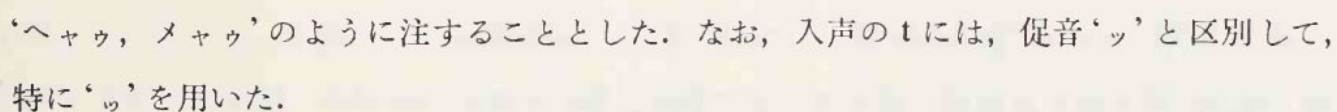


Fig. 1. *Nippo Jisho*, Preface

A sample entry with ♪ looks as follows:

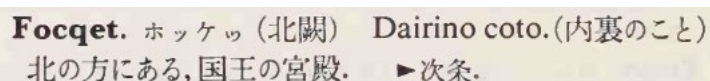


Fig. 2. A sample entry with ♪

As we observe, the character is used with katakana.

4. MOTIVATION FOR ENCODING

It is desirable to have a method of expressing the distinctions of the earlier forms in Japanese while writing in kana unambiguously. Some steps toward this goal were already taken by creating the Small Kana Extension block (starting with the proposal L2/16-334 of 2016); the encoded characters included the small versions of kana letters wi, we, wo (ヰヱヲ), which allow to spell the syllables with labialized vowels. Spelling the final -t seems a logical next stage. The proposal implies creating a small version of the existing character U+1B06A ヰ HENTAIGANA LETTER TU-2.

- The existing character U+1B06A cannot serve its purpose, as it is full-sized and thus denotes the syllable /tu/. The usage of smaller versions of characters to denote the consonantal part without vocalic is widely attested, say, in Ainu spellings. Furthermore, the 日葡辞書 spellings themselves use a smaller version (Fig. 1).
- The existing “small tu” kana symbols, neither hiragana つ nor katakana ツ, also cannot fulfill the purpose, as 日葡辞書 renders the character under question as different from both; also, they denote different phonemes, final -t as opposed to the geminating /Q/, and the need of the exact transcription is not met. As an example, the following can be given (Fig. 3):



Fotnet. ホツネツ (発熱) 何かの病気の熱が出ること、
あるいは、熱をもつこと。 ¶ Fotnet suru. (発熱する) 同上。

Fig. 3. 発熱 in Vocabvlario

Spelling, say, ホツネツ here is, first, ungrammatical and can be understood, if anything, as [fon:et] instead of [fotnet].

- The proposed character would not be specified for hiragana or katakana; it can work with both, as exemplified in Fig. 3 above.

5. PROPOSED CHARACTER

HENTAIGANA LETTER TU-2	ヰ	derived from U+1B06A
------------------------	---	----------------------

1B11F(?);HENTAIGANA LETTER SMALL TU-2;Lo;0;L;;;;;N;;;;;

BIBLIOGRAPHY

1. Frellesvig, Bjarke. *A history of the Japanese language*. Cambridge University Press, 2010.
2. Sim, Cheon Hyeong (2016-11-04), Hiragana and Katakana (Small Letters). Unicode Document L2/16-334.
3. Sugahara, Norio [菅原範夫著] キリシタン資料を視点とする中世国語の研究. Tokyo: Musashino Shoin, 2000.

FONT USAGE

The text is typeset in EB Garamond. The Japanese characters are typeset in Source Han Serif (Japanese flavour). The kana characters missing in SHS are given in BabelStone Han.

**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:	Proposal to Encode a Kana Character for Transcription of Late Middle Japanese
2. Requester's name:	<i>Alexander Zapryagaev</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual Contribution</i>
4. Submission date:	<i>2019-11-22</i>
5. Requester's reference (if applicable):	
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 type="checkbox"/>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>	
Name of the existing block:	<i>Kana Extended-A</i>	
2. Number of characters in proposal:	<i>1</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?	<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>Alexander Zapryagaev</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Alexander Zapryagaev rudetection@gmail.com</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 type="checkbox"/>	

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? If YES explain _____	<i>No</i>
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: _____	<i>No</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: _____	<i>Yes</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference: _____	<i>Rare</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: _____	<i>Yes</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: _____	<i>No</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? _____	
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: _____	<i>No</i>
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: _____	<i>No</i>
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: _____	<i>Yes</i> <i>Yes</i> <i>In proposal</i>
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: _____	<i>No</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) _____ _____	<i>No</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference: _____	<i>No</i>