

**Doc Type:** Working Group Document  
**Title:** Proposal to Add Tamil Nukta Character  
**Source:** Martin Hosken  
**Status:** Individual contribution  
**Action:** For consideration by UTC and WG2  
**Date:** 2015-10-26

**Executive Summary** This proposal is to add a character U+0BBC TAMIL SIGN NUKTA. To the Tamil block of Unicode. The Unicode character data for it will be:

0BBC;TAMIL SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;

The representative glyph will be a lower single dot following Devanagari. It's sorting and shaping behaviour will follow Devanagari.

**Introduction** The Tamil language does not make use of a Nukta. But there are a few minority languages that use the Tamil script that need a Nukta. The following table shows examples of their use:

Language	Example	Unicode
Chetti (ctt)	பீயிக்கு	0BAA 0BBC 0BC0 0BAF 0BBF 0B95 0BCD 0B9E
Betta Kurumba (xub)	ஒண்டெ	0B92 0BA3 0BCD 0B9F 0BBC 0BC6 0BBC
Jennu Kurumba (xuj)	நட்து	0BA8 0B9F 0BBC 0BA4 0BBC
Irula (iru)	பெல	0BAA 0BBC 0BC6 0BB2

There are two key points to notice about the Tamil Nukta from these examples:

- There are different styles of nukta: single dot, as used in Chetti; single ring as used in Betta and Jennu Kurumba and double dot as used in Irula.
- The nukta may occur under a vowel, as in the Betta Kurumba example, where the vowel U+0BC6 TAMIL VOWEL SIGN E is reordered in front of the U+0B9F TAMIL LETTER TTA and both take a nukta.

**Rationale** While there are nukta letters encoded in other blocks, there is a need for a nukta character in the Tamil block to ensure script consistency across a run containing this diacritic.

**Representative Glyph** One of the styles of the Tamil nukta needs to be chosen for the representative glyph. Since Chetti is a more dominant language of the region, the relevant communities have decided that the default glyph should be the single dot as used for Chetti. This form is also the most prevalent nukta, being used in Devanagari, Bengali, Gurmukhi, Gujarati and Oriya, with only Kannada having a different form.

## Examples

## ஸ்வர அக்ஷர

அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ

## வஞ்சுன அக்ஷர

க	க	ங	
ச	ஜ	ஞ	
ட	ட	ண	
த	த	ந	ன
ப	ப	ம	
ய	ர	ற	வ
ல	ள	ஷ	ஸ
ஹ	க்ஷ	ழ	

Illustration 1: Chetti alphabet from spelling book

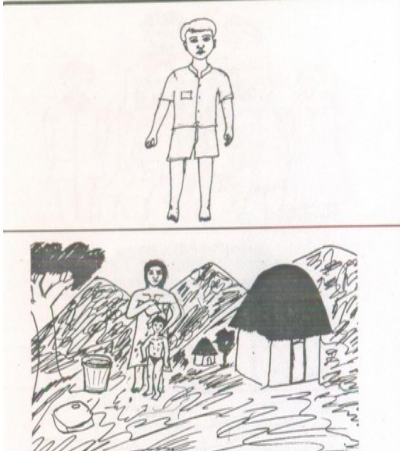
## நங்கள் சமுதாயத பற்றிட்டுள்ளா பிவற

நங்கள் சமுதாயா சுமாராயிட்டு மொத்த ஆறாயிர ஜனத்தொகெ இததீடு விகிதாச்சார ஏகதேச சம ஆப்புது நங்க இப்பத்திநாக்கு தரவாடு உ குலதெய்வங்ஙனம், உட்டு. இதனொளெ ஐது காராய்மெ காரும் உட்டு

Illustration 2: Sample from chetti newsletter where in desperation they used a double nukta, despite their style preference

பேரு கடக்க குருவ  
ரோனுங்க குருவனு  
| பேரு எலிசபெத்து

Illustration 3: Irula sample text with nukta use marked



## நங் பென்னு

ஒன் மெத்தொல் கேத்தன் அம்பெ ஒண்டெ  
உடுகன் இத்தெ.

அதெ தனப்பி நீராட்டிசானி, மண்டொட்டய நலா  
தித்தாட்டு நல் குப்பாயெ உன்னு ஸ்கோல்க்கா  
போப்பா.

Illustration 4: Betta Kurumba story book

மருதலாய் கொண்டணரார்வாண்டார்  
மருதலனை பாத்து சொல்லுகக்குள்.  
ன. ரண்டவல்லமணுசாரநேந்து

Illustration 5: Irula story showing alternative style of nukta

0BB0

Tamil

0BFF

	0B8
0	ர
1	ற
2	ல
3	ள
4	ழ
5	வ
6	ஶ
7	ஷ
8	ஸ
9	ஹ
A	
B	
C	ஃ
D	
E	ஃ
F	ஃ

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>.**

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

**Please read Principles and Procedures Document (P & P) from <http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.**

**See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.**

**A. Administrative**

1. Title:	<i>Tamil Nukta</i>	
2. Requester's name:	<i>Martin Hosken</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>10/Sep/2015</i>	
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:		<i>X</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>X</i>
Name of the existing block:	<i>Tamil</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 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?		
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. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?		
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	<i>NLCI</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?		<i>no</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?		<i>yes</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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> - Form number: N3102-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)

**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>Chetti community: NIRMAAN</i>  	<i>yes</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>no</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	  	<i>common</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>addition to existing BMP block</i>	<i>yes</i> <i>yes</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>no</i>
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 an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	  <i>This document</i>	<i>yes</i> <i>yes</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>This document</i>   	<i>yes</i> <i>yes</i> <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 character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	   	<i>no</i>