TO: **Unicode Technical Committee**

FROM: **Deborah Anderson, Script Encoding Initiative, UC Berkeley**

SUBJECT: **Proposal for encoding the Toto script in the SMP of the UCS**

DATE: **27 September 2019**

The Toto (txo) language has a population of only 1500 living in a single jungle village in India near Bhutan. The script for Toto was designed by Dhaniram Toto who is an elder in the Toto community. The script was officially launched in the community on 22nd May 2015. Until the development of this script, there was very little interest in language development. Since the development of their own script there is renewed interest in writing their own language. Having their own script has granted the Toto a sense of status. Because of this, it should be called the Toto script. Currently, there is limited use in the community. There is one NGO multi-lingual school which teaches the script in a half-hour weekly class. Work on a primer has begun but is not completed.

The script supports the 30 phonemes found in the language.

Tone

Tone is only used when the lack of it would confuse two words (the same could be said of vowel length). When tone is used it can either be rising or falling, and it is pronounced across the entire word (or phrase) - most easily heard in the final syllable. Tone is carried in the lexical stem but generally heard most in the suffix morphemes. A character for rising tone is included in this proposal. Falling tone is not marked. The tone marker appears only on vowels, and it is currently placed on the first vowel of the stem. (Figures 2,9)

Vowel Length

Another supra-segmental that is only used when required for contrast is vowel length. Vowel length is indicated by doubling the vowels (Figure 9).

Some words have a vowel sound that is repeated with a very short gap between. In these words an apostrophe (’) is used between the vowels (Figure 8). It is recommended that implementations use U+02BC MODIFIER LETTER APOSTROPHE to represent this.

Breathiness

Breathiness is only marked on vowels. Front vowels all have contrasting breathy versions, but not back vowels. Mid and close front vowels also have rounded versions that contrast. We have chosen to encode the breathy vowels rather than encode a combining mark. This means decompositions are not needed. (Figures 3, 4, 7, 10).

For those who are interested wish to understand more of the background, the following information may be useful. Toby Anderson (pc) believes that breathiness is a feature of the vowels and not of the consonants. He provides four pieces of evidence to support this:

1. There is at least one word that starts with a breathy vowel /æ̤puwɑ/ “decay” (Latin script orthography: aehpuwa, Toto script: 𞊬𞊐𞊥𞊜𞊭). /hæpuwɑ/ “gone” (Latin script orthography: haepuwa, Toto script: 𞊞𞊫𞊐𞊥𞊜𞊭) is not completely contrastive, but it is still a useful contrast with the breathy vowel.
2. No words have been found which end in an aspirated consonant (though closed syllables are possible).

3. If these are interpreted as breathy vowels, then there is a clean distribution: front vowels have breathy counterparts and back vowels do not. If one were to interpret these as aspirated consonants, then every consonant must have an aspirated counterpart, but the aspirated ones are never used at the end of a word or before a back vowel, which seems much less likely (though a linguist who wrote a grammar on Dhimal, the most closely related language to Toto, seemed to go with this unlikely interpretation for that language).
4. The numerals for 2 and 7 are minimal pairs on breathiness. "two" is /ni̤/ and "seven" is /ni/. The noun classifier for people is /-tʃo/ attached to the stem of each numeral up to 10. When attached to "two" this suffix is softened and pronounced [-ʃo] which doesn't happen on the "seven" This is clearly a phonological process caused by the breathiness. It is more likely for features to influence the neighbouring segment than a segment one step removed, so it's more likely that the breathiness is a feature of the vowel which is adjacent to the affected segment, rather than a feature of the consonant which is one step removed.

Diphthongs

There is some disagreement on whether there are diphthongs in the language. In the Toto script they are not written, and thus the phonology should not be an issue.

However, for those who are interested, the following information may be helpful. There are no glides or diphthongs in the language, although the front rounded vowels are often spoken as diphthongs of other vowels in normal speech. Basumatary and Wikipedia (see references) call them diphthongs. Samy does not discuss diphthongs.

In surface representation diphthongs are used, but in every instance one of two things is happening (pc with Toby Anderson):

1. It is a matter of interpretation between the vowel [i] and the consonant [j] or between the vowel [u] and the consonant [w]. To give an example the Toto word for “Betel Nut” is pronounced [guɑi] which can also be interpreted as [guwɑj] (Latin script orthography: guway, Toto script: 𞊕𞊣𞊜𞊨𞊛). Though T. Anderson has written this two different ways in IPA, it is pronounced exactly the same. So T Anderson questions whether Toto has diphthongs or not. The word for "flower" has the phonemes /mɑjbe/ (Latin script orthography: maybe, Toto script: 𞊖𞊨𞊛𞊑𞊤).
2. The front-rounded vowels in natural speech are generally pronounced as a diphthong consisting of the back-rounded vowel followed by the front-unrounded vowel of the same closeness. However in careful speech the underlying front-rounded vowel is clear. For example, the Toto word for "above" is naturally pronounced [toetɑ] but the underlying phonology is /tøtɑ/ which becomes more clear when you remove the locative suffix [-tɑ] and say the bare Toto word for "up" [tø]
(Latin script orthography: toeta, Toto script: 𞊒𞊥𞊒𞊨).

Other example words might be “village” [loi] /løj/ (Latin script orthography: loey, Toto script: 𞊘𞊥𞊛) and “hill” [jægoe] /jægø/ (Latin script orthography: yaegoe, Toto script: 𞊛𞊧𞊕𞊥). T. Anderson has not found any diphthong in the language that doesn't fit into one of the two categories above. Regarding words fitting into the first category they could be interpreted either way, but he has chosen to interpret them as not having diphthongs because of Occam's razor (“that in explaining a thing no more assumptions should be made than are necessary”). Why maintain the extra complexity of having diphthongs when there is an interpretation without them? According to T. Anderson, it seems much more likely that a language has /w/ and /j/ and often uses them alongside vowels in ways that could be interpreted as diphthongs, than that the language has diphthongs.

Character Names

On the final page is a chart of the characters and names of characters for the Toto script.

Toto language speakers would mostly pronounce them with the vowel [ɒ] (the vowel in the British English word "pot") following the consonant. This is because the government schooling is in Bangla medium, and Bangla names its own letters in that fashion - which in turn is because that's the inherent vowel (the default vowel used when no vowel marking is put on a consonant in a written word). However, if you look at the Unicode names for the Bangla letters you'll see they are written with an 'A' following the consonant, even though that's not how they are pronounced. The reason for this is that when Bangla is transcribed into Latin letters then 'A' is used for the inherent vowel and the reason for this is that it is conforming to how other Indic languages are transcribed - languages that have other pronunciations for the inherent vowel. This Toto writing system is not syllable-based and doesn't have an inherent vowel.

Digits

There are currently no script-specific digits. A column including 10 free slots may be needed in the future if the community decides to encode script-specific digits.

Punctuation

Latin script punctuation is currently being used with the Toto script. An apostrophe is also used to mark a glottal (Figure 8).

Potential issues

Since rising tone is only marked on vowels, there should not be any confusion on the similarity of the marks on TOTO LETTER NA and TOTO LETTER NGA.

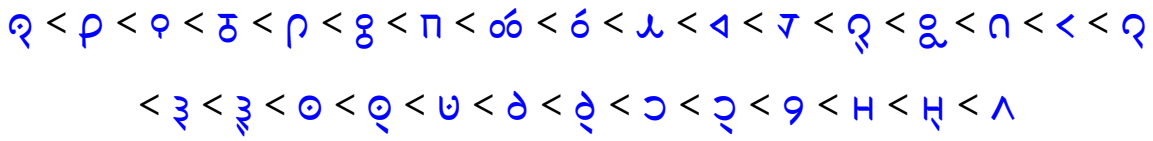
The mark that is on TOTO LETTER W is not a breathiness mark.

Unicode Character Properties

```
1E290;TOTO LETTER PA;Lo;0;L;;;;;N;;;;;
1E291;TOTO LETTER BA;Lo;0;L;;;;;N;;;;;
1E292;TOTO LETTER TA;Lo;0;L;;;;;N;;;;;
1E293;TOTO LETTER DA;Lo;0;L;;;;;N;;;;;
1E294;TOTO LETTER KA;Lo;0;L;;;;;N;;;;;
1E295;TOTO LETTER GA;Lo;0;L;;;;;N;;;;;
1E296;TOTO LETTER MA;Lo;0;L;;;;;N;;;;;
1E297;TOTO LETTER NA;Lo;0;L;;;;;N;;;;;
1E298;TOTO LETTER NGA;Lo;0;L;;;;;N;;;;;
1E299;TOTO LETTER SA;Lo;0;L;;;;;N;;;;;
1E29A;TOTO LETTER CHA;Lo;0;L;;;;;N;;;;;
1E29B;TOTO LETTER YA;Lo;0;L;;;;;N;;;;;
1E29C;TOTO LETTER WA;Lo;0;L;;;;;N;;;;;
1E29D;TOTO LETTER JA;Lo;0;L;;;;;N;;;;;
1E29E;TOTO LETTER HA;Lo;0;L;;;;;N;;;;;
1E29F;TOTO LETTER RA;Lo;0;L;;;;;N;;;;;
1E2A0;TOTO LETTER LA;Lo;0;L;;;;;N;;;;;
1E2A1;TOTO LETTER I;Lo;0;L;;;;;N;;;;;
1E2A2;TOTO LETTER BREATHY I;Lo;0;L;;;;;N;;;;;
1E2A3;TOTO LETTER IU;Lo;0;L;;;;;N;;;;;
1E2A4;TOTO LETTER BREATHY IU;Lo;0;L;;;;;N;;;;;
1E2A5;TOTO LETTER U;Lo;0;L;;;;;N;;;;;
1E2A6;TOTO LETTER E;Lo;0;L;;;;;N;;;;;
1E2A7;TOTO LETTER BREATHY E;Lo;0;L;;;;;N;;;;;
1E2A8;TOTO LETTER EO;Lo;0;L;;;;;N;;;;;
1E2A9;TOTO LETTER BREATHY EO;Lo;0;L;;;;;N;;;;;
1E2AA;TOTO LETTER O;Lo;0;L;;;;;N;;;;;
1E2AB;TOTO LETTER AE;Lo;0;L;;;;;N;;;;;
1E2AC;TOTO LETTER BREATHY AE;Lo;0;L;;;;;N;;;;;
1E2AD;TOTO LETTER A;Lo;0;L;;;;;N;;;;;
1E2AE;TOTO LETTER RISING TONE;Mn;230;NSM;;;;;N;;;;;
```

Collation

The community has not yet expressed a sort order. At this point, the dictionaries mentioned in the reference section are using the sort order used in the chart:



Figures

Figure 1. “Rising Sun” song in Toto script

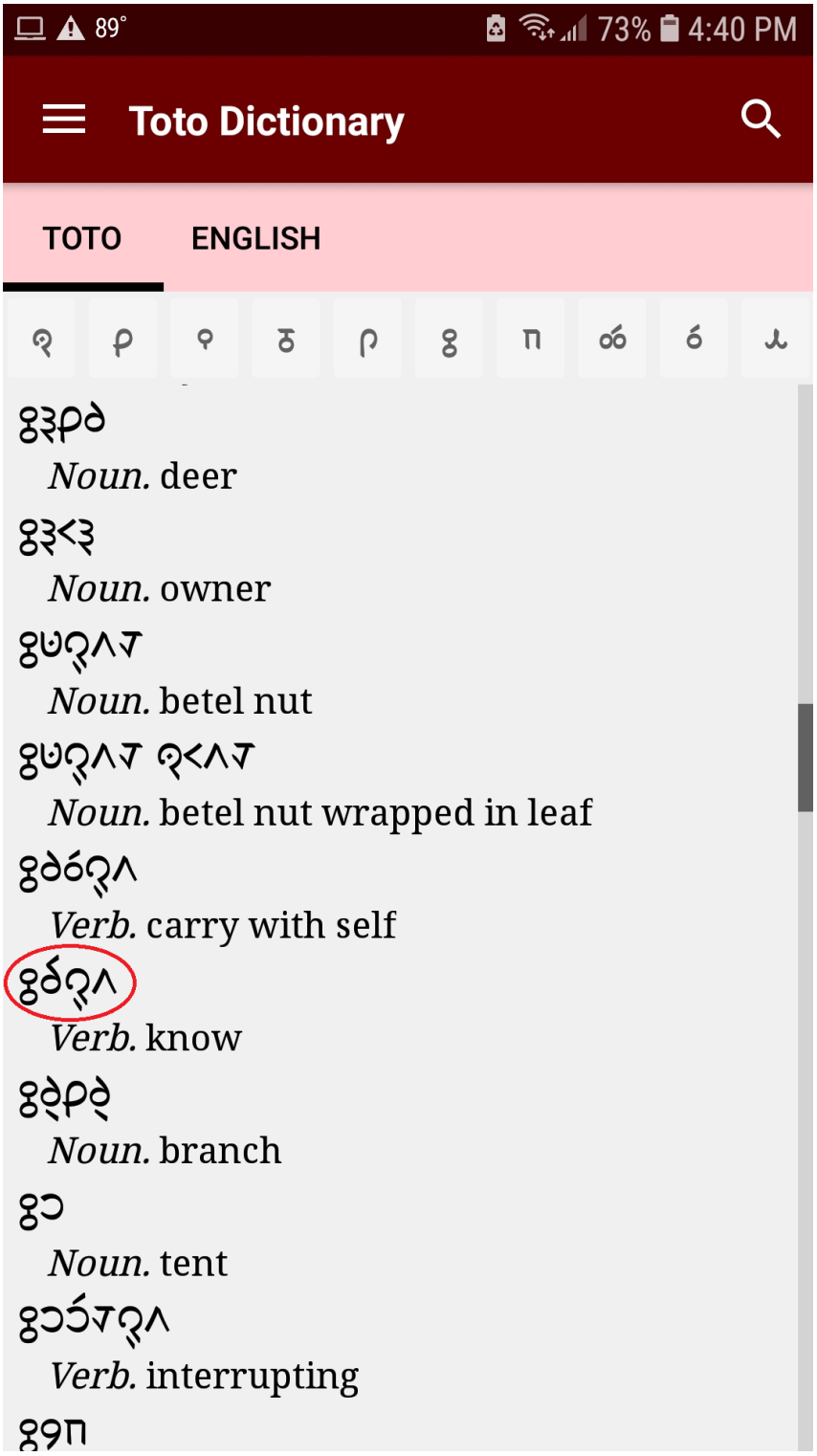


Figure 2. Android app of Toto dictionary. (Because the font does not have OpenType support, combining marks sometimes will be too close to the base character. This is an undesirable feature of non-smart font support.)

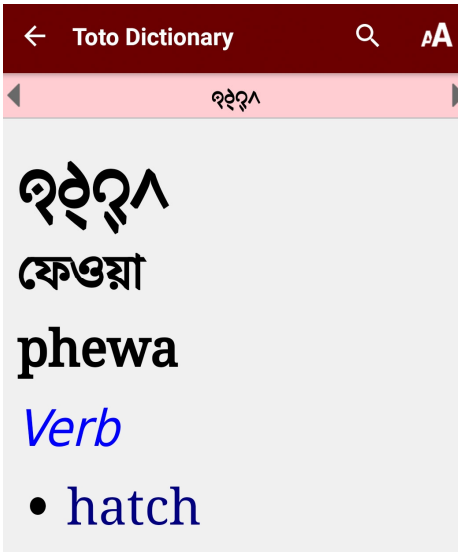


Figure 3. Android app of Toto dictionary. Breathiness is marked on following vowel (phe).

Figure 4. Android app of Toto dictionary. Breathiness is marked on following vowel (the).

Figure 5. “Story of the Bees” by Prakash Toto. Toto Writer’s Guide. TOTO SIGN RISING TONE is marked in red. (Because the font does not have OpenType support, combining marks sometimes will be too close to the base character. This is an undesireable feature of non smart-font support.)

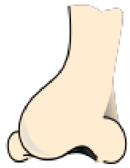
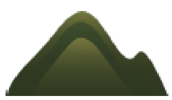
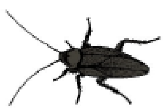
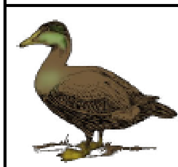
	[illegible]	ন নাবৌ	**n** naboe
	[illegible]	ঙ ঙোকা ং	**ng** ngoka
	[illegible]	স সা	**s** sa
	[illegible]	চ চেরেং	**ch** chereng
	[illegible]	য় য়্যাগী	**y** yaegoe
	[illegible]	ওয় ওয়্যাতী	**w** waeti
	[illegible]	জ জুসী	**j** jusi
	[illegible]	হ হাংসা	**h** hangsa

Figure 6. Example from Toto Writer's Guide demonstrating use of Toto script, Bengali script and Latin script in the Toto language.

[illegible]

জিয়া সো কীয়া লৌয়তা নীনা

jhiya so kiya loeyta nina

Figure 7. Example from Toto Writer's Guide demonstrating use of Toto script, Bengali script and Latin script in the Toto language. Breathiness is written on the following vowel.

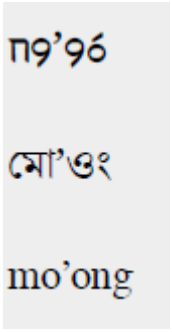


Figure 8. Example from Toto Writer's Guide demonstrating use of Toto script, Bengali script and Latin script in the Toto language. Vowel length with glottal.

Thinking	Coming out
ꠢꠖꠋ নোওয়া nowa	ꠢꠖꠋ নোঃওয়া nówa

Carry on head	Measure
ꠢꠖꠖꠋ নো-ওয়া no-wa	ꠢꠖꠖꠋ নোঃ-ওয়া nó-wa

“Thinking” /nowɑ/ ꠢꠖꠋ (short vowel, falling tone)

“Coming out” /↗nowɑ/ ꠢꠖꠋ (short vowel, rising tone)

“Carry on head” /no:wɑ/ ꠢꠖꠖꠋ (long vowel, falling tone)

“Measure” /↗no:wɑ/ ꠢꠖꠖꠋ (long vowel, rising tone)

Figure 9. Example from Toto Writer's Guide demonstrating four-way contrast (last four lines are not from the writer's guide, they are explanatory)

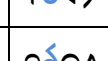
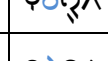
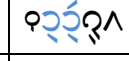
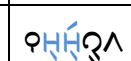
water	/ti/	
carrying	/↗t̤i:wɑ/	
adopting	/↗ty:wɑ/	
laying egg	/↗t̤y:wɑ/	
six	/tutʃo/	
walking	/↗tewɑ/	
sweet	/t̪ewɑ/	
ending	/↗tø:wɑ/	
hanging up	/↗t̤ø:wɑ/	
throwing	/↗to:wɑ/	
crossing river	/↗tæ:wɑ/	
slipping	/↗t̪æ:wɑ/	
feeding	/tɑ:wɑ/	

Figure 10. All vowels in contrast with breathy vowels (where there is a breathy version).

References

Anderson, Toby and Abhishek Toto (compilers). 2018. Toto Writer's Guide. [illegible]. Unpublished.

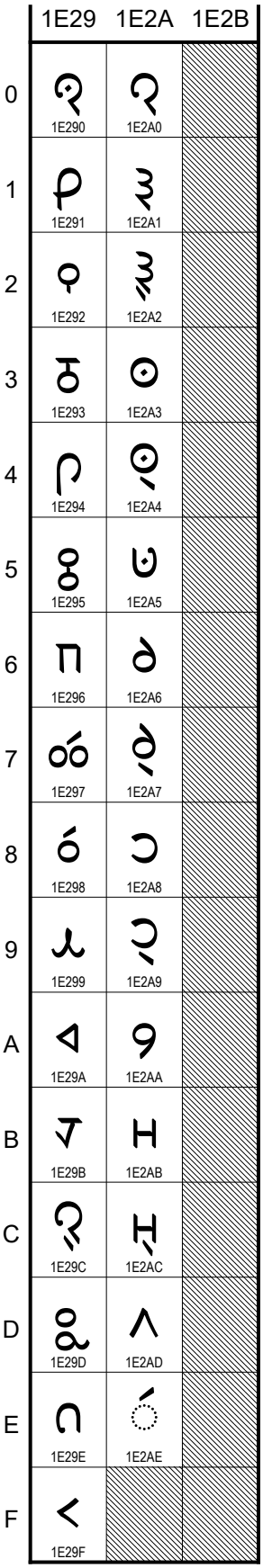
Android app of a 680 word dictionary here https://play.google.com/store/apps/details?id=in.threestrands.txo.lexicon.

Basumatary, Chibiram June 2014. *The Phonological Study of Toto Language*. LANGUAGE IN INDIA. Volume 14:6June2014ISSN 1930-2940 http://languageinindia.com/june2014/chibiramtotophonology1.pdf .

Samy, P.Perumal. *A Linguistic Description of Toto Language Spoken in West Bengal*. https://www.academia.edu/30313328/A_Linguistic_Description_of_Toto_Language_Spoken_in_West_Bengal

Toto Dictionary (https://toto.webonary.org/). Out-of-date compared to the Android app.

Wikipedia article on Toto. https://en.wikipedia.org/wiki/Toto_language.



Basic Consonants

1E290 TOTO LETTER PA
1E291 TOTO LETTER BA
1E292 TOTO LETTER TA
1E293 TOTO LETTER DA
1E294 TOTO LETTER KA
1E295 TOTO LETTER GA
1E296 TOTO LETTER MA
1E297 TOTO LETTER NA
1E298 TOTO LETTER NGA
1E299 TOTO LETTER SA
1E29A TOTO LETTER CHA
1E29B TOTO LETTER YA
1E29C TOTO LETTER WA
1E29D TOTO LETTER JA
1E29E TOTO LETTER HA
1E29F TOTO LETTER RA
1E2A0 TOTO LETTER LA

Basic Vowels

1E2A1 TOTO LETTER I
1E2A2 TOTO LETTER BREATHY I
1E2A3 TOTO LETTER IU
1E2A4 TOTO LETTER BREATHY IU
1E2A5 TOTO LETTER U
1E2A6 TOTO LETTER E
1E2A7 TOTO LETTER BREATHY E
1E2A8 TOTO LETTER EO
1E2A9 TOTO LETTER BREATHY EO
1E2AA TOTO LETTER O
1E2AB TOTO LETTER AE
1E2AC TOTO LETTER BREATHY AE
1E2AD TOTO LETTER A

Various signs

1E2AE TOTO SIGN RISING TONE

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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 for encoding the Toto script in the SMP of the UCS*
2. Requester's name: *Deborah Anderson, Script Encoding Initiative, UC Berkeley*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *27 September 2019*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Y*
   (or) More information will be provided later: *N*

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Y*
      Proposed name of script: *Toto*
   b. The proposal is for addition of character(s) to an existing block: *N*
      Name of the existing block:
2. Number of characters in proposal: *31*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary x 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? *Y*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Y*
   b. Are the character shapes attached in a legible form suitable for review? *Y*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   author
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Open Font License
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Y*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Y*
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)? *Y*

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? N
   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.)? Y
   If YES, with whom? *Toto script users*
   If YES, available relevant documents: *See proposal*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Y
   Reference: *See proposal*
4. The context of use for the proposed characters (type of use; common or rare) Y
   Reference: *See proposal*
5. Are the proposed characters in current use by the user community? Y
   If YES, where? Reference: *See proposal*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? N
   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)? Y
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? N
   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? Y
   If YES, is a rationale for its inclusion provided? Y
   If YES, reference: *See proposal*
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? N
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? Y
    If YES, is a rationale for such use provided? Y
    If YES, reference: *See proposal*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? N
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? N
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? N
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: