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://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:** Proposal to encode a Tifinagh punctuation character and update TUS documentation
2. Requester's name: Lorna A. Priest, Jon Coblentz
3. Requester type (Member body/Liaison/Individual contribution): Individual contribution
4. Submission date: 5 May 2008
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): No
        Proposed name of script:
    b. The proposal is for addition of character(s) to an existing block: Yes
        Name of the existing block: Tifinagh
2. Number of characters in proposal: 1
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. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 1
    Is a rationale provided for the choice? No
        If Yes, reference:
5. Is a repertoire including character names provided? Yes
    a. If YES, are the names in accordance with the "character naming guidelines" Yes
        in Annex L of P&P document?
    b. Are the character shapes attached in a legible form suitable for review? Yes
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? SIL International
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
7. 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
8. 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)?
    Yes, suggested character properties are included (see section D)
9. 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.

---

[1] Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   No

    If YES explain _____

2. Has contact been made to members of the user community (for example: National Body,   Yes
    user groups of the script or characters, other experts, etc.)?

        If YES, with whom?   script experts: Andrew Savage, Christian Grandoullier

        If YES, available relevant documents: _____

3. Information on the user community for the proposed characters (for example:
    size, demographics, information technology use, or publishing use) is included?

    Reference: See comments in Section E

4. The context of use for the proposed characters (type of use; common or rare)   common

    Reference: See comments in Section E

5. Are the proposed characters in current use by the user community?   Yes

    If YES, where? Reference: See comments in Section E

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
    in the BMP?   Yes

        If YES, is a rationale provided?

        If YES, reference: should be kept with Tifinagh block

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   No
    character or character sequence?

        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?   No

        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)   No
    to an existing character?

        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?   No

        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   No
    control function or similar semantics?

        If YES, describe in detail (include attachment if necessary) _____

13. Does the proposal contain any Ideographic compatibility character(s)?   No

        If YES, is the equivalent corresponding unified ideographic character(s) identified? _____

        If YES, reference: _____

D.1. Proposed Characters

D.2. Character Names

	2D7
0	⵰
1	
2	
3	
4	
5	
6	
7	
8	
9	
A	
B	
C	
D	
E	
F	

2D70   TIFINAGH TAZARAST

D.3. Unicode Character Properties

`2D70;TIFINAGH TAZARAST;Po;0;L;;;;;N;;;;;`

E. Other Information

E.1 Punctuation

Section D.4 (p. 32) of the Unicode proposal for the Tifinagh script (n2739R.pdf) states that "No specific Tifinagh punctuation marks are known." The footnote (in the original proposal) to this states:

> "Prasse, [20] p. 152, however, mentions the presence in Ahaggar of a separator ⊂ « à l'intérieur duquel s'écrit la dernière lettre de chaque mot phonétique ». Its usage does not seem systematic; letter 23 (a) in [10] makes use of a similar symbol. Savage [21] also shows a few recent samples of this separator. No trace of this separator was found in recent works and we have thus preferred not to code this character at this stage."

The document goes on to say:

> "Ircam has decided that the conventional punctuation signs used in the Latin script shall be used in Tifinagh: " " (space), ".", ",", ";", ":", "?", "!", "…", etc. Consequently, this proposal does not encode any Tifinagh punctuation sign."

What was not known at the time n2739R was written, is that the word division character is actually in current use, and we believe it should be encoded. U+2D70 TIFINAGH TAZARAST has traditionally been a "word break" as often "spaces" were not used to separate words - a message was simply a long string of Tifinagh characters, of which the "Tazarast" is used to indicate word separation. In a world more and more influenced by other written languages (French or Arabic), Tifinagh users tend to revert to the convention of blank space to indicate word break. For example, in Algeria a teacher of Tifinagh uses blank spaces and doesn't use this character. However, in Mali, many of the older traditional Tifinagh writers use it, and it is also used in Niger where it can be used in the place of a full stop as well as a comma. In the 03143-tifinagh-samples.pdf (Savage, 2003. pp. 10-12). Savage provided examples of usage in handwritten right-to-left texts. These examples were from Algeria (thus we know of usage in Niger, Mali and Algeria). Figure 1 (from 03143-tifinagh-samples.pdf) illustrates the use of the word division character. Because of the right to left text the character is mirrored. This particular example is from "a couple of decades ago" and illustrates long term usage of this character. Figures 2 and 3 illustrate usage in two languages in Niger where the glyph is oriented for a left to right texts.

In regards to bidi properties, Tifinagh *can* be written right to left. The script has been encoded in Unicode as "strong left to right" (TUS, p 458). In keeping with this, we have suggested a bidi value of "L" for the Tazarast. Regarding mirroring, if the script is used right to left then the Tazarast would need to be mirrored. However, we have not suggested Mirrored=Y because TUS has said it is a left to right script. If this needs to change, we are happy to change it to 'Y'.

In regards to the name for this character. Təzərəzt, pl. šizəraz (also written tazərəzt) comes from the root r-z (ərəz: to cut, to stop) (in the Tamajeq dialect in Niger). It is a substantive built on the causative form (i.e. make a cut, make a stop). It means: end, limit. It is something that marks a stop; it could even be something physical marking a limit, an end, like an object, a gate, a property landmark; something that draws limits.

This same name is also used in the Tayərt (Agadez) dialect, the other major dialect in Niger; it is written tezzərəzt, pl. tyəzraz. It is also used in Algeria. We are unaware of whether this term is used in the other main dialect in Mali. It is likely to exist there too as the root is a Təmajəq/Berber root.

Figure 1. TIFINAGH TAZARAST (Savage, 2003. p. 11)

Figure 2. TIFINAGH TAZARAST (SIM, 2004. p. 1)

Figure 3. TIFINAGH TAZARAST (Mission Baptiste. p. 1)

E.2 Changes to 13.4 in TUS.

We would also like to suggest updating the documentation on pp.457-458 of the TUS 5.0 with the following.

E.2.1 Bi-consonants

The bi-consonantal characters in Tifinagh indicate where there is a consonant cluster, and therefore where there is *not* a vowel to be pronounced between the two consonants. The Unicode Standard does not assign independent code points for the bi-consonants. Instead, bi-consonants can be achieved by placing the ZERO WIDTH JOINER between the two consonants. The glyph for U+200D ZERO WIDTH JOINER shown in the code charts is arbitrary and is not actually rendered directly; the dotted box around the glyph indicates that special rendering is required. While the Unicode encoding for a bi-consonant is thus three codepoints, the sequence displays with a single glyph. Figure 13.x illustrates the use of the bi-consonants. Note that this is by no means a comprehensive list of bi-consonants. Note also that there are glyph variants for these bi-consonants depending on the region.

Figure 13.x Using ZERO WIDTH JOINER to create bi-consonants

⊏ 2D4E	+	ZWJ 200D	+	+ 2D5C	➡
| 2D4F	+	⋮ 200D	+	⋮ 2D3E	➡
| 2D4F	+	ZWJ 200D	+	+ 2D5C	➡
◯ 2D54	+	ZWJ 200D	+	+ 2D5C	➡
⊙ 2D59	+	ZWJ 200D	+	+ 2D5C	➡

E.2.2 Contextual shaping

Besides bi-consonants, there is sometimes contextual shaping of characters. This occurs when "ll" (ⵍⵍ), "ln"(ⵍⵏ), "nl" (ⵏⵍ) or "nn" (ⵏⵏ) appear together. This contextual shaping can occur in order to distinguish between characters when like pairs appear next to each other. To write this, the 2nd character is shifted vertically or it can be slanted. Figure 13.y illustrates the use of contextual shaping.

Figure 13.y Contextual shaping

‖ 2D4D	+	‖ 2D4D	➡	‖\\
‖ 2D4D	+	| 2D4F	➡	‖\
| 2D4F	+	‖ 2D4D	➡	\\\\
| 2D4F	+	| 2D4F	➡	\\\

F. References

Andries, P. 2004-06-06. *Proposal to add the Tifinagh Script.* N2739R ISO/IEC JTC 1/SC 2/WG 2 N2739. http://anubis.dkuug.dk/jtc1/sc2/wg2/docs/n2739.pdf.

Mission Baptiste – Agadez. In press. *Ayuba*. Mission Baptiste: Agadez, Niger

Savage, Andrew. 2003. Tifinagh Unicode Propositions, Stanthorpe, http://www.unicode.org/L2/L2003/03143-tifinagh-samples.pdf

The Unicode Consortium. *The Unicode Standard*, Version 5.0.0, defined by: The Unicode Standard, Version 5.0 (Boston, MA, Addison-Wesley, 2007. ISBN 0-321-48091-0)

SIM. 1997, 2004. *Isălan n-Ălkher win iktăb Luk* (The Gospel of Luke in the Tamajeq Language). SIM, Niamey, Niger. (http://www.westafricanscripts.com/ accessed 2008-01-25)