

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://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 additional Tifinagh characters

2. Requester's name: Lorna A. Priest, Jon Coblentz, Andrew Savage

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 11 July 2008

5. Requester's reference (if applicable): L2/08-198R

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: 2

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): 3 (one joiner)

Is a rationale provided for the choice? Yes

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" in Annex L of P&P document? Yes

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? If YES explain _____	No
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? <u>script experts: Andrew Savage, Christian Grandoullier</u> If YES, available relevant documents: _____	Yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: <u>See comments in Section E</u>	
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>See comments in Section E</u>	common
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>See comments in Section E</u>	Yes
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: <u>should be kept with Tifinagh block</u>	Yes
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? _____ 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? If YES, is a rationale for its inclusion provided? _____ If YES, reference: <u>See comments in Section E.</u>	Yes Yes
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: <u>See comments in Section E.</u>	Yes Yes
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: _____	No Yes
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) <u>See comments in Section E</u>	Yes
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? _____ If YES, reference: _____	No

D.1. Proposed Characters

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

D.2. Character Names

Punctuation

2D70 TIFINAGH SEPARATOR MARK
=Tazarast

Various signs

2D7F TIFINAGH CONSONANT JOINER

- suppresses inherent vowel, functions to indicate the previous character and the following character form a bi-consonant
- shape shown is arbitrary and is not visibly rendered

D.3. Unicode Character Properties

```
2D70;TIFINAGH SEPARATOR MARK;Po;0;L;;;;N;;;;;  
2D7F;TIFINAGH CONSONANT JOINER;Mn;9;NSM;;;;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 C « 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 SEPARATOR MARK 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.

In regards to the alias for this character. Təzərəzt, pl. šizəraz (also written tazərəzt) comes from the root r-z (əraz: to cut, to stop) (in the Təwəlləmmət 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.

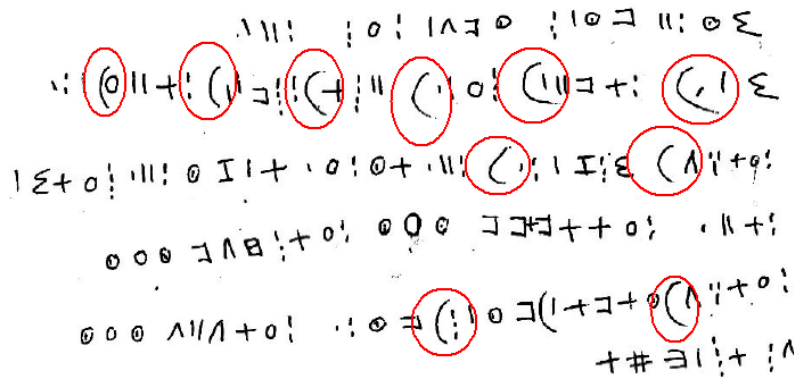


Figure 1. TIFINAGH SEPARATOR MARK (Savage, 2003. p. 11)

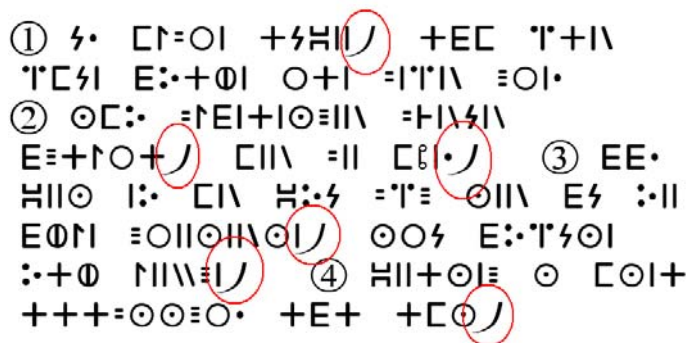


Figure 2. TIFINAGH SEPARATOR MARK (SIM, 2004. p. 1)

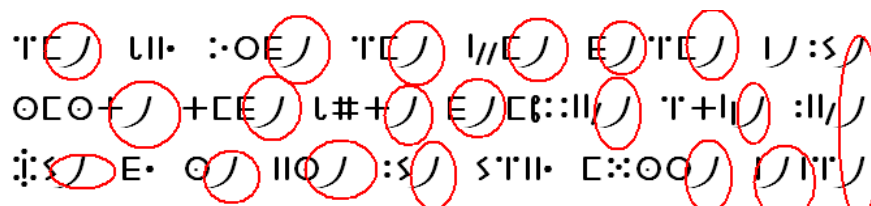


Figure 3. TIFINAGH SEPARATOR MARK (Mission Baptiste. p. 1)

E.2 Various signs

E.2.1 The Use of Tifinagh Bi-consonantal Characters

Tifinagh is the ancient and original script of the Tuareg people of the Sahara desert. The Tuareg language is related to the Semitic languages, and like those languages (including Arabic and Hebrew) it lends itself to a “consonants-only” script.

The majority of the languages of the world could not be written conveniently with a consonantal (minimum vowel representation) script, however due to the unique syllabic structure of the Semitic and related languages, this is not only possible, but preferable.

One of the features of the syllabic structure of the Tuareg language is that there are never more than two consonants in a cluster eg. *Tefert*, *temarwelt*, *ilsa*, *talamt*, *enta*, *tajrist* etc, as opposed to multiple consonant clusters in other languages, such as in English, eg. *Stride*, *clumps*, *thirsts* etc. Another feature is that there are no vowel clusters.

These two features lend themselves perfectly to the use of bi-consonantal characters (sometimes called bi-graphs).

There are many different variations in the way Tifinagh is written. Some of the basic variations include character shape and character orientation (left to right, or right to left). Other, more ‘specialised’ variations include vowel representation (which we are not addressing in this proposal as we do not believe there is any general consensus, as yet) and usage (total, partial or non-usage) of bi-consonantal characters. Even if additional vowel characters are encoded in the future, this would not have any affect on how the TIFINAGH CONSONANT JOINER is used.

The character shape and character orientation variations tend to be region specific, having developed in relatively isolated environments in much the same way as spoken dialect variation develops.

The usage of bi-consonantal characters is a device to help disambiguate words that are written with limited vowel representation. A similar device is used in the Arabic writing system which helps to understand the value of the Tuareg bi-consonantal characters. This device is called the ‘sukun’ (a small circle) which is placed over a consonant, as in the second word [ənkər] in Table 1, which looks like this: انكر. That sukun signifies that there is *no* vowel following that consonant (the “n”), ie. that the “n” and the “k” form a consonant cluster. If that ‘sukun’ was not shown, there would be several additional possible renderings for that word, ie. you could pronounce it with different vocalic possibilities. As you can imagine, the sukun greatly limits the number of possible renderings of that string of three consonants “nkr” and hence makes deduction of the correct meaning readily accessible.

	Meaning	Phonetic	Arabic script	Tifinagh
1.	‘get up’ IMP. habitual	[nâk:ər]	ناكر /nagr/	l:○ /nkr/
2.	‘get up’ IMP non hab.	[ənkər]	انكر /’nkr/	ḷ:○ /nkr/

Table 1. Two Tuareg words: written in both the Arabic script and the Tifinagh script

The bi-consonantal characters in Tifinagh perform exactly the same function, ie. they indicate where there is a consonant cluster, and therefore where there is *not* a vowel to be pronounced between the two consonants.

Without using the bi-consonantal characters, each of the two example words in the above table would be written identically in Tifinagh, ie. l:○, and there would be potential confusion regarding the correct ‘reading’ of the word. However, by using the bi-consonantal character ḷ:, ie. ”nk”, there can only be one ‘reading’ of that word. And consequently, if we consistently use the bi-consonantal characters, we can deduce that there *is* a vowel to be pronounced between the “n” and the “k” of our first example word since we did *not* make use of a bi-consonantal character in that instance.

One can see that there is not an automatic combination of two consonantal characters to give a bi-consonantal character: the two consonants can be represented separately (side by side) or as a single bi-consonantal character depending on the presence or not of a vowel.

The following chart lists contrastive pairs of vocalic and non-vocalic clusters. An examination of these examples clearly shows why bi-consonants are needed in the script.

Non-vocalic clusters (using bi-consonants)				Vocalic clusters (no bi-consonant use)			
Script		Examples		Script		Examples	
Tifinagh	Latin	Tifinagh	Latin	Tifinagh	Latin	Tifinagh	Latin
ⵎⵔ	mb	ⵎⵔ	imbal	ⵎⵔ	m _{/v} b	ⵎⵔⵔ	šemeber
ⵔⵓ	rt	+ⵎⵔⵔ	təməddurt	ⵔⵓ	r _{/v} t	ⵔⵓⵔ	əstizarat
ⵔⵓ	lt	+ⵎⵔⵔ	tədhəlt	ⵔⵓ	l _{/v} t	ⵔⵓⵔ	šilat
ⵔⵓ	št	ⵔⵓⵔ	aštərot	ⵔⵓ	š _{/v} t	+ⵔⵓⵔ	tawšetən
ⵔⵓ	mt	+ⵎⵔⵔ	təryəmt	ⵔⵓ	m _{/v} t	+ⵔⵓⵔ	təmatay
ⵔⵓ	ft	ⵔⵓⵔ	annafat	ⵔⵓ	f _{/v} t	+ⵔⵓⵔ	təbbəffet
ⵔⵓ	ms	+ⵎⵔⵔ	təmsay	ⵔⵓ	m _{/v} s	ⵔⵓⵔ	əmosan
ⵔⵓ	nd	ⵔⵓⵔ	iyyanda	ⵔⵓ	n _{/v} d	ⵔⵓⵔ	šinad
ⵔⵓ	nf	ⵔⵓⵔ	infəs	ⵔⵓ	n _{/v} f	+ⵔⵓⵔ	tənafut
ⵔⵓ	ng	ⵔⵓⵔ	əngəm	ⵔⵓ	n _{/v} g	ⵔⵓⵔ	ənəgmar
ⵔⵓ	ny	ⵔⵓⵔ	inyal	ⵔⵓ	n _{/v} y	ⵔⵓⵔ	ilatanay
ⵔⵓ	nj	ⵔⵓⵔ	šinjar-net	ⵔⵓ	n _{/v} j	ⵔⵓⵔ	dak-za-najjawwab
ⵔⵓ	nk	ⵔⵓⵔ	jīwanken	ⵔⵓ	n _{/v} k	ⵔⵓⵔ	əmənokāl
ⵔⵓ	ns	+ⵎⵔⵔ	təsənsəq	ⵔⵓ	n _{/v} s	+ⵔⵓⵔ	təmənesen
ⵔⵓ	nt	ⵔⵓⵔ	əntanay	ⵔⵓ	n _{/v} t	ⵔⵓⵔ	əddənət
ⵔⵓ	nz	ⵔⵓⵔ	izakkənzar	ⵔⵓ	n _{/v} z	ⵔⵓⵔ	ənezər
ⵔⵓ	st	ⵔⵓⵔ	əstizarat	ⵔⵓ	s _{/v} t	+ⵔⵓⵔ	təmosat

Table 2. Tifinagh bi-consonant example chart (Tawallammat dialect)

Table 3 illustrates pairs of words which are exactly the same in the Tifinagh script were it not for the bi-consonants. The words have different meanings.

Tifinay word	Bi-cons. used	Transliteration	English Gloss (simplified)
ⵜⵉⵎⵓⵏ	ⵜ	inbal	“he buried”
ⵉⵎⵓⵏ	n/a	anabal	“burial”
ⵜⵉⵎⵓⵏ	ⵜ	ərtək	“to fall vertically”
ⵉⵎⵓⵏ	n/a	aratak	“the fact of f. vert.”
ⵜⵉⵎⵓⵏ	ⵜ	əstəy	“to chase”
ⵉⵎⵓⵏ	n/a	asatay	“the fact of chas.”
ⵜⵉⵎⵓⵏ	ⵜ	tastəq	“song, way of sing.”
ⵉⵎⵓⵏ	n/a	tasetəq	“infection”
ⵉⵎⵓⵏ	ⵉ	ənyəl	“to pour, spread”
ⵉⵎⵓⵏ	n/a	aneyəl	“copy, matrix”

Table 3. Tifinagh pairs of words with and without bi-consonant (Tawallammat dialect)

Table 4 consists of six pairs of words comparing the use and non-use of bi-consonantal characters. The first five pairs of words are taken from the “Dictionnaire touareg-français” of Charles de Foucauld. Scanned segments of the entries for those words, along with their Tifinagh representation, are included.

Regarding these scanned sections, please note the following:

1. The transcription (Latin characters) in the table does not exactly match that of de Foucauld. This is of no consequence.
2. The Tifinagh representation of de Foucauld is oriented from right to left, whereas in the table it is left to right.
3. Not all the bi-consonantal characters used in the table are the same as those used by Foucauld, as they were unavailable in the font. This is of no consequence as variant characters are very common.
4. Bi-consonantal characters used (in the table and by de Foucauld) are the following:

nk = ⵏⵓ

rt = ⵔ or ⵓ

st = ⵔ or ⵓ

nd = ⵏⵉ

nt = ⵏⵓ

5. The sixth word includes a direct object suffix, and so is not available for comparison from the de Foucauld dictionary.
6. All words are represented (for comparison’s sake) in the Arabic script, showing how the sukun functions in the same way as the Tifinagh bi-consonantal characters, ie. indicating where there is *not* a vowel to be pronounced between the consonants, ie. a consonant cluster.
7. Arabic and Tifinagh representations are transcribed in Latin characters for reference.
8. The final column of the chart is the reference to the page number of de Foucauld’s dictionary.

E.2.2 ZWJ vs TIFINAGH CONSONANT JOINER

Section D.8 (pp.34-36) of the Unicode proposal for the Tifinagh script (n2739R.pdf) discusses the “vast repertoire² of Tuareg ligatures (sometimes called “biconsonants”)” (p. 34). It concludes:

“This proposal neither encodes nor prescribes any ligature. It leaves the choice of ligatures to the font chosen at rendering time. Some fonts may be deprived of any ligature; others may include only some ligatures for a given geographical variant. Users may use the usual convention of inserting a U+200D ZERO WIDTH JOINER (resp. U+200C ZERO WIDTH NON-JOINER) to encourage (resp. discourage) the formation of a “biconsonant” at rendering time.” (p. 36).

Although we have listed some bi-consonants in the chart above, there are ever increasing numbers of them as is referenced by the original Tifinagh proposal – “vast repertoire...” If one were to encode each bi-consonant there would likely be an ever growing number that needed encoding. Rather than proposing the addition of each bi-consonant, we are proposing the addition of U+2D7F TIFINAGH CONSONANT JOINER. This would be different than the ZWJ in that it would not be default ignorable. It would have the Mn (Mark, Non-Spacing) property rather than Cf (Other, Format). We are aware of at least one font implementation which uses the ZWJ to form bi-consonants. We do not believe that using the ZWJ is the right decision. ZWJ is default ignorable and bi-consonants in certain languages are only formed when there is no inherent vowel. In searching and analysis, these bi-consonants are considered different from when two characters appear together. If the ZWJ is used, they are processed as the same characters, because the ZWJ is default ignorable. In fact, they should be processed separately. TUS (p. 537) states that the ZWJ is “meant only for overriding the normal behavior of the text.” Because the bi-consonants are not used just as “normal behavior”, but indicate *different meaning to the text* we believe that the ZWJ is inappropriate. For example l: “nk” (with inherent vowel) has a different meaning than ḥ: “nk” (no vowel). While some dialects use the bi-consonants as optional, there are a number of dialects which use the bi-consonants in a consistent manner; in those dialects it would be considered a spelling error if the bi-consonant was not used.

An advantage to *not* using ZWJ and using the TCJ is that it doesn't have to be checked for in domain names. Bi-consonants should not look like sequences of consonants, they should appear in their ligated form, this will automatically maintain this distinction.

For these reasons we believe the TIFINAGH CONSONANT JOINER should be encoded so that the bi-consonants will be processed separately in searching, analysis and eventually in domain names.

Although the TIFINAGH CONSONANT JOINER does not exist in the script, there is precedent for introducing such a character in the script in the Coeng encoding model (Everson).

E.2.3 Rendering of TIFINAGH CONSONANT JOINER

It is clear that if U+2D7F is used in text, the bi-consonant should be rendered. Since TIFINAGH CONSONANT JOINER is part of the spelling of a word, it must always impact rendering either by creating a bi-consonant or by being rendered as a mark of some kind to show that a TIFINAGH CONSONANT JOINER is part of the character sequence. If a font does not implement the bi-consonant, there should be a visible display of 2D7F. For example, 2D54 + 2D7F + 2D5C should render as: ☉ if the font has implemented the bi-consonant and as ○ ḥ: + if the font has not implemented the bi-consonant.

² Aghali-Khalil (1994, p. 118) indicates awareness of 82 bi-consonants although he does not list them. We have provided a listing in section F.1 of a number of bi-consonants.

U+2D7F cannot occur before *or* after anything but a Tifinagh consonant. These are:

(⊖ 2D31) (⊕ 2D32) (⊗ 2D33) (⊗ 2D34) (⊗ 2D35) († 2D36) (∧ 2D37) (∨ 2D38) (∃ 2D39) (⊙ 2D3A) (⊙ 2D3B) (⊙ 2D3C) (⊙ 2D3D) (⊙ 2D3E) (⊙ 2D3F) (⊙ 2D40) (⊙ 2D41) (⊙ 2D42) (⊙ 2D43) (⊙ 2D44) (⊙ 2D45) (⊙ 2D46) (⊙ 2D47) (⊙ 2D48) (⊙ 2D49) (⊙ 2D4A) (⊙ 2D4B) (⊙ 2D4C) (⊙ 2D4D) (⊙ 2D4E) (⊙ 2D4F) (⊙ 2D50) (⊙ 2D51) (⊙ 2D52) (⊙ 2D53) (⊙ 2D54) (⊙ 2D55) (⊙ 2D56) (⊙ 2D57) (⊙ 2D58) (⊙ 2D59) (⊙ 2D5A) (⊙ 2D5B) (⊙ 2D5C) (⊙ 2D5D) (⊙ 2D5E) (⊙ 2D5F) (⊙ 2D60) (⊙ 2D61) (⊙ 2D62) (⊙ 2D63) (⊙ 2D64) (⊙ 2D65) (⊙ 2D6F)

E.2.4 Input (keyboarding) of Tifinagh script

Regarding the keyboarding of consonant clusters as bi-consonantal characters, the choice needs to remain with the keyboarder himself for each individual word. It simply does not work, for example, to automatically convert keystroke “n” followed by keystroke “k” to the bi-consonantal character “nk”, as often those two Tifinagh characters “n” and “k” will need to remain separate, indicating that there is a vowel between them (though not written) that needs to be pronounced. Refer to the Tifinagh rendering of the two words in Table 1.

Regarding the bi-consonants themselves, it has been said there are two types of consonant clusters that they represent, i.e. those clusters with “n” as the first consonant and those with “t” as the final consonant. This is generally true, but there are a considerable number of exceptions, some of which can be seen in Section F.

E.2.5 Sorting

TIFINAGH SEPARATOR MARK should be sorted after 2D6F TIFINAGH MODIFIER LETTER LABIALIZATION MARK. By default, the usage of 2D7F (TIFINAGH CONSONANT JOINER) should not change the sort order (see section F.2). However, it should be possible for the sort order to be tailored for languages where bi-consonants are sorted separately from the regular consonants.

E.3 Changes to 13.4 in TUS.





















The documentation on pp.457-458 of the TUS 5.0 should be updated. We would suggest the following.

E.3.1 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 consonants³. The Unicode Standard does not assign independent code points for the bi-consonants. Instead, bi-consonants can be achieved by placing the TIFINAGH CONSONANT JOINER between the two consonants. The glyph for U+2D7F TIFINAGH CONSONANT 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 functions as if it had been encoded as a single character. Figure 13.x illustrates the use of the bi-consonants. It may also be possible to have tri-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.

³ There is the possibility of tri-consonants, however they have traditionally been called bi-consonants which is why we are using this term. We would also accept the term “ligature”.













Figure 13.x Using TIFINAGH CONSONANT JOINER to create bi-consonants

	+		+		➡	
2D4E		2D7F		2D5C		
	+		+		➡	
2D4F		2D7F		2D3E		
	+		+		➡	
2D4F		2D7F		2D5C		
	+		+		➡	
2D54		2D7F		2D5C		
	+		+		➡	
2D59		2D7F		2D5C		

E.3.2 Contextual shaping

Besides bi-consonants, there is sometimes contextual shaping of characters. This occurs when "ll" (lll), "ln" (lll), "nl" (lll) or "nn" (ll) 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. Extra Information

F.1 Chart of various Bi-Consonants

Here are listed some of the bi-consonants in use for various dialects. This is not a comprehensive list and is provided solely for those who are unfamiliar with the script to see the varieties used.

	transcription	Bi- tri- consonants	CG	DE	AS	PA	Andries (2005)						
							0	1	2	3	D	E	F
1.	bt	𐤁𐤕					𐤁𐤕	𐤁𐤕					
2.	dj	𐤃𐤍		𐤃𐤍									
3.	dt	𐤃𐤕					𐤃𐤕						
4.	ft	𐤕𐤃	𐤕𐤃		𐤕𐤃		𐤕𐤃						
5.	gt	𐤍𐤕					𐤕𐤍						
6.	ġt	𐤍𐤕					𐤕𐤍						
7.	lt	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕		𐤕𐤕	𐤕𐤕	𐤕𐤕				
8.	mb	𐤕𐤁	𐤕𐤁	𐤕𐤁							𐤕𐤁	𐤕𐤁	𐤕𐤁
9.	ms	𐤕𐤓	𐤕𐤓										
10.	mt	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕		𐤕𐤕	𐤕𐤕	𐤕𐤕				
11.	nb	𐤕𐤁					𐤕𐤁						
12.	nd	𐤕𐤃	𐤕𐤃	𐤕𐤃	𐤕𐤃		𐤕𐤃	𐤕𐤃	𐤕𐤃	𐤕𐤃			
13.	nđ	𐤕𐤃					𐤕𐤃						
14.	nf	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕								𐤕𐤕
15.	nft	𐤕𐤕𐤕	𐤕𐤕										
16.	ng	𐤕𐤍					𐤕𐤍	𐤕𐤍					
17.	nġ	𐤕𐤕	𐤕𐤕	𐤕𐤕			𐤕𐤕						
18.	ny	𐤕𐤕	𐤕𐤕				𐤕𐤕	𐤕𐤕					
19.	ngh	𐤕𐤕		𐤕𐤕									
20.	nj	𐤕𐤕	𐤕𐤕	𐤕𐤕									𐤕𐤕
21.	nk	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕	𐤕𐤕					
22.	nkn	𐤕𐤕𐤕				𐤕𐤕							
23.	ns	𐤕𐤓	𐤕𐤓										
24.	nš	𐤕𐤕											𐤕𐤕

	transcription	Bi- tri- consonants	CG	DE	AS	PA	Andries (2005)						
							0	1	2	3	D	E	F
25.	nty	lC		±									
26.	nt	l+	┐	└	T		†	T				4	┐
27.	nz	lʔ	└	┐								┐	└
28.	rd	OE		⊖									
29.	rf	OJL		⊖									
30.	rh	O:		⊖									
31.	rk	O:		⊖									⊕
32.	rg	Oʔ		⊖									
33.	rgh	O:		⊖									
34.	rm	OC		⊖									
35.	rn	OI		⊖									O'
36.	ry	O:		⊖									
37.	rš	OC		⊖									
38.	rt	O+	⊕	⊕	⊕		⊕	⊕	+O	+□			
39.	rty	OC		⊖									
40.	rw	O:		⊖									
41.	rz	OX		⊖									
42.	sn	OI		⊖	⊖								
43.	st	O+	⊕	⊕	⊕		+O	+□	⊕				
44.	št	C+	⊖				+Z	+Z	+Z				
45.	wt	:+		±	±								
46.	yt	ζ+		ζ	ζ								
47.	zt	#+			±		±	+±					

F.2 Sort Order

This is page 1382 of the third of the four volume Charles de Foucauld dictionary. The dictionary is presented by order of word roots (usually a three consonantal root), and on this page you see the root “NKR” (in the small box). Roots are presented *without* bi-consonantal characters, but then words listed after each root are written, firstly in Latin script, then in Tifinagh *with* the use

of bi-consonantal characters. On this page you can see seven words listed under that root NKR (in the smaller boxes). You will notice that the usage of bi-consonantal characters does not change the sort order at all – the bi-consonantal characters are considered simply as two individual consonants. It is unclear why the last word entry is repeated.

- 1.382.
- ānKerneKer 0:10:1 sm. nv. prim; φ (pl. inKerneKēren 10:10:1); ḏax ēnKerneKēren || fait de se lever gā et lā || at. l. s. c. ā c. du prim.
- neKiret (Ta.1) + 0:1 m. prim; conj. 56 "ḏegigēt (Ta.1) p (inKaret, ienKaret, ēd inKiret, our inKaret) || souffler bruyamment (par suite d'excs de nour = riture) (respirer bruyamment et péniblement en poussant des sorts de rou = flements (par suite d'excs de nourriture)) (le suj. étant une p. ou un an.) || v. 1:1 meheKket (Ta.1).
- sonneKiret (Ta.1) + 0:10 va. f. l; conj. 147 "seḏḏegigēt (Ta.1) p (isneKaret, iesneKaret, ēd isenneKiret, our isneKaret) || faire souffler bruyamment.
- tēnKīrīt (Ta.7) + 0:1 m. f. 16⁶; conj. 256 "tēḏḏigigēt (Ta.7); (itēnKīrīt, our itenKirit) || souffler bruyamment hab.
- sēnKīrīt (Ta.7) + 0:10 va. f. 1. 16⁶; conj. 256 "tēḏḏigigēt (Ta.7); (isēnKīrīt = rīt, our isenKirit) || faire hab. souffler bruyamment.
- ānKīrī 30:1 sm. nv. prim; φ (pl. inKīrīten 1+0:1); ḏax ēnKīrīten || fait de souffler bruyamment.
- āsenneKīrī 30:10 sm. nv. f. l; φ (pl. isenneKīrīten 1+0:10); ḏax senneKīrīt = tēn || fait de faire souffler bruyamment.
- 0:1 tounKax 0:1 * va. prim; conj. 71 "ouksaḏ"; (ienKoux, ienKoux, ēd iounKax, our ienKoux) || renier (désavouer [sa parole, son écrit, son acte, une p.]; déclarer contre la vérité n'avoir rien de commun avec [une parole, un écrit, un acte, une p.]; déclarer ne plus rien avoir de commun avec [une p, une ch, un acte, une religion]) || p. ext. "avoir en horreur extrême (détester, abhorrer, exécuter) [une p, un an, une ch.] || d. le s. "avoir en horreur extrême", est syn. de kenheh || très peu us.
- tounKart + 0:1 t sf. nv. prim; (pl. tounKārūn 10:1+) || fait de renier || at. l. s. c. ā c. du prim.
- ēneKīr 0:1 sm. n. d. e. prim; φ (pl. ineKīren 10:1; fs. tēneKīrt + 0:1+; fs. tēneKīrūn 10:1+); ḏax āneKīr (ēneKīr), ḏax neKīren, ḏax tāneKīrt (tēneKīrt), ḏax tneKīrūn || scélérat (hom. scélérat) || p. ext. la fem. tēneKīrt signifie "fem. prostituée". Ce sens est très-fréquent; aussi le mot tēneKīrt ne se prononce-t-il pas devant des p. ā qui on doit le respect || le masc. ēneKīr peut s'employer soit sérieusement, soit en plaisantant.
- ēmesneKīr 0:10:1 sm. n. d. e. prim; φ (pl. imesneKīren 10:10:1; fs. tēmesneKīrt = neKīrt + 0:10:1+; fs. tēmesneKīrūn 10:10:1+); ḏax āmesneKīr (ēmesneKīr), ḏax mesneKīren, ḏax tāmesneKīrt (tēmesneKīrt), ḏax tmesneKīrūn || scélérat (hom scélérat) || s'emploie rarement en plaisantant, presque touj. sérieusement.
- ēneKīr 0:1 sm. φ (pl. ineKīrān 10:1); ḏax āneKīr (ēneKīr), ḏax nokīrān || scélératesse (crime; acte exécrable).
- elmenKōur 0:1:1 sm. (pl. elmenKōūren 10:1:1) || m. s. q. le pr.
- elmenKōur 0:1:1 sm. (pl. elmenKōūren 10:1:1; fs. telmenKōurt + 0:1:1+; fs. telmenKōūrūn 10:1:1+) || hom. exécuté (h. abhorré de tout le monde)

G. References

- Aghali-Zakara, M. 1994. "Graphies berbères et dilemme de diffusion: interactions des alphabets latin, ajami et tifinagh, *Etudes et Documents berbères*, 11, 1994 : 107-121.
- 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>.
- _____. 2005. *Inventaire des oeils de la police pan-berbère Hapax Berbère*. <http://hapax.qc.ca/polices/inventaire-des-oeils.pdf> (accessed 19-May-2008)
- de Foucauld, Charles Eugène. 1951. *Dictionnaire touareg-français, dialecte de l'Ahaggar*. 4 vols. [Paris]: Imprimerie nationale de France.
- Everson, Michael. 2002-03-31. *On the suitability of the COENG encoding model for Khmer*. ISO/IEC JTC1/SC2/WG2 N2412. <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2412.pdf>
- Mission Baptiste – Agadez. In press. *Ayuba*. Mission Baptiste: Agadez, Niger
- Savage, Andrew. 2000. *Writing Tuareg Vowels. Advantages and Disadvantages of the Three Script Options: Arabic, Tifinagh and Roman*. Submitted in partial fulfillment of the requirements for the degree of Master of Letters in Linguistics from the University of New England, New South Wales, Australia.
- _____. 2003. Tifinagh Unicode Propositions, Stanthorpe, <http://www.unicode.org/L2/L2003/03143-tifinagh-samples.pdf>
- SIM. 1997, 2004. *Isālan n-Ālkher win iktāb Luk* (The Gospel of Luke in the Tamajeq Language). SIM, Niamey, Niger. (<http://www.westafricanscripts.com/> accessed 2008-01-25)
- 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)