

To: UTC

From: Deborah Anderson, Script Encoding Initiative, UC Berkeley

RE: Tifinagh Summary Document (based on L2/10-278 by Paul Anderson)

Date: 8 August 2010

This document is a summary of L2/10-278 “Evolution of the Tifinagh script in Unicode,” which asks for direction in how to deal with regional variants in the Tifinagh script in Unicode. Some additional information is provided, which draws in part on information from the original Tifinagh proposal by Patrick Andries (N2739 = L2/04-142).

Executive Summary: There are a variety of glyph shapes used for writing the regional forms of the Tifinagh script, and, in Tuareg Tifinagh, different notational systems for vowels that vary from what is encoded in Unicode. Tuareg Tifinagh also has contextual shaping (for LL and NN), and a wide variety of consonantal ligatures, which do not occur in neo-Tifinagh (in Morocco). Because of these factors, it is not easy to write the regional varieties of Tifinagh with a font using the current Tifinagh encoding.

The earlier proposal by Patrick Andries for Tifinagh (N2739) only encoded those shapes for Tuareg consonants that were not already covered by current Neo-Tifinagh consonants and their glyph variants. The Tuareg vowel notational marks were not proposed, as no one system seemed to prevail. Tuareg ligatures were to be left up to the font.

The author of L2/10-278 asks the UTC for guidance on whether to:

(a) unify the Tuareg variants with the currently encoded Tifinagh forms (eschewing the Tuareg-specific characters that are encoded), handling contextual shaping in the font (for LL and NN), and using fonts to show the regional shapes;

or

(b) continue encoding additional characters as needed for unsupported regional forms (such as for Algerian Neo-Tifinagh or separately encode YAL and YAN), and use the Tuareg-specific forms already encoded.

The choice will affect spelling, since option (b) will rely on Tuareg-specific characters, rather than using already encoded Tifinagh characters.

I. Background

1. Document L2/10-278 differentiates different “styles” for Tifinagh, grouped into the following general three categories:

a. Traditional Tuareg Tifinagh

Description: Widely known and used as basis for literacy programs, but “rarely used other than for short inscriptions and letters.”

Features: Several consonant letters are composed of dots (see figure 1); very limited vowel marking (and if vowels are marked, there can be variation in how they are written); often

contains consonantal ligatures; clusters of LL and NN are leaned. For regional forms, see columns H, G, D, Y, W, and T in Table 1, below. (This table is referred to in L2/10-278 as “Table 1a”, but is not included there. It is attributed to Karl Prasse 1972.)

Current support in Unicode fonts: Tuareg Tifinagh letters are not yet in widespread use, partly due to lack of ligature support. Regional forms are included in Patrick Andries’ fonts.

b. Modern vowelized Tuareg Tifinagh

Description: Builds on Traditional Tuareg Tifinagh (a), above, rather than neo-Tifinagh (c), below. This group includes a variety of different repertoires, put forward by various groups or individuals, including SIL, APT [Association for the Promotion of Tifinagh], Rissa Ixa, Hawad (latter two appear to be less commonly used).

Features: Like Traditional Tuareg Tifinagh, has context-dependent letter leaning for LL and NN, filled dots for vowels YA and YU, but new vowel letters are added to the regional form of consonants (see figure 1 for regional forms; figure 2 shows different vowel notations).

Current support in Unicode fonts: Tuareg Tifinagh letters do not yet have widespread support, due to lack of support for ligatures and Tuareg-specific vowels. Regional forms are included in Patrick Andries’ fonts.

c. Neo-Tifinagh (sometimes referred below to as “IRCAM”¹)

Description :Refers to the writing systems used to represent the Maghreb Berber dialects. First variant proposed by Académie Berbère (AB) on the basis of Tuareg Tifinagh; spread to Morocco and Algeria, especially in Kabyle. Term also covers other variants that appear to complement or correct AB proposal (N2739, by P. Andries). Morocco has adopted the IRCAM version as the official standard.

Features: In Morocco: vowels YA and YU often depicted with open circles, but punctiform letters are avoided; YAL and YAF are connected with an intervening line: (YAL) ⵢⵏ and (YAF) ⵢⵏ.

In Algeria: Some variation in repertoire (see figs. 3a and 3b in L2/100-278), but solid dots are used for vowels; there is no linking between the vertical lines of YAL and YAF; contextual leaning of LL and NN. Only certain Algerian-specific forms have been encoded; the letters for ‘y’ and for ‘tch’, for example, are not yet encoded as separate characters. ²

Current support in Unicode fonts: Ebrima in Windows 7 covers Tifinagh as it appears in Unicode codechart. At present, Algerian-style writing is done with non-Unicode fonts that map Latin letters to neo-Tifinagh forms.

2. Summary of Current Unicode coverage for Tifinagh:

- 31 letters from IRCAM for Neo-Tifinagh in Unicode 4.1
- 8 IRCAM-Extended characters (for words that appear in literary works and to cover regional and foreign sounds) in Unicode 4.1

¹ The national governmental Berber body of Morocco

² Note that in the font HapaxB DzN Ancien, the ‘tch’ form is mapped to 25DE YACH, ‘y’ to 2D62 YAY.

- 11 modern Tuareg letters (identified in character names by “TUAREG”, “AHAGGAR”, “AYER”, or “TAWELLEMET”) in Unicode 4.1.
These do not have widespread use, probably due in part to lack of font support for consonantal ligatures and lack of Tuareg-specific vowels. (p.12)
- 2 additional characters are added in 6.0: Tifinagh separator mark and Tifinagh consonant joiner

II. Question: Unification or Disunification?

Problem:

“Although Unicode encodes many variant Tifinagh letters, these do not cover all those needed for different attested styles of Tifinagh writing. Even where existing letters are sufficient...the different ways of writing Tifinagh still cannot all be shown correctly by any one font, because of the differences in the customary appearance of the letter and differing solutions for sequences of letters with vertical lines...

Since the Tifinagh character mappings from the IRCAM set to the variant repertoires are one to one, and different fonts could capture the variety, the original reason for encoding the variant letters individually seems to have been the greatly differing glyphs, or to allow different Tifinagh styles to be written side-by-side by the same font (not actually possible exactly, as we have seen).” (p.11)

Two options to be able to handle the different styles of Tifinagh writing:

a. Unification

“There is no need for displaying regional repertoires side by side in the same font; two language variants in the same document would either use the same Tifinagh writing style or different fonts as appropriate.”

- Since the unity of the script is easily visible, with each region differing only in a handful of characters, and the repertoire in each region represents almost the same set of sounds, the script could be generally unified using the Unicode Tifinagh characters with ‘unmarked’ names (i.e., “TIFINAGH LETTER <X>” .³ “Superfluous variant letters could be marked as ‘to be unified’ (essentially deprecated).”
- Use fonts to show correct form for region. Typographic features, such as letter leaning, would be handled by fonts, not by disunifying characters. Tuareg fonts would need to map their existing typographical effects onto the new shared character set.
- No need to add new characters,⁴ except perhaps for phonetic modifier letters.
- For Tuareg vowels, diacritics would additionally be used. Almost all vowel systems can be written using the unified set with suitable font support.⁵

³ He suggests making one exception: use TIFINAGH LETTER BERBER ACADEMY YAJ for ‘dj’, and not TIFINAGH LETTER YAJ, which is generally used for ‘g’.

⁴ For example, for the APT Niger “missing” forms, the glyphs would be mapped to Tifinagh letter equivalents: schwa to YEY, ‘u’ to YU, and ‘tch’ to YACH; hence they would not be separately encoded.

- Only one keyboard needed for all regions. Local repertoires would not be constrained by keyboard installed there but by font availability.
- Can quickly support various repertoires (since new characters would not be proposed), and the fonts can ensure typographic features for repertoires are consistent.
- Webpages or documents for a particular language would need to be rendered in the appropriate regional font (which is true already for display of Tuareg biconsonants, filled vowel dots, and leaned letters). To see the correct repertoire, users would set the font used by the chat program or set the OS font by locale. Would need to rely on page-embedded or locally-installed fonts.
- The fall-back font would be the neo-Tifinagh IRCAM reference glyphs, which would apply to typing Tuareg into a web form not specialized for Tuareg. “Users would know most of these as they would look the same and the rest would soon become familiar. Confusion would be unlikely in general because for each regional form the corresponding IRCAM letter is either near-identical or looks unlike any of the other regional letters [with some exceptions].”
- This approach stretches the definition of unification: the characters’ glyphs would be substantially different from each other visually.

b. Disunification

- Use currently encoded Tifinagh characters and propose currently un-supported regional forms (such as Algerian neo-Tifinagh ‘y’ and ‘tch’), perhaps dis-unifying letters which have significant variations in typography. Possibly encode YAN, YAL, YA, and YU, which have different shapes than the representative glyphs for neo-Tifinagh.
- Overall set of characters would be large, and additional letters would need to be added later after they have been proposed and approved.
- A large table of Tuareg ligatures is needed in fonts to cover each combination of letter variants for biconsonants.
- Each regional repertoire would need its own keyboard and other tools.
- Would be able to type documents in Neo-Tifinagh and Tuareg together, without having to change fonts.

⁵ With one exception: a character of Rissa Ixa, which has a stand-alone character (^) for short ‘a’. A font would need to provide shapes and stacking ability for Hawad’s system (Ḫ, Ḫ̣, X, Ṣ, Ṣ̣).

Figure 1 (Source: <http://www.mondeberbere.com/langue/tifinagh/tableau1.htm>)

TABLEAU 1										
[- Historic forms -] [- Trad. Tuareg Tifinagh Regional Styles -] [Acad. B.]										
V	L/or	L/oc	Sah	H	G	D	Y	W	T	AB
b	⊙	⊙ ⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
d	⌒	⌒	⌒	⌒ ⌒	⌒	⌒	⌒	⌒	⌒	⌒
ḍ			⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
f	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
g	⌒	⌒	⌒*	⌒	⌒	⌒	⌒	⌒	⌒	⌒
h	≡	≡	≡	⋮	⋮	⋮	⋮	⋮	⋮	⋮
ḥ	⌒?									⌒
ḥ				⋮	⋮	⋮				⌒
k	⌒	⌒	⌒	⋮	⋮	⋮	⋮	⋮	⋮	⋮
l	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
m	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
n	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
q	÷?	÷	⌒	⋮	⋮	⋮	⋮	⋮	⋮	⌒
y	÷?	÷	≡	⋮	⋮	⋮	⋮	⋮	⋮	⌒
r	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
s	⌒	⌒	⌒	⊙	⊙	⊙	⊙	⊙	⊙	⊙
c	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒		⌒
ṣ	⌒	⌒	⌒		⌒					⌒
t	+ X	X	+	+	+	+	+	+	+	+
ṭ	⌒	⌒	⌒	⌒	⌒	⌒		+	⌒	⌒
w	=	=	⌒	⋮	⋮	⋮	⋮	⋮	⋮	⌒
y	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒		⌒
z	⌒	⌒	⌒	⌒	⌒	⌒	⌒		⌒	⌒
ṣ	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒	⌒
j	⌒	⌒								⌒
a			⌒	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i				(⌒)	(⌒)	(⌒)				⌒
u				(⋮)	(⋮)	(⋮)				⋮

Comparaison entre le libyque (occidental et oriental), le tifinagh saharien, six variations du système touarègue et le néo-tifinagh développé par l'Académie berbère

- Les signes du libyque occidental, n'étant pas déchiffrés, sont rangés aux côtés des signes du libyque oriental les plus similaires.
- Les signes dont la valeur n'est pas établie sont marqués d'un ?
- Les caractères sont orientés horizontalement de gauche à droite.
- Dans la colonne Sah, * veut dire que deux signes ou trois forment un seul caractère.
- Le tableau 1, ainsi que toutes les précisions qui vont avec, hormis la colonne AB, sont tirés de Prasse (1972)

V = valeur
L/or = Libyque oriental
L/oc = Libyque occidental
Sah = Tifinagh saharien
H = Hoggar

G = Ghat
D = Adrar
Y = Ayer
W = Iwelmedan
T = Tanslemt
AB = Néo-tifinagh de l'Académie berbère

Figure 2: Sampling of some notations used for Tuareg vowel additions

D. 2.4.1 Tuareg vowels additions

Several individual and associations have attempted to overcome the lack of written in the Tuareg notation. Several competing and isolated systems have been advocated:

- i. APT (•, <, ⋈, --, ∪, ∥), see [7] ;
- ii) SIL (⤴, ⤵, ⤶, ⤷ or ⤸, ⤹, ⤺ or ⤻), see [24] ;
- iii) Hawad (ⵍ, ⵍ̣, ⵍ̤, ⵍ̥, ⵍ̦), see [17] ;

Source: “Proposal to add the Tifinagh Script” by Morocco, Canada, France (prepared by Patrick Andries) <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2739>.