

Title: Unification of the Tifinagh script - considerations
Source: Paul Anderson
Status: Individual Contribution
Action: For consideration by UTC
Date: 07-Apr-2010

This document aims to outline some of the problems facing the Tifinagh script in Unicode and to suggest possible future directions.

At the moment, the Tifinagh script is neither completely unified, despite having a theoretical basic alphabet almost perfectly suited to writing both Northern Berber and Tuareg, nor adequately disunified for writing both neo-Tifinagh and Tuareg with the same font, because of typographic considerations – and not all regional variants are supported.

Two possible responses would be:

- to disunify by adding further characters as required for unsupported language variants; perhaps also to add Tuareg versions of letters such as YAN/YAL/YA/YU having different typographical conventions; and perhaps to mandate font support for ligatures.
- to unify on the basic Tifinagh set with 'unmarked' names (TIFINAGH LETTER X) by marking existing additional variant letters outside the basic set as 'to be unified'. TIFINAGH CONSONANT JOINER and TAZARAST would be unaffected.

Background – Tifinagh addition to the UCS

When the Tifinagh letters were first added to the UCS, the vowelled neo-Tifinagh repertoire developed by Morocco's IRCAM for Moroccan variants of Northern Berber was used as the base. The Tifinagh script was declared the official script for these variants in Morocco. The IRCAM letters were called simply “TIFINAGH LETTER X” where X corresponds to the sound represented by the letter.

Other attested neo-Tifinagh, for example from the Academie Berbère, extended IRCAM letters, and Tuareg Tifinagh were given simple names too, where they extended the basic set without duplication of sound, for example U+2D36 TIFINAGH LETTER YAJ, U+2D32 TIFINAGH LETTER YABH. These letters were not needed by IRCAM in the basic set either because a different orthographic convention is used to represent these sounds in Morocco (digraphs), or because the sounds are considered to be dialectal or foreign variants and are written by convention with a different letter.

Then, other attested neo-Tifinagh and Tuareg variant letters were added, with names such as U+2D4C TIFINAGH LETTER TUAREG YAZH.

Analysis

The naming convention for Tifinagh letters suggests a basic alphabet with 'unmarked' letter names (TIFINAGH LETTER X, with no further precision) that accurately and almost completely covers Northern Berber and Tuareg requirements for representing sounds. However, only Moroccan IRCAM-style writing uses a subset of these basic letters exclusively. Tuareg styles mix additional variant letters with the basic set, often having different letters from IRCAM's for the same sounds. Some letter variants needed for other Northern (Algerian) use and for some Tuareg styles (such as APT Niger) are missing from Unicode. Extra vowels required for Tuareg are also missing, which is why the basic letter set is just short of complete.

Northern neo-Tifinagh shares many Unicode characters with Tuareg. In some cases this mixing of the basic set and variant letters for Tuareg causes difficulties, because while neo-Tifinagh fonts can choose for the basic set whether to follow innovations such as linked bars for YAL and YAN, and hollow dots for YA and YU, Tuareg fonts can not. The hollow dots for vowels are incorrect for Tuareg and also do not match the variant consonants formed with filled dots, such as Tuareg YAK. Contextual shaping works differently for combinations of the letters TIFINAGH LETTER YAL and TIFINAGH LETTER YAN when writing neo-Tifinagh and when writing Tuareg (and even differs between APT and SIL Tuareg). In neo-Tifinagh the vertical lines within a letter are linked, so no confusion can arise when letters are clustered, but in Tuareg (and Algerian Northern Berber) usage the lines must be 'leaned' to show which letter they belong to (see [1]).

Writing Tuareg is only possible in Unicode in the traditional style without vowels. The ligatures needed for this (see [1]) are as yet only rarely even partially implemented in fonts (and even then not for use with the TIFINAGH CONSONANT JOINER but for the earlier mechanism of ZERO WIDTH JOINER), and the required mapping table is large because of the number of variant Tuareg letters. Also, if the font used is neo-Tifinagh, the traditional 'nf' ligature looks the same as neo-Tifinagh YAF alone.

The result is that although the two sets are encoded partially dis-unified, they still cannot both be shown correctly by the same font. The situation is rather like for traditional and simplified Chinese, except that the Tifinagh character mappings from basic to variant are one to one - so that the original reason for disunification seems rather to have been visual glyph differences.

There are many different versions of the Tifinagh alphabet in use in different regions, both in the North and for Tuareg. However, not counting the extra vowels of Tuareg, where different systems have been suggested, it is possible to tabulate all these versions side by side, correlated by sound – and historically this is what all researchers have intuitively done. There are some gaps depending on the phonology of the language variants, and some letters are used in different areas to represent different sounds because of historical sound shifts, but in general the unity of the alphabet is easily visible. Each region differs only in a handful of positions from the others, and each region requires letters to represent almost the same repertoire of sounds. There is a clear shared superset corresponding to the Unicode basic Tifinagh with 'unmarked' names. Although regional glyphs do differ in some positions, there is no need for displaying different regional sets side by side in the same font (except for neo-Tifinagh and Tuareg, which is currently impossible anyway), and in each case the same logical letter is intended, for example TIFINAGH LETTER YAT.

There is one exception – APT Niger Tuareg encodes emphatic letters as the unemphatic letter plus combining dot below, for example YAT plus COMBINING DOT BELOW instead of standalone letter YATT. That poses no problem for unification of the script, however.

Implications of two possible solutions

Disunification:

- Characters required for writing unsupported regional variants, such as Algerian Northern Berber or APT Niger Tuareg, would continue to be added.
- Each regional repertoire would need its own keyboard (although the layouts could be equivalent).
- A large table of Tuareg ligatures would be required in fonts to cover each combination of letter variants.
- Any regional text would be readable in any font in the form that it was written. However, this would only be completely true for Tuareg displayed in a neo-Tifinagh font if several existing letters were disunified for typographical reasons (see analysis section), and ligatures made with TIFINAGH JOINER were fully supported by the font.

Unification:

- There would be no need to add new characters. This is true even for alternative Tuareg vowel systems (with the exception of one character for Rissa Ixa's system, which has a standalone letter for short 'a' with no diacritic – see [2]). The base letters for diacritics in most Tuareg vowel systems, if taken to be visual forms of YI/YAY, YU/YAW, YA, YEY, can be shown in the required form with a suitable font.
- The font would be responsible for showing the correct visual repertoire for a region, and standard fonts would simply render the reference glyphs (so no change required to standard fonts). This is already true for Tuareg effects such as filled dots, contextual shaping and ligatures, so unification would have the same result as now, but without the need to mix basic and variant letters. Tuareg fonts would need to map their existing typographical effects onto the basic letter set (as well as retaining them for the variant letters for backwards compatibility).
- Only one keyboard layout would be sufficient for all regions – important for these language variants which have been poorly supported with tools. Local repertoires would not be constrained by the keyboard installed there, but by font availability.
- To see the correct repertoire, users would set the font used by their chat program, or set the operating system font by locale. For word processing, there would be a few commonly installed fonts per region as now (currently these are often non-Unicode). Pdf exchange allows font embedding. Web pages would either take advantage of common regional fonts assumed to be installed, or embed the font on the page – the technology to do this is now mature. Currently, web pages in Tuareg Tifinagh are vanishingly rare; in neo-Tifinagh they are rare and assume that the correct Unicode or non-Unicode font is installed. In recent versions of Windows, there is now (as yet buggy) support for Tifinagh in the default fonts, making neo-Tifinagh use in documents and on the web more likely.
- The situation not covered would be users typing Tuareg into a web form on a page like Yahoo mail not specialized for Tuareg. The letters would appear as neo-Tifinagh. This is somewhat the case already, however (see analysis section). Workarounds are to send the message as an attachment or override the browser fonts.

References

[1] Morocco-Canada-France (prepared by P. Andries) - Proposal to add the Tifinagh Script, n2739.pdf (Unicode proposal)

[2] Naïma LOUALI, LES VOYELLES TOUAREGUES ET L'ALPHABET TIFINAGH : EVALUATION DE QUELQUES PROPOSITIONS RECENTES, PHOLIA (Laboratoire de Phonétique et Linguistique Africaine) Volume 8, 1993