

# Unification

## **A Proposal to Ease the Migration of Ge'ez Software to The Unicode Standard**

*by Daniel Yacob*

A fundamental void in the field of Computational Ethiopics is scheduled to be filled in 1996. Launched in cyberspace and landing in the very real space of Quebec City, Canada; the International Standards Organization (ISO) will meet in August to give the Ge'ez writing system a computer standard.

The history of the Ethiopians and others involved in the three-year effort to bring about the standard is recorded in email and world wide web exchanges for those who may wish to review it. It will not be the purpose of this article to make a tour of those events -the failures and the victories that will become the standard. Nor will it be the purpose of this article to describe the standard itself -its merits, weaknesses, and the work left unfinished. Indeed, after a character coding standard has arrived, the work, theory, and emotions that created it are soon forgotten and of little value to the person seeing it for the first time. To developers who cannot read Fidel in three of the companies already applying the standard, the standard is arbitrary.

In this article we will look forward into the near term. Into a time of transition that brings about new problems that need to be bridged to reach the solutions offered by the ISO/Unicode standard. The issues we will address would be those that arise independent of the characteristics for any standard devised for the Fidel.

Given an arbitrary character coding standard for an arbitrary writing system; what is the writing system offered? Primarily, the writing system acquires a lowest-common-denominator state for information interchange between applications running on a given OS type, and across OSs and computer architectures. This means application users can write a document in Word Perfect using Agafari fonts on a PC and import the document into MS Word on a Macintosh using AlexEthiopian fonts and worry only about a minor loss to formatting. The qualitative text remains the same.

Secondly, the script lexicon is given an agreed upon sorting order. This effects precedence of alphabetic characters prior to punctuation, or numerals before punctuation and letters, etc. The ordering of punctuation elements (i.e. comma preceding colon preceding quotation) is fixed for better or for worse. Member inclusion and sort order are the governing concerns in the design of a character coding system.

Thirdly, the script is given legitimacy. With an internationally recognized standard for a given script, the script will be perceived by developers as one to be considered for support by their product. With a standard, the script is acknowledged as one that ``belongs" in computer environments. Considered for a moment how seriously a nation's football team is taken if it is not recognized and sanctioned by the International Football Federation. No one outside of the country will even know they exist. Nor do they get to ``play the game". In this respect, Ge'ez has just now entered the world's stadium.

What does a writing system lose when a standard is introduced? The answer to this question is an important one to Computational Ethiopics, as important as the consequences are painfully severe. In the absence of a recognized standard for Ge'ez script, many standards developed -one for each vendor selling fonts. The thousands of pages composed with one Ethiopic font or software is no more readable by the new standard than they were with any of the other coexisting standards. This impacts all vendors equally and is the first

sense in which the standard unifies the industry. United by this problem; can we be united in bringing about its resolution?

The work required of a font or application provider to adjust their Fidels to the ISO/Unicode standard will generally be minor. To update old documents to the new system is considerably more demanding. This is primarily because the traditional solution for bringing Fidel's generous lexicon into PC environments has been to divide the Fidel into 2 -> 8 font sets. Unicode now lets us use a single font. Hence a document that a word processor believes to contain two or more fonts, must be made to believe now holds only a single font after the character codes have been updated accordingly. During the time of transition there would also be a need to be able to export documents to the older, non-Unicode system.

Only a person who has spent a few hundred or a few thousand hours hand editing fonts can fully appreciate the nature and burden of the task. For some font providers it will still be a few hundred hours of work to properly update their packages for the ISO/Unicode system. The updating process, of the fonts, will have unique requirements for each vendor. The resulting work depends also on the individual artistic sensibilities of the font composer.

For each font provider to also take on the added burden of writing a document transcriber neither serves the user nor business communities. The users loose when they wait an uncertain amount of time for the converter to become available and are unable to import their documents to the new system. Such converters would likely come at a cost required for the vendor to recoup his time invested. It is conceivable a market balance could shift if customers migrate to the vendor who provides a forward and backward system converter first. This shift cannot happen if a converter is made available for all font packages simultaneously.

The task of writing a document transcriber for a given font package is a modest amount of work. To modify the same transcriber for a second or third font package consumes trivial time relative to the initial effort. The question now becomes: Who would be willing to do this initial and additional work on behalf of both the public and the software industry? Who could be trusted to include all software vendors and not exclude selected competitors? Who could do this in the non-profit arena, keep the source code in the public domain, and still benefit from the time and effort invested?

The best choice, in the mind of this author, would be for a department at Addis Ababa University to head such a project. Perhaps at either the School for Information Studies in Africa (SISA) or the Mathematics Department. Launched as a student project; the faculty as well as the students stand to benefit. The faculty has the opportunity to experience distance management if they coordinate the project with concurrent efforts and other interested groups and individuals in the Internet community. The students benefit as they learn about multi byte transformation methods and from working on a group software project.

Such a conversion resource might best be written in a versatile computer language such as Visual Basic - which many of the Microsoft applications are capable of running as macros. Were the "LibEth" ANSI C Ethiopic library migrated from Unix to DOS platforms in the form of a DLL, much time and energy is saved immediately. LibEth also offers I/O in other useful encoding systems for Ge'ez; such as Unicode, JIS, JUNET, and SERA (7-bit email safe) phonetic transcription. LibEth resources could be applied with a VB approach. Indeed, even these kinds of decision-making problems become a part of the learning experience.

There would naturally be other unforeseen benefits that come out of organizing and executing a cooperative project between academia and the business community. A project who's goal was for the public good and

who's progress was kept in the public view. What some of these gains might be could not become apparent until after the work began and was reaching its fruition.

To conclude this informal proposal lets itemize now the expected benefits:

1. The public receives a utility to convert documents between the font systems of different vendors.
2. The public receives a utility to convert documents to and from present and forth coming computer systems.
3. The public receives an ``Email-Safe", and human readable document converter.
4. Font and software vendors are freed to focus on other aspects of the migration to ISO/Unicode.
5. Students and Faculty and AAU acquire experience in multi byte character coding principles, and (optionally) Internet project management that may be reapplied in future efforts.
6. The department responsible gains acclaim and notoriety for putting an essential Ethiopics tool on every PC with Ethiopic software.

What is left is for the parties interested, including the font providers, the coders, shareware distributors, and the institutions, to initiate communication and formalize the project.