Encoding SignWriting
in the Universal Character Set

Michael Everson
Dublin, Ireland
August 15, 1999

The Universal Character Set or UCS (International Standard ISO/IEC 10646-1 and computer industry implementation Unicode) is a solution to the many problems which earlier computer coding systems have had when dealing with multilingual text. 8-bit computer systems such as Mac OS 8 and Windows 95 had to resort to complex solutions for switching between character sets (usually understood by the user as "code pages" or "fonts") even to do such simple things as display German and Polish text in the same document. The UCS presents a single coding system for text that can handle, to put it simplistically, all the letters of all the alphabets of all the languages of the world. Use of this standard is being given the highest priority by such companies as Microsoft, Apple, and Sun. Today, the UCS handles writing systems as different as Latin, Greek, Cyrillic, Armenian, Chinese, Bengali, Tibetan, Arabic, Hebrew, Korean, Mongolian, and Cherokee. Plans to extend the repertoire to include scripts like Egyptian Hieroglyphs, Phoenician, Aramaic, Tagalog, and Blissymbols are being considered by the Technical Committees responsible for ISO/IEC 10646 and the Unicode Standard.

SignWriting, as a new writing system uniquely suited to the representation of the world's Sign Languages, has recently come to the attention of these Technical Committees. Michael Everson, Irish national representative to the committee responsible for ISO/IEC 10646, is a director of an Irish company dedicated to providing computing solutions for lesser-used languages. He has spearheaded discussions with the Deaf Action Committee to explore the possibility of encoding SignWriting in the UCS.

The advantage of having SignWriting encoded in the UCS are many. As the corpus of Sign Language literature, currently written using the DAC's SignWriter 4.0 program, grows, it becomes increasingly important that users have the ability to exchange data safely (with no corruption) and to search, sort, and otherwise

manipulate this data in the same way that users of spoken languages do. With the increasing availability of SignWriting text on the internet, this is becoming even more relevant. Since it is clear that the Universal Character Set will be the means for encoding and exchanging text, supported by all software vendors, it is important that work begin now to prepare SignWriting for encoding in the UCS.

SignWriting has a number of features in common with other writing systems encoded in the UCS. It also has unique features, such as the semantically-relevant reversal (flopping) and rotation of base characters, and their positioning in both vertical and horizontal contexts, which will require a concerted effort on the part of programmers involved with the SignWriter program, experts in UCS implementation, and the user community. Preliminary explorations of the effort required indicate that 6-10 man-years of effort will be required. Funding for this effort is urgently required if SignWriting is to take its place among the world's writing systems in the UCS. This will benefit all users of SignWriting in the countries currently using SignWriting and will in addition pave the way for expansion of SignWriting into new communities, as it is well-established that SignWriting is a robust writing system suitable for representing Sign Languages worldwide.

A project team needs to be set up to analyze SignWriting as it is currently implemented (in MS-DOS and the new Java version of SignWriter 5.0 which is presently under development) with a view to ensuring that existing data can be migrated to the UCS platform, and to prepare the proposal for UCS encoding of SignWriting itself. This preparation will include at a minimum Valerie Sutton, Michael Everson, and a programmer familiar with SignWriter 5.0. A prototype SignWriter 6.0b will result, making use of the UCS Private Use Zone to encode SignWriting characters. (The Private Use Zone is an area defined by hexadecimal characters U+0000E000 - U+0000F8FF; the prototype algorithms and programs can be modified when SignWriting is actually encoded in Plane 1 (U+00010000 - U+0001FFFF) of the UCS.) One deliverable of the Project Team not previously mentioned will be a TrueType font for SignWriting. This will greatly facilitate printing of SignWriting (better glyph shapes, improved user control of type size, etc.).

For more information, contact:

Michael Everson
everson@evertype.com
https://www.evertype.com

Valerie Sutton
Sutton@SignWriting.org