ISO/IEC JTC1/SC2/WG2 N1641

DATE: 1997-09-18

DOC TYPE:	Expert contribution
TITLE:	**Proposal to encode Tengwar in Plane 1 of ISO/IEC 10646-2**
SOURCE:	Michael Everson, EGT (IE)
PROJECT:	JTC1.02.18.02
STATUS:	Proposal.
ACTION ID:	FYI
DUE DATE:	--
DISTRIBUTION:	Worldwide
MEDIUM:	Paper and web
NO. OF PAGES:	5

A. Administrative

1. Title	Proposal to encode Tengwar in Plane 1 of ISO/IEC 10646-2
2. Requester's name	Michael Everson
3. Requester type	Expert request
4. Submission date	1997-09-18
5. Requester's reference	
6a. Completion	This is a complete proposal.
6b. More information to be provided?	No

B. Technical -- General

1a. New script? Name?	Yes. Tengwar
1b. Addition of characters to existing block? Name?	No.
2. Number of characters	93
3. Proposed category	Category B.1
4. Proposed level of implementation and rationale	Level 2 and Level 3
5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes
6a. Who will provide computerized font?	Michael Everson, Everson Gunn Teoranta
6b. Font currently available?	Michael Everson, Everson Gunn Teoranta
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	No
8. Does the proposal address other aspects of character data processing?	Yes

C. Technical -- Justification

1. Contact with the user community?	Yes. There are several Internet discussion lists and web sites.
2. Information on the user community?	Tengwar enjoys both scholarly and popular use.
3a. The context of use for the proposed characters?	Used to write Quenya, Sindarin, English, and other languages.
3b. Reference	
4a. Proposed characters in current use?	Yes
4b. Where?	By scholars and enthusiasts.
5a. Characters should be encoded entirely in BMP?	No. Positions U+0001 CC00 - U+0001 CC7F are proposed for the encoding.
5b. Rationale	Accordance with the Roadmap.
6. Should characters be kept in a continuous range?	Yes
7a. Can the characters be considered a presentation form of an existing character or character sequence?	No.
7b. Where?	
7c. Reference	
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	No
8b. Where?	
8c. Reference	
9a. Combining characters or use of composite sequences included?	Yes
9b. List of composite sequences and their corresponding glyph images provided?	Characters identified on the code table with dotted circles are combining characters.
10. Characters with any special properties such as control function, etc. included?	Yes. The numerals have right-to-left properties while the rest of the script has left-to-right.

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	
4. Assigned category and assigned priority/time frame	
Other Comments	

The Tengwar script was invented by the philologist and author J. R. R. Tolkien as part of the mythological world he created, and was widely popularized through his work, *The Lord of the Rings*, *The Silmarillion*, etc. Along with a family of artificial languages and a large corpus of etymological data describing their relationships, the Tengwar script has attracted the attention of a large community of linguists and other enthusiasts interested in this expression of Tolkien's expertise in historical and comparative linguistics. The Tengwar shouldbe treated as a Category D (Attested Extinct) alphabet: there is a relatively limited corpus, and a relatively small (but existent) scholarly body studying it. In order to provide a standard Tengwar character coding for such scholars and enthusiasts, it has been suggested that this character set be included into the Unicode standard and ISO 10646.

8 columns are reserved to encode the Tengwar. The last column is currently unused, and is reserved for future discoveries in the Tolkien manuscripts. Character names derive from Tolkien's published writings; as usual, long vowels are written double.

General Principles of the Tengwar script

The Tengwar script is a system of consonantal signs without strictly fixed values; their glyphic structure comprises a matrix of potential phonetic relationships, rather than a set of fixed relationships between sound and character. The primary letters (U+xx00 - U+xx17) are formed of a *telco* 'stem' and a *l™va* 'bow'; raising the stem might indicate spirantization of a consonant, or doubling the bow might indicate voicing. Consonants are modified by *tehtar* 'signs', described below.

A series of "stemless consonants" have been encoded. STEMLESS OORE is used as DIGIT ZERO; STEMLESS VILYA is used as DIGIT ONE; STEMLESS ANNA is used as a vowel in the mode of Beleriand; STEMLESS VALA is as yet unattested, but is included here because of the inherent structure of the script.

Tengwar are written from left-to-right. Tengwar numerals are written from right-to-left (the least significant digit is on the left). The DECIMAL BASE MARK and DUODECIMAL BASE MARK are applied to the digits to indicate what the arithmetic base is used; the DUODECIMAL LEAST SIGNIFICANT DIGIT MARK is used on the least significant digit in a duodecimal expression. The numeric marks are not generally considered optional.

No positional variants of the letters exist. Like Arabic, the script is founded on calligraphic handwriting, and many ligatures may be required for high-quality rendering -- though unligatured forms may often be acceptable. No ligatures are encoded here.

Vowels and Other Marks of Pronunciation

Non-spacing marks, generically called *tehtar* 'signs', indicate vowels or other modifications of consonantal letters. *Tehtar* are placed above or below consonants, or atop "carriers" when no consonant is present in the required position. The occurrence of a character in the *tehtar* range, depicted with relation to a dashed circle, constitutes an assertion that this character is intended to be applied via some process to the consonantal character that precedes it in the text stream. General rules for applying non-spacing marks are given in Section 2.5 of the Unicode Standard. In ISO 10646, Level 2 encoding is intended. See the remarks on Modes below.

The SHORT CARRIER simply bears the vowel *tehta*; the LONG CARRIER indicates that the vowel was long; this can also be done by doubling the vowel sign.

Modes

The morphological structure of a language determines the "mode" in which the Tengwar script is used for it. For instance, the *tehtar* are placed above or below the *preceding* consonant in languages in which words tend to end in a vowel; but they are placed above or below the *following* consonant in languages in which words tend to end in a consonant (compare Quenya *nelde* 'three', *neltildi* 'triangle' with Sindarin *neled* and *nelthil*.). In accordance with Unicode specifications, however, the *tehtar* are encoded as non-spacing characters, and so must *follow* the consonant over which they appear. For Sindarin, this requires that the logical order of backing store does not reflect its true syllabic structure. For instance, the Quenya examples here are encoded NUUMEN-ACUTE-ALDA-ACUTE (**n-e-ld-e**), and NUUMEN-ACUTE-LAMBE-TINCO-AMATICSE-ALDA-AMATICSE (**n-e-l-t-i-ld-i**); the Sindarin encoded NUUMEN-LAMBE-ACUTE-ANDO-ACUTE (**n-l-e-d-e**), and NUUMEN-LAMBE-ACUTE-THUULE-LAMBE-AMATICSE (**n-l-e-th-l-i**). English is generally written according to a Sindarin-type mode; Italian would be written according to a Quenya-type mode. This inconsistency of phonetic representation and encoding in the backing store is a function of the script's unique representation of modalities which must be reckoned with apart from the character set itself. Smart inputting methods, such as are used for some Southeast Asian Brahmic scripts, could solve the problem for Sindarin-type mode inputting. In the mode of Beleriand, where the *tehtar* are not used, but full vowels, the Sindarin examples are written: OORE-YANTA-LAMBE-YANTA-ANDO (**n-e-l-e-d**) and OORE-YANTA-LAMBE-THUULE-SHORT CARRIER-LAMBE (**n-e-l-th-i-l**). Mapping software for conversion of standard-mode and Beleriand-mode Sindarin will be requisite.

Punctuation

Tengwar punctuation characters are considered to be unique to the script and are coded in the Tengwar block. Som composition of punctuation occurs in Tengwar: DOUBLE PUSTA can be followed by SECTION MARK, LONG SECTION MARK, PUSTA, and DOUBLE PUSTA.

Sometimes word space is not used; word separation may be achieved in that case with U+200B, ZERO WIDTH SPACE. Hyphenation is not used; words may be broken *before* any LETTER.

Encoding Structure

The Tengwar block is divided into the following ranges:

```
U+xx01 -> xx17 Consonants
U+xx18 -> xx33 Miscellaneous letters
U+xx40 -> xx4F Vowel signs
U+xx50 -> xx53 Punctuation
U+xx54 -> xx55 unassigned
U+xx56 -> xx57 Additional vowel signs
U+xx58 -> xx59 unassigned
U+xx5A         Additional vowel sign
U+xx5B         unassigned
U+xx5C -> xx5D Miscellaneous letters
U+xx5E -> xx5F unassigned
U+xx60 -> xx61 Punctuation
U+xx62 -> xx6B Numerals
U+xx6C -> xx6E Numeric modifiers
U+xx6F -> xx7F unassigned
```

```
U+xx00  TENGWAR LETTER TINCO
U+xx01  TENGWAR LETTER PARMA
U+xx02  TENGWAR LETTER CALMA
U+xx03  TENGWAR LETTER QUESSE
U+xx04  TENGWAR LETTER ANDO
U+xx05  TENGWAR LETTER UMBAR
U+xx06  TENGWAR LETTER ANGA
U+xx07  TENGWAR LETTER UNGWE
U+xx08  TENGWAR LETTER THUULE (suule)
U+xx09  TENGWAR LETTER FORMEN
U+xx0A  TENGWAR LETTER HARMA (aha)
U+xx0B  TENGWAR LETTER HWESTA
U+xx0C  TENGWAR LETTER ANTO
U+xx0D  TENGWAR LETTER AMPA
U+xx0E  TENGWAR LETTER ANCA
U+xx0F  TENGWAR LETTER UNQUE
U+xx10  TENGWAR LETTER NUUMEN
U+xx11  TENGWAR LETTER MALTA
U+xx12  TENGWAR LETTER NOLDO (ngoldo)
U+xx13  TENGWAR LETTER NWALME (ngwalme)
U+xx14  TENGWAR LETTER OORE
U+xx15  TENGWAR LETTER VALA
U+xx16  TENGWAR LETTER ANNA
U+xx17  TENGWAR LETTER VILYA (wilya)
U+xx18  TENGWAR LETTER ROOMEN
U+xx19  TENGWAR LETTER ARDA
U+xx1A  TENGWAR LETTER LAMBE
U+xx1B  TENGWAR LETTER ALDA
U+xx1C  TENGWAR LETTER SILME
U+xx1D  TENGWAR LETTER SILME NUQUERNA
U+xx1E  TENGWAR LETTER AARE (aaze, esse)
U+xx1F  TENGWAR LETTER AARE NUQUERNA (aaze n., esse n.)
U+xx20  TENGWAR LETTER HYARMEN
U+xx21  TENGWAR LETTER HWESTA SINDARINWA
U+xx22  TENGWAR LETTER YANTA
U+xx23  TENGWAR LETTER UURE
U+xx24  TENGWAR LETTER HALLA
U+xx25  TENGWAR LETTER SHORT CARRIER
U+xx26  TENGWAR LETTER LONG CARRIER
U+xx27  TENGWAR LETTER ANNA SINDARINWA
U+xx28  TENGWAR LETTER EXTENDED THUULE
U+xx29  TENGWAR LETTER EXTENDED FORMEN
U+xx2A  TENGWAR LETTER EXTENDED HARMA
U+xx2B  TENGWAR LETTER EXTENDED HWESTA
U+xx2C  TENGWAR LETTER EXTENDED ANTO
U+xx2D  TENGWAR LETTER EXTENDED AMPA
U+xx2E  TENGWAR LETTER EXTENDED ANCA
U+xx2F  TENGWAR LETTER EXTENDED UNQUE
U+xx30  TENGWAR LETTER STEMLESS OORE (digit zero)
U+xx31  TENGWAR LETTER STEMLESS VALA
U+xx32  TENGWAR LETTER STEMLESS ANNA
U+xx33  TENGWAR LETTER STEMLESS VILYA (digit one)
U+xx34  (This position shall not be used)
U+xx35  (This position shall not be used)
U+xx36  (This position shall not be used)
U+xx37  (This position shall not be used)
U+xx38  (This position shall not be used)
U+xx39  (This position shall not be used)
U+xx3A  (This position shall not be used)
U+xx3B  (This position shall not be used)
U+xx3C  (This position shall not be used)
```

```
U+xx3D  (This position shall not be used)
U+xx3E  (This position shall not be used)
U+xx3F  (This position shall not be used)
U+xx40  TENGWAR SIGN THREE DOTS ABOVE
U+xx41  TENGWAR SIGN THREE DOTS BELOW
U+xx42  TENGWAR SIGN TWO DOTS ABOVE
U+xx43  TENGWAR SIGN TWO DOTS BELOW
U+xx44  TENGWAR SIGN AMATICSE (dot above)
U+xx45  TENGWAR SIGN NUNTICSE (dot below)
U+xx46  TENGWAR SIGN ACUTE (andaith, long mark)
U+xx47  TENGWAR SIGN DOUBLE ACUTE
U+xx48  TENGWAR SIGN RIGHT CURL
U+xx49  TENGWAR SIGN DOUBLE RIGHT CURL
U+xx4A  TENGWAR SIGN LEFT CURL
U+xx4B  TENGWAR SIGN DOUBLE LEFT CURL
U+xx4C  TENGWAR SIGN NASALIZER
U+xx4D  TENGWAR SIGN DOUBLER
U+xx4E  TENGWAR SIGN TILDE
U+xx4F  TENGWAR SIGN BREVE
U+xx50  TENGWAR PUSTA (putta, stop)
U+xx51  TENGWAR DOUBLE PUSTA (putta)
U+xx52  TENGWAR EXCLAMATION MARK
U+xx53  TENGWAR QUESTION MARK
U+xx54  TENGWAR SECTION MARK
U+xx55  TENGWAR LONG SECTION MARK
U+xx56  TENGWAR SIGN LONG CARRIER BELOW
U+xx57  TENGWAR SIGN DOUBLE ACUTE BELOW
U+xx58  TENGWAR SIGN RIGHT CURL BELOW
U+xx59  (This position shall not be used)
U+xx5A  TENGWAR SIGN LEFT CURL BELOW
U+xx5B  (This position shall not be used)
U+xx5C  TENGWAR SIGN LEFT FOLLOWING SILME
U+xx5D  TENGWAR SIGN RIGHT FOLLOWING SILME
U+xx5E  (This position shall not be used)
U+xx5F  (This position shall not be used)
U+xx60  (This position shall not be used)
U+xx61  (This position shall not be used)
U+xx62  TENGWAR DIGIT TWO
U+xx63  TENGWAR DIGIT THREE
U+xx64  TENGWAR DIGIT FOUR
U+xx65  TENGWAR DIGIT FIVE
U+xx66  TENGWAR DIGIT SIX
U+xx67  TENGWAR DIGIT SEVEN
U+xx68  TENGWAR DIGIT EIGHT
U+xx69  TENGWAR DIGIT NINE
U+xx6A  TENGWAR DUODECIMAL DIGIT TEN
U+xx6B  TENGWAR DUODECIMAL DIGIT ELEVEN
U+xx6C  TENGWAR DECIMAL BASE MARK
U+xx6D  TENGWAR DUODECIMAL BASE MARK
U+xx6E  TENGWAR DUODECIMAL LEAST SIGNIFICANT DIGIT MARK
U+xx6F  (This position shall not be used)
U+xx70  (This position shall not be used)
U+xx71  (This position shall not be used)
U+xx72  (This position shall not be used)
U+xx73  (This position shall not be used)
U+xx74  (This position shall not be used)
U+xx75  (This position shall not be used)
U+xx76  (This position shall not be used)
U+xx77  (This position shall not be used)
U+xx78  (This position shall not be used)
U+xx79  (This position shall not be used)
U+xx7A  (This position shall not be used)
```

```
U+xx7B  (This position shall not be used)
U+xx7C  (This position shall not be used)
U+xx7D  (This position shall not be used)
U+xx7E  (This position shall not be used)
U+xx7F  (This position shall not be used)
```

	xx0	xx1	xx2	xx3	xx4	xx5	xx6	xx7
0	p	ɩɔ	λ	ɔ	◌̈	·		
1	p	ɩɔ	d	ɔ	◌̈	:		
2	q	cɑ	λ	c	◌̈	ſ	ᴄ	
3	q	ɩɔ	o	c	◌	β	ᴄᴄ	
4	pɔ	ɪɔ	ǀ		◌	~	ʟ	
5	pɔ	ɪɔ	ǀ		◌	~	ʦ	
6	cq	ɑ	ǀ		◌	◌	ʦ	
7	cq	ɑ	cr		◌	◌	ʃ	
8	b	ʒ	þ		◌	◌	ʄ	
9	b	ʒ	þ		◌		ʄɔ	
A	d	⊂	q		◌	◌	ʃ	
B	d	ꟻ	q		◌		ʒ	
C	bɔ	Ꮯ	þɔ		◌̄	ᶘ	◌̇	
D	bɔ	ꝑ	þɔ		◌	ᶘ	◌	
E	cd	ʒ	cq		◌		◌	
F	ɑd	ꝫ	ɑq		◌			