Artificial Scripts X.X Tengwar

X.X Tengwar (draft by Michael Everson, 2001-03-07)

Tengwar 1xx00–1xx7F

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 community of linguists and other enthusiasts interested in this expression of Tolkien's expertise in historical and comparative linguistics (*tengwar* is plural of Quenya *tengwa* 'letter'; the Sindarin singular and plural are **têw* (or **teiw*) and *tîw*). The Tengwar should be treated as a Category D (Attested Extinct) script: there is a relatively limited corpus, and a relatively small scholarly body studying it. In order to provide a standard Tengwar character coding for both scholars and enthusiasts (many of whom create new texts with the Tengwar), it has been suggested that this character set be included in Plane 1 of the UCS. Eight columns are reserved to encode the Tengwar. Character names mostly 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+1xx00-U+1xx17) are formed of a telco 'stem' and a lúva 'bow'; raising the stem might indicate spirantization of a consonant (Noldorin psố toron 'brother', bố thoron 'eagle'), or doubling the bow might indicate voicing (Noldorin psố doron 'oak'). Consonants are modified by tehtar 'signs', described below. A series of "stemless consonants" could be encoded. U+1xx31 2 STEMLESS VALA is used in some modes for English [w]; U+1xx32 c STEMLESS ANNA is used as a vowel in the mode of Beleriand; *STEMLESS OORE, and *STEMLESS VILYA are as yet unattested, but could be included because of the inherent structure of the script. On the other hand, this may not be appropriate, as U+1xx71 2 DIGIT ZERO is essentially *STEMLESS OORE and U+1xx72 c DIGIT ONE is essentially *STEMLESS VILYA.

Tengwar are written from left-to-right. Tengwar digits are written from right-to-left (the least significant digit is on the left). The <u>DECIMAL BASE MARK and DUODECIMAL BASE MARK</u> are applied to the digits to indicate what the arithmetical base is used; U+1xx7E of 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, at least in duodecimal notation.

A related script written in vertical columns (the Sarati script of Rúmil, from which Tengwar was derived according to Tolkien's mythology) appears to be as separate from Tengwar as Aramaic is separate from Mongolian. It is not encoded here.

Rendering. Non-spacing marks, generically called *tehtar* 'signs' (singular *tehta*), indicate vowels (*ómatehtar*) 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. Most of these marks, however, are *not* combining

characters, because it cannot be said that they should always *follow* the character to which they are applied, if logical (phonetic) order is to be applied to this script.

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 (*i.e.* with a CV structure); but they are placed above or below the *following* consonant in languages in which words tend to end in a consonant (*i.e.* with a CVC structure). Compare Quenya $\cancel{m} \le nel\underline{de}$ 'three', $\cancel{m} \ne nelt\underline{ildi}$ 'triangle' with Sindarin $\cancel{m} \ne neled$ 'three' and $\cancel{m} \ne nelt\underline{hil}$ 'triangle'. English is generally written according to a Sindarin-type mode; Japanese would be written according to a Quenya-type mode.

Tengwar shaping behaviour is considered to be similar to that of Arabic, in that certain sequences of characters combine automatically to produce certain glyphs. Fonts must be identified as pertaining to one of the two modes. In a Quenya font, the sequence $C \subset \mathbb{R}^{1} \times \mathbb$

In the mode of Beleriand, the *tehtar* are not generally used to indicate vowels. Instead, individual *tengwar* represent them. In this mode, GITDC SIGNORIMATILLION is written just as presented: [S][I][L][M][A][R][I][L][I][O][N]. Note that the DOUBLER of L ll is a true combining character here as above in the Quenya and Sindarin modes.

True ligatures also occur in Tengwar; these should be represented with the U+200D zero-width joiner in plain text. Examples (in a Sindarin-mode font): Old English $\text{price} \hat{p}$ \hat{p} \hat{m} $\hat{m$

Spelling variants

A certain variation in spelling can occur, but the model remains sound. For the word English *animal*, the following spellings would obtain:

Roman	Tengwar	Representation	Cursive glyphs (mode)
animal	حشن أ	ح + [°] + \ \ ا + \ \ + \ \ + \ ا	[XA][NI][MA][L] (Quenya)
animal	څڟڟ		[AN][IM][AL] (Sindarin)
animal	ם ומ	c + a + a + a + a + b	[A][N][I][M][A][L] (either)

Artificial Scripts X.X Tengwar

For some lexical purposes the variation would have to be taken into account, though this is not much different from what has to be done in English with spellings like façade/facade, colour/color, or æsthetic/aesthetic/esthetic.

In the mode of Beleriand, "a dot is sometimes placed over and 1 to mark these as separate letters, and not a curl or a stem belonging to an adjacent letter" (Krieg 1978). Examples are ccq cs aglar 'glory' and pcdcasaps nachaered 'to a remote distance' (in the second example the dot shows that cd is ach rather than cd gh). The true combining character U+0307 COMBINING DOT ABOVE should be used for these instances. Likewise, when in the mode of Beleriand the acute is used atop a tengwa used as a vowel, the combining character U+0301 COMBINING ACUTE ACCENT should be used. It is only when a tehta is vocalic that the cursivity model applies. The user has to know which of the dots applies.

Punctuation. Tengwar punctuation characters are considered to be unique to the script and are coded in the Tengwar block. Some composition of punctuation occurs in Tengwar. The following combinations are attested:

- FULL STOP
- · PUSTA
- : DOUBLE PUSTA
- : DOUBLE PUSTA+ PUSTA
- :: DOUBLE PUSTA + DOUBLE PUSTA
- :~ DOUBLE PUSTA+ SECTION MARK:≈ DOUBLE PUSTA+ DOUBLE SECTION MARK
- : TRIPLE PUSTA
- ∴ QUADRUPLE PUSTA (not ·:· PUSTA + DOUBLE PUSTA + PUSTA)
- ⊕ QUINTUPLE PUSTA (not ·:· PUSTA + TRIPLE PUSTA + PUSTA)
- S EXCLAMATION MARK
- **S** EXCLAMATION MARK + PUSTA

Sometimes word space is not used; word separation may be achieved in that case with U+200B, ZERO-WIDTH SPACE. Hyphens are not used at linebreaks; words may be broken before any syllable. One has to know the language to know where the syllable boundaries are.

The SHORT CARRIER 1 simply bears the vowel *tehta*; the LONG CARRIER 1 indicates that the vowel is long; this can also be done by doubling the vowel sign (so i = e, $j = \hat{e}$, $j = \hat{e}$). In Quenya fonts, these are coded j + j, etc.; in Sindarin fonts, they are coded j + j, etc.

Sources for characters

The characters in this proposal are attested (explicitly or in technical description) in the following sources. Page numbers follow the UCS code positions given.

Tolkien 1965a xx53:6 Tolkien 1965b xx67:59 Tolkien 1965c xx30:319, xx32:319 Tolkien 1965d

```
xx00:396, xx01:396, xx02:396, xx03:396, xx04:396, xx05:396, xx06:396, xx07:396,
   xx08:396, xx09:396, xx0A:396, xx0B:396, xx0C:396, xx0D:396, xx0E:396,
   xx0F:396, xx10:396, xx11:396, xx12:396, xx13:396, xx14:396, xx15:396, xx16:396,
   xx17:396, xx18:398, xx19:398, xx1A:398, xx1B:398, xx1C:398, xx1D:398,
   xx1E:398, xx1F:398, xx20:396, xx21:396, xx22:396, xx23:396, xx24:396, xx25:396,
   xx26:396, xx27:396, xx28:396, xx29:396, xx2A:396, xx2B:396, xx2C:400,
   xx2D:401, xx2E:399, xx42:400, xx45:400, xx4E:400, xx4F:400, xx50:400
Tolkien & Swann 1967
   xx40:v, xx43:v, xx44:v, xx46:v, xx48:v, xx4A:v, xx4C:v, xx51:v, xx60:v, xx61:ix,
   xx63:vi, xx64:vi
Tolkien 1977
   xx35:4, xx4B:4
Tolkien 1989
   xx34:ii
Tolkien 1992 (pictures)
   xx31, xx36
Tolkien 1992a (epilogue Sauron defeated)
   xx55, xx62, xx65
Tolkien 1992b (notion club Sauron defeated)
   xx41:326, xx49:325, xx4D:326, xx52:324,. xx57:326
Used positions (not yet sourced)
   xx2F
   xx33, xx37
   xx54, xx56,
   xx66, xx68
   xx70, xx71, xx72, xx73, xx74, xx75, xx76, xx77, xx78, xx79, xx7A, xx7B, xx7C,
   xx7D
```

Unused positions:

xx38, xx39, xx3A, xx3B, xx3C, xx3D, xx3E, xx3F, xx47, xx58, xx59, xx5A, xx5B, xx5C, xx5D, xx5E, xx5F, xx69, xx6A, xx6B, xx6C, xx6D, xx6E, xx6F, xx7E, xx7F

Encoding Structure (*This needs to be revised when the code table is fixed*)

The Tangyar block is divided into the following ranges:

The Tengwar block is divided into the following ranges:

```
U+xx01 \rightarrow xx3E
                          Tengwar
U+xx2E \rightarrow xx30
                          unassigned
U + xx31 -> xx32
                          Tengwar
U+xx33
                          unassigned
U + xx34 -> xx39
                           Tengwar
U+xx3A \rightarrow xx3F
                          unassigned
U + xx40 -> xx46
                          Tehtar
U+xx47
                          unassigned
U + xx48 -> xx57
                          Tehtar
U+xx58 \rightarrow xx5F
                          unassigned
U + xx60 -> xx68
                          Punctuation
U+xx69 \rightarrow xx6F
                          unassigned
U+xx70 \rightarrow xx7D
                          Digits
U+xx7E
                          Numeric modifiers
U+xx7F
                          unassigned
```

Artificial Scripts X.X Tengwar

Unicode Character Properties

(These have not been sorted out yet.)

Spacing letters, category "Lo", bidi category "L" (strong left to right)

xx00-xx37, xx40-xx46, xx48-xx4F, xx52-xx54

Non-spacing marks, category "Mn", bidi category "ON" (other neutral); combining priorities in parentheses:

xx50-xx51, xx55-xx57, xx7D (xxxxxxx)

Symbols, category "Po", bidi category "L" (strong left to right) xx60-xx68

Numbers, category "Po", bidi category "R" (strong right to left) xx70-xx7C

References

- Krieg, Lawrence J. 1978. "The Tengwar of Fëanor", in Jim Allan, ed. *An introduction to Elvish*. Frome: Bran's Head. Pp. 236–XX.
- Tolkien, Christopher. 1992a. "The Epilogue" in *Sauron defeated*. (The History of Middle Earth; 9) London: HarperCollins. Pp. 114-135. ISBN 0-261-10240-0
- Tolkien, Christopher. 1992b. Notes to "The Notion Club Papers (Part Two)" in *Sauron defeated*. (The History of Middle Earth; 9) London: HarperCollins. Pp. 307-327. ISBN 0-261-10240-0
- Tolkien, J. R. R. 1965a. Title page to *The Fellowship of the Ring*. Boston: Houghton Mifflin. ISBN 0-395-08254-4
- Tolkien, J. R. R. 1965b. "The Ring inscription", in *The Fellowship of the Ring*. Boston: Houghton Mifflin. P. 59. ISBN 0-395-08254-4
- Tolkien, J. R. R. 1965c. "The West-Gate inscription", in *The Fellowship of the Ring*. Boston: Houghton Mifflin. P. 319. ISBN 0-395-08254-4
- Tolkien, J. R. R. 1965d. "Appendix E: Writing and spelling" in *The return of the King*. Boston: Houghton Mifflin. Pp. 391-404. ISBN 0-395-08256-0
- Tolkien, J. R. R., and Donald Swann. 1967. *The road goes ever on: a song cycle*. Boston: Houghton Mifflin. ISBN 0-395-08258-7
- Tolkien, J. R. R. 1977. Title page to *The Silmarillion*. Boston: Houghton Mifflin. ISBN 0-395-25730-1
- Tolkien, J. R. R. 1989. Title page to *The Lost Road and other writings*. London: Unwin Paperbacks. (The History of Middle Earth; 5) ISBN 0-04-440398-4
- Tolkien, J. R. R. 1992. *Pictures by J. R. R. Tolkien*. London: HarperCollins. Plate 48. ISBN 0-261-10258-3
- Tolkien, J. R. R. 19XX. "SXXX" in *The Return of the King*. Boston: Houghton Mifflin.

Annex A. Discussion of the Tengwar number system.

Numerals.

The tengwar numerals were found and presented to Quettar by Christopher Tolkien (see Quettars 13 and 14). The following is adapted therefrom.

For lists, series and the like, the tengwar from 1 to 24 (p, p, ..., a) were used, just as we use (a), (b), A sign such as a dot or bar above could be used, e.g. p or p.

For numbers proper, the Eldar used both a decimal and a duodecimal system, the Dwarves used a duodecimal system, and the Men of the West in the Third Age used mainly a decimal system. The digits used were as follows:

In all systems the numbers are written with the unit digit at the left. The digits were usually marked, either by a dot over each digit, or by a line drawn above the number, for a decimal system; in a duodecimal system, the dot or line was written below. The mark could be omitted where no confusion would arise. In a duodecimal system, a circle could be used (in place of the dot) to mark the twelves digit; and the symbol o or 9 could be used for 12 (though how, it is not clear!).

Examples:

(Italic numbers are used for reversed numbers, roman for normal numbers.)

Decimal:

$$\frac{\dot{2}\dot{c} = 01 = 10}{c}$$
 غَطْرَ $\dot{c} = 091 = 190$
قطر خالت $\frac{\dot{2}\dot{c}}{c} = 091 = 190$

$$93 \times 3 \times \times = 70 \hat{11} 2 \hat{10} 3 \hat{2}_{12} = 23 \hat{10} 2 \hat{11} 0 7_{12} = 6930 871$$

The only numerals we have for the runes are three digits found in the Book of Mazarbul in Pictures by J. R. R. Tolkien: A = 3, W = 4, Y = 5.

THE TENGWAR NUMERALS

(For numeration, especially in lists or series, the letters up to II 24 (see The Return of the King p. 396, table of the Tengwar) were commonly used, with or without some sign, as p. p.)

The numerals of the Feanorian script were arranged in triads:

The smaller value was placed on the left: thus jt (01) = 10. th (24)

When the use of the numerals as such was evident, and no confusion could arise, the dots could be omitted. Often a long series of numerals in the midst of words was marked by a line drawn above, as:

to be read as 'six million, nine hundred & thirty thousand, eight hundred & seventy one.

CRT after JRRT 16 May 1981

TABLE XX - Row xx: TENGWAR

	xx0	xx1	xx2	ххЗ	xx4	xx5	xx6	xx7
0	p	כמ	8		*	ै	•	С
1	р	മ	B	2	*	ੁ	•	τ
2	q	ccı	T	c	••	\	•	T
3	9	Œ	5		••	~	•:•	ac
4	po	a	G	Γ	•	`	•••	
5	ည	מ	9	cr	•	ं	S	ट
6	cq	a	\G	ч	/	ं	B	च्च
7	q	a	3	q		ុ	J	J
8	b	þ	λ	Д	//		~	ط
9	Ь	þ	d	2	"		≈	ط
Α	d	d	A	ညဒ	P			9
В	d	þ	0		b			3
С	bo	ြာ	1		d			0
D	pa	ည	J		J			
Е	ccl	ccl	1		r			
F	ದ	व्य			J			

G = 00 P = 01

TABLE XX - Row xx: TENGWAR

dec	hex	Name	dec	hex	Name
000	00	TENGWAR LETTER TINCO	089	59	(This position shall not be used)
001 002	01 02	TENGWAR LETTER PARMA TENGWAR LETTER CALMA	090 091	5A 5B	(This position shall not be used) (This position shall not be used)
003	03	TENGWAR LETTER CALMA TENGWAR LETTER QUESSE	092	5C	(This position shall not be used) (This position shall not be used)
004	04 05	TENGWAR LETTER ANDO TENGWAR LETTER UMBAR	093	5D 5E	(This position shall not be used) (This position shall not be used)
006	06	TENGWAR LETTER ANGA	095	5F	(This position shall not be used)
007	07 08	TENGWAR LETTER UNGWE TENGWAR LETTER THUULE (suule)	096 097	60 61	TENGWAR PUSTA (putta, stop) TENGWAR DOUBLE PUSTA (putta)
009	09	TENGWAR LETTER FORMEN`	098	62	TENGWAR TRIPLE PUSTA (putta)
010	0A 0B	TENGWAR LETTER HARMA (aha) TENGWAR LETTER HWESTA	099 100	63 64	TENGWAR QUADRUPLE PUSTA (putta) TENGWAR QUINTUPLE PUSTA (putta)
012	0C	TENGWAR LETTER ANTO	101	65	TENGWAR EXCLAMATION MARK
013 014	0D 0E	TENGWAR LETTER AMPA TENGWAR LETTER ANCA	102	66 67	TENGWAR QUESTION MARK TENGWAR PARENTHESIS MARK
015	0F	TENGWAR LETTER UNQUE	104	68	TENGWAR SECTION MARK
016 017	10 11	TENGWAR LETTER NUUMEN TENGWAR LETTER MALTA	105 106	69 6A	TENGWAR DOUBLE SECTION MARK (This position shall not be used)
018	12	TENGWAR LETTER MALTA TENGWAR LETTER NOLDO (ngoldo)	107	6B	(This position shall not be used) (This position shall not be used)
019 020	13 14	TENGWAR LETTER NWALME (ngwalme) TENGWAR LETTER OORE	108	6C 6D	(This position shall not be used) (This position shall not be used)
021	15	TENGWAR LETTER VALA	110	6E	(This position shall not be used)
022 023	16 17	TENGWAR LETTER ANNA TENGWAR LETTER VILYA (wilya)	111 112	6F 70	(This position shall not be used) TENGWAR DIGIT ZERO
024	18	TENGWAR LETTER EXTENDED TINCO	113	71	TENGWAR DIGIT ONE
025 026	19 1A	TENGWAR LETTER EXTENDED PARMA TENGWAR LETTER EXTENDED CALMA	114 115	72 73	TENGWAR DIGIT TWO TENGWAR DIGIT THREE
027	1B	TENGWAR LETTER EXTENDED QUESSE	116	74	TENGWAR DIGIT FOUR
028 029	1C 1D	TENGWAR LETTER EXTENDED ANDO TENGWAR LETTER EXTENDED UMBAR	117 118	75 76	TENGWAR DIGIT FIVE TENGWAR DIGIT SIX
030	1E	TENGWAR LETTER EXTENDED ANGA	119	77	TENGWAR DIGIT SEVEN
031	1F 20	TENGWAR LETTER EXTENDED UNGUE TENGWAR LETTER ROOMEN	120 121	78 79	TENGWAR DIGIT EIGHT TENGWAR DIGIT NINE
033	21	TENGWAR LETTER ARDA	122	7A	TENGWAR DUODECIMAL DIGIT TEN
034 035	22 23	TENGWAR LETTER LAMBE TENGWAR LETTER ALDA	123 124	7B 7C	TENGWAR DUODECIMAL DIGIT ELEVEN TENGWAR DUODECIMAL DIGIT TWELVE
036	24	TENGWAR LETTER SILME	125	7D	TENGWAR DUODECIMAL LEAST SIGNIFICANT DIGIT MARK
037 038	25 26	TENGWAR LETTER SILME NUQUERNA TENGWAR LETTER AARE (aaze, esse)	126 127	7E 7F	(This position shall not be used) (This position shall not be used)
039	27	TENGWAR LETTER AARE NUQUERNA (aaze n., esse n.)	121	''	(This position shall not be used)
040 041	28 29	TENGWAR LETTER HYARMEN TENGWAR LETTER HWESTA SINDARINWA			
042	2A	TENGWAR LETTER YANTA			
043	2B 2C	TENGWAR LETTER UURE TENGWAR LETTER LONG CARRIER			
045	2D	TENGWAR LETTER HALLA			
046 047	2E 2F	TENGWAR LETTER SHORT CARRIER (This position shall not be used)			
048	30	(This position shall not be used)			
049 050	31 32	TENGWAR LETTER STEMLESS VALA TENGWAR LETTER STEMLESS ANNA			
051	33	(This position shall not be used)			
052 053	34 35	TENGWAR LETTER LIGATING SHORT CARRIER TENGWAR LETTER ANNA SINDARINWA			
054	36	TENGWAR LETTER OPEN ANNA			
055 056	37 38	TENGWAR LETTER REVERSED PARMA TENGWAR LETTER REVERSED FORMEN			
057	39	TENGWAR LETTER TALL STEMLESS VALA			
058 059	3A 3B	TENGWAR LETTER MH (This position shall not be used)			
060 061	3C 3D	(This position shall not be used)			
062	3E	(This position shall not be used) (This position shall not be used)			
063 064	3F 40	(This position shall not be used) TENGWAR SIGN THREE DOTS ABOVE			
065	41	TENGWAR SIGN THREE DOTS BELOW			
066 067	42 43	TENGWAR SIGN TWO DOTS ABOVE TENGWAR SIGN TWO DOTS BELOW			
068	44	TENGWAR SIGN AMATICSE (dot above)			
069 070	45 46	TENGWAR SIGN NUNTICSE (dot below) TENGWAR SIGN ACUTE (andaith, long mark)			
071	47	(This position shall not be used)			
072 073	48 49	TENGWAR SIGN DOUBLE ACUTE TENGWAR SIGN DOUBLE ACUTE BELOW			
074	4A	TENGWAR SIGN RIGHT CURL			
075 076	4B 4C	TENGWAR SIGN RIGHT CURL BELOW TENGWAR SIGN LEFT CURL			
077	4D	TENGWAR SIGN LEFT CURL BELOW			
078 079	4E 4F	TENGWAR SIGN DOUBLE RIGHT CURL TENGWAR SIGN DOUBLE LEFT CURL			
080	50	TENGWAR SIGN NASALIZER			
081 082	51 52	TENGWAR SIGN DOUBLER TENGWAR SIGN TILDE			
083	53	TENGWAR SIGN BREVE			
084 085	54 55	TENGWAR SIGN GRAVE TENGWAR SIGN YANTA ABOVE			
086	56	TENGWAR SIGN THREE INVERTED DOTS ABOVE			
087 088	57 58	TENGWAR SIGN LONG CARRIER BELOW (This position shall not be used)			
		(50 mon 50 mon)			