

Writing the Arabic language with *Tengwar* characters

Introduction

This document discusses my ideas for a “mode” suitable for writing the Arabic language with J.R.R. Tolkien’s *Tengwar* characters. A brief explanation by Tolkien of this alphabet can be found in Appendix E of his book *The Lord of the Rings: The Return of the King*, and there is also abundant information available on-line. (See the HTML web-site version of this document for links to various sites, and downloadable Tengwar Arabic fonts.)

The Arabic Tengwar mode described here is a thought-experiment which shows how a number of annoying ambiguities, incompletenesses, and cases of lack of visual distinctiveness in the Arabic alphabet (as it is commonly used) might be remedied by using a non-cursive left-to-right alphabet, but while retaining a number of the basic structural features of current Arabic orthography – and without resorting to Latin script (since the Latin alphabet is not suitable to write the Arabic language unless a large number of diacritic marks and/or new symbols were added, and since many Arabs would regard the use of the Latin alphabet to write Arabic as inappropriate cultural imperialism). In addition, the Tengwar Arabic mode has features which make it useful in transcribing not just standard written Arabic, but also Arabic colloquial dialects (something which is very difficult to do with the current Arabic-script orthography).

Current Arabic Orthography

Here is a very brief discussion of some of the difficulties with current Arabic orthography:

First, Arabic letters can take on radically different shapes depending on the surrounding context of the letters (if any) which occur preceding and following (see the horizontal row at the bottom of Table 1). And many of these contextual Arabic letter-forms are only distinguished from contextual forms of completely different Arabic letters (which represent completely different sounds) by means of detached diacritical dots which aren’t part of the basic letter-shape at all (see the vertical column in Table 1). So a particular form of an Arabic letter (such as the medial form of *yāʔ*) often resembles forms of other Arabic letters (such as the medial forms of the letters for “th”, “t”, “n”, and “b”) more than it does the other forms of the same letter (such as the initial, final, and isolated forms of *yāʔ*) – and often only the presence of the diacritical dots serves to distinguish letters which represent radically different sounds (such as [b] and [n]):

Table 1

Medial contextual forms of several different letters:	ⲉ	θ
	Ⲇ	t
	Ⲏ	n
Contextual variants of the letter <i>yāʔ</i> :	ⲃ	b
→	ي ي ي ي	y

Secondly, all the letters of the Arabic alphabet are originally basically consonantal, but some of them have come to be used to write long vowels. Therefore the letters which can write both consonant and long vowel sounds are significantly ambiguous – unless the **full** panoply of Arabic-script diacritics is used (which it very seldom is, outside of Qur’ans, dictionaries, and children’s textbooks). In the case of the letter *ʔalif* (which is used to write both the glottal stop consonant sound and the long [a] vowel sound), the ambiguities are considered to be so severe that a whole complicated series of diacritics has been developed exclusively (or almost exclusively) for use with the letter *ʔalif* (see Table 2) – a special series of diacritics which is separate from the general system of diacritics denoting short vowels and syllable structure. However, even these special diacritics shown in Table 2 unfortunately do not completely resolve all the ambiguities connected with the letter *ʔalif* without the presence of the rarely-used general vowel diacritics – and the special diacritics of Table 2 are themselves quite often omitted in practice when writing Arabic.

Table 2



(For a partial explanation of these particular diacritics, see the listing in Appendix 1 below.)

Other ambiguities of current Arabic-script orthography will be referred to below, in the discussion of how these problems are resolved in the Tengwar orthography.

Listing of Tengwar characters to write Arabic

The following three pages each contain three diagrams showing the use of Tengwar “primary” letters and “additional” letters (as defined by Tolkien) – together with other modified letters and diacritics – to write the Arabic language. After each Tengwar character is placed first the corresponding symbol of Arabic-script orthography, and then a transcription of the relevant sound (using conventional phonetic and/or Semitological symbols which are often used in linguistic works, but which will not be explained here). If a Tengwar character corresponds to a sound which is not used in the basic vocabulary of standard written Arabic (but is used in Arabic dialects and/or the pronunciation of foreign words borrowed into Arabic), then the corresponding Arabic orthography symbol and “broad phonetic” transcription are shown in [blue](#).

Each page uses a different Tengwar font – first, a basic font (Tengwar Quenya); second, a “calligraphic” font (Tengwar Elfica), and last a “sans-serif” font (Tengwar Elfetica). Note that some of the letters in the second diagram on each page were devised by myself for this Arabic Tengwar mode, and will not be found in standard Tengwa fonts available over the Internet. Instead, special fonts with the Arabic-specific Tengwar symbols can be downloaded from the website.

(Turn page to see diagram.)

“Primary” letters

p ت t	پ پ p	ق ك k	ق ق q
د d	ب b	ج g	—
ث θ	ف f	خ χ	ح ḥ
ذ ð	ف v	غ γ	ع ʿ
ن n	م m	—	—
ر r	و w	ا إطباق Emphasis	ؤى إاء ؟

“Additional” letters, etc.

ل l	س s	ز z	ه h
ي y	ج j	ط t	ظ z, ð
ش š	ص ṣ	ض d	أ Silent ʔalif in various contexts
ا ā	و ū	ي ī	آ ʔā
ى ā	ة t ^{or} silent	ال al, l or doubling of the following consonant (التعريف)	

Diacritics

ا a	ي i	و u	د doubling of consonant
او aw	اي ay	e	o
o	e	تنوين (إعراب) Vowel + word-final indefinite -n	

“Primary” letters

ط ت t	پ p	ق ك k	ق q
ظ د d	ب b	ج g	—
ث θ	ف f	خ χ	ح ḥ
ذ ð	ظ v	غ γ	ع ʿ
ن n	م m	—	—
ر r	و w	ا ^{إطباق} Emphasis	ؤىءأء ?

“Additional” letters, etc.

ل l	س s	ز z	ه h
ي y	ج j	ط t	ظ z, ð
ش š	ص ṣ	ض d	أ ^{Silent ʔalif in various contexts}
ا ā	و ū	ي ī	آ ?ā
ى ā	ة ^{t^{or} silent}	ال al, l ^{or doubling of the following consonant}	

(*Definite Article* التعريف)

Diacritics

ا̇ a	ا̇ i	ا̇ u	ا̇ ^{doubling of consonant}
ا̇aw	ا̇ay	ا̇ e	ا̇ o
ا̇ - ā	ا̇ - ē	ا̇ (إعراب) ^{تنوين}	ا̇ ^{Vowel + word-final indefinite -n}

“Primary” letters

ط	ت	t	پ	پ	p	ك	ك	k	ق	ق	q
م	د	d	ب	ب	b	ج	ج	g	—	—	—
ث	ث	ḥ	ف	ف	f	خ	خ	ḫ	ح	ح	ḥ
ذ	ذ	ḏ	ف	ف	v	غ	غ	ḡ	ع	ع	ʿ
ن	ن	n	م	م	m	—	—	—	—	—	—
ر	ر	r	و	و	w	إطباق	إطباق	u	ؤ	ؤ	ʾ
						Emphasis	Emphasis				

“Additional” letters, etc.

ل	ل	l	س	س	s	ز	ز	z	ه	ه	h
ي	ي	y	ج	ج	j	ط	ط	ṭ	ظ	ظ	ẓ, ḏ
ش	ش	š	ص	ص	ṣ	ض	ض	ḍ	أ	أ	Silent ʾalif in various contexts
ا	ا	ā	و	و	ū	ي	ي	ī	آ	آ	ʾā
ى	ى	ā	ة	ة	t ^{or} silent	ال	ال	al, l			or doubling of the following consonant
											(Definite Article التعريف)

Diacritics

(Vowel diacritics would not ordinarily be used with this font.)



doubling of consonant

— Semivowels ([y] and [w] sounds) and long high vowels (long [i] and [u] sounds) are strictly distinguished in the Tengwar Arabic mode, so that that no single Tengwar symbol can write **both** a long [i] vowel and a following [y] simultaneously (as the letter *yāʔ* sometimes does in current Arabic-script orthography). Therefore, the *nisbah* adjective suffix must always be written with a sequence of long [i] Tengwa + [y] Tengwa (as shown in the case of the words ʕarabīyah and qadhhdhāfī above). This also has the effect of distinguishing certain pairs of forms which are not ordinarily distinguished in current Arabic orthography (unless the full panoply of diacritics is used):

Examples of orthographic contrasts created by vowel/semivowel distinctions of Tengwar Arabic mode:

في $\text{b} \text{ʔ} \text{ʕ}$ fīya “in me” يدعو $\text{ʕ} \text{p} \text{ʔ} \text{ʕ} \text{w}$ yadʕuwa “calls”
(subjunctive)

في $\text{b} \text{ʔ}$ fī “in” يدعو $\text{ʕ} \text{p} \text{ʔ} \text{ʕ} \text{u}$ yadʕū “calls”
(indicative)

القاضي “the judge” $\text{ʕ} \text{q} \text{ʕ} \text{ʕ} \text{ʕ}$ (nominative & genitive) $\text{ʕ} \text{q} \text{ʕ} \text{ʕ} \text{ʕ}$ (accusative)

— Those words which in current Arabic-script orthography still retain remnants of the archaic tendency not to write long [a] sounds with the letter *ʔalif* are regularized in the Arabic Tengwar mode:

Regular use of long [a] Tengwar symbol:

الله $\text{ʕ} \text{ʕ} \text{ʕ} \text{ʕ}$ allāh “God” هذا $\text{ʕ} \text{h} \text{ʕ} \text{ʕ}$ hāḏā “this”

— The combination of the tengwar glottal-stop symbol with a short [i] vowel diacritic over it doesn’t have any inherently special status in Tengwar Arabic mode, but it may be used with greater frequency than other consonant letter + short-vowel diacritic combinations if there is a desire to preserve the information signalled in the current orthography by *ʔalif* with *hamzah* below (at or near the beginning of a word), or by *yāʔ* with *hamzah* next to a short [i] vowel (in the middle of a word):

Use of Tengwar short-i diacritic to preserve *hamzah* vowel information:

الإسلام $\text{ʕ} \text{ʔ} \text{ʕ} \text{ʕ} \text{w}$ al-ʔislām “Islam”
 دوائر $\text{p} \text{w} \text{ʕ} \text{ʕ} \text{w}$ dawāʔir “circles”

— Since *hamzat-ul-waʕl* (a deletable short vowel at the beginning of a word’s main stem, or the definite article) and *hamzat-ul-qatʕ* (a glottal-stop consonant sound at the beginning of a word’s main stem) are written with completely different Tengwar letters, the contrast between them is strictly obligatory in Tengwar Arabic mode (as opposed to current Arabic orthography, where the contrast is indicated by the presence of a somewhat optional *hamzah* diacritic). The following table shows the contrast between a verbal noun or *maʕdar* belonging to the Stem IV or *yufʕilu* verbal derivation (with true glottal stop) and a verbal noun belonging to the Stem VIII or *yaftaʕilu* verbal derivation (with *hamzat-ul-waʕl*):

Obligatory distinction between *hamzat-ul-waṣl* and *hamzat-ul-qaṭʿ*:

إضراب ṭḥṛṇṇ ʔidrāb “a strike”

اضطراب ṭḥṫṛṇṇ idṭirāb “disorder”

— One goal of Tengwar Arabic mode is to make sure that the boundary between the main stem of a word and any attached syntactic clitic prefixes is always completely unambiguous. These syntactic prefixes are *wa-* “and”, *fa-* “and”, *bi-* “in, by”, *li-* “to, for”, *ka-* “like”, *la-/l-* intensive sentence particle, *sa-* future tense marker, *ʔa-* interrogative particle, and *al-* “the”. In current Arabic orthography, attaching one (or more) of these prefixes to a word often changes the visual appearance of what it attaches to (due to the cursive contextual nature of Arabic script), and it is not always immediately obvious what is part of a syntactic clitic prefix, and what is part of the word itself (the main stem).

In Tengwar Arabic mode, the definite article *al-* “the” is always written with a special ligature symbol (combining the Tengwar “short carrier” symbol, which writes *hamzat-ul-waṣl*, with the Tengwar letter for the [l] sound), as seen in various tables above. This means that the definite article is written unambiguously – and since no other clitic prefix can come after the definite article, therefore whenever a word has a definite article, the boundary between clitic prefixes and the main stem always falls immediately **after** the definite article.

Similarly, any *hamzat-ul-waṣl* which is not part of the definite article always occurs at the beginning of a word’s stem, so that if the Tengwar “short carrier” symbol occurs near the beginning of an orthographic word (and is not part of the definite article ligature), then the boundary between the syntactic clitic prefixes and the main stem always falls immediately **before** the “short carrier” symbol.

For these reasons, the “short carrier” symbol is NEVER deleted when syntactic prefixes are added to a word – unlike in current Arabic orthography, where the letter *ʔalif* which marks *hamzat-ul-waṣl* is sometimes deleted:

**Retention of Tengwar “short carrier” symbol
(*hamzat-ul-waṣl*) after syntactic clitic prefixes:**

باليڊ ṇḫṛṇṇ bilyad (bi-ʿl-yad) “by the hand”

لليڊ ḫḫṛṇṇ lilyad (li-ʿl-yad) “for the hand”

بسم ṇṇṇṇ bism (bi-ʿsm) “in the name of”

لله ḫḫḫṛṇṇ lillāh (li-ʿllāh) “for God”

So if a word contains one or more syntactic clitic prefixes, but no definite article or *hamzat-ul-waṣl* is present, then in Tengwar Arabic mode a small high hyphen is inserted between the last prefix and the main stem of the word – as in the example form *fa-l-takun* shown above – in order to make clear

The above text displays a version of Tengwar Arabic mode with all optional diacritics added, including word-final inflectional short vowels (*ʔiʔrāb*). There are fewer diacritics in this text than in a fully-diacriticized version of this same passage in current Arabic-script orthography, for several reasons: First, since long vowels are indicated completely unambiguously with basic Tengwar letters, Tengwar vowel diacritics are needed only to indicate short vowels and diphthongs. Second, the lack of any short-vowel diacritic over a Tengwar “short carrier” symbol or definite article ligature near the beginning of a word indicates that a deletable word-initial vowel has in fact been deleted (without any need for a special *hamzat-ul-waʕl* diacritic). Third, the assimilation of the [l] consonant of the definite article to a following “sun letter” (coronal consonant) can be assumed, so that the doubling of the consonant after the definite article does not have to be explicitly marked. And fourth, there are almost no silent letters in Tengwar Arabic mode (except for the “short carrier” symbol, which is set aside for that purpose) – so if full diacritics are being used, then any Tengwar letter which does not have a short vowel diacritic over it, is not directly followed by any long-vowel Tengwar ᚢᚣᚣᚣ , and is not itself a “short carrier” symbol or long-vowel Tengwar ᚢᚣᚣᚣᚣ , will almost certainly represent a consonant sound which is not followed by a vowel sound. Therefore there is no real need for a special symbol to mark consonants not followed by vowels (such as the *sukūn* diacritic of current Arabic orthography).

In general terms, one can say that all of the contrasts written by the special diacritics of Table 2 above are handled by the use of separate base letters in the Tengwar Arabic mode. Similarly, the Arabic-script entities *yāʔ* and *ʔalif maqṣūrah* – the contrast between which is marked only by the somewhat-optional presence of consonantal diacritic dots in current orthography (see Appendix 1) – are written by distinct base letters in the Tengwar Arabic mode (e.g. *على* *ʕalā* “on” ᚩᚣᚣ vs. *علي* – either *ʕalayya* “on me” ᚩᚣᚣᚣ or *ʕalīy* “Ali” ᚩᚣᚣᚣᚣ).

Pruning and streamlining the apparatus of diacritics makes it likelier that some useful disambiguating diacritics will be used in ordinary text (i.e. outside of pedagogical works and sacred scriptures).

Here is an interlinear version of the above passage (with current Arabic-script orthography and Tengwar on alternate lines). Note that the Arabic-script text is partially-diacriticized (NOT fully diacriticized), and uses standard Modern Arabic orthography (not special Qur’anic orthography):

سورة الفاتحة

ḡḡnr ḡḡnr

① بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

1. ḡḡnr ḡḡnr ḡḡnr ḡḡnr

② الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

2. ḡḡnr ḡḡnr ḡḡnr ḡḡnr

③ الرَّحْمَنِ الرَّحِيمِ ④ مَالِكِ يَوْمِ الدِّينِ

3. ḡḡnr ḡḡnr 4. ḡḡnr ḡḡnr ḡḡnr

⑤ إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ

5. ḡḡnr ḡḡnr ḡḡnr ḡḡnr

⑥ أَهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ

6. ḡḡnr ḡḡnr ḡḡnr ḡḡnr

صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ

7. ḡḡnr ḡḡnr ḡḡnr ḡḡnr ḡḡnr

⑦ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ

ḡḡnr ḡḡnr ḡḡnr ḡḡnr ḡḡnr

Fonts and Links

Go to the website for links to various Tengwar and Tolkien linguistics sites, and to download fonts containing special Arabic-mode characters for use with the “Tengwar Quenya” font by Daniel Steven Smith, the “Tengwar Elfica” font by Enrique Mombello, and the “Tengwar Elfetica” font (a revised version of the “Elfetica” font by Ronald Kyrmse). This diagram shows basic information about the glyphs in the three special Arabic Tengwar fonts:

Font Glyphs			Keystroke	Description
QuenyArabic	ElficArabic	ElfeticArabic		
Ʒ	Ʒ	Ʒ	comma	Alternative [z] glyph
Ṛ	Ṛ	Ṛ	1	<i>tāʔ marbūṭah</i>
Ɔ	Ɔ	Ɔ	3	[ǰ],[z]
ᶆ	ᶆ	ᶆ	l	[ṣ]
ƶ	ƶ	ƶ	`	<i>ʔalif maqṣūrah</i>
Ɔ	Ɔ	Ɔ	a	[t]
Ɔ	Ɔ	Ɔ	j	[d]
ᶆ	ᶆ	ᶆ	l	Long [i]
Ṛ	Ṛ	Ṛ	m	Definite article ligature
ᶆ	ᶆ	ᶆ	n	<i>ʔalif maddah</i>
ᶆ	ᶆ	ᶆ	y	Long [u]
J	J	J	tilde	Long [a]

In addition, the “Tengwar-QuenyArabic” font contains larger and more distinct forms of the vowel diacritics, for use with the characters of the TengwarQuenya and Tengwar-QuenyArabic fonts. Different widths of the circumflex (short-[a]) diacritic are mapped to keystrokes C, D, E, #; different widths of the overdot (short-[i]) diacritic are mapped to keystrokes B, G, T, %; and different widths of the left curl (short-[u]) diacritic are mapped to keystrokes M, J, U, &. Also, for the foreign/dialect vowels, different widths of the acute accent (short-[e]) diacritic are mapped to keystrokes V, F, R, \$, and different widths of the right curl (short-[o]) diacritic are mapped to keystrokes N, H, Y, ^.

Arabic letter and name	Sound	Tengwar	Arabic letter and name	Sound	Tengwar
ا <i>Palif</i>	long [a]	∟	ط <i>tāʔ</i>	[t]	Ɔ
	(silent)	ı	ظ <i>zāʔ</i>	[ð],[z]	Ɔ
آ <i>Palif maddah</i>	[ʔā]	Ɔ	ع <i>ʕayn</i>	[ʕ]	Ɔ
أ <i>Palif hamzah</i>	[ʔ]	ɑ	غ <i>ghayn</i>	[ɣ]	Ɔ
إ <i>Palif hamzah (below)</i>	[ʔi]	ɑ,ɑ̇	ف <i>fāʔ</i>	[f]	Ɔ
ء <i>hamzat-ul-waṣl</i>	(silent)	ı	ڤ	[v]	Ɔ
ب <i>bāʔ</i>	[b]	Ɔ	ق <i>qāf</i>	[q]	Ɔ
ت <i>tāʔ</i>	[t]	Ɔ	ك <i>kāf</i>	[k]	Ɔ
ث <i>thāʔ</i>	[θ]	Ɔ	ل <i>lām</i>	[l]	Ɔ
ج <i>jīm</i>	[j]	Ɔ	م <i>mīm</i>	[m]	Ɔ
	[g]	Ɔ	ن <i>nūn</i>	[n]	Ɔ
ح <i>ḥā</i>	[ħ]	Ɔ	ه <i>hāʔ</i>	[h]	λ
خ <i>khā</i>	[χ]	Ɔ	ة <i>tāʔ marbūṭah</i>	[t] or silent	Ɔ
د <i>dāl</i>	[d]	Ɔ	و <i>wāw</i>	[w]	Ɔ
ذ <i>dhāl</i>	[ð]	Ɔ		long [u]	Ɔ
ر <i>rāʔ</i>	[r]	Ɔ	ؤ <i>wāw hamzah</i>	[ʔ]	ɑ
ز <i>zāy</i>	[z]	Ɔ	ي <i>yāʔ</i>	[y]	^
س <i>sīn</i>	[s]	Ɔ		long [i]	Ɔ
ش <i>šīn</i>	[š]	Ɔ	ئ <i>yāʔ hamzah</i>	[ʔ]	ɑ,ɑ̇
ص <i>ṣād</i>	[ṣ]	Ɔ	ى <i>Palif maqṣūrah</i>	long [a]	Ɔ
ض <i>ḍād</i>	[ḍ]	Ɔ	ء <i>hamzah</i>	[ʔ]	ɑ

Appendix 2 - Use of Tengwar to write Arabic dialects

In addition to the [g] Tengwar and the [e] and [o] vowel diacritics discussed above, other symbols appropriate for transcribing Arabic colloquial dialect sounds can be derived by placing the two-dots diacritic under a velar or sibilant letter, or by placing the “emphasis” *ṣiḥbāq* Tengwar letter before or over any other appropriate Tengwar letter. Here are some examples of Tengwar letters or combinations which could be used to transcribe Arabic colloquials:

ḡ [č] ʒ [ž] ʔ or ḡ [ʔ] Ʒ [G] (voiced uvular stop)

Note that the Tengwar “emphasis” letter modifying the Tengwar letter for [l] is quite different from the Tengwar “emphasis” letter **merged with** the Tengwar letter for [l]. The former writes a velarized [l] sound, while the latter writes the letter *ḍād*.

Here’s one hypothetical example of the use of Tengwar symbols to transcribe a colloquial Arabic dialect – the following table shows a translation of the special symbols of Nizar Habash’s “Palestinian Arabic Spelling Standardization Project” from Arabic script to Tengwar:

PASS:	Sound:	Tengwar:
وْ	long [o]	ó
يْ	long [e]	á
چ	[g]	Ʒ
اَ	Adverbial final [n]	ĩ
قْ	Variable reali- zations of <i>qāf</i>	Ƴ

Also, the original Tengwar system described by J.R.R. Tolkien in Appendix E to *The Lord of the Rings: The Return of the King*, together with the implementation choices and slight modifications made by me for the Tengwar Arabic mode, will suggest a number of ideas for transcribing those non-Arabic languages which are currently written with Arabic script into Tengwar – for example, Ʒ could be used to write an [ŋ] sound, or the Tengwar “emphasis” letter ʔ could be used as a retroflexion diacritic in Indic languages such as Urdu (instead of a velarization/pharyngealization diacritic, as in the case of Arabic).