

# Unicode request for open q and x with stroke

Kirk Miller, kirkmiller, gmail.com

Ophir Lifshitz, me, ophir.li

Denis Moyogo Jacquerye, moyogo, gmail.com

2025 May 09

This request is for a Latin open *q*, <ç>, and *x* with a stroke, <ꜥ>, both of which are used in the transcription of Arabic.

The letter <ç> has been used for the Egyptian Arabic reflex of Classical Arabic ق *qoph*, and in that context it contrasts with the letter U+0295 <ʁ> for the reflex of Classical Arabic ه *hamza*. In the colloquial language both are pronounced as a glottal stop, but /ʁ/ undergoes elision in some contexts where /ç/ does not. The use of <ʁ> here differs from its IPA value as a voiced pharyngeal fricative; the Egyptian voiced pharyngeal, ځ *ayin*, is transcribed instead with U+01B9 <ɣ̣>. This odd use of <ʁ> and <ɣ̣> for IPA [ʁ] and [ʁ̣], perhaps motivated by the shapes of the Arabic letters ه *hamza* and ځ *ayin*, is found in descriptions of Arabic such as Palmer (1955), Mitchell (1956, 1960, 1962) and Wise (1975). The reason that the glyph <ç> graphically resembles both <q> and <ʁ> is presumably that it is pronounced like <q> in Modern Standard Arabic and like <ʁ> in colloquial Egyptian Arabic in the diglossic situation that characterizes modern Egypt.

<ç> has also been used to transcribe orthographic *q* in the Nguni languages (Figure 5).

The second proposed letter, <ꜥ>, has been used to transcribe Arabic ځ *kha*.

In the Soviet Latinization (*Yanalif*) period, the casing pair <Ꜧ, ꜥ> was used for Avar /q̄x:/ (previously Arabic-script ځ, now Cyrillic ɣ̣), and continues in use for Avar Latin orthography in online dictionaries. The letter was also used for other languages of Dagestan as well as for Shona (Doke 1931).

In electronic documents and online, these letters are substituted with Cyrillic characters, specifically with U+0481 <Ꜧ> CYRILLIC SMALL LETTER KOPPA and U+04FF <ꜥ> CYRILLIC SMALL LETTER HA WITH STROKE. However, it is the opinion of SEW that the Latin letters should be encoded separately from the Cyrillic.

## Characters

ç U+1DF86 LATIN SMALL LETTER OPEN Q

Ꜧ U+1DF87 LATIN CAPITAL LETTER X WITH STROKE

ꜥ U+1DF88 LATIN SMALL LETTER X WITH STROKE

## Properties

1DF86;LATIN SMALL LETTER OPEN Q;Ll;0;L;;;;;N;;;;;

1DF87;LATIN CAPITAL LETTER X WITH STROKE;Lu;0;L;;;;;N;;;1DF88;

1DF88;LATIN SMALL LETTER X WITH STROKE;Ll;0;L;;;;;N;;;1DF87;;1DF87

## Collation

Given that the letter <ϙ> derives graphically from *q* and is currently substituted with Cyrillic *koppa*, a reflex of the Greek ancestor of Latin *q*, it should probably be collated with Latin *q* rather than with the pharyngeal ʕ.

## Confusables

These characters may be confused with Cyrillic characters:

1DF86 LATIN SMALL LETTER OPEN Q

→ U+0481 <ϙ> CYRILLIC SMALL LETTER KOPPA

1DF87 LATIN CAPITAL LETTER X WITH STROKE

→ U+04FE <X> CYRILLIC CAPITAL LETTER HA WITH STROKE.

1DF88 LATIN SMALL LETTER X WITH STROKE

→ U+04FF <x> CYRILLIC SMALL LETTER HA WITH STROKE.

## Annotations

The denarius sign <ϙ> is an epigraphic currency symbol, and in addition has graphic variants that are not appropriate for the letter, and so is not unifiable with X-stroke.

10196 ROMAN DENARIUS SIGN

→ 1DF86 LATIN CAPITAL LETTER X WITH STROKE

1DF86 LATIN CAPITAL LETTER X WITH STROKE

→ 10196 ROMAN DENARIUS SIGN

## References

Ekab Al-Shawashreh (2016) *Aspects of grammatical variation in Jordanian Arabic*. PhD thesis, University of Ottawa.

Avar Me (2016–2021) *Аварский язык. Аварско-Русский и Русско-Аварский словарь с примерами*. <https://avar.me/>.

Akimov, trans. (1935). *Ci Lenin*. Дагполиграфтрестра Printers, Makhachkala.

- C. M. Doke (1931) *A Comparative Study in Shona Phonetics*. University of Witwatersrand.
- Martin Hinds & El-Said Badawi (1986) *A Dictionary Of Egyptian Arabic*. Librairie du Liban.
- Harry Johnston (1919) *A comparative study of the Bantu and semi-Bantu languages*, vol. 1. Oxford.
- Adel Refaat Mahfouz (2021a) Variation in the Cairene Slum Vernacular: A Sociophonetic Study. *Asian EFL Journal Research Articles*, 28(1.4).
- (2021b) Lexical Variation in the Cairene Slum Vernacular and Its Impact on the Prestigious Variety. *International Journal of Linguistics, Literature and Translation*, 4(4).
- T. F. Mitchell (1978) *An introduction to Egyptian colloquial Arabic*. Clarendon Press.
- (1960) Prominence and Syllabication in Arabic. *Bulletin of the School of Oriental and African Studies* 23(2).
- (1980) *Colloquial Arabic*. Hodder and Stoughton.
- F. R. Palmer (1955) The 'Broken Plurals' Of Tigrinya. *Bulletin of the School of Oriental and African Studies* 17(3).
- Hilary Wise (1975) *A transformational grammar of spoken Egyptian Arabic*. Blackwell for the Philological Society.

# Chart

1DF00

Latin Extended-G

1DFFF

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ƒŋ	ɣ	ɔ̣	ɔ̤	Ǻ	ǻ	ω	ı	Ǽ					ɥ	ɔ̧	ɭ
1	ƒ	ɮ	ɔ̣	ɔ̤	Ǻ	ǻ	ɣ	ı	Ǽ					ɣ	ɔ̧	ɭ
2	ɔ̣	ɔ̤	ɮ	ɮ̣	ə	ɣ	ʃh	ɔ̣	ɣ̣					ɮ	ɮ̣	ɮ̣
3	ɣ	ɮ̣					ɮ̣	ɮ̣	ɮ̣							
4	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
5	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
6	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
7	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
8	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
9	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
A	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
B	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
C	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
D	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
E	ɮ̣					ɮ̣	ɮ̣	ɮ̣								
F	ɮ̣					ɮ̣	ɮ̣	ɮ̣								

# Figures

i) **ق/q** and **ق**  
 The **ق** of literary Arabic is associated with two distinct sounds in Egyptian Arabic. The first of these (symbolized by **ق**) is of more widespread occurrence and is similar to the glottal stop (ق). The second (symbolized by **q**) is similar to its counterpart in literary pronunciation and occurs commonly only in 'āmmiyyat al-muthaqqafīn. One good reason for distinguishing between ق and **ق** is that the former may elide while the latter never elides. Compare, for example, *yilit fi-sm-i* (or *fi-قism-i*) (he made a mistake in my name) with *yilit fi-قism-i* (he made a mistake in my share).

**Figure 1.** Hinds & Badawi (1986: xvii). The letter <ق> here resembles a modified metal sort for <q> with the right side of the bowl filed off. It contrasts lexically with ق (green).

اتعامل **قitgaamil** /vi/ 1 to manage, conduct oneself (in relation to another). **baqa** **mif** **قarfa titgamli maqa l-xaddaama** you mean you can't cope with the maid?

**Figure 2.** Hinds & Badawi (1986: 602). <ق> for Egyptian Arabic ق q contrasts with <ق> for glottal stop (a substitution for <ق> used by other authors of Arabic dictionaries) and with <ق> for the pharyngeal [ق].

<p>ق اع see ق اع : قاع <b>qaaε, qaaεa, قaaεa</b></p> <p>ق اف  <b>qaaf</b> fem n, pl -aat/ twenty-first letter of the Arabic alphabet. <b>gabal qaaf</b> name of legendary mountain, said to lie at the end of the earth and to be the place upon which Adam descended after being expelled from the Garden.  <b>قaaf: bi-yitkallim bil-قaaf</b> /deris/ he speaks in a pompous way.</p>	<p>ق ب ح  <b>قabaaha</b> /n/ 1 obscenity. <i>kalaam قabaaha</i> [children] dirty words. 2 rudeness.  <b>قabiih</b> /adj/ ugly.  <b>قabiih</b> /adj pl قubaha, قubaah/ obscene (especially in language). <i>walad قabiih</i> foul-mouthed kid. <i>film قabiih</i> blue movie.  <b>قabaayih</b> /pl n/ dirty tricks, low acts.  <b>قabah</b> /el/ 1 more/most obscene or foul-mouthed. 2 more/most ugly.</p>
---	--

**Figure 3.** Hinds & Badawi (1986: 682). Colloquial [ق] (red) contrasting with formal [q] (green) pronunciations of the letter ق.

generation speakers. For example, the (q) sound is pronounced as [ç] and [g] in urban and Horani dialects respectively. Yet, FallaHi speakers who originally came from Tirat Haifa pronounce (q) as [q] (El-Salman 2003) and those who came from villages around Jerusalem pronounce it as [k]

Figure 4. Al-Shawashreh (2016: 46). <ç> for Jordanian Arabic. <ç> is used for the pharyngeal.

Island ...	Şe-tl'ake. Si-oli (74 b)	Isi-çiti	Isi-çiŋgi	Isi-hleŋga
Ivory ...	Lω-naka	Ama-bamba	U-pondω lwe n-dhlωvu	Ulu-pondvω lwe n-dhlωvu
Knee ...	Le-ñōle.	I-dωω	I-dωω	Li-dv'ωω,

Figure 5. Johnston (1919: 304). <ç> for orthographic q in Xhosa *isiqithi* and Zulu *isiqihingi* ‘island.’

/juhruf ʔil 'le:la/ lit. to itch the night, is used v  
/kabbar 'moxxak/ lit. enlarge your brain, is u  
/ji wirr fi 'wida:n/ lit. to urinate in one's ears,

Figure 6. Mahfouz (2021b: 7). Latin <x> for Arabic خ.

and T.V. anchor Ahamed Shobier uses [d] instead of [d] in his tweet “h̄isi:n ɪf-faha:t juçlIn çan xudu:çu lıçamalıjja gırahıjja qari:ban. Lit. Hussien Al-Shahaat is going under the knife soon” (Shobier, 2020)

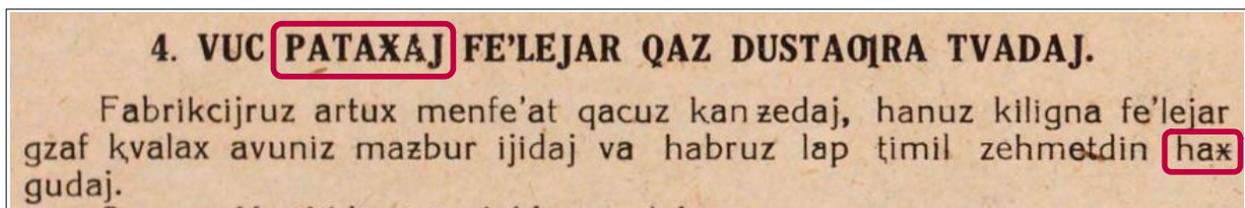
Figure 7. Mahfouz (2021a: 369).

нартичирахъ نارتیچیراخ narticirax	нартичирахъ нартичирахъалъ нартичирахъалъул нартичирахъал	керосиновая лампа; нартичирахъ бакизе зажечь керосиновую лампу; нартичирахъалъубе нарт тлезе налить керосин в керосиновую лампу
хъихъидизе хъихъидизи xixidize	хъихъидизе хъихъидула хъихъидила хъихъидана хъихъиде	накричать, поднимать голос на кого-л.; ццидаль хъихъидизе накричать со злостью см. хъичизе

Figure 8. Avar Me (2016–2021). Latin <x> for Avar ɣ / хъ in a modern dictionary.

1932-1938																										
Sahnazarov H. Avar alif																			Maḥac-Xala	1935						
a	b	c	cc	ç	çç	d	e	g	oɟ	h	ḥ	ḥ	i	j	k	kk	k	kk	l	l	ll	ṭ	m	n	o	p
f	q	q	r	s	ss	s	ss	ş	şş	t	ṭ	u	v	x	xx	x	ḫ	z	z	zz	z	'				

**Figure 9.** Wikimedia editor Geoalex (2008). A rare modern token of a capital Latin letter <X>, on an Avar alphabet chart from Wikimedia. Here the city of Makhachkala is spelled in Avar as *Maḥac-Xala*.



**Figure 10.** Akimov (1935: 14). The letter <Ḫ, ḫ> used in the Lezgian alphabet during the Soviet Latinization period. Lezgian is another literary language of Dagestan. The remaining Avar and Lezgian letters are supported by Unicode, given the allographic equivalence of cedilla <ṭ> etc. and tick <ṭ> etc. across publishers.

**ISO/IEC JTC 1/SC 2/WG 2**  
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS**  
**FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646.**  
**Please fill all the sections A, B and C below.**  
Please read Principles and Procedures Document (P & P) from [std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html) for guidelines and details before filling this form.  
Please ensure you are using the latest Form from [std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html).  
See also [std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html) for latest Roadmaps.

**A. Administrative**

1. **Title:** Unicode request for open q and x with stroke

2. Requester's name: Kirk Miller, Ophir Lifshitz, Denis Jacquerye

3. Requester type (Member body/Liaison/Individual contribution): individual

4. Submission date: 2025 May 09

5. Requester's reference (if applicable): \_\_\_\_\_

6. Choose one of the following:

This is a complete proposal: yes

(or) More information will be provided later: \_\_\_\_\_

**B. Technical – General**

1. Choose one of the following:

a. This proposal is for a new script (set of characters): \_\_\_\_\_  
Proposed name of script: \_\_\_\_\_

b. The proposal is for addition of character(s) to an existing block: yes  
Name of the existing block: Latin Extended-G

2. Number of characters in proposal: 3

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		<input type="checkbox"/>	G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?

a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? yes

b. Are the character shapes attached in a legible form suitable for review? yes

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Kirk Miller

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Kirk Miller

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? yes

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? yes

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? no

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at [www.unicode.org](http://www.unicode.org) for such information on other scripts. Also see Unicode Character Database ([www.unicode.org/reports/tr44/](http://www.unicode.org/reports/tr44/)) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>no</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>no</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>transcription</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>See figures</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>no</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>no</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>no</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>no</i>