

PROPOSAL TO ENCODE TWO QURANIC ARABIC CHARACTERS (Replaces L2/22-153)

Rikza F. Sh.

ريكزاف.ص.

rikzafsh@gmail.com

17 October 2022

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

1. Introduction

The Quran is a holy book for Moslems that was revealed by Allah ﷻ to the Prophet Muhammad ﷺ through the intercession of the angel Jibrīl ؑ. At first, the Quran was only memorized by the companions of the Prophet ﷺ and was not written into a complete book. Until the caliph Uthman bin ‘Affan ؓ codified and established a standard version, which becomes the basis for writing most of the Qurans that exist today.

At that time, the Quran text was only written with the body letter without any dots or other diacritical marks as in today’s Quran. This is because the Arabic Moslems can easily read it depend on the context. The additional dots and other diacritical marks are later added to ease the Moslems who can’t speak Arabic to recite the Quran properly.

However, this additional diacritical marks may be different across different Qurans. For example, the Madinah mushaf has a different style of additional diacritical marks compared to the Indonesian mushaf. Although this style difference is only orthographic and not affected on recitation nor the meaning of the Quran.

Currently, most of the characters needed to type the Quran text are well-encoded in Unicode. However, there are several special characters used in various mushaf that not yet encoded.

2. Proposal History

L2/22-153 Proposal to Encode Three Quranic Arabic Characters

A proposal which was written by the same author of this document, was submitted in July 2022 with the aim to encode ARABIC RAISED SMALL ALEF, ARABIC SMALL LOW NOON, and ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW.

L2/22-128 Recommendations to UTC #172 on Script Proposals

A document submitted in July 2022 by Deborah Anderson, Ken Whistler, Roozbeh Pournader, and Peter Constable. The document reviews L2/22-153 in section 4c along with several recommendations. In sum, the document recommends the author of L2/22-153 revise the proposal, removing RAISED SMALL ALEF and putting it in a separate document, and change the codepoint for other two characters as noted in the comments.

This document

This document is a revised version of L2/22-153 as recommended in L2/22-128.

3. Character to be encoded

3.1. ARABIC SMALL LOW NOON

This character has similar shape and function with U+08D9 ARABIC SMALL LOW NOON WITH KASRA, but without the kasra below the letter noon. This character is used in *Mushaf Standar Indonesia* (The standard mushaf published in Indonesia) contrast with U+08D9. The small low

noon character is used to mark noon wasl when there is a tanwin (U+064B..064D) before the letter alef wasla¹. Here are the rules of adding noon wasl in Indonesian standard mushaf:

- If the harakat tanwin is DAMMATAN (U+064C) or KASRATAN (U+064D), then those tanwin become regular DAMMA (U+064F) and KASRA (U+0650), and the low noon is written with kasra (U+08D9).



Figure 1. Showing with U+08D9 ARABIC SMALL LOW NOON WITH KASRA preceded with U+064F ARABIC DAMMA. From [3] p.333

- If the harakat tanwin is fathatan (U+064B), the tanwin is not changed into regular fatha (U+064E), and the low noon is written without kasra. Except when fathatan is placed over the letter teh marbuta (U+0629), fathatan becomes regular fatha, and the low noon is written with kasra.



Figure 2. Showing ARABIC SMALL LOW NOON without kasra preceded with U+064B ARABIC FATHATAN. From [3] p.335

- If the harakat tanwin is in the end of ayah with waqf sign صلے (U+06D6) or لا (U+06D9), then the low noon is written without kasra.

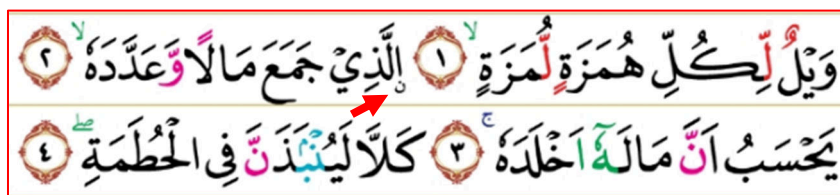


Figure 3. Showing ARABIC SMALL LOW NOON without kasra preceded with sign the end of ayah with waqf sign لا (U+06D9) ARABIC SMALL HIGH LAM ALEF. From [3] p.601

The small low noon without kasra character is also attested in other mushafs, but it usually doesn't have contrast use with the low noon with kasra.

¹ The wasla symbol is not written in Indonesian mushaf, so it will be appeared as regular alef



Figure 4. Showing ARABIC SMALL LOW NOON without kasra. From [6] p.33

3.2. ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW

This character has similar shape with U+08CB ARABIC SMALL HIGH YEH BARREE WITH TWO DOTS BELOW. However, U+08CB is a non-spacing mark, and the ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW is a spacing mark like U+06E6 ARABIC SMALL YEH. This character is used in Mushaf Standar Indonesia to mark unwritten yeh in Uthmanic rasm.

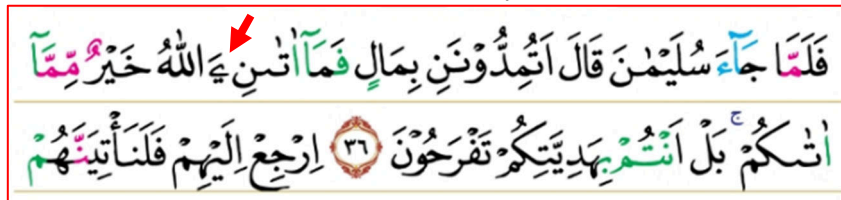


Figure 5. Showing ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW and with fatha above. From [3] p.380

The unwritten yeh in Uthmanic rasm is usually written with small yeh in modern mushafs. However, the shape and position of small yeh could be different across different mushafs. Here's the comparison:

Verse	Madinah Mushaf [2]	Indo-Pak Mushaf [4]	Indonesian Mushaf [3]	Tripoli Mushaf [5]
7:196	إِنَّ وَلِيَّيَ اللَّهِ	إِنَّ وَلِيَّيَ اللَّهِ	إِنَّ وَلِيَّيَ اللَّهِ	إِنَّ وَلِيَّيَ اللَّهِ
25:49	لِنُحْيِيَ بِهِ	لِنُحْيِيَ بِهِ	لِنُحْيِيَ بِهِ	لِنُحْيِيَ بِهِ
46:33	يُحْيِي الْمَوْتَى	يُحْيِي الْمَوْتَى	يُحْيِي الْمَوْتَى	يُحْيِي الْمَوْتَى
27:36	أَمْ آتَيْنَا اللَّهَ	فَمَا آتَيْنَا اللَّهَ	أَمْ آتَيْنَا اللَّهَ	أَمْ آتَيْنَا اللَّهَ

As we see in the table, The Madinah Mushaf used the SMALL YEH with the shape like ے U+06D2 ARABIC YEH BARREE. But the position of SMALL YEH is depend on the word's final letter. If the word ends with ى U+06CC ARABIC LETTER FARSI YEH, the SMALL YEH inserted before ى and it's placed above the base letter. The tatweel (U+0640) could be used to stretch character and U+06E7 ARABIC SMALL HIGH YEH is inserted above the tatweel. If the word ends

with any letter except ى, the position of small yeh is inserted after the word's final letter as a spacing mark, so U+06E6 ARABIC SMALL YEH should be used.

In the Indo-Pak Mushaf, the shape of SMALL YEH is same with Madinah Mushaf. But the position of SMALL YEH is always above the word's final letter regardless of what final letter it is. So, U+06E7 ARABIC SMALL HIGH YEH should be used.

The Indonesian Mushaf used the SMALL YEH with the shape like ے U+06D2 ARABIC YEH BARREE but it has two dots below. The SMALL YEH is always inserted after the word's final letter as a spacing mark regardless of what final letter it is. Currently, this character is not encoded in Unicode. So, I think we need to encode this character.

The Tripoli Mushaf used two different shapes of SMALL YEH. If the word ends with ى U+06CC ARABIC LETTER FARSI YEH, the shape of SMALL YEH is like ے U+06D2 ARABIC YEH BARREE but it has two dots below the base line and inserted before ى. So, ے U+0885 ARABIC TATWEEL WITH TWO DOTS BELOW and U+06E7 ARABIC SMALL HIGH YEH could be used. If the word ends with any letter except ى, the shape of SMALL YEH is like ى and it's placed above the word's final letter. So, U+08CA ARABIC SMALL HIGH FARSI YEH should be used.

As ے U+0885 ARABIC TATWEEL WITH TWO DOTS BELOW encoded as an atomic character, I think ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW should be encoded as an atomic character too, and not represented by sequence U+06E6 ARABIC SMALL YEH and combining two dots below. And although it is a "small" yeh, the size of those two dots are as big as dots in regular letter yeh ي.

4. Character Data

Unicode character properties

```
10EFB; ARABIC SMALL LOW NOON; Mn; 220; NSM; ; ; ; ; N; ; ; ; ;
10EC5; ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW; Lm; 0; AL; ; ; ; ; N; ; ; ; ;
```

Annotation

The following annotation could be updated in Namelist.txt

```
10EFB; ARABIC SMALL LOW NOON
  x (arabic small low noon with kasra - 08D9)
10EC5; ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW
  x (arabic small yeh - 06E6)
  x (arabic small high yeh barree with two dots below - 08CB)
```

5. Reference

[1] Pedoman Pentashihan Mushaf Al-Qur'an, Lajnah Pentashihan Mushaf Al-Qur'an Badan Litbang dan Diklat Kementerian Agama Republik Indonesia.

[2] Mushaf Madinah

المصحف الشريف بالمدينة المنورة

[3] Qur'an Kemenag v2.3 beta, Mushaf Standar Indonesia. Published by Ministry of Religion of the Republic of Indonesia c.q. Lajnah Pentashihan Mushaf Al-Qur'an.

[4] Qurān Majīd. Published by Taj Company Limited.

قران مجيد تاج كيني لميٹڈ

[5] Mushaf Tripoli Qaloon

مصحف الجماهيرية برواية الإمام قالون والرسم العثماني على ما اختاره الحافظ أبو عمرو الداني

[6] Al Quran Al Kareem (Noori Usmaan) Qaish Zaadah

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646².**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <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 <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to Encode Two Quranic Arabic Characters	
2. Requester's name:	<i>Rikza F. Sh.</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>17 October 2022</i>	
5. Requester's reference (if applicable):	<i>N/A</i>	
6. Choose one of the following:		
This is a complete proposal:		YES
(or) More information will be provided later:		NO

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):		NO
Proposed name of script:	<i>N/A</i>	
b. The proposal is for addition of character(s) to an existing block:		YES
Name of the existing block:	<i>Arabic Extended-C</i>	
2. Number of characters in proposal:		2
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?		YES
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?	<i>Rikza F. Sh.</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Rikza F. Sh. (rikzafsh@gmail.com)</i>	
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?	<i>YES</i>	
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)?	<i>See proposal</i>	YES

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 <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<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.

² Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>This document is a revised version of L2/22-153</i>	YES
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>N/A</i> <i>N/A</i>	NO
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Moslems in Indonesia</i>	YES
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>See proposal</i>	Common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Indonesian Standard Mushaf published by Ministry of Religion of the Republic of Indonesia</i>	YES
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>N/A</i> <i>N/A</i>	NO N/A
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		NO
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>N/A</i> <i>N/A</i>	NO N/A
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>N/A</i> <i>N/A</i>	NO N/A
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>See proposal</i>	YES YES
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>See proposal</i> <i>N/A</i>	YES YES N/A
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>N/A</i>	NO N/A
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>N/A</i>	NO N/A