

Proposal to encode three Arabic symbols

Authors: Khaled Hosny and Roozbeh Pournader (WhatsApp)




Date: September 9, 2020

Background

We found three textual symbols commonly used in Arabic script that are not yet encoded in Unicode. One marks the end of texts, and the two others denote prices in Egyptian pounds or in its subdivision, in piastres. Encoding these will help with digitizing existing texts, as well as typesetting future material.

Proposal

Encode the following three characters in the Unicode Standard:

Glyph	Code	Name
	061D	ARABIC END OF TEXT MARK
	0890	ARABIC POUND MARK ABOVE
	0891	ARABIC PIASTRE MARK ABOVE

Proposed properties

UnicodeData.txt:

```
061D;ARABIC END OF TEXT MARK;Po;0;AL;;;;;N;;;;;
0890;ARABIC POUND MARK ABOVE;Cf;0;AN;;;;;N;;;;;
0891;ARABIC PIASTRE MARK ABOVE;Cf;0;AN;;;;;N;;;;;
```

ArabicShaping.txt:

```
0890; ARABIC POUND MARK ABOVE; U; No_Joining_Group
0891; ARABIC PIASTRE MARK ABOVE; U; No_Joining_Group
```

SentenceBreakProperty.txt:

```
061D ; STerm # Po ARABIC END OF TEXT MARK
0890..0891 ; Format # Cf [2] ARABIC POUND MARK ABOVE..ARABIC PIASTRE MARK ABOVE
```

Proplist.txt:

```
0890..0891 ; Prepend_Concatenation_Mark # Cf [2] ARABIC POUND MARK ABOVE..
ARABIC PIASTRE MARK ABOVE
```

The end of the text symbol

The end of the text symbol (علامة ختام الخطاب in Arabic, literally “end of speech sign”) is used at the end of texts where a period would typically go. It is added after the last sentence.

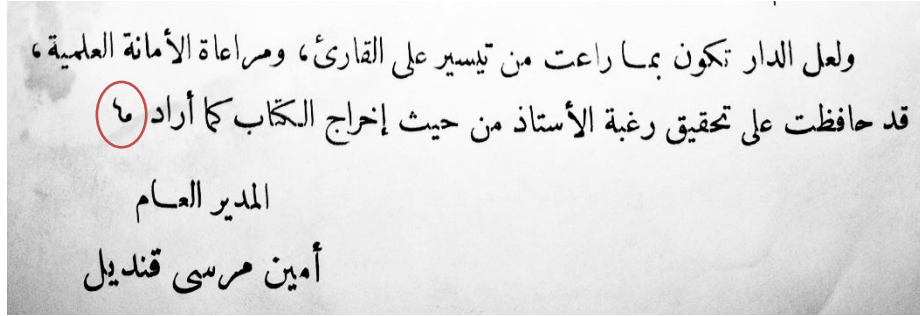


Figure 1. Diwan of Sahim, Egyptian National Library printing press, 1950.

(ديوان سحيم، مطبعة دار الكتب المصرية، ١٣٩٦هـ، ١٩٥٠م)

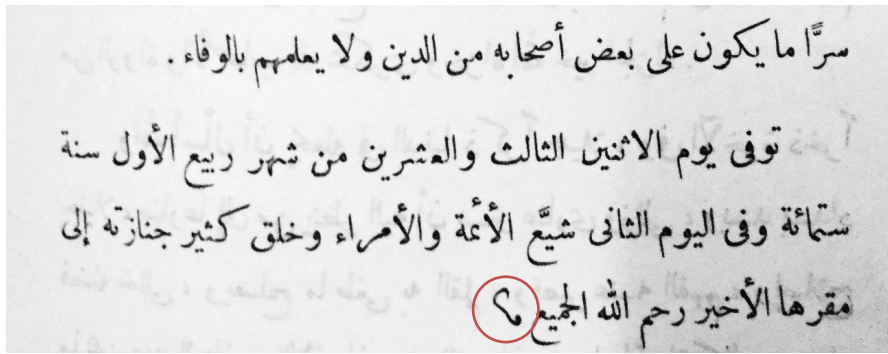


Figure 2.

نيل المرام: التعليق على عمدة الأحكام لتقي الدين أبي محمد عبد الغني المقدسي الجماعيلي الحنبلي، حسن سليمان النوري وعلوي عباس المالكي، مطبعة مصحف مكة.

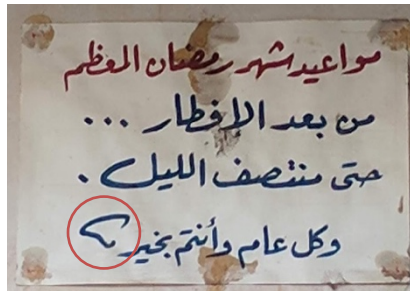


Figure 3. Handwritten sign, Cairo, Egypt, 2018.

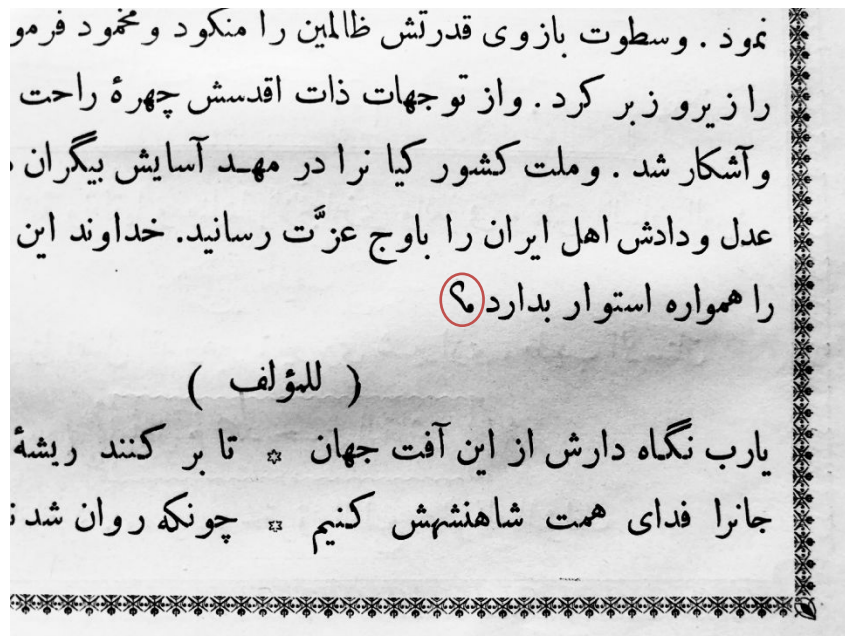


Figure 6. A rare example of the character appearing in Persian text in a book published in Egypt. Courtesy of Borna Izadpanah. From Mirza Fazlullah Rahbar, *Hakim Rahbar Book*, Cairo, probably 1930s.

(کتاب حکیم رهبر، میرزا فضل الله رهبر نیریزی شیرازی، المطبعة العصرية بمصر، قاهرة، احتمالاً دهه ۱۹۳۰ م.)

Egyptian currency signs

Egyptian pound (جنية مصری) is informally written in Arabic as a usually dotless (but occasionally dotted) head of *jeem* above the amount, in a manner similar to other existing Arabic subtending or supertending marks at U+0600..U+0605. Egyptian piastre (قرش مصری) is written using a mirrored version of the same symbols. Both symbols are commonly used in advertising and price tags, as well as in handwritten texts, but to the best of our knowledge they are not used in official material.

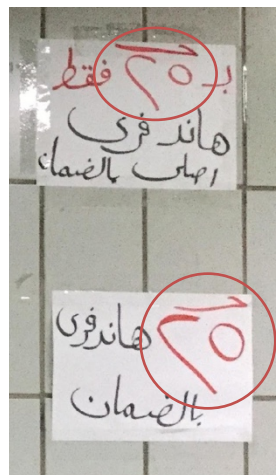


Figure 7. Handwritten sign using Egyptian pound symbol, Cairo, Egypt, 2019.



Previous experience is an asset
 d your C.V to
 package will be
 offered to the selected candidate

١٩٩٩/٦/١١

البيرة مع الباعة والاكشاك والكتبات

المؤسسة العربية الحديثة تقدم:
 * العددان ١٢٥ ، ٩ من سلسلة:
 رحل المستقبل (٣٠٠) «عملية النيل»
 رحل المستقبل (٣٠٠) «معرض المواهب الأولى»
 * العدد ١٢٥ من سلسلة:
 ملف المستقبل (٣٠٠) «كرة النار»
 * العدد ٢٨ من سلسلة:
 كوكتيل (٣٠٠) «قلعة الأبرار»
 للدكتور / نبيل فاروق



Figure 8. Various examples of the currency symbols seen in advertisement or price tags.



Figure 9. Fruit vendor in Cairo. Photographed by Tinou Bao in 2007. Source: <https://www.flickr.com/photos/57621379@N00/2066076908>



Figure 10. Bookstore in Cairo. Photographed by Hatem Moushir in 2017. Source: [https://commons.wikimedia.org/wiki/File:Soor el Azbakeya , photo by Hatem Moushir 1.jpg](https://commons.wikimedia.org/wiki/File:Soor_el_Azbakeya_ ,_photo_by_Hatem_Moushir_1.jpg)

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

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:	<i>Proposal to encode three Arabic symbols</i>
2. Requester's name:	<i>Khaled Hosny and Roozbeh Pournader</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Liaison contribution</i>
4. Submission date:	<i>August 12, 2020</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(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):	<i>No</i>
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>
Name of the existing block:	<i>“Arabic” and “Arabic Extended-B”</i>
2. Number of characters in proposal:	<i>3</i>
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"/>	E-Minor extinct <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested 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?	<i>Yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>Yes</i>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Khaled Hosny</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Khaled Hosny</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>
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>Yes</i>
	<i>Character properties are provided</i>

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.

¹ 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>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>Yes</i> <i>Khaled Hosny, one of the authors, is an expert living among the user community</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>Yes</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Common</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes</i> <i>In Egypt and other Arabic-speaking countries</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>Yes</i> <i>Yes</i> <i>Similarity to existing character at U+0600..U+0605</i> <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>Yes</i> <i>See proposal</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>