Proposal to encode two historical Persian letters used for *ezafe*

Authors: Ebrahim Byagowi and Roozbeh Pournader
Date: 2025-05-09

Proposal

Encode the following two characters:

codepoint	glyph	name
U+10EC0	ڮ	ARABIC LETTER DOTLESS HALF YEH
U+10EC1	ڮۣ	ARABIC LETTER HALF YEH

Background

This is a proposal to encode two new characters in the Arabic Extended-C block in order to represent letters used in Early New Persian as well as modern scholarly publications.

These two proposed letters are the historical forms of the mark put after a word-final silent *heh* to represent *ezafe* (pronounced /je/ in modern Iranian Persian). Over the centuries, these merged into a single dotless version, became smaller and moved over the *heh*, and then merged with *hamza* above (U+0654) to produce ۀ. (One hypothesis is that the merger with *hamza* above may have happened due to limitations in typesetting environments, others blame the scribes.)

Evidence of the usage of the characters in both manuscripts and modern scholarly text is provided. There is contrastive usage between the dotted form and the dotless form in at least one manuscript, and that contrast is reproduced in modern scholarly works about the language and the writing system of the era. The significance of the contrast is not known to us and is probably lost in time. Since Unicode doesn't unify Arabic letters that differ in dots, and the contrast has been kept in modern scholarly works, two characters are proposed.

1

We wish to thank Mahmoud Arasteh, Theo Beers, Mir Salar Razavi, and Yusef Saadat for providing us with additional samples from manuscripts.

Character properties

```
10EC0;ARABIC LETTER DOTLESS HALF YEH;Lo;0;AL;;;;;N;;;;;
10EC1;ARABIC LETTER HALF YEH;Lo;0;AL;;;;;N;;;;;
```

The characters only appear in isolated form (typically after a final or isolated *heh*), and do not change shape contextually. Here are the proposed data for ArabicShaping.txt:

```
10EC0; DOTLESS HALF YEH; U; No_Joining_Group
10EC1; HALF YEH; U; No_Joining_Group
```
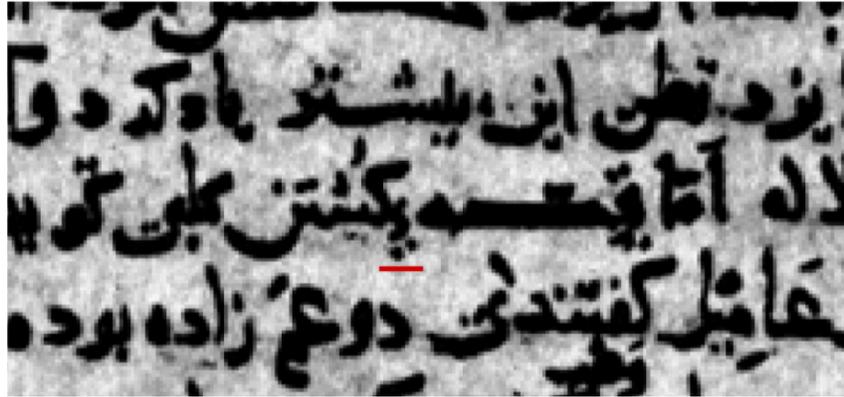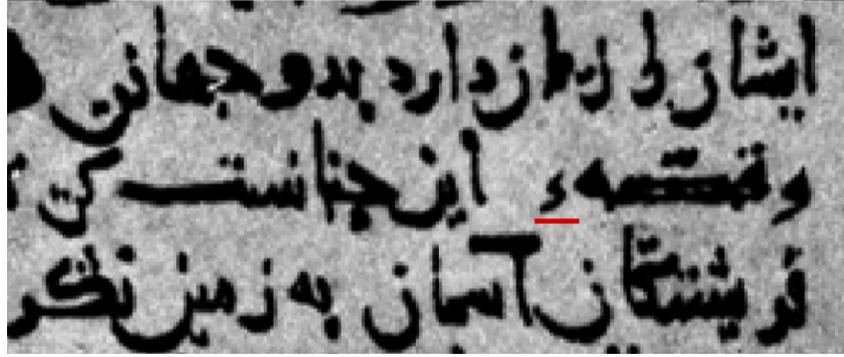
Confusables:
The dotless version is confusable with a stand-alone *hamza* (U+0621) as well as the isolated form of kashmiri *yeh* (U+0620).

Samples

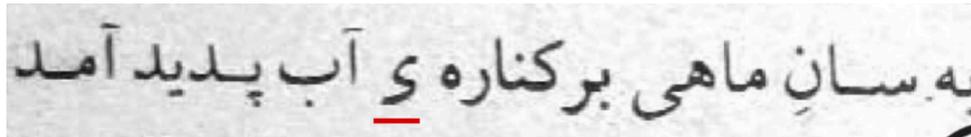Source: ۴۹ ویژگی‌های املائی دستنویس تفسیر قرآن پاک، حسین داوری آشتیانی، صفحهٔ

Modern reproduction

2

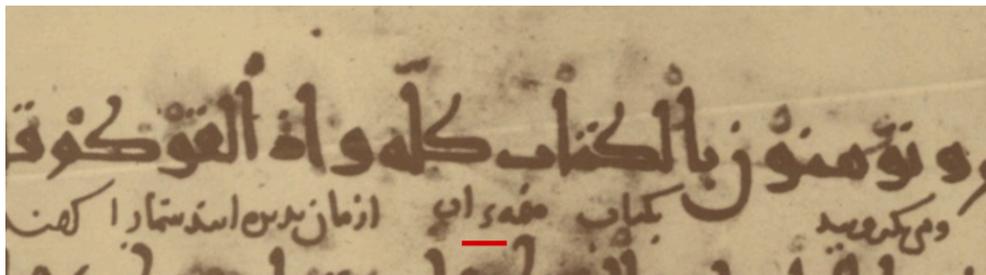Source: نسخهٔ خطی تفسیر قرآن پاک، صفحات ۴۲ و ۹

Historical manuscript

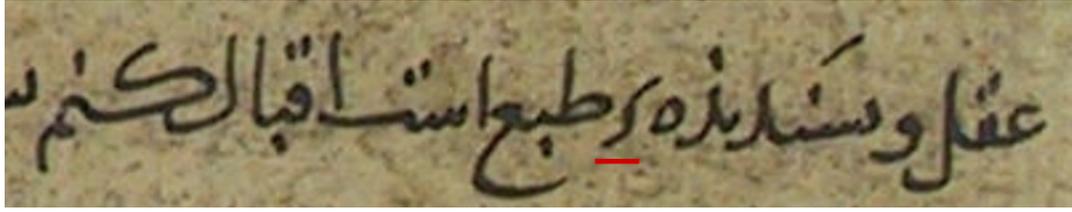Source: نقد و تصحیح متون، نجیب مایل هروی، ۱۳۶۹ شمسی، آستان قدس رضوی

Modern reproduction

Source: قرآن قدس، جلد ۱، صفحهٔ ۲۰

Historical manuscript

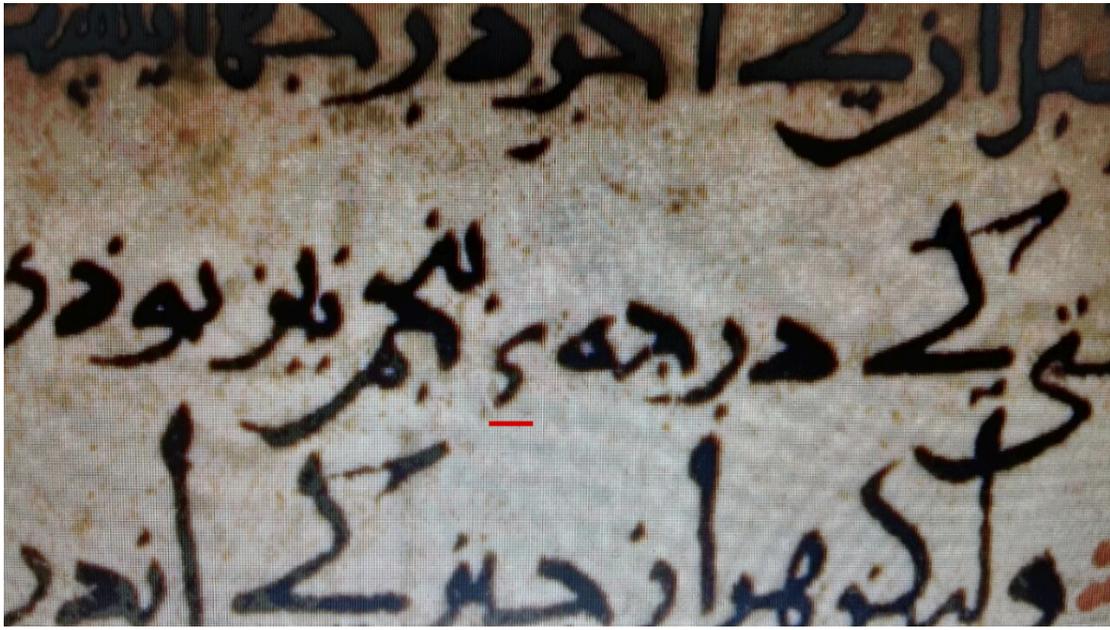Source: ۱۷۲۷ کلیله و دمنۀ بهرامشاهی. استانبول، کتابخانۀ جار الله افندی (جزئی از کتابخانۀ سلیمانیه)، نسخۀ شمارۀ

مورّخ ۵۵۱ ه.ق.، برگ شمارۀ ۲۴ ورسو

Historical manuscript

Source: الابنیه عن حقائق الادویه، ابومنصور موفق هروی، به خط اسدی طوسی

Historical manuscript

۲۸۰

۴ ـ اتکاء به قواعد منطقی دستور زبان فارسی .

۵ ـ پیروی از شیوهٔ خط عربی در موارد مشترك میان دو زبان فارسی و عربی.

از این جمله اصل اول هرگز ما را به جائی نمی‌رساند. کسانی که با نسخه های خطی قدیم فارسی و خطوط خوشنویسان معروف آشنائی دارند می‌دانند که اولاً درهیچ موردی خط فارسی قاعدهٔ ثابت و واحد نداشته است ؛ و ثانیاً در طی قرون متعددی که آثار ادبی زبان فارسی کتابت شده‌است تحول و تغییر بسیار در شکل حروف و شیوهٔ نوشتن کلمات روی داده‌است . بنابراین از نسخه‌هـای قـدیم نمی‌توان‌حکم‌واحدی استخراج کرد . برای مثال می‌توان شیوهٔ کتابت کسرهٔ اضافه را در حالی که کلمه به هاء بیان حرکت (یا هاء غیر ملفوظ) ختم شده باشد ذکر کرد .

دراین مورد ابتدا بعد از کلمه شکل « ی » نوشته می‌شد . مانند : « خانهٔی من » . بعد ، برای آنکه نشان بدهند که این حرف یای اصلی نیست و تنها برائر التقاء دو کسره چنین تلفظ می‌شود ، در کتابت شکل آن را اندکی تغییر دادند و سرحرف « ی » را نوشتند . به این صورت :« خانهٔی من » پس از آن شاید بـرای آنکه حروف و کلمات کمتر جا بگیرد این علامت را بالای حرف هاء قرار دادند و آن را به شکل همزه در آوردند و این صورت از آن حاصل شد : « خانهٔ من ».

بنابراین ، شکل نـوشتن این حـرف پیوسته در تحول بـوده‌است و صورت نخستین شاید ناقص‌ترین وجه آن باشد .

در موارد فصل و وصل حروف و کلمات هم کمتر نسخه‌ای را می‌توان پیدا کرد که در سراسر آن قاعدهٔ واحدی برای نوشتن یك وجه مراعات شده باشد ، تا آنجاکه گاهی در یك سطر کاتب کلمهٔ واحد را به چند صورت نوشته است . یك مثال این مورد ، حرف « به » اضافه است که‌گاهی به کلمهٔ بعد وصل شده و گاهی جدا نوشته شده‌است .

با توجه به این نکته‌ها در می‌یابیم‌که اتکاء به شیوهٔ نـوشتن کاتبان قدیم مـا را به جائی نمی‌رساند و مشکلی را حل نمی‌کند . پس باید درپی اصول دیگر رفت.

Source: زبان‌شناسی و زبان فارسی، پرویز ناتل خانلری
Modern reproduction

5
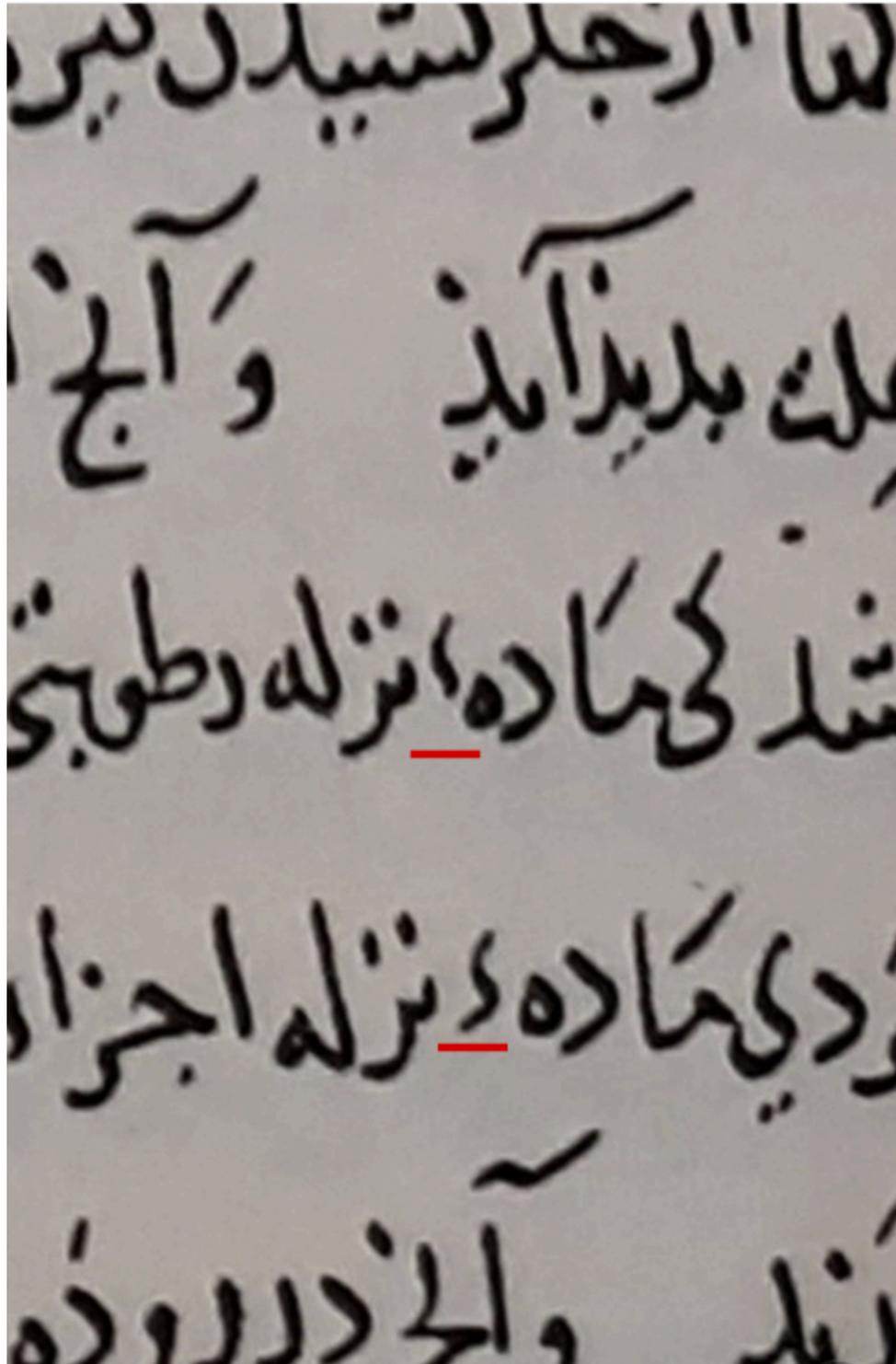
Source: نسخهٔ خطی ذخیرهٔ خوارزمشاهی، قرن ششم هجری

Historical manuscript

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
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*.
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)

A. Administrative

1.**Title:**	**Proposal to encode two historical Persian letters used for ezafe**
2. Requester's name:	*Ebrahim Byagowi and Roozbeh Pournader*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual Contribution*
4. Submission date:	*2025-05-09*
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:	*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: *N/A*

    b. The proposal is for addition of character(s) to an existing block: *Yes*

        Name of the existing block: *Arabic Extended-C*

2. Number of characters in proposal: *2*

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

A-Contemporary		B.1-Specialized (small collection)		B.2-Specialized (large collection)	
C-Major extinct	*X*	D-Attested extinct		E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols	

4. Is a repertoire including character names provided?   *Yes*

    a. If YES, are the names in accordance with the "character naming guidelines"   *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?

        Lorna Evans, SIL

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

        SIL

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 issue

    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)?   *Yes*

        See proposal

8. 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 UAX#44: 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?   *No*

    If YES explain

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.)?   *Yes*

        If YES, available relevant documents: *Proposers are members of the user community*

3. Information on the user community for the proposed characters (for example:

    size, demographics, information technology use, or publishing use) is included?   *Yes*

        Reference:

4. The context of use for the proposed characters type of use; common or rare)   *rare*

    Reference:

5. Are the proposed characters in current use by the user community?   *Yes*

    If YES, where? Reference: *Scholarly publications*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely

    in the BMP?   *No*

        If YES, is a rationale provided?

            If Yes, reference:

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?   *Yes*

8. Can any of the proposed characters be considered a presentation form of an existing

    character or character sequence?   *No*

    If YES, is a rationale for its inclusion provided?

If Yes,
reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either

existing characters or other proposed characters? | *No*

If YES, is a rationale for its inclusion provided?

If Yes,
reference:

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? | *No*

If YES, is a rationale for its inclusion provided?

If Yes,
reference:

11. Does the proposal include use of combining characters and/or use of composite
sequences? | *No*

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:

12. Does the proposal contain characters with any special properties such as

control function or similar semantics? | *No*

If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility characters? | *No*

If YES, are the equivalent corresponding unified ideographic characters identified?

If Yes,
reference: