

Proposal to encode COMBINING DOT BELOW LEFT for Syriac *

Ben Yang 楊偉堅 †¹,
Sebastian Kenoro Kiraz ܫܝܫܝܘܢ ܩܝܪܐܝܝܢ ܩܝܪܐܝܝܢ ‡², and
George A. Kiraz ܓܝܘܪܓܝܐ ܩܝܪܐܝܝܢ §³

¹*PanLex—The Long Now Foundation*
²*Rutgers Preparatory School & Beth Mardutho: The Syriac Institute*
³*Institute for Advance Study, Princeton & Beth Mardutho: The Syriac Institute*

02020-01-07

1 Introduction

The Syriac writing system is used to write the Classical Syriac language, as well as many other varieties of Aramaic spoken as native languages by (in alphabetical order) Aramean, Assyrian, Chaldean, and Syriac communities around the world.

Syriac potentially uses a very large number of diacritics, many of which take the forms of dots. The specific dots being discussed in this document are used to disambiguate between words that differ in pronunciation and/or semantics that would otherwise be written identically. Investigations have revealed that two diacritic dots used to disambiguate certain words are not yet supported.

The diacritics in question take the form of a dot placed above or below a word, in between two letters (hereafter referred to as the “above-between-dot” and “below-between-dot”). It is notable distinct from the similar diacritic of a dot placed above or below a specific letter (hereafter referred to as the “above-dot” and “below-dot”, respectively), which are currently supported. All four types of dots are used for disambiguating between words that would be written in the same fashion, but each are used for different classes of disambiguation. For example, the above-dot can be used to mark the third-person masculine possessive suffix, the below-dot can be used to mark unpronounced consonants, the above-between-dot can mark present tense verbs, and the below-between-dot can mark past tense verbs.

All four of these dots are used in all three styles of Syriac: ʿEstrangēlā, Western (Serṭā) and Eastern (Maḏḏhāyā).

For the above-dot and below-dot, the Unicode Core Spec recommends using U+0307 COMBINING DOT ABOVE and U+0323 COMBINING DOT BELOW, respectively. Keeping with the model of using script-agnostic diacritics when possible, we are proposing that U+1DF8 COMBINING DOT ABOVE LEFT be indicated in the Core Spec as the appropriate code point to represent the above-between-dot. However, there is no equivalent code point that could represent the below-between-dot, thus we also propose COMBINING DOT BELOW LEFT be added to the Unicode Standard.

*This proposal is a finalized proposal for the preliminary proposal by Sebastian Kiraz and George Kiraz *Expanding the “Syriac Supplement” U+0860 Block for historical glyphs and marks* (document L2/19-338 in the Unicode Document Register).
†ben@panlex.org
‡skiraz22@rutgersprep.org
§gkiraz@gorgiaspress.com

2 Request

- This proposal requests the addition of one new character in the Combining Diacritical Marks Supplement block with the following name and code point:

– U+1DFA ◌ COMBINING DOT BELOW LEFT

with the following annotation:

– used as disambiguation dot between two characters in Syriac

- Additionally, this proposal requests the following changes to be made to the Unicode Core Spec, section 9.3 Syriac:

– In **Table 9-12: Miscellaneous Syriac Diacritic Use**, replace the row beginning with “U+0307, U+0323” with the following:

U+0307, U+0323, U+1DF8, U+1DFA	These points are used for various purposes—grammatical, phonological, and otherwise. They differ typographically and semantically from the qushshaya, rukkakha points, and the dotted vowel points. If the point should occur above or below a single letter, U+0307 or U+0323 should be used. If the point should occur between two letters (above or below), U+1DF8 or U+1DFA should be used after the first letter.
-----------------------------------	--

- This proposal also requests the following changes to the **Combining Diacritical Marks Supplement** Code Chart:

– Change the glyph for U+1DF8 COMBINING DOT ABOVE LEFT to ◌ (the current glyph in the chart is specific to Typicon Cyrillic. This version is more script-neutral).

– Change heading for U+1DF8 from “Typicon marks” to “Miscellaneous mark”.

– Add the following annotations to U+1DF8:

* used as disambiguation dot between two characters in Syriac

* used in Typicon Cyrillic, where it takes a more square appearance

3 Justification

- Figures 1 and 2 (from page 12 and 14 of “Lives of the Eastern Saints” by John of Ephesus) shows COMBINING DOT ABOVE LEFT and COMBINING DOT BELOW LEFT (red arrows) contrasted with COMBINING DOT ABOVE and COMBINING DOT BELOW (blue arrows), each placed on SYRIAC LETTER MIM (U+0721).
- Figure 3 (from Ephrem’s Madroshe by Assemani) shows additional examples of COMBINING DOT ABOVE LEFT and COMBINING DOT BELOW LEFT.

4 Alternative Encoding Possibilities

The between-dots could be encoded as unique Syriac diacritics, like U+0743 SYRIAC TWO VERTICAL DOTS ABOVE. However, the current encoding model favors using common, script-agnostic combining diacritics whenever possible for Syriac diacritics.

On first glance, it may appear that U+0740 SYRIAC FEMININE DOT would be a good candidate to represent the above-between-dot. However, in addition to differing in semantics, the SYRIAC FEMININE DOT takes the form of a dot to the upper left only in the Western (Serṭā) form of the Syriac script. In the ’Estrangēlā and Eastern (Maḏnḥāyā) forms of Syriac, the SYRIAC FEMININE DOT takes the form of two dots below the letter. (Of course, COMBINING DOT BELOW LEFT would also still need to be proposed).

مَدَّةٌ وَفَاتَا هَدَا مَدَّةٌ : ٥٥
 وَفَاتَا وَفَاتَا وَفَاتَا : ٥٦
 مَدَّةٌ مَدَّةٌ : ٥٧
 مَدَّةٌ وَفَاتَا * مَدَّةٌ مَدَّةٌ : ٥٨
 مَدَّةٌ : مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٥٩
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٠
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦١
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٢
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٣
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٤
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٥

N
 qu
 co
 ler
 sci
 qu
 fin
 rec
 vel
 cit
 vit
 ve

مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٦
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٧
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٦٨
 مَدَّةٌ مَدَّةٌ مَدَّةٌ * : ٦٩
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٠
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧١
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٢
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٣
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٤
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٥
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٦
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٧
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٨
 مَدَّةٌ مَدَّةٌ مَدَّةٌ : ٧٩
 مَدَّةٌ مَدَّةٌ مَدَّةٌ * : ٨٠

I
 E

Figure 3: COMBINING DOT ABOVE LEFT and COMBINING DOT BELOW LEFT (red)

5 Character Data

5.1 Glyphs

- U+1DF8 COMBINING DOT ABOVE LEFT



- U+1DFA COMBINING DOT BELOW LEFT



5.2 Character Properties

- U+1DF8 COMBINING DOT ABOVE LEFT (changes):
 - Script_Extensions: Cyril Syrc
- U+1DFA COMBINING DOT BELOW LEFT (new character):
 - In UnicodeData.txt format:
 - * 1DFA;COMBINING DOT BELOW LEFT;Mn;218;NSM;;;;;N;;;;;
 - Canonical_Combining_Class: 218
 - Script_Extensions: Syrc
 - All other properties identical to U+1DF8 COMBINING DOT ABOVE LEFT

6 Sources

- John of Ephesus. Lives of the Eastern Saints, from Patrologia Orientalis, 1923, edited by R. Griffin and F. Nau (<https://archive.org/details/patrologiaorient17pariuoft/>)
- Ephraem, Syrus, Saint, 303-373. Sancti patris nostri Ephraem Syri Opera omnia quae exstant Graece, Syriace, Latine (<https://contentdm.lib.byu.edu/digital/collection/CUA/id/24129>)

7 Acknowledgements

We would like to thank the participants of the Unicode Script Ad-Hoc, and especially Debbie Anderson and Roozbeh Pournader, for their assistance in editing this proposal.

**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:	Proposal to encode COMBINING DOT BELOW LEFT for Syriac
2. Requester's name:	Ben Yang, Sebastian Kenoro Kiraz, and George A. Kiraz
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution
4. Submission date:	02020-01-06
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):	NO	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	YES	
Name of the existing block:	COMBINING DIACRITICAL MARKS SUPPLEMENT	
2. Number of characters in proposal:	1	
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?	Ben Yang	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Ben Yang, ben@panlex.org, OFL	
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 <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	NO
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:	YES George A. Kiraz
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	YES see proposal
4. The context of use for the proposed characters (type of use; common or rare) Reference:	common see proposal
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	YES see proposal
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:	NO
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	N/A
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:	NO
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:	NO
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:	NO
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:	NO
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)	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	NO