

Proposal to Encode Characters for the Babylonian Vocalization System of Hebrew

From: Aleksandr Andreev, University of Oslo, Faculty of Theology
(aleksandr.andreev@gmail.com); Christa Müller-Kessler, Friedrich Schiller University Jena,
Institute of Oriental Studies, Indo-European Studies, and Prehistoric and Early-History Archaeology
Date: May 27, 2025

Introduction

The Babylonian vocalization system of Hebrew is an ancient system of diacritical marks used primarily by Jewish communities in Babylonia (present-day Iraq) between the 6th and 10th centuries CE to indicate vowel sounds and other phonological features in Biblical and liturgical texts. It is distinct from the Tiberian system of vocalization that ultimately became the dominant tradition. In addition to Hebrew, this vocalization system was also used extensively for Jewish Babylonian Aramaic, the primary language of the Babylonian Talmud and other rabbinic and liturgical texts of the region. Its application to both Hebrew and Aramaic makes the system a key resource for understanding the phonology and linguistic history of Jewish communities in Babylonia. Despite its cultural and linguistic significance, the Babylonian vocalization system has not been fully encoded in the Unicode Standard, limiting scholarly access and digital preservation. This proposal seeks to encode the Babylonian vocalization marks in Unicode to support academic research, manuscript digitization, and the broader preservation of Jewish literary heritage.

The following supralinear vocalization marks are used in the system, demonstrated over the letter bet (U+05D1 Hebrew Letter Bet):

Patah	Qamatz	Qamatz Alternative	Tzere	Hiriq	Holam	Qubutz	Shva Mobile
							

Usage Examples

The following examples are offered from Edward M. Cook, *A Glossary of Targum Onkelos According to Alexander Sperber's Edition*, Leiden–Boston: Brill, 2008 (Studies in the Aramaic Interpretation of Scripture 6).

, emph. ¹, pl. abs.  *n. m.* father, ancestor
[DNWSI 4 #1; Vogt 1; ATTM 1:503; DJA 29; DJPA 31; DJBA 72; LSp 1;
DSA 1; LS2 1; MD 1.]

, emph.  *n. f.* ear
[DNWSI 26 s.v. ^{zn} #2; ATTM 1:505, 2:342; DJPA 36; DJBA 85; DSA 9;
LSp 3; LS2 6; MD 342.]

אֹדוּדַי, emph. אֹדוּדַיְאָ adj. another (f.)

[Vogt 5, ATTM 1:508, 2:344 s.v. אָדוּדַי; DJA 30; DJPA 38; cf. DJBA 105; DSA 20 s.v. אָדוּדַי; LSp 53 s.v. ḥwrn, MD 137 s.v. hurintia.]

אֵיוּבָא, emph. אֵיוּבָא n. m. hyssop

Ex 12:22; Lev 14:4, 6, 49, 51-52; Num 19:6, 18.

[DJPA 48 (targumic only); DSA 15 (SamTg only); cf. LS2 193 s.v. אֵיוּבָא.]

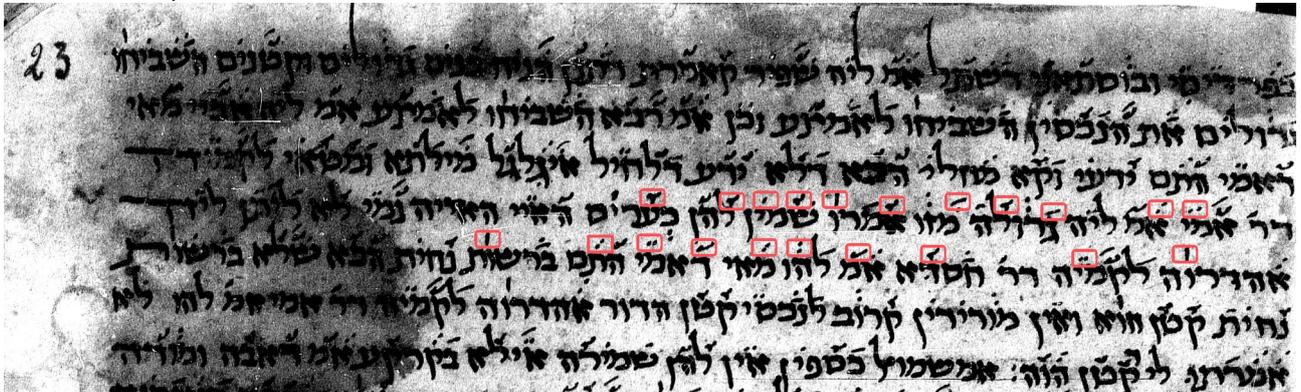
This example is from Hans Bauer and Pontus Leander, *Historische Grammatik der Hebräischen Sprache des Alten Testaments*, Bd. 1: Einleitung, Schriftlehre, Laut- und Formenlehre, Halle, 1922, p. 102:

x Die in den verschiedenen Punktationssystemen verwendeten Vokalzeichen sind die folgenden:

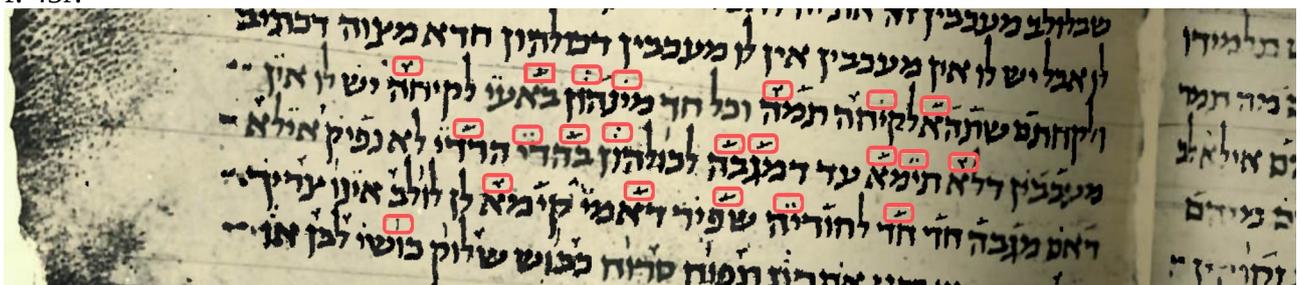
Palästinisch	◌ֶ ◌ֵ	◌ִ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ
Tiberisch	◌ֶ ◌ֵ	◌ִ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ
Babylonisch	◌ֶ ◌ֵ	◌ִ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ	◌ֶ ◌ֵ

Trotz aller Verschiedenheit im einzelnen ist unverkennbar, daß diese

The following example is from *Halachot Gedolot*, a work on Jewish law attributed to Simeon Kayyara (8th century CE), preserved in the manuscript Paris, Bibliothèque nationale de France, Hébreu 1402, f. 23r:



The following example is from *Halachot Pesukot*, a work on Jewish law attributed to Yehudai Gaon (8th century), preserved in the manuscript Toronto, Thomas Fisher Rare Book Library, MS. F4655, f. 45r:



Discussion

Since the the Hiriq is visually and typographically the same as U+0307 Combining Dot Above, the Tzere – as U+0308 Combining Diaeresis, and the Shva Mobile – as U+0304 Combining Macron, it is proposed not to encode these as separate characters, but to use the existing codepoints. The rationale for this is from the encoding model for Classical Syriac, where U+0308 Combining Diaeresis is used to typeset the double dots (*seyome*).

On the other hand, the Holam visually looks like U+0594 Hebrew Accent Zaqef Qatan, however they are functionally different since Zaqef Qatan is a cantillation mark. Thus, it is proposed not to unify these characters.

Proposed Characters

Thus, the following characters are proposed for encoding in the Hebrew block of Unicode. The representative glyphs are presented over the dotted circle to indicate their appearance over an arbitrary base character.



U+05C8 Combining Babylonian Point Patah



U+05C9 Combining Babylonian Point Qamatz



U+05CA Combining Babylonian Point Qamatz Alternative



U+05CB Combining Babylonian Point Holam



U+05CC Combining Babylonian Point Qubutz

Proposed Character Properties

The General Category for the proposed characters is to be set to Mn (Nonspacing Mark).

The Combining Class property is to be set to 230 (Above)

The Bidirectional Class is to be set to NSM (Nonspacing Mark)

The following entries are proposed for addition to UnicodeData.txt:

```
05C8;COMBINING BABYLONIAN POINT PATAH;Mn;230;NSM;;;;;N;;;;;
05C9;COMBINING BABYLONIAN POINT QAMATZ;Mn;230;NSM;;;;;N;;;;;
05CA;COMBINING BABYLONIAN POINT QAMATZ
ALTERNATIVE;Mn;230;NSM;;;;;N;;;;;
05CB;COMBINING BABYLONIAN POINT HOLAM;Mn;230;NSM;;;;;N;;;;;
05CC;COMBINING BABYLONIAN POINT QUBUTZ;Mn;230;NSM;;;;;N;;;;;
```

**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 Characters for the Babylonian Vocalization System of Hebrew

2. Requester's name: Aleksandr Andreev and Christa Müller-Kessler

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

4. Submission date: May 27, 2025

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: Hebrew

2. Number of characters in proposal: 5

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

A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input type="checkbox"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input checked="" 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?
Aleksandr Andreev, aleksandr.andreev@gmail.com

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
The font is called Nehemiah, and is a fork of Ezra SIL: <https://github.com/typiconman/Nehemiah>

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?	Yes
If YES explain	<i>There was a proposal in 2004 that did not advance for reasons unknown to us: https://www.unicode.org/wg2/docs/n2839.pdf</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.)?	Yes
If YES, with whom?	<i>Gilad Almosnino, liaison between Standards Institute of Israel and Unicode</i>
If YES, available relevant documents:	<i>Zoom meeting</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	N/A
Reference:	<i>Characters are archaic and used only by scholars</i>
4. The context of use for the proposed characters (type of use; common or rare)	N/A
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	<i>Printed grammars and dictionaries of Hebrew and Jewish Babylonian Aramaic, scholarly editions of Biblical and liturgical texts</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	Yes
If YES, reference:	<i>Should be placed in Hebrew block</i>
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?	Yes
If YES, is a rationale for such use provided?	Yes
If YES, reference:	<i>The characters are diacritical marks</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	N/A
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:	