Proposal to Encode Characters for the Babylonian and Palestinian Vocalization Systems of Hebrew

From: Aleksandr Andreev, University of Oslo, Faculty of Theology (aleksandr.andreev@gmail.com); Christa Müller-Kessler, Friedrich Schiller University Jena, Institut für Orientalistik, Indogermanistik, Ur- und Frühgeschichtliche Archäologie; Avi Shmidman, Bar-Ilan University, Department of Hebrew Literature, and The Academy of the Hebrew Language; Benjamin Outhwaite, University of Cambridge, Genizah Research Unit; Kim Phillips, Institute for Hebrew Bible Manuscript Research and University of Cambridge, Genizah Research Unit.
Date: Oct 22, 2025

Introduction

The Babylonian and Palestinian supralinear vocalization systems of Hebrew are two ancient systems of diacritical marks used primarily by Jewish communities in Babylonia (present-day Iraq) and Palestine between the 6th and 10th centuries CE to indicate vowel sounds, cantillation instructions, and other phonological features in Biblical, Rabbinic, and poetic Hebrew texts. They are distinct from the Tiberian system of vocalization that ultimately became the dominant vocalization tradition for Hebrew. The Babylonian and Palestinian vocalization systems are attested widely in hundreds of manuscripts from the Cairo Genizah. The Babylonian system is also attested across hundreds of pages of medieval European manuscripts as well. It should be emphasized that these alternate systems are *not* simply alternate sets of diacritic marks representing the same effective vocalization as in the Tiberian system; rather, in many instances, each of these two alternate systems maintain divergent pronunciations of Hebrew words, reflecting alternate phonological norms. Thus, the diacritic marks preserved by these systems provide indispensable evidence for researching and reconstructing the linguistic history of the Hebrew language.

It should be noted that the Babylonian 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 Babylonian diacritics all the more crucial for understanding the phonology and linguistic history of Jewish communities in Babylonia.

Nevertheless, despite their cultural and linguistic significance, these two vocalization systems have not yet been encoded in the Unicode Standard, limiting scholarly access and digital preservation. This proposal seeks to encode the Babylonian and Palestinian vocalization marks in Unicode to support academic research, manuscript digitization, and the broader preservation of Jewish literary heritage.

Part 1: The Babylonian Vocalization System for Hebrew

Most Common Babylonian Points

The primary supralinear vocalization marks used in the Palestinian tradition are (from left to right): Qamats, Patah, Tsere, Hiriq, Holam, and Qubuts. Here they are given in the classical reference of Paul Kahle, "Die Zeichen für die Vokale," in: *Historische Grammatik der Hebräischen Sprache des Alten Testaments* von Hans Bauer und Pontus Leander. Bd. 1: Einleitung, Schriftlehre, Laut- und Formenlehre, Halle, 1922, p. 102:

			e			
Palästinisch	⌣ å	— a	⌣ æ / ⌣ e	⌣ i	⌣ o	⌣ u
Tiberisch	⌣ å	— a	⌣ æ	⌣ e	— i	⌣ o	⌣וּ u
Babylonisch	⌣ å	⌣ } ä	— e	⌣ i	⌣ o	⌣ u

Trotz aller Verschiedenheit im einzelnen ist unverkennbar, daß diese

Note that Patah has an alternative dotted form, which is sometimes used alongside the main form in order to indicate an alternate meaning. This alternative form and alternative dotted forms of Qamats and Qubuts are discussed below.

These marks are used in academic literature, for example in 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).

אַב, emph. אַבָּא,[1] pl. abs. אַבָהָן *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.]

אדן, emph. אוֹדְנָא *n. f.* ear

    [**DNWSI** 26 s.v. *ʾzn* #2; **ATTM** 1:505, 2:342, **DJPA** 36*;* **DJBA** 85; **DSA** 9;
    LSp 3; **LS2** 6; **MD** 342.]

אוֹחֲרִי, emph. אוֹחֲרָנְתָא *adj.* another (*f.*)

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

איזוב, emph. אֵיזוֹבָא *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.
    ܙܘܦܐ.]

These marks are found in manuscripts. 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:
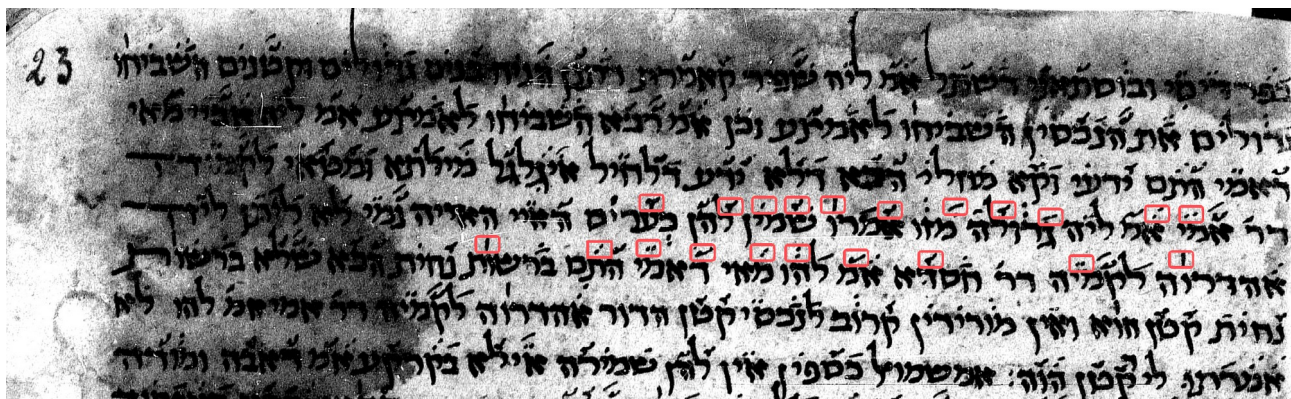
This example is from *Halachot Pesukot*, a work on Jewish law attributed to Yehudai Gaon (8ᵗʰ century), preserved in the manuscript Toronto, Thomas Fisher Rare Book Library, MS. F4655, f. 45r:
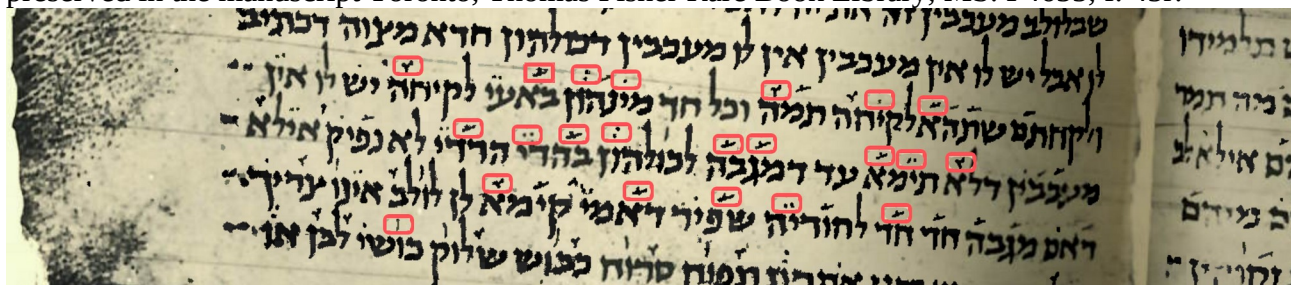
Thus, a total of seven principal Babylonian points (including Dotted Patah) are proposed for encoding.

Additional Babylonian Points

In addition to the primary Babylonian points mentioned in the 1922 work cited above, over the course of the subsequent 100 years, a comprehensive survey of all extant Hebrew manuscripts has identified a set of additional Babylonian points. These marks are detailed most comprehensively in Israel Yeivin, *The Hebrew Language Tradition as Reflected in the Babylonian Vocalization*, Jerusalem: The Academy of the Hebrew Language, 1985 (in Hebrew), and are summarized and updated in Geoffrey Khan's entry "Vocalization, Babylonian" in the *Encyclopedia of Hebrew Language and Linguistics* (Leiden: Brill, 2013).

First of all, in addition to the alternative Dotted Patah specified above, we also have similar dot-based alternatives for Qamats and for Qubuts, as per the following table from Khan (ibid.), p. 954, with Dotted Qamats on the third line (boxed in red) and Dotted Qubuts on the last line (boxed in red):

Tiberian	Babylonian lines	Babylonian dots
אִ ʾi	אִ ʾi	אִ ʾi
אַ ʾa	אַ ʾa	אַ ʾa
אָ ʾā̊	אָ ʾā̊	אָ ʾā̊
אֵ ʾē	אֵ ʾē	אֵ ʾē
אֹ ʾō	אֹ ʾō	אֹ ʾō
אֻ uʾ	אֻ ʾu	אֻ ʾu

An example of the Alternative Qamats is found in the following snippet from Ms. Cambridge T-S NS 246.26.16, containing a Masoretic note. The word "שלאוריתא" contains two qamats marks; in both cases, the qamats is indicated by a pair of dots, one below the letter and one above the letter.
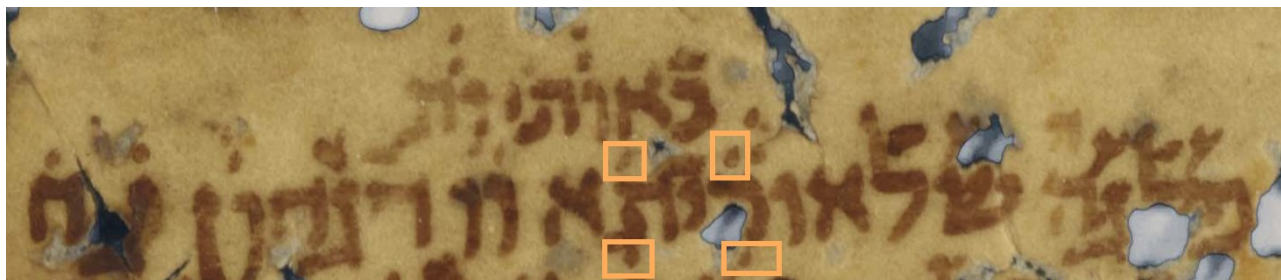
In the following Cairo Genizah fragment (Ms. Cambridge T-S B 4.36), an Alternative Qubuts symbol (right) is used adjacent to an Alternative Patah symbol (left):

Additionally, there is a set of miniature Hebrew letters written above the primary text in order to differentiate between *shin* and *sin* and in order to indicate the presence or absence of gemination (*dagesh* / *mappiq* / *rafe*), as detailed in the following paragraph from Khan (ibid.), page 956; manuscript attestations for all five of these items follow afterward.

The following diacritical signs are written over specific letters in the manuscripts. The letter *shin* is distinguished from *šin* by writing a miniature שׁ above the former and a miniature ס above the latter. The consonantal pronunciation of final ה is expressed by writing a miniature ה above the letter, which corresponds in function to Tiberian *mappiq*. *Dagesh* is expressed by a miniature ג above a letter. The Babylonian term for *dagesh* was דיגשא *digša*. The first letter of this term (ד *d*) was not used, since this was already in use as an accent sign, demonstrating that the accent signs were developed before the consonantal diacritics (→ Biblical Accents: Babylonian). *Rafe* is marked by placing over the letter a miniature letter ק *q*, representing the Babylonian term קיפיא *qipya*.

The use of the miniature 'ס' (samekh) to attest a "sin" pronunciation is attested twice in the following snippet from a Cairo Genizah fragment (Ms. Cambridge T-S NS 246.26.7) containing liturgical poetry:
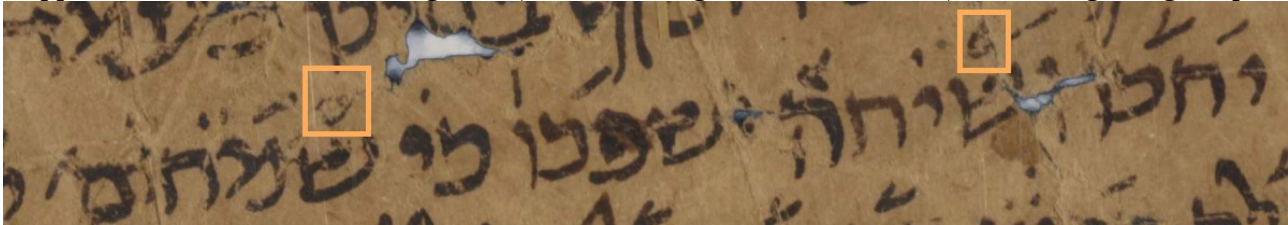
The use of the miniature 'שׁ' (shin) to attest a "shin" pronunciation is attested in the following Cairo Genizah snippet, from Ms. Cambridge T-S NS 246.26.18a:
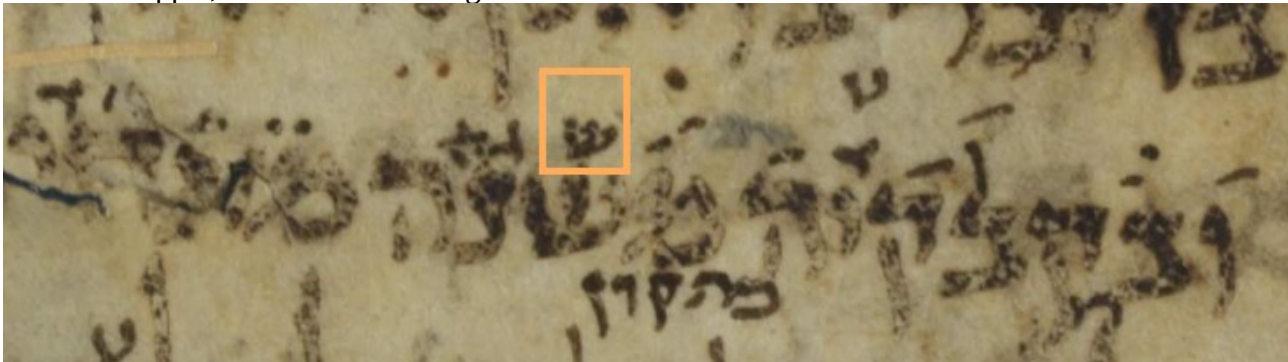
The use of miniature 'ק' (qof) to indicate *rafe* (lack of gemination) is attested in the following snippet, from Ms. Cambridge T-S A39.21 (a Biblical fragment; the snippet here contains Pss 77:13):
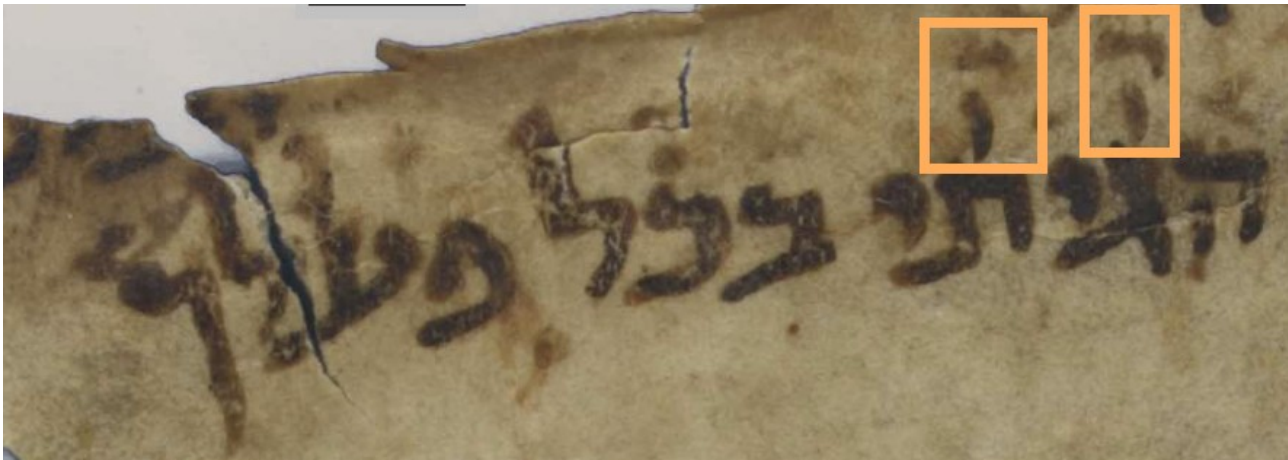
The use of the miniature 'ג' (gimel) to indicate *dagesh* (gemination) is attested, for example, in Ms. Cambridge T-S NS 246.26.7, which contains a liturgical poem for the Day of Atonement; the gemination is indicated above the letter 'ט' (tet) in the word 'מיטת':

Finally, the use of a miniature 'ה' (he) to indicate *mapiq* is attested, for example, in Ms. Cambridge T-S B 2.7 (a Biblical fragment), on the last letter of the word "עלותה" (Hos. 2:17):
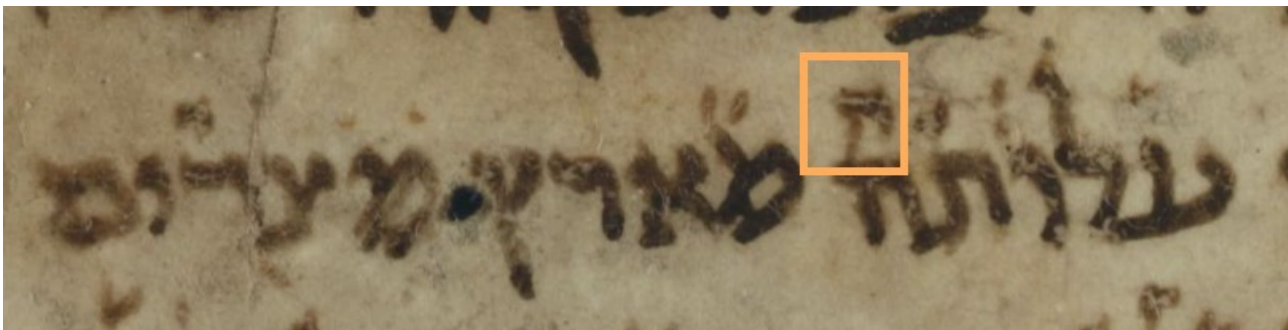

In manuscripts with Babylonian vocalization we also find the use of a "compound system", wherein a horizontal bar (named *hitfa*) is used together with the other Babylonian diacritics to indicate whether the syllable is closed or whether it is a short syllable. The *hitfa* may occur above or below another diacritical mark. These two uses of the *hitfa* mark are described in the following two paragraphs by Khan (ibid.), page 956:

The *ḥitfa* sign is placed over the vowel sign in unstressed syllables closed by *dagesh* in Tiberian, viz., אָ̄ *a*, אִ̄ *i*, אֻ̄ *u*. Examples: יֻבַּ̄קֹּשׁ *yəbaqqēš* (Tiberian יְבַקֵּשׁ *yəḇaqqēš*) 'he seeks', מֻגִ̇דֹּו *məḡiddō* (Tiberian מְגִדֹּו *məḡiddō*) 'Megiddo', וּמִכֻּלָּ̊ם *u-mik-kullå̊m* (Tiberian וּמִכֻּלָּם *u-mik-kullå̊m*) 'and from all of them (m)'. These signs are used also on the final vowel of a word that is followed by a word beginning with a consonant with *dagesh* in *dəḥiq* constructions, reflecting the shortening of the vowel (→ Vowel Length), e.g., הִשָּׁ̄בְעָ̊ה לִי *hiššå̊ḇə‘å̊ llī* (Tiberian הִשָּׁבְעָה לִי *hiššå̊ḇ‘å̊ llī*) 'swear to me' (Gen. 21.23) (Yeivin 1985:338–339).

The *ḥitfa* sign was placed under the vowel sign in unstressed syllables closed by *shewa* in Tiberian, in closed syllables at the end of a word, and where Tiberian has *ḥateph* signs, viz., אֻ̠ *u*, אֵ̠ *e*, אָ̠ *å̊*. Examples: בְּקִ̠רְבֹּו *bə-qirbō* (Tiberian בְּקִרְבֹּו *bə-qirbō*) 'within it (m)', וַתֵּ̠רֹד *wat-tēreḏ* (Tiberian וַתֵּרֶד *wat-tēreḏ*) 'and she came down', אֱ̠נֹושׁ *’enōš* (Tiberian אֱנֹושׁ *’ĕnōš*) 'man', וַיָּ̠קֹם *way-yå̊qåm* (Tiberian וַיָּקָם *way-yå̊qåm*) 'and he arose'.

For examples of these two usages of *hitfa*, see the following manuscript snippet (from Ms. Cambridge T-S B 4.23), in which we have highlighted examples of *hitfa* either above or below another Babylonian diacritic:
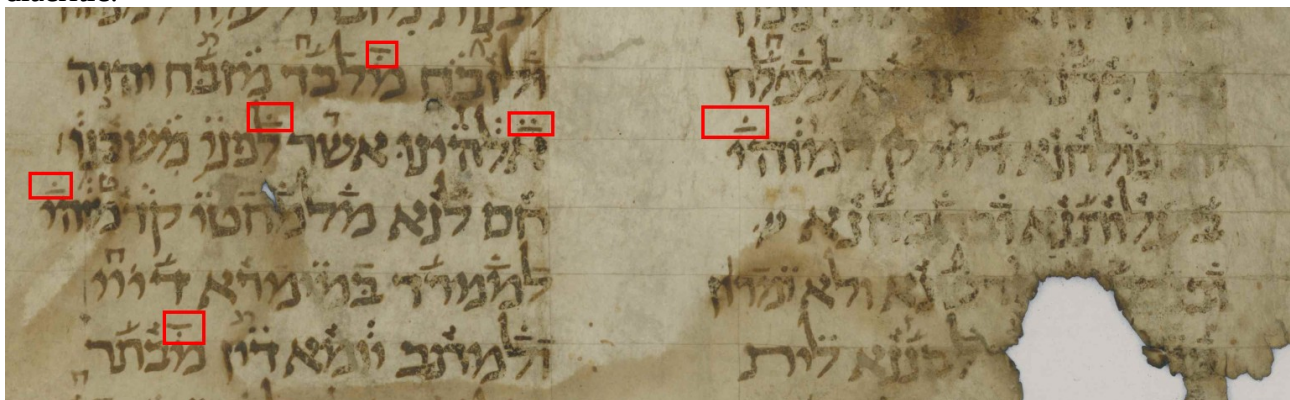
We propose to use a standalone *hitfa* (without any other diacritical mark) to encode the Shva Mobile. The Babylonian tradition literature refers to this horizontal bar as *hitfa* both when it combines with another diacritic mark to indicate a short or closed syllable and when it represents Shva Mobile.

Finally, although the original Babylonian tradition did not include a "segol" vowel at all, later manuscripts with Babylonian supralinear pointing do include indications for "segol". Khan (ibid., p. 55) notes four ways in which the segol may be represented:

> Yemenite reading tradition. Some later manuscripts use an additional sign corresponding to *seghol* in an attempt to represent Tiberian pronunciation. Various signs are used for this, including the *ṣere* sign א̇, a combination of *pataḥ* and *ṣere*, inverted Tiberian *seghol*, and the sign א̈ (the sign א̇ being used for *pataḥ*).

Of these four options, three are already covered in our proposed set of marks (e.g. the last option, with the two dots on a diagonal is the mark that we noted above as "Alternative Patah"). However, the "inverted Tiberian segol" does not overlap with any of the other Babylonian points and it requires a separate mark. An example of this mark follows here, from Ms. London, British Library, Or. 2373, f. 9v (a Biblical fragment; shown here is Pss 25:1):
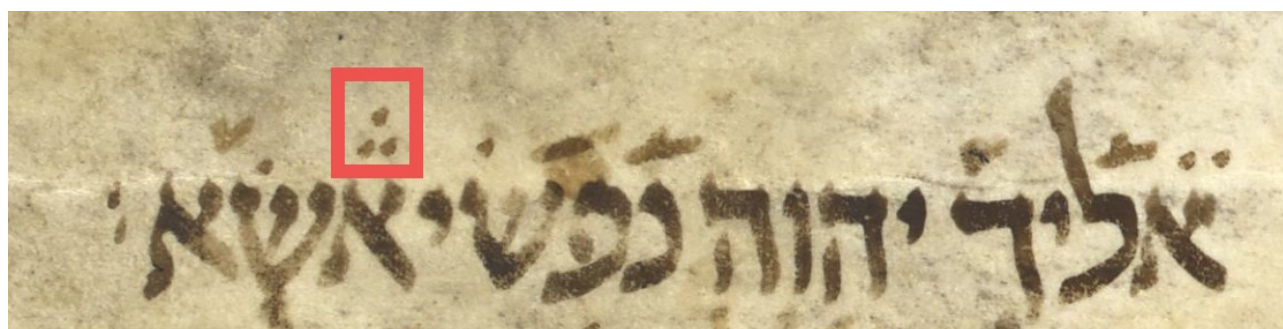
In summary this subsection proposes encoding eight additional marks for the Babylonian system: the Alternative Qamats and Qubuts marks, as well as the combining characters Digsha, Hitfa, Mapiq, Qifya, Sin Mark, and Shin Mark.

Babylonian Accents

Biblical Hebrew texts often contain "accents" as well: these additional diacritic marks are cantillation marks which preserve the Biblical reading tradition, indicating how the text is to be divided prosodically when reading it aloud, and how it is to be musically performed. The existing Hebrew Unicode block already contains a full set of Hebrew accents according to the Tiberian system (Unicode codepoints U+0591 to U+05AE, all defined as "Hebrew Accent" characters). This proposal seeks to encode an additional 13 supralinear accents used in the Babylonian system. These 13 accents are documented in the chart below, reproduced from Ronit Shoshany, *Biblical Accentuation in the Babylonian Tradition,* Jerusalem: The Ben-Zvi Institute, 2022, pages xv–xvi.

Accent	Role	Tiberian equivalent
silluq (no sign)		*silluq*
ʾatnaḥ 𐤀	Divides the whole verse.	in most cases – *ʾatnaḥ*
shin 𐤔	In the early strata it divides both the *silluq* and the *ʾatnaḥ* domains, whereas in the later strata it only divides the domain of the *ʾatnaḥ*	in many cases – *segol*
zayin 𐤆	Divides both the *silluq* and the *ʾatnaḥ* domains, and sometimes even divides the whole verse.	in many cases – *zaqef*
dalet 𐤃/𐤃/𐤃	The final accent before the *silluq*. In the early Babylonian stratum there is an additional accent, also marked with a *dalet*, which divides the accent *taw* in a number of manuscripts.	as the final accent before the *silluq*, usually – the *ṭipḥa* before the *silluq*
rimya 𐤓	The final accent before the *ʾatnaḥ*.	usually – the *ṭipḥa* before the *ʾatnaḥ*
nun 𐤍	In the early strata, divides the accents *shin* and *ḥet*, but in the later strata it is only employed in the domain of the accent *ḥet*.	in the later strata, usually – *munnaḥ legarmeh*

nun-yod ‏	The final accent in the domain of the accent *shin*.	in the later strata, usually – *zarqa*
taw ‏	In the early strata it is the main divider in the domain of the accent *zayin*, and in the later strata it is the final accent in the domains of the accents *zayin*, *dalet* and *rimya*.	
slanted-*nun* ‏	In the early strata it is the final divider of the accent *zayin*. It disappears in the later strata and only occurs in cases of transformation of a repeated *ḥet*.	
ḥet ‏	In the early strata it is the main divider of the accent *shin* and also the divider of the accent *taw*. In the later strata it is a secondary divider after the accent *zayin* in both the *silluq* and the *ʾatnaḥ* domains; a main divider in the *shin*, *zayin*, *dalet* and *rimya* domains; and a divider of the accent *taw*.	as a main divider in the *shin*, *zayin*, *dalet* and *rimya* domains, usually – *reviaʿ*
yod ‏	Unique to the c stratum. It replaces the accent *taw* as the final divider of the accent *zayin*.	usually – *pashṭa* or *yetib*
ṭet ‏	Divides the accents *ḥet* and slanted-*nun*.	
half-ṭet ‏	Unique to the early strata. It divides the accents *nun* and *nun-yod* and sometimes also divides the accent *ḥet*.	

These marks have Hebrew names, which are discussed *ibid.* (pages 61 to 64; in Hebrew), and form the basis for the proposed names of the characters (see our list of proposed characters, below). Note that while the Accent Shidya consists of a combining Hebrew Letter Shin, this is functionally and visually different from the Combining Shin Mark proposed for encoding in the previous section.

The usage of all of these 13 accents is highlighted in six manuscript snippets below; the first is from Ms. Cambridge T-S A 39.9, the second is from Ms. Cambridge T-S B 4.23, the third is from Ms. Cambridge

T-S AS 62.160, the fourth is from the Codex Babylonicus Petropolitanus (Ms. Russia, Evr. I. B 3), the fifth is from Ms. Oxford heb. d. 49, folio 6, and the sixth is from Ms. Berlin Or. Qu. 680.
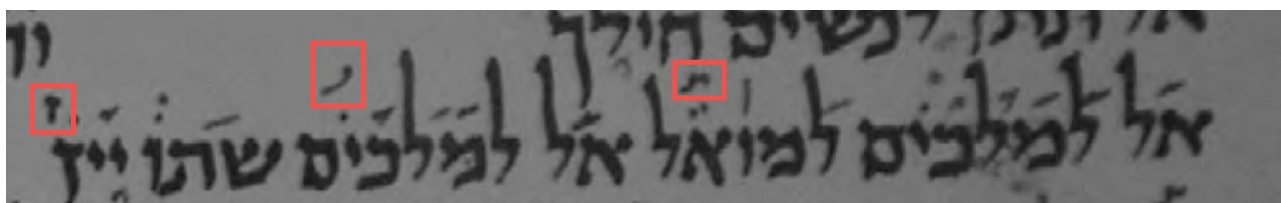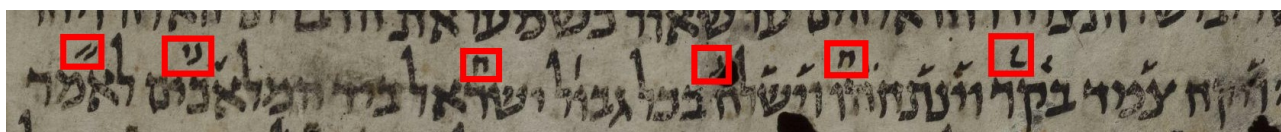
Part 2. The Palestinian Vocalization System for Hebrew

Palestinian Points

The six primary supralinear marks in the Palestinian tradition are Qamats, Patah, Segol, Hiriq, Holam, and Qubuts, as attested in the chart from Kahle, "Die Zeichen", p. 102, discussed at the beginning of Part 1 and reproduced here with the relevant marks highlighted:

All six of these marks are attested in this manuscript snippet (Ms. Cambridge T-S NS 249.7):
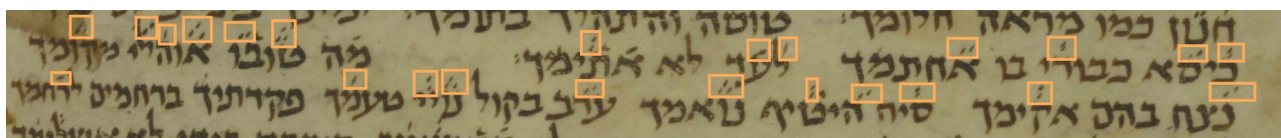
As indicated in the same chart, the "ae/e" sound is sometimes separated into two distinct sounds, with a seventh symbol used for the "e" (*tsere*) sound:

This is attested, for instance, in the Cairo Genizah fragment Ms. Cambridge T-S 12.195:
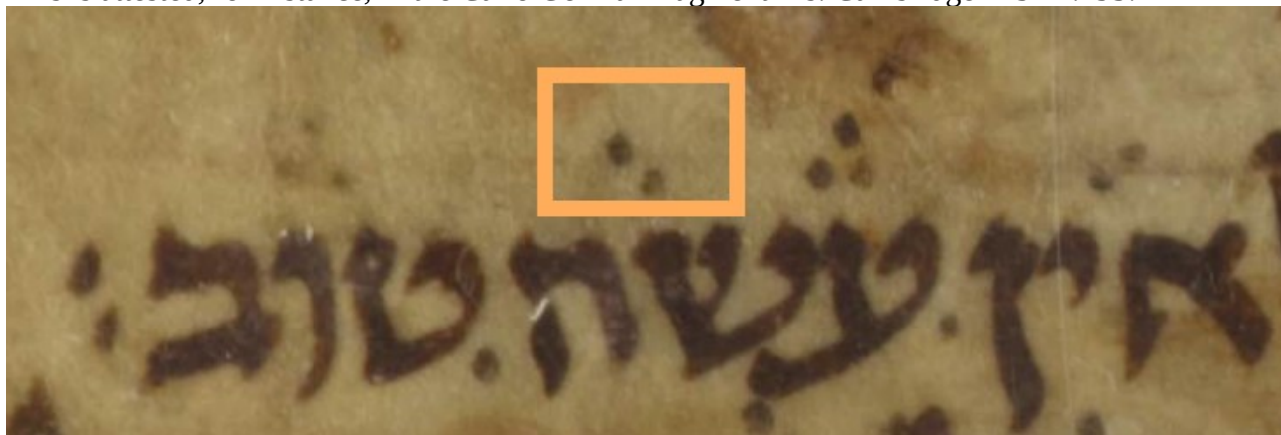
Additionally, the Palestinian tradition uses supralinear marks to specify *dagesh* (gemination) and *rafe* (lack of gemination), as described in the following paragraph by Shai Heijmans (from his entry "Vocalization, Palestinian" in the aforementioned *Encyclopedia of Hebrew Language and Linguistics*, page 964):

The signs בּ, בּ, and בּ (the latter two having probably developed from the first) are used as *dagesh*, e.g., משמׁרים (Tiberian מְשַׁמְּרִים *məšammərīm*) 'conserve (mpl)', or as *mappiq*, e.g., אחריתֹה (Tiberian אַחֲרִיתָ֖הּ *ʾaḥărīṯå̄h*) 'her end', וענׁותֹך (Tiberian וְעַנְוָתְךָ *wə-ʿanwå̄ṯḵå̄*) 'and your (ms) condescension'. In some cases the signs can be interpreted as counterparts of the Tiberian mobile *shewa* (Fassberg 1990:46–47). The sign בֹ corresponds to the Tiberian *rafe*; it can also be used to denote a *mater lectionis*.

The following Cairo Genizah fragment (Ms. Oxford heb. e. 30 folio 49) demonstrates the use of both the *rafe* mark (on the right), as well as the *dagesh* mark (on the left). This snippet is a transcription of Isa. 47:9; the highlighted marks indicate that the penultimate letter of the first word is not geminated, while the first letter of the second word is to be geminated.
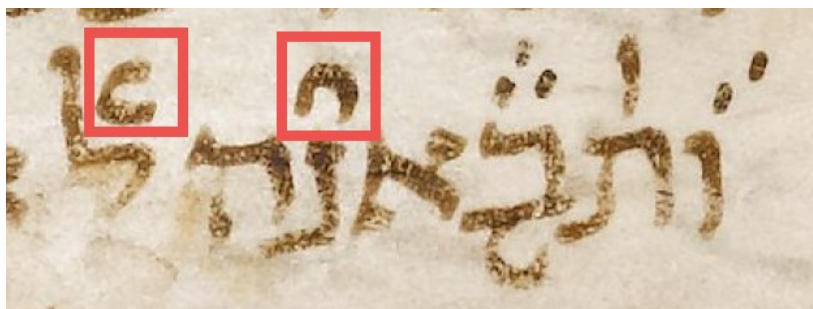
In summary, we propose encoding nine marks of the Palestinian system: the seven vowel marks and the *dagesh* and *rafe* symbols.

Palestinian Accents

The Palestinian system also uses a set of accents (cantillation marks). The most comprehensive study to date of this accent system is that of E. J. Revell, *Biblical Texts with Palestinian Pointing and their Accents*, Missoula, Montana: Scholars Press, 1977.

The primary set of 10 Palestinian accents is provided in a chart in Revell's book on page 37:

In addition to these 10 accent characters, an eleventh accent (*munah*/*mehuppakh*) is attested in other manuscripts, as he notes on page 63:

At first glance, a number of these might seem to be standard diacritics that already exist in other Unicode blocks. However, the need to encode these accents as separate characters derives from the fact that they can have varying graphical appearances, many of which diverge widely, as we will discuss below. On the one hand, clearly, only a single Unicode point should exist for each one of these functional accents, rather than for each of the possible graphic realizations. For instance, although the *zaqef* can be realized via three very different graphemes, we propose only a single Unicode point for *zaqef*, because functionally these three graphemes are all equivalent, and in practice, scholars will simply utilize different fonts to select the relevant grapheme for any given manuscript publication. At the same time, though, it would not be appropriate to use an existing character such as U+0307 Combining Dot Above for *zaqef* – even though the "prototypical" glyph looks similar to an upper dot –

because the alternate forms for *zaqef* include an angled line or a vertical bar and it would not make sense to produce fonts in which e.g. U+0307 is realized as an angled line or as a vertical bar. Furthermore, our approach allows for correct processing of multilingual text.

Therefore we propose to encode 11 accents of the Palestinian cantillation system. As the prototypical glyph for each accent that will be displayed in the Unicode codecharts, we use the graphic representation found in the chart provided by Revell, above. Graphical variation of these characters will be handled at the font level.

Excursus: Graphical Variants of Palestinian Accents

In this subsection we discuss the 11 proposed accents one by one, providing literature and manuscript attestations for each graphical variant of each mark. This subsection is not necessary for understanding the proposal as a whole.

The *zaqef*

The *zaqef* has three forms: *zaqef* 1, which is a high dot placed above the letter, as noted by Revell (ibid.), page 63:

The second form, *zaqef* 2, is a vertical bar placed above the letter, as in the chart from Revell (ibid.), page 37, noted above.

Zaqef 3 is an angled line placed above the letter, as noted by Revell (ibid.), page 128:
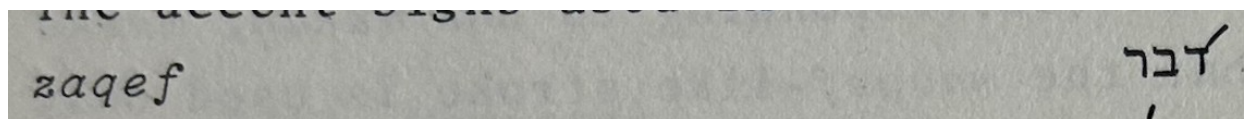
Zaqef 1 can be seen in Cambridge Genizah Ms. T-S A43.1, which is a shorthand manuscript of the Hebrew Bible, at Jer. 27:4, above the letter "ת" (tav) (abbreviated form of אתם):
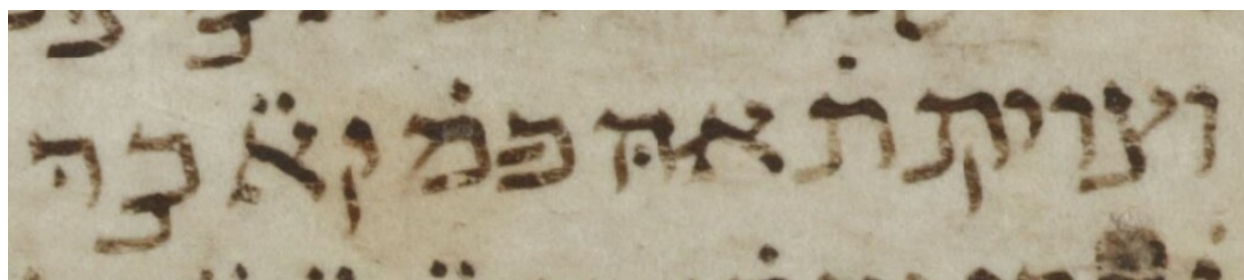
Zaqef 2 can be found in Cambridge Genizah Ms. T-S 16.383 at Judg. 7:12, above the letter "פ" (pe) of מספר:
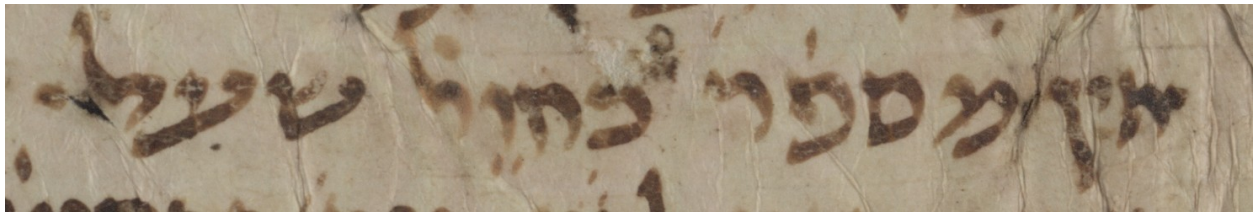
Zaqef 3 can be found in Cambridge Genizah Ms. T-S AS 63.61, a scroll of the Torah, at Num. 10:35 on the letter "א" (alef) of איביך:
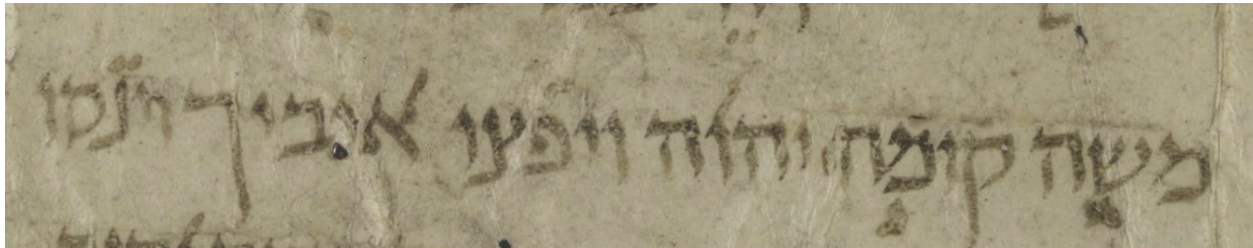
The *revia*

Revia is a horizontal line above the letter, usually placed to the left, as noted *ibid*., page 43:
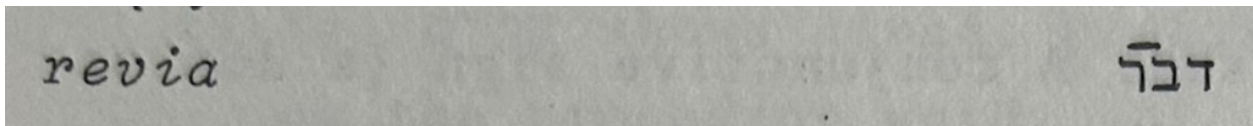
An example can be seen in Strasbourg BNU MS. 4065/33, at Ex. 21:34, above the "ו" (vav) of בור:
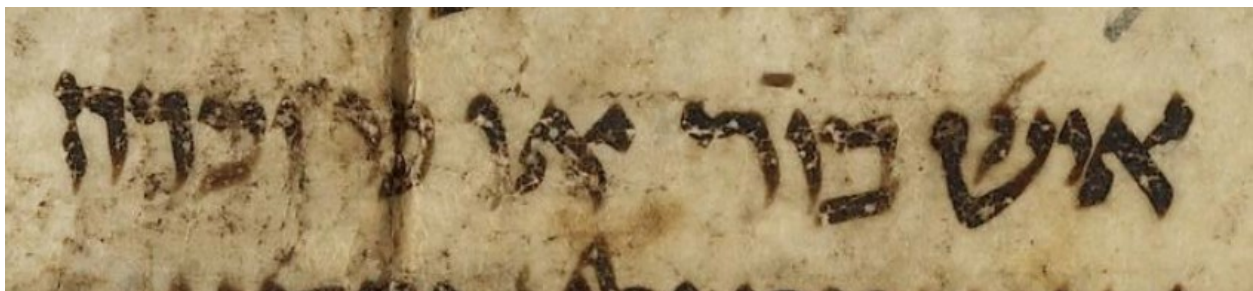
The *pashta*

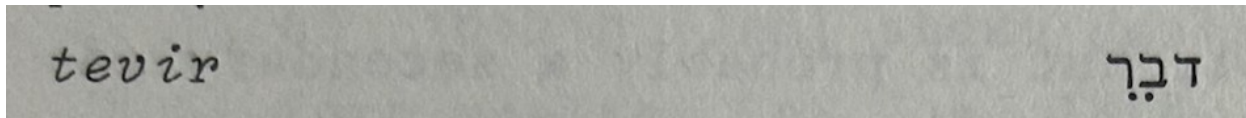*Pashta* is represented by two horizontal dots placed above the letter, as noted *ibid*., page 63:

Two examples can be found together in Cambridge Genizah Ms. T-S A43.1, at Jer. 27:6, on the fifth and sixth letters on the line ("כ" [kaf] and "ת" [tav] = abbreviated forms of the Hebrew words אנכי נתתי):
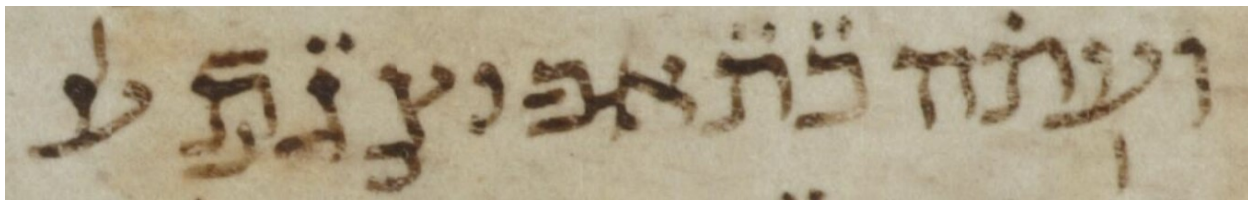
The *tevir*

Tevir takes the form of two dots under the letter, as noted *ibid*., page 43:

An example, showing its placement to the left, can be found in Strasbourg BNU MS. 4065/33, at Ex. 22:2, under the "שׁ" (shin) of השמש:

The *zarqa/legarme*

In the above chart, the *zarqa/legarme* is represented by a vertical bar, similar to the Tiberian *paseq*. However, in other manuscripts, it is represented as a lower dot to the left of the letter, as per Revell (*ibid*.), page 63:

And in still other manuscripts, it is represented by a postpositive hook (*ibid*., page 121):

An example of the former is from the Cairo Genizah fragment Ms. Cambridge T-S A 43.1, displaying Isa. 58:6. Here the dot is positioned to the left of the letter "ז" (zayin), representing the second word of the verse:
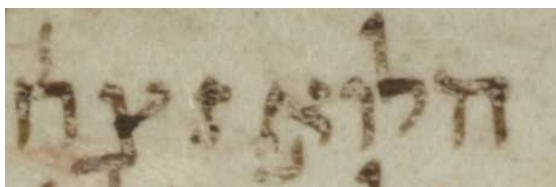
An example of the latter is in the Cairo Genizah fragment Ms. Cambridge T-S AS 63.61, displaying Num. 10:10. Here the *zarqa* mark is post-positioned to the letter "ם" (mem), at the end of the word ובמועדיכם:
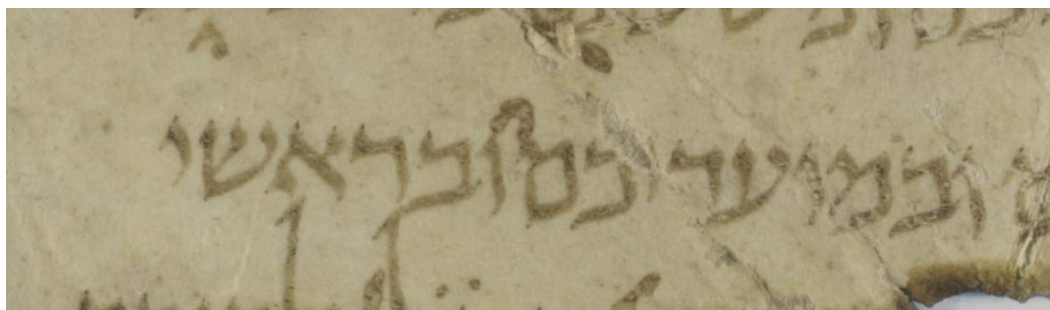
The *pazer*

The *pazer* is like a *háček*, as noted in Revell, page 57:

It is usually tilted, as in this example from Cambridge Genizah Ms. T-S 20.59, above the "ר" (resh) of ישראל at Ez. 14:4:

The *tifha/atnah/segolta/silluq*

The *tifha/atnah/segolta/silluq* (hereafter just *tifha*) has two forms: *tifha* 1, which is a dot marked, usually postpositively, below the letter, as noted in Revell (*ibid*.), page 37:
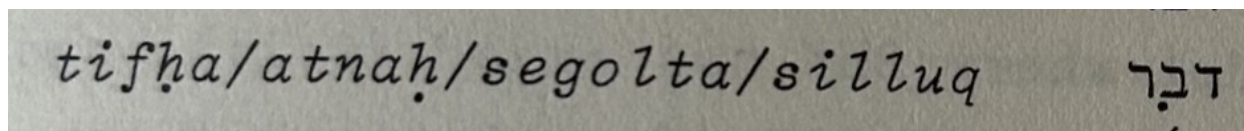
Tifha 2 is a small backwards-facing "c" sign placed postpositively beneath the letter, as noted in Revell (*ibid*.), page 121:
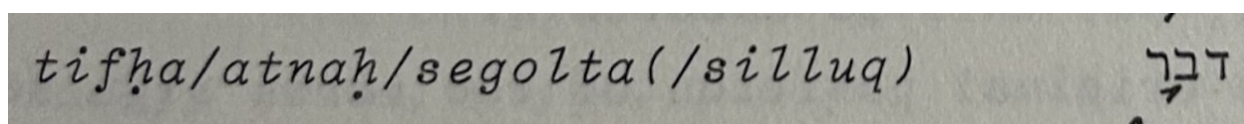
An example of *tifha* 1 can be seen in Cambridge Genizah Ms. T-S Misc.2.25, under the "ו" (vav) of נחום at Nahum 1:1:
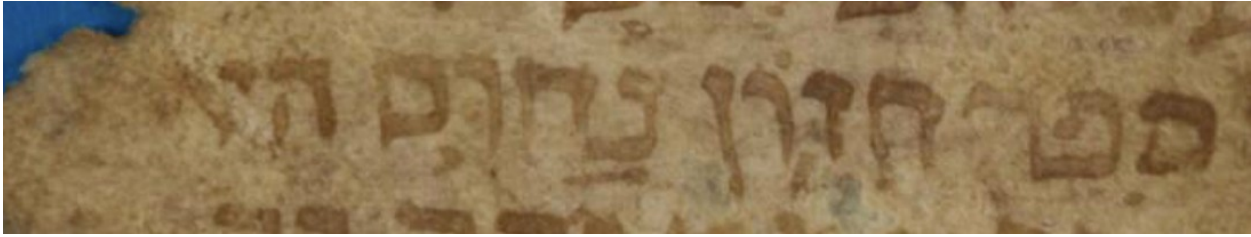
An example of *tifha* 2 can be seen in Cambridge Genizah Ms. T-S 16.383, under the "ת" (*tav*) of נתתיו at Judg. 7:9:
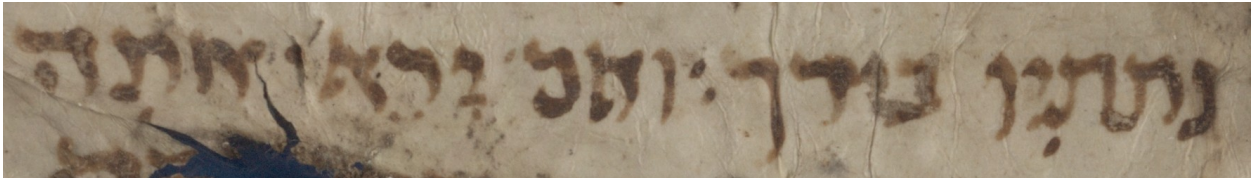
The *geresh*

The *geresh* has two forms. *Geresh* 1 is a backwards sloping line placed above the letter, as noted in Revell (*ibid.*), page 43 (showing the placement above the letter or prepositively):
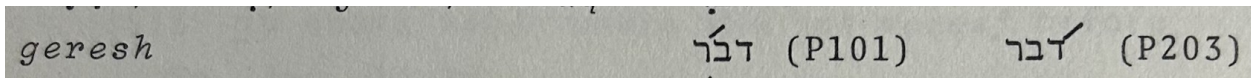
And *geresh* 2 is a double backwards-sloping line, as noted in Revell (*ibid.*), p. 128:
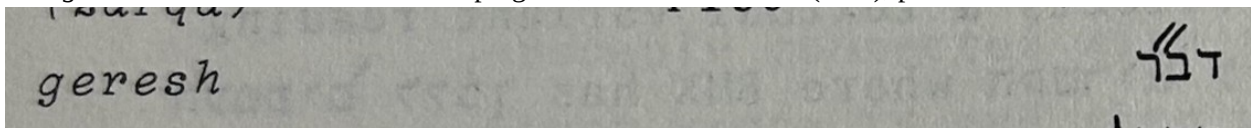
An example can be seen in Cambridge Genizah Ms. T-S AS 63.61, above the "ו" (*vav*) of the Divine Name in Num. 10:33:
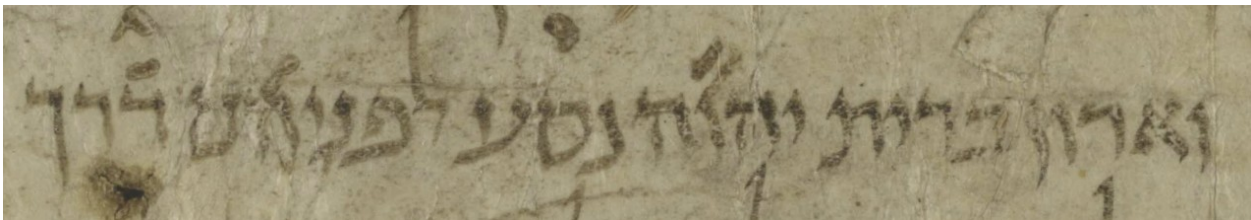
The *telisha*

The *telisha* (also known as the *telisha gedola*, to distinguish it from the *telisha qetanna* below) has two forms. *Telisha* 1 is represented by an inverted breve, often angled or squeezed, and placed to the right of the letter, as noted in Revell (*ibid.*), page 37.

telisha   דּבׄר

Telisha 2 is represented by two vertical dots placed post-positively above the letter, as noted in Revell (*ibid*.), page 43.

telisha   דּבׄר
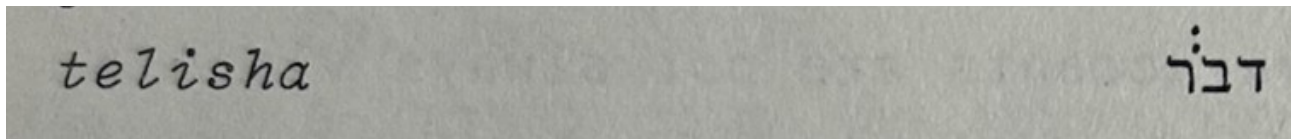
An example of *telisha* 1 is from Cairo Genizah Ms. Cambridge T-S AS 63.41, which is a scroll of the book of Daniel, with both Tiberian and Palestinian signs. The *telisha* is marked to the top right of the letter "י" (yod) at the beginning of ישחית (Dan. 9:26):
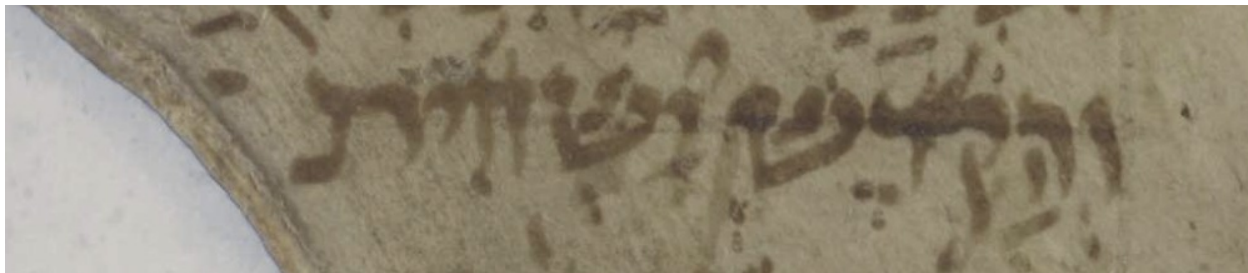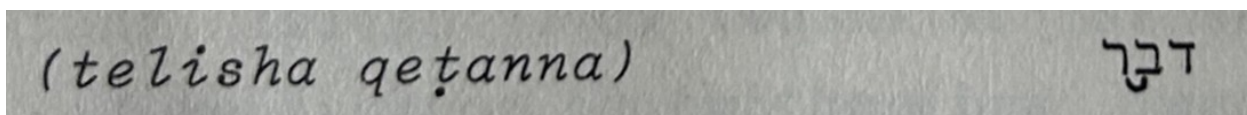

Telisha 2 can be found in an Oxford Genizah Ms., Bodl. MS Heb d29/17b, a fragment with both Palestinian and Tiberian signs, above the "שׁ" (shin) of אשר at Josh. 15:8:
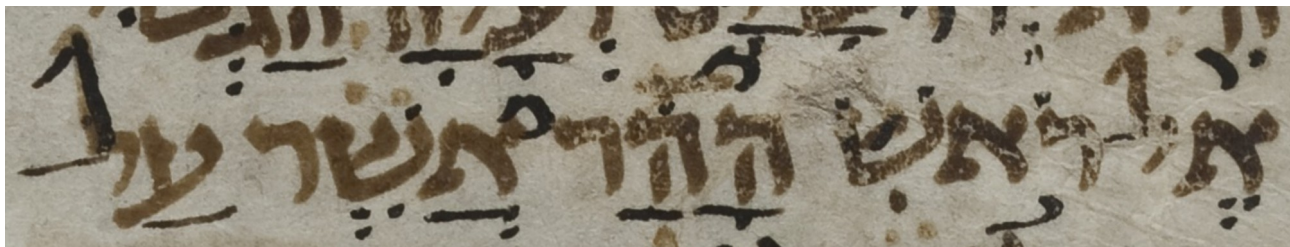
The *telisha qetanna*

Telisha qetanna is marked in the form of a breve-like shape placed below and to the left of the letter, as noted in Revell (*ibid*.), page 57:

(telisha qeṭanna)   דּבֶ֔ר

It can be found, for example, in the shorthand Hebrew Bible manuscript Cairo Genizah Ms. Cambridge T-S NS 249.6, at 1 Chron. 2:53, underneath and to the left of the "א" (alef) in האתלי:
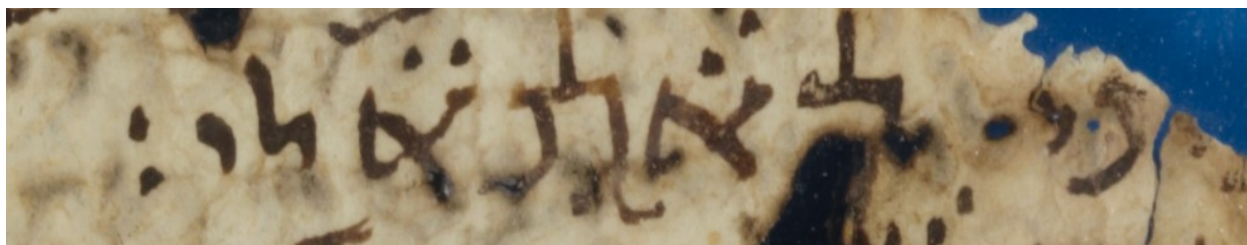
The *munah/mehuppak*

The *munah/mehuppak* sign takes the form of a vertical line under the letter, as noted in Revell (*ibid*.), page 63:


It often intersects with the letter's baseline or even sits inside the open base of a letter. Such an example of *mehuppak* can be found in the shorthand Hebrew Bible manuscript Cairo Genizah Ms. Cambridge T-S NS 249.5, at Ez. 36:5, underneath and slightly inside the "ח" (het) (at the beginning of the line here) which stands for the word בשמחת:

Section 3: Summary of Proposed Characters and Technical Information

The existing Hebrew block in Unicode has eight unassigned codepoints U+05C8 to U+05CF that follow the already assigned points for accents and marks and precede the points assigned to the Hebrew letters. We propose to assign these codepoints to the principal marks of the Babylonian system as well as the *hitfa* and *segol* in the following manner:

|	
	U+05C8 Babylonian Point Patah **Annotation**: = *Miftah Puma, Pitha*
	U+05C9 Babylonian Point Qamats **Annotation**: = *Miqpats Puma, Metsaf Puma, Imtsa*
	U+05CA Babylonian Point Tsere

|	
◌	U+05CB Babylonian Point Hiriq
◌	U+05CC Babylonian Point Holam **Annotation**: = *Holem, Halma*
◌	U+05CD Babylonian Point Qubuts
◌	U+05CE Babylonian Point Hitfa **Annotation**: May be used by itself for Shva Mobile or combined with another point to indicate syllable length.
◌	U+05CF Babylonian Point Segol

The Unicode Hebrew block has 11 unassigned codepoints following the Hebrew ligatures and punctuation marks, starting at U+05F5 and up to and including U+05FF, the end of the block. We propose to assign these codepoints to the seven principal Palestinian points and the three alternative Babylonian points, leaving U+05FF unassigned. This also leaves four codepoints U+05EB to U+05EE following the Hebrew letters unassigned and available for future use.

|	
◌	U+05F5 Palestinian Point Patah
◌	U+05F6 Palestinian Point Qamats
◌	U+05F7 Palestinian Point Tsere
◌	U+05F8 Palestinian Point Hiriq
◌	U+05F9 Palestinian Point Holam
◌	U+05FA Palestinian Point Segol
◌	U+05FB Palestinian Point Qubuts

ּ	U+05FC Babylonian Point Dotted Patah
ִ	U+05FD Babylonian Point Dotted Qamats
ּ	U+05FE Babylonian Point Dotted Qubuts

For encoding the remaining characters, we propose creating a Hebrew Extended block of 32 slots U+1xx00 through U+1xx1F at a place in the SMP to be determined by the UTC. There we propose to encode the additional Babylonian and Palestinian points and the accents. The last two points U+1xx1E and U+1xx1F remain unencoded.

	Additional Babylonian and Palestinian Points
∩	U+1xx00 Palestinian Point Rafe
⊂	U+1xx01 Palestinian Point Dagesh
ג	U+1xx02 Babylonian Point Digsha
ה	U+1xx03 Babylonian Point Mapiq
ס	U+1xx04 Babylonian and Palestinian Point Sin Mark
ק	U+1xx05 Babylonian Point Qifya
ש	U+1xx06 Babylonian and Palestinian Point Shin Mark
	Babylonian Accents
ז	U+1xx07 Babylonian Accent Ziqpa
ד	U+1xx08 Babylonian Accent Dihya

ֶ	U+1xx09 Babylonian Accent Nigda
ֺ	U+1xx0A Babylonian Accent Tivra
ֻ	U+1xx0B Babylonian Accent Teres
˅	U+1xx0C Babylonian Accent Rimya
˄	U+1xx0D Babylonian Accent Sihpa
ֻ	U+1xx0E Babylonian Accent Shidya
ֻ	U+1xx0F Babylonian Accent Nihla
ֹ	U+1xx10 Babylonian Accent Yetiv
ֻ	U+1xx11 Babylonian Accent Half-Tet
ֻ	U+1xx12 Babylonian Accent Slanted Nun
ֻ	U+1xx13 Babylonian Accent Hazer
Palestinian Accents	
׀	U+1xx14 Palestinian Accent Zaqef
—	U+1xx15 Palestinian Accent Revia
••	U+1xx16 Palestinian Accent Pashta

	U+1xx17 Palestinian Accent Tevir
	U+1xx18 Palestinian Accent Zarqa **Annotation:** = *Legarmeh*
	U+1xx19 Palestinian Accent Pazer
	U+1xx1A Palestinian Accent Tifha **Annotation:** = *Atnah, Segolta, Silluq*
	U+1xx1B Palestinian Accent Geresh
	U+1xx1C Palestinian Accent Telisha Gedolah
	U+1xx1D Palestinian Accent Telisha Qetanna
	U+1xx1E Palestinian Accent Munah **Annotation:** = *Mehuppakh*

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 as defined below for each character.
The Bidirectional Class is to be set to NSM (Nonspacing Mark)
The following entries are proposed for addition to UnicodeData.txt:

```
05C8;BABYLONIAN POINT PATAH;Mn;230;NSM;;;;;N;;;;;
05C9;BABYLONIAN POINT QAMATS;Mn;230;NSM;;;;;N;;;;;
05CA;BABYLONIAN POINT TSERE;Mn;230;NSM;;;;;N;;;;;
05CB;BABYLONIAN POINT HIRIQ;Mn;230;NSM;;;;;N;;;;;
05CC;BABYLONIAN POINT HOLAM;Mn;230;NSM;;;;;N;;;;;
05CD;BABYLONIAN POINT QUBUTS;Mn;230;NSM;;;;;N;;;;;
05CE;BABYLONIAN POINT HITFA;Mn;230;NSM;;;;;N;;;;;
05CF;BABYLONIAN POINT SEGOL;Mn;230;NSM;;;;;N;;;;;
05F5;PALESTINIAN POINT PATAH;Mn;230;NSM;;;;;N;;;;;
05F6;PALESTINIAN POINT QAMATS;Mn;230;NSM;;;;;N;;;;;
05F7;PALESTINIAN POINT TSERE;Mn;230;NSM;;;;;N;;;;;
05F8;PALESTINIAN POINT HIRIQ;Mn;230;NSM;;;;;N;;;;;
05F9;PALESTINIAN POINT HOLAM;Mn;230;NSM;;;;;N;;;;;
05FA;PALESTINIAN POINT SEGOL;Mn;230;NSM;;;;;N;;;;;
05FB;PALESTINIAN POINT QUBUTS;Mn;230;NSM;;;;;N;;;;;
05FC;BABYLONIAN POINT DOTTED PATAH;Mn;230;NSM;;;;;N;;;;;
05FD;BABYLONIAN POINT DOTTED QAMATS;Mn;1;NSM;;;;;N;;;;;
```

```
05FE;BABYLONIAN POINT DOTTED QUBUTS;Mn;230;NSM;;;;;N;;;;;
1xx00;PALESTINIAN POINT RAFE;Mn;230;NSM;;;;;N;;;;;
1xx01;PALESTINIAN POINT DAGESH;Mn;230;NSM;;;;;N;;;;;
1xx02;BABYLONIAN POINT DIGSHA;Mn;230;NSM;;;;;N;;;;;
1xx03;BABYLONIAN POINT MAPIQ;Mn;230;NSM;;;;;N;;;;;
1xx04;BABYLONIAN AND PALESTINIAN POINT SIN MARK;Mn;230;NSM;;;;;N;;;;;
1xx05;BABYLONIAN POINT QIFYA;Mn;230;NSM;;;;;N;;;;;
1xx06;BABYLONIAN AND PALESTINIAN POINT SHIN MARK;Mn;230;NSM;;;;;N;;;;;
1xx07;BABYLONIAN ACCENT ZIQPA;Mn;230;NSM;;;;;N;;;;;
1xx08;BABYLONIAN ACCENT DIHYA;Mn;230;NSM;;;;;N;;;;;
1xx09;BABYLONIAN ACCENT NIGDA;Mn;230;NSM;;;;;N;;;;;
1xx0A;BABYLONIAN ACCENT TIVRA;Mn;230;NSM;;;;;N;;;;;
1xx0B;BABYLONIAN ACCENT TERES;Mn;230;NSM;;;;;N;;;;;
1xx0C;BABYLONIAN ACCENT RIMYA;Mn;230;NSM;;;;;N;;;;;
1xx0D;BABYLONIAN ACCENT SIHPA;Mn;230;NSM;;;;;N;;;;;
1xx0E;BABYLONIAN ACCENT SHIDYA;Mn;230;NSM;;;;;N;;;;;
1xx0F;BABYLONIAN ACCENT NIHLA;Mn;230;NSM;;;;;N;;;;;
1xx10;BABYLONIAN ACCENT YETIV;Mn;230;NSM;;;;;N;;;;;
1xx11;BABYLONIAN ACCENT HALF-TET;Mn;230;NSM;;;;;N;;;;;
1xx12;BABYLONIAN ACCENT SLANTED NUN;Mn;230;NSM;;;;;N;;;;;
1xx13;BABYLONIAN ACCENT HAZER;Mn;230;NSM;;;;;N;;;;;
1xx14;PALESTINIAN ACCENT ZAQEF;Mn;230;NSM;;;;;N;;;;;
1xx15;PALESTINIAN ACCENT REVIA;Mn;228;NSM;;;;;N;;;;;
1xx16;PALESTINIAN ACCENT PASHTA;Mn;230;NSM;;;;;N;;;;;
1xx17;PALESTINIAN ACCENT TEVIR;Mn;218;NSM;;;;;N;;;;;
1xx18;PALESTINIAN ACCENT ZARQA;Mn;224;NSM;;;;;N;;;;;
1xx19;PALESTINIAN ACCENT PAZER;Mn;230;NSM;;;;;N;;;;;
1xx1A;PALESTINIAN ACCENT TIFHA;Mn;218;NSM;;;;;N;;;;;
1xx1B;PALESTINIAN ACCENT GERESH;Mn;230;NSM;;;;;N;;;;;
1xx1C;PALESTINIAN ACCENT TELISHA GEDOLAH;Mn;232;NSM;;;;;N;;;;;
1xx1D;PALESTINIAN ACCENT TELISHA QETANNA;Mn;218;NSM;;;;;N;;;;;
1xx1E;PALESTINIAN ACCENT MUNAH;Mn;220;NSM;;;;;N;;;;;
```

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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 and Palestinian Vocalization Systems of Hebrew*

2. Requester's name: *Aleksandr Andreev, Christa Müller-Kessler, Avi Shmidman, Benjamin Outhwaite, Kim Phillips*

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

4. Submission date: *October 22, 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 + new Hebrew Extended block*

2. Number of characters in proposal: *49*

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 *C* 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"
      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 will be ready when codepoints are approved by UTC: 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   *This is a revision of a proposal we submitted in May 2025.*
There was also a proposal in 2004 that did not advance for reasons unknown to us:
https://www.unicode.org/wg2/docs/n2839.pdf

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?   *Gilad Almosnino, liaison between Standards Institute of Israel and Unicode*

        If YES, available relevant documents: *Zoom meeting*

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: *Characters are archaic and used only by scholars*

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: *Printed grammars and dictionaries of Hebrew and Jewish Babylonian*
Aramaic, scholarly editions of Biblical and liturgical texts

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: *Should be placed in Hebrew block*

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: *The characters are diacritical marks*

    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: