To: Unicode Script Encoding Working Group (SEWG)
Author: Ben Denckla
Subject: Adding "*Paseq* not *Legarmeh*" to Hebrew
Date: 6 August 2025 (12th of Av, 5785)

I propose adding a "*Paseq* not *Legarmeh*" code point to the Hebrew block, disunifying the existing code point U+05C0 (Hebrew Punctuation *Paseq*). This distinction is needed in some types of Hebrew Bible publishing.

1. Background

There is an existing code point U+05C0 with name "Hebrew Punctuation *Paseq*" and annotation "= legarmeh." It encodes two vertical bars, *paseq* and *legarmeh*. These bars are semantically distinct, though traditionally they have been graphically identical. Their usual meanings are as follows:

- *Paseq* (פסק). Calls for a slight pause between the words it separates.[1] There are about 450 *paseq* bars in the Hebrew Bible.

- *Legarmeh* (לגרמיה). Modifies the musical motif of the word that precedes it (from conjunctive to disjunctive). It modifies the motifs of the marks *munaḥ, shalshelet, mehuppak,* and *azla*.[2] There are about 1,750 *legarmeh* bars in the Hebrew Bible.

In recent decades, some Hebrew Bible publications have distinguished these two bars, with one shape dedicated to *paseq* and another shape dedicated to *legarmeh*. For example, this can be seen in Numbers 3:38 in Feldheim's Simanim Tanakh:

This close-up more clearly shows the *paseq* to be a skinny box rather than a simple bar:


---

[1] Yeivin #283.

[2] Yeivin #277.

Though only recent publications make these distinctions graphically, these distinctions have been well-established, semantically, for at least a thousand years in the theory and practice of the cantillation of the Hebrew Bible.

No manuscript distinguishes *paseq* from *legarmeh* in the way that some recent publications do, but some manuscripts distinguish *paseq* from *legarmeh* using marginal notes. Yeivin[3] observes that "Some [manuscripts] — particularly those with expanded Tiberian pointing — mark every case of the vertical stroke as *paseq* (פס, פ), or *legarmeh* (לגר, לג)."[4]

For example, in Numbers 5:22 in the "Erfort 3" manuscript,[5] there is a marginal פס note referencing a (faint) *paseq* between אָמֵן and אָמֵן: in the following phrase (*paseq* in green):

<div dir="rtl">

אָמֵן | אָמֵן׃

</div>
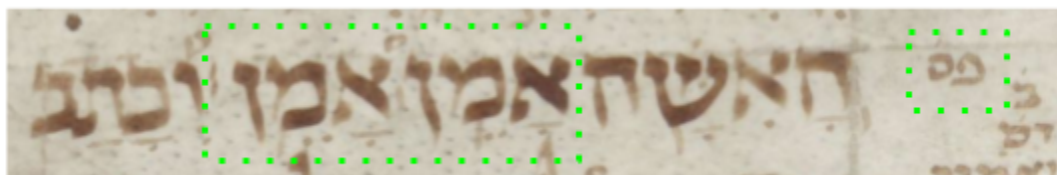
In Numbers 4:26 in that same "Erfort 3" manuscript,[6] there are two marginal לג notes referencing two *legarmeh* marks that have mostly or wholly faded away but presumably appeared after מָסַךְ and after פֶּתַח in the following phrase (*legarmeh* marks in green) (I have included וְאֶת־ but made it gray because it is on the previous line):

<div dir="rtl">

וְאֶת־מָסַךְ | פֶּתַח |

</div>

The older, more authoritative manuscripts rarely distinguish *paseq* from *legarmeh*, perhaps because they were intended for a more expert audience, for whom these distinctions were obvious in all but a few cases. As Yeivin notes,[7] "In some [manuscripts] in which [the distinction between *paseq* and *legarmeh*] is not done systematically, the [vertical] stroke is [distinguished]

---

[3] Yeivin #280.

[4] A dot above the last letter of a word indicates that it has been abbreviated, e.g. פ or פס for פסק and לג or לגר for לגרמיה. This is analogous to how, in English, we might use "*p.*" or "*pas.*" as an abbreviation for *paseq* and "*leg.*" or "*legar.*" as an abbreviation for *legarmeh*.

in cases where confusion is likely, as in the [Masorah parva] of [the Leningrad Codex (μL)], where the note לגר is given against the two cases where this accent precedes *pazer* [Daniel 3:2 and Neḥemiah 8:7], and the note פסק is given at [Isaiah] 42:5." Here are those three cases in μL (Dan. 3:2[8], Neḥ. 8:7[9], and Isa. 42:5[10]):
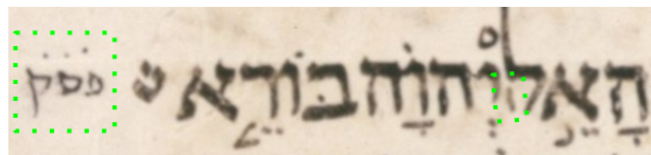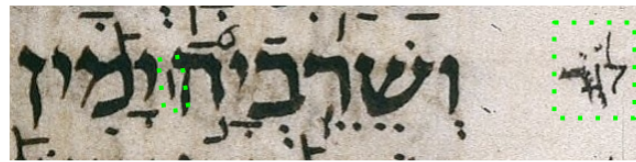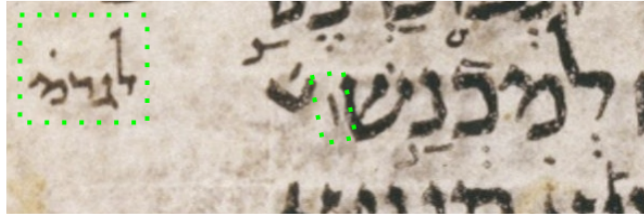
Manuscript precedent is not needed to justify modern typographic innovations. Indeed, modern innovations like *qamats qatan*, *sheva na*, and *dagesh ḥazaq* have no manuscript precedent. But we can see there is manuscript precedent for distinguishing *paseq* from *legarmeh*, although manuscripts make this distinction using marginal notes rather than differing glyphs (shapes).

---

[7] Yeivin #280.

[8] Page 439B, column 3, line 2. (The note happens to be לגרמ rather than לגר.)

[9] Page 458B, column 3, about halfway down. The לגר note is messy, with the *gimel* overlapping the *resh*, and with an unexpected dot above the *gimel*. Perhaps the scribe originally wrote only לג and then decided to supplement that with a ר. The note appears to be in a slightly different "hand" than the other Masorah parva notes on the page.

[10] Page 236A, column 3, about ⅔ of the way down, right after the blank line. It is slightly surprising that פסק appears with dots over each of its three letters, i.e. appears as פֿסֿקֿ, because with all three letters present, this is no longer an abbreviation.

2. Proposed code point

I propose a code point named "Hebrew Punctuation *Paseq* not *Legarmeh*" (abbrev. PNL). (The "not" in this name can be thought of as a shorter form of "as opposed to" or "distinct from.") This name gives the code point a meaning that avoids the problems with U+05A2 (*Atnaḥ Hafukh*). It is widely agreed that U+05A2 should have been added with a name (and attendant meaning) like "*Galgal*" or "*Yeraḥ Ben Yomo* not *Atnaḥ Hafukh*." See Peter Constable's "Attachment 1" to L2/05-259 (WG2 N2987). The principle is the following:

> **When a code point *x* is disunified by adding a code point *y*,**
>
> **it is *y* that should have the new glyph (shape), not *x*.**

This normative (prescriptive) principle applies not only to the representative glyphs used in Unicode documentation, but also to fonts that have the following goals:

- "Unified" texts, i.e. texts using only *x*, should look acceptable. I.e. the glyph for *x* should be suitable for the unified (generic) use of *x*. In our case, this means that the glyph for U+05C0 should be suitable for use as *paseq/legarmeh*.
- "Disunified" texts, i.e. texts using both *x* and *y*, should not only look acceptable, but also show a distinction between *x* and *y*. I.e.:
  - The glyph for *x* should be suitable for the disunified (specific) use of *x*. In our case, this means that the glyph for U+05C0 should be suitable for use as *legarmeh*.
  - The glyph for *y* should be suitable for its one and only use. In our case, this simply means that the glyph for PNL should be suitable for use as PNL.
  - The glyphs for *x* and *y* should be different.

Not all fonts will need to satisfy all the goals listed above. But we should disunify a code point in a way that makes it possible to satisfy all those goals. Toward that end, we chose PNL as the new code point, despite the (name-motivated) temptation to choose *Legarmeh*.[11] We chose PNL because it is likely that fonts will want to give *legarmeh* rather than *paseq* a glyph that is also suitable for (unified) *paseq/legarmeh*.

In other words, it is likely that fonts will want *legarmeh* rather than *paseq* to retain its appearance when the two meanings are distinguished. This is supported by section 7 of this document, "Examples." There, most glyphs used for *paseq* are not suitable for (unified) *paseq/legarmeh*, but all glyphs used for *legarmeh* are suitable for use as (unified) *paseq/legarmeh*.

I propose the following annotations for the existing and new code points (trying to mimic the wording of the annotation currently used for U+05B8 (*Qamats*)):

---

[11] Giving in to such name-motivated temptation is what caused the problems with U+05A2 (*Atnaḥ Hafukh*).

U+05C0 Hebrew Punctuation *Paseq*

> • used either generically (for *paseq* or *legarmeh*), or as *legarmeh* in texts that distinguish it from *paseq*
>
> → U+05CA Hebrew Punctuation *Paseq* not *Legarmeh*

U+05CA Hebrew Punctuation *Paseq* not *Legarmeh*

> • used as *paseq* in texts that distinguish it from *legarmeh*
>
> → U+05C0 Hebrew Punctuation *Paseq*

The suggestion above assumes that the new code point is assigned location U+05CA, but the exact location is of course not important. The suggestion above assumes that the annotation for U+05C0 (*Paseq*) can be updated, as seems to have been the case for U+05B8 (*Qamats*) when U+05C7 (*Qamats Qatan*) was introduced.

Proposing only one new code point implicitly proposes that the existing code point, U+05C0, would then do "double duty":

- In some texts ("unified" texts), U+05C0 would continue to represent both *paseq* and *legarmeh*, just as it currently does.

- In other texts ("disunified" texts), U+05C0 would start to represent only *legarmeh*, while the new code point would represent only *paseq*.

Despite its ambiguity, this "double duty" seems to be the preferred way to disunify code points in the Hebrew block, judging from the way various code points were added or are planned to be added: U+05C7 (*Qamats Qatan*), U+05BA (*Holam Haser* for *Vav*), U+05C8 (currently named "Heavy *Sheva*"), and U+05C9 (currently named "Heavy *Dagesh*").

3. Properties

The properties of PNL would be the same as U+05C0 (*Paseq*).

4. Collation

The collation of PNL would be the same as U+05C0 (*Paseq*).

5. References

Yeivin, Israel and Revell, E. J. *Introduction to the Tiberian Masorah*. United States: Scholars Press, 1980.

6. Acknowledgments

Thanks to Seth (Avi) Kadish for his help preparing this proposal.

7. Examples

The following 5 publishers use a *paseq* distinct from *legarmeh*:

1. Jewish Publication Society (JPS) (paper)

2. פלדהיים (Feldheim) (paper)

3. על־התורה (Al-Hatorah) (web)

4. Hebrew Wikisource (web)

5. Ben Denckla (me) (web)

Example 1 is from 1 Samuel 14:47 in an upcoming JPS book. It is a phrase having both *paseq* and *legarmeh*, in that order (here *paseq* is shorter than *legarmeh*, and positioned higher relative to the baseline):

סָבֶיב ׀ בְּכׇל־אֹיְבָיו בְּמוֹאָב ׀ וּבִבְנֵי־עַמּוֹן וּבֶאֱדֹום

For the remaining examples, we'll see how different publishers represent the *paseq* and *legarmeh* (in that order) in Numbers 3:38.

Example 2a is from Feldheim's Simanim Tanakh (already shown in the first section of this document, "Background"):

אֹהֶל־מוֹעֵד ׀ מִזְרָחָה מֹשֶׁה ׀ וְאַהֲרֹן וּבָנָיו

Example 2b is from Feldheim's Simanim Tiqqun (here the font's *paseq* and *legarmeh* are distinguished as they are in the Tanakh, although the font's letters are quite different from those of the Tanakh):

וְהַחֹנִים לִפְנֵי הַמִּשְׁכָּן קֵדְמָה לִפְנֵי אֹהֶל־מוֹעֵד ׀
מִזְרָחָה מֹשֶׁה ׀ וְאַהֲרֹן וּבָנָיו שֹׁמְרִים מִשְׁמֶרֶת

Example 3 is from the Tanakh at the core of the [Al-Hatorah Mikraot Gedolot](#) (here *paseq* is gray whereas *legarmeh* is black):

אֹהֶל־מוֹעֵד ׀ מִזְרָחָה מֹשֶׁה׀ וְאַהֲרֹן וּבָנָיו

Example 4 is from Hebrew Wikisource's Tanakh (here *paseq* is shorter and thinner than *legarmeh*, and *paseq* is gray whereas *legarmeh* is black):

אֹהֶל־מוֹעֵד׀מִזְרָחָה מֹשֶׁה׀ וְאַהֲרֹן וּבָנָיו

The use of gray in the above two examples merits some comment. If a PNL code point is added, I suggest that styling such as size and stroke weight be "pushed down" inside the font. But, despite recent color font standards,[12] I don't suggest that color be "pushed down" inside the font to get a gray PNL. I suggest that, as is normally the case, color be controlled outside of the font, in document styling.

This does not remove the need for a PNL code point, even in the case of Al-Hatorah, where color is the **only** way *paseq* is distinguished from *legarmeh*. The power of a PNL code point can be seen in a use-case like the following. Assume that Al-Hatorah starts using PNL, styling it gray, but retains the Al-Hatorah "look" by using a font with the same glyph for PNL as for the existing code point, U+05C0 (*Paseq*). Now, imagine a user copies a phrase from Al-Hatorah and pastes it **without formatting** into a document using a font with a Feldheim-style skinny box for PNL.

---

[12] Such as [OpenType-SVG](#).

After the paste, "magically," the expected effect is achieved: the PNL from Al-Hatorah now appears as a skinny box.

Example 5 is from my Tanakh called "MAM with doc" (here *paseq* is represented using U+2016 (Double Vertical Line)):

אֹֽהֶל־מוֹעֵ֑ד ‖ מִזְרָ֫חָה מֹשֶׁ֣ה| וְאַֽהֲרֹ֧ן וּבָנָ֛יו

8. Workarounds vs. a new code point

Some of the *paseq* "looks" shown above can be achieved with styling that offers control of size, stroke weight, and color. Perhaps even the skinny box "look" can be achieved with outline styling.[13] Other "looks" can be achieved by using a non-Hebrew code point, as I do with U+2016 (Double Vertical Line). These options show that the lack of a PNL code point is easier to work around than the lack of a diacritic like *sheva na* or *dagesh ḥazaq*. (It is not easy to work around the lack of a diacritic because in most environments it is hard or impossible to style a diacritic or use a non-Hebrew code point as a diacritic.)

I urge the SEWG to recommend to the UTC that PNL be added, despite the existence of these workarounds. They do not remove the need for a new code point. Indeed, workarounds, though painful, even exist for the new diacritics that were recently recommended to the UTC and provisionally assigned: *sheva na* ("Heavy *Sheva*") and *dagesh ḥazaq* ("Heavy *Dagesh*"). Those workarounds include briefly switching into a different font every time there is a grapheme cluster that needs to include a new diacritic.

The existence of a workaround, even an easy one, should be only a minor factor in deciding whether to add a code point. The existence of a workaround is relevant only to the narrow role of Unicode as a mechanism to achieve graphical results by any means. This "by any means" attitude is typical in (and appropriate to) paper-only publishing. But much paper-only publishing didn't benefit that much from the advent of Unicode. It got along quite well before Unicode, in the old days of code pages and font-specific encodings.

So, workarounds remove the need for a new code point only in the kind of publishing that doesn't really need Unicode to begin with.

We no longer live in a world where paper is king. This new world, where paper has been "dethroned," is where Unicode shines. A notable example is that if a user can copy and paste text, it matters what means were used to encode that text. See the copy-and-paste use-case I

---

[13] E.g. via -webkit-text-stroke in CSS.

describe in section 7, "Examples." That use-case is not specific to Al-Hatorah, though I happened to describe it to make a point about Al-Hatorah.

On the web, and indeed almost everywhere except on paper, Unicode has an important role as a mechanism to **encode not only graphics but semantics**. These two roles, graphics and semantics, are sometimes in tension, but not always. For example, semantic encoding can make it easier to achieve a variety of graphical results through one uniform mechanism: font selection.[14] Contrast this with achieving graphical results through a hodge-podge of ad hoc workarounds, some of which are environment-specific, e.g. specific to Adobe InDesign and therefore only relevant to paper publishing.

In summary, the SEWG should recommend disunification if someone from a user community (often associated with a particular culture) can make the following case:

- A semantic distinction exists between two marks.
- This distinction is reflected graphically in some publications, though not necessarily in a single, standard way.

If a good such case is made, the SEWG should recommend disunification, almost irrespective of the existence of workarounds or their ease of use. I hope I have made such a case here.

9. ISO Proposal Summary

(The ISO proposal summary forms will appear here but the Template for Character Additions advises that these need not appear here in this preliminary version of the proposal.)

---

[14] Or selection from among a font's stylistic sets.

10. Supporters

Prof. Avi Shmidman
Consultant member of [The Academy of the Hebrew Language](#)

Rabbi David E. S. Stein
Independent Hebrew Editing and Publishing Professional

Dr. Joshua R. Jacobson
Author of *Chanting the Hebrew Bible: The Art of Cantillation*

Scott-Martin Kosofsky
[The Philidor Company](#)

Raphaël Freeman, MISTD
[Renana Typesetting](#)

Seth (Avi) Kadish
Creator and maintainer of [MAM](#) (*Miqra* According to the *Masorah*)

Hillel Novetsky
Publisher of [Al-Hatorah](#)

Maxim Iorsh
Maintainer of the [Culmus Project](#) for Hebrew fonts

(End of document.)