

To: Unicode Script Encoding Working Group
 Author: Ben Denckla
 Subject: Re-documenting ZARQA and ZINOR
 Date: 16 May 2025

This document suggests edits to the documentation of the following two code points:

- U+0598 HEBREW ACCENT ZARQA
- U+05AE HEBREW ACCENT ZINOR

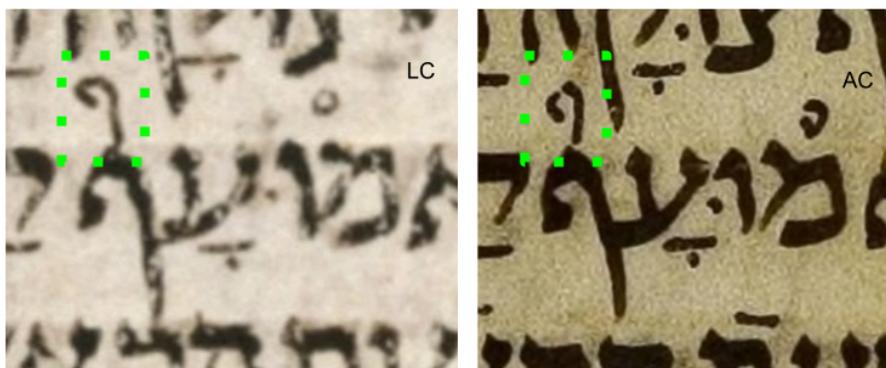
This document suggests edits to the following three areas of documentation:

- Annotations in the names list
- [Unicode Technical Note #27](#): Known Anomalies in Unicode Character Names
- The Unicode *Core Spec* [Chapter 9 section 1.1](#), “Hebrew: U+0590–U+05FF”

Background

There are two cantillation systems in the Hebrew Bible. The poetic system is used for Psalms, Proverbs, and most of Job. The prose system is used everywhere else. The prose accent *zarqa* is placed above the last letter of its word. Horizontally, it is either centered on its letter or placed to the left of its letter’s center. The poetic accent *tsinor* has the same shape and placement as *zarqa*.

Below are examples of *zarqa* in Isaiah 8:23 in the two most important Hebrew Bible manuscripts. (*Zarqa* appears inside a dotted green rectangle.)



We see that in the manuscripts, *zarqa* has a shape somewhat like a shepherd’s crook. Or, if you like, it has a shape somewhat like a final *pe* (פ), which happens to be the letter it appears over in the examples above. For *zarqa*, fonts sometimes use a shape like that used in the manuscripts,

but more often, fonts use a shape somewhat like a tilde (~).¹ Here are examples of both, in the plain Ezra SIL font and its SR variant:

מוֹעֵף מוֹעֵף

I will let these examples of *zarqa* serve for *tsinor* as well because these marks look the same in any given manuscript or printed edition.

The third and final accent we will discuss is the poetic accent *tsinorit*. It has a shape similar to or the same as *zarqa/tsinor*. Like *zarqa/tsinor*, it is placed above its letter. But unlike *zarqa/tsinor*, its letter is rarely if ever the last of its word. Also unlike *zarqa/tsinor*, it is never placed to the left of its letter's center. As examples, here are the seven (!) cases of *tsinorit* in Psalm 68 in the Aleppo Codex (AC):



We see that in the AC, *tsinorit* looks like a *zarqa/tsinor* twisted about 30 degrees clockwise and placed above-center or above-right. For comparison, here are those same seven words shown in the Ezra SIL font:

¹ Or, if you like, the shape is somewhat like the mathematical symbol “inverted lazy ‘S’” (∞).

מִזְשִׁיב בְּפֶרֶשׁ מְלָכִים
 עֲלִית שְׁבִית הָאֵל
 הַלִּיכֹת

As in all printed editions, in Ezra SIL, *tsinorit* is centered on its letter. I.e. no printed editions have it above-right as it sometimes is in manuscripts.

We now return to *zarqa* to discuss a complexity not mentioned yet. Sometimes, in addition to having a *zarqa* on its last letter, a word also has a *zarqa*-shaped mark earlier. For example, here is a word from 2 Chronicles 19:2 in the Leningrad Codex (LC) and rendered in the Ezra SIL font:



This earlier mark shows which syllable has the main stress, if that stress is not final. We call this earlier mark a **zarqa stress helper**. We must admit that this term is our invention: I am not aware of any standard term for this, either in English or Hebrew.

In the manuscripts, *zarqa* stress helper is optional and that option is taken very rarely. For example, in the LC, it appears only twice.² In contrast, in many printed editions *zarqa* stress helper is not optional: it is used in all 180 or so *zarqa* words that have non-final stress.

A stress helper for *tsinor* is a perfectly valid idea, but it is rarely if ever used.³ It could be used in about 55 words.

Having established this background, we are now ready to make our suggestions.

² *Zarqa* stress helper appears in the LC only in 2 Samuel 3:8 and 2 Chronicles 19:2.

³ This is likely due to the lack of poetic chant in most Jewish ritual today.

Suggested edits to annotations in the names list

Here are my suggested edits (gray indicates no change):

U+0598 HEBREW ACCENT ZARQA

current	suggested
= tsinorit, zinorit; tsinor, zinor	= tsinorit, zinorit
<ul style="list-style-type: none"> • This character is to be used when Zarqa or Tsinor are placed above, and also for Tsinorit. 	<ul style="list-style-type: none"> • Name should have been TSINORIT; 05AE's preferred stress helper is itself, not this.
→ (hebrew accent zinor - 05AE)	

U+05AE HEBREW ACCENT ZINOR

current	suggested
= tsinor; zarqa	
<ul style="list-style-type: none"> • This character is to be used when Zarqa or Tsinor are placed above left. 	<ul style="list-style-type: none"> • Name should have been spelled TSINOR; its preferred stress helper is itself, not 0598.
→ (hebrew accent zarqa - 0598)	

These edits do the following:

1. They mention the naming problems covered in TN 27.
2. They remove graphical prescriptions.
3. They promote 05AE as its own stress helper (closely related to #2).

The current annotations define these two code points graphically, based on their placement relative to their base letter. My suggested edits define these two code points semantically rather than graphically. The main implications of a semantic definition are as follows:

- *Tsinorit* is only encoded with U+0598, and it is the only thing U+0598 encodes.
- Regardless of the user’s desires for how they should be placed, all four of the following are only encoded with U+05AE, and they are the only things encoded by U+05AE:
 - *Zarqa*
 - *Tsinor*
 - *Zarqa* stress helper
 - *Tsinor* stress helper (valid, but rarely if ever used)

The user’s desires for how those marks should be placed should be satisfied not by choice of code point but by doing one of the following:

- Choose a font that has, by default, the behavior you want.
- Choose a font that has, via a stylistic set, the behavior you want.
- Override the font’s behavior in desktop publishing software.

Pashta and *qadma* are, for these purposes, analogous to *zarqa/tsinor* and *tsinorit*. The *pashta* stress helper’s shape and placement often looks just like *qadma*, just as the *zarqa* stress helper’s shape and placement often looks just like *tsinorit*.

But, Unicode wisely defines *pashta* and *qadma* semantically rather than graphically. It merely gives them semantic names and makes no graphic prescriptions in their annotation. Unicode does not prescribe that U+05A8 is to be used not only for *qadma* but also for above-center *pashta*. Nor does Unicode prescribe that U+0599 is to be used only for above-left *pashta*. Unicode should stop making graphic prescriptions about *zarqa/tsinor* and *tsinorit* and instead treat them as it does *pashta* and *qadma*.

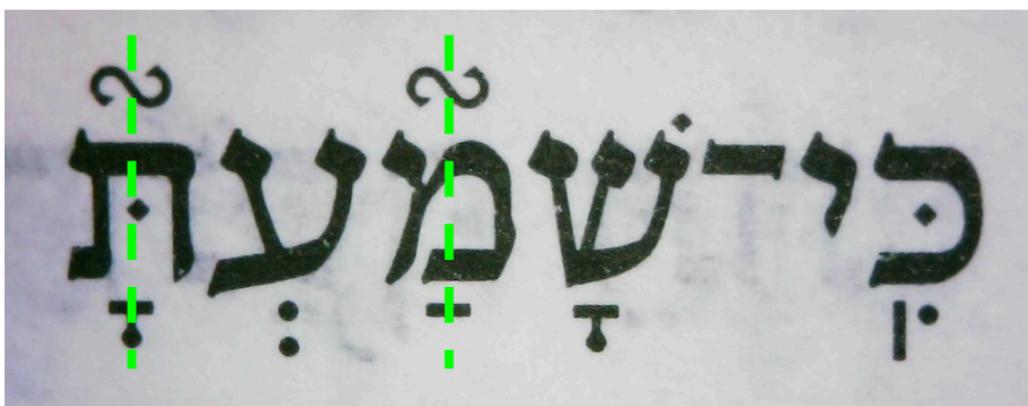
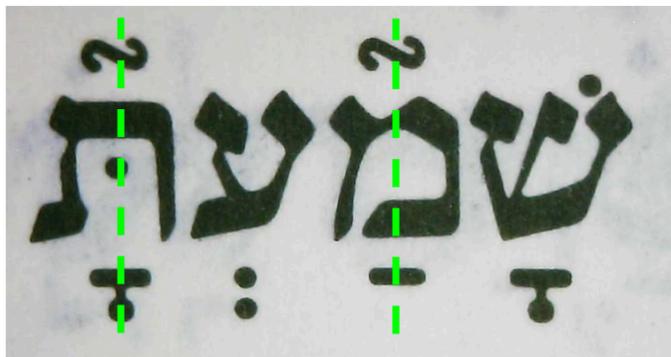
The pitfalls of defining these two code points graphically are the usual ones:

- It prescribes a particular font implementation.
- It complicates applications that need semantic information.

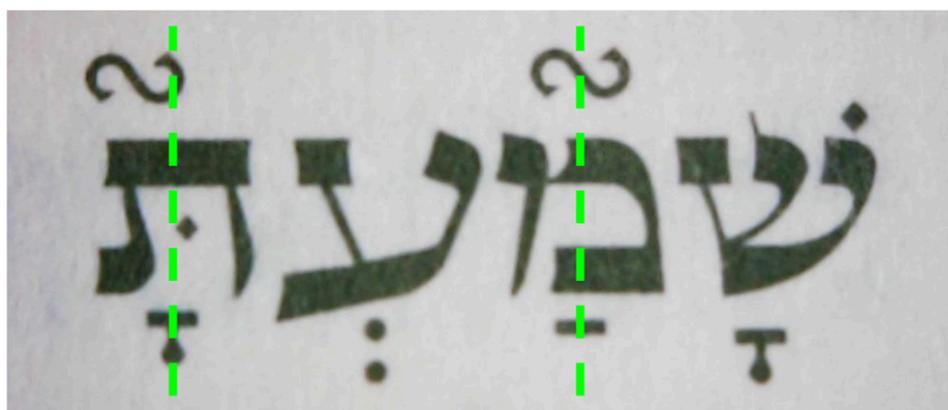
The graphical definition of these code points implies a particular placement of their graphemes relative to their base letters. As such, it seems to assume that the following linkage between semantics and graphics is universal (SH = stress helper):

semantics	graphics
<i>zarqa, tsinor</i>	above-left
<i>zarqa</i> SH, <i>tsinor</i> SH, <i>tsinorit</i>	above-center

That particular linkage is popular, but not universal. In particular, some publications use above-center placement for **both** *zarqa* and its stress helper. For example here is a word from Genesis 3:17 in a Koren *Tanakh* and in Feldheim’s Simanim *Tanakh*:



I have added dashed green lines to more clearly show the alignments of *zarqa* and its stress helper, relative to their base letters (*tav* and *mem* respectively) and, perhaps even more usefully, relative to their vowel marks (*qamats* and *pataḥ* respectively). For reference, here is the same word in a Simanim *Tiqqun* showing a *zarqa* that is above-left:



But, perhaps the definition of these code points is solely graphical, implying no linkage to semantics. In that case we could perhaps think of the code points as follows:

- U+0598 above-center tilde
- U+05AE above-left tilde

Accordingly, in such a solely graphical interpretation, we'd code each of the first two examples using two center tilde marks and the third example using a center tilde and a left tilde. This does not seem like Unicode, as the encoding would vary depending on typographic taste.

The graphical definition of these code points not only specifies the placement of their graphemes; it also implicitly specifies that they should have matching shapes. While matching shapes have always been used in printing, Unicode should not prescribe matching shapes, for the following reasons:

- Some manuscripts use different shapes for these marks, and therefore future publications may want to respect these manuscript differences.
- A general trend in Jewish Bible and prayerbook publishing has been to distinguish marks that have, historically, been indistinct. Distinctions have emerged in *qamats*, *sheva*, and *dagesh*. Similarly, a distinction should be allowed to emerge with respect to *tsinorit*, especially since there is manuscript precedent for this distinction, and a code point already present!
- Although Unicode documentation mentions that *atnaḥ hafukh* and *yeraḥ ben yomo* are often printed with the same shape, it by no means prescribes that fonts mimic this.
- Although *pashta* and *qadma* almost always have the same shape,⁴ Unicode by no means prescribes that they must have the same shape. Unicode is similarly non-prescriptive towards other pairs of accents whose shapes are often indistinct. These pairs are *mahapakh-yetiv* and *tipeḥa-dehi*.

⁴ The Simanim *Tiqqun* is a notable exception here, having distinct shapes for *pashta* and *qadma*.

Suggested edits to the body of TN 27

current	<p>U+0598 HEBREW ACCENT ZARQA: Perhaps should have been called <i>Hebrew accent tsinnorit</i>. May also be used for <i>zarqa</i> when shown on [an] accented non-final letter. See Appendix A.</p> <p>U+05AE HEBREW ACCENT ZINOR: Should have been called <i>Hebrew accent zarqa</i> (= <i>tsinor</i>). See Appendix A.</p>
suggested	<p>U+0598 HEBREW ACCENT ZARQA: Should have been called <i>Hebrew accent tsinorit</i>. U+05AE's preferred stress helper is itself, not this. See Appendix A.</p> <p>U+05AE HEBREW ACCENT ZINOR: Should have been spelled <i>tsinor</i>, or perhaps should have been called <i>Hebrew accent zarqa</i>. Its preferred stress helper is itself, not U+0598. See Appendix A.</p>

TN 27 currently says that U+0598 “May also be used for *zarqa* when shown on [an] accented non-final letter.” In other words it says that U+0598 may be used as a stress helper for U+05AE. Nothing explicit is said one way or another as to whether, instead, U+05AE may be used as a stress helper for itself. Presumably it can.

Above I suggest edits in which TN 27 is revised to do the following:

- Explicitly deprecate the use of U+0598 as a stress helper for U+05AE.
- Explicitly promote the use of U+05AE as a stress helper for itself.

For more details, see the section above containing my suggested edits to annotations in the names list.

Also, above I suggest edits that slightly change the language describing the errors in the names of these code points.

Suggested edits to Appendix A of TN 27

Appendix A is entitled “Notes on Zarqa and Zinor” and was written by Mark Shoulson. Here is the current version:

There are two separate cantillation systems in the Hebrew Bible. One is used for Psalms, Proverbs and (most of) Job, (the "poetic" books, hence the "poetic system"), and the other is used everywhere else. The two systems have structural similarities and share some graphemes, but not all. In modern printing the accents have roughly the same shape; old

manuscripts actually had them written slightly differently. In the prose system there is an accent called ZARQA, which is postposed (on or to the left of the last letter), and in the poetic system there is one called TSINOR (and also *zarqa* and vice-versa; each of these has many names) which has the same shape and placement and even an analogous function in the structure of the cantillations. There is another accent, only in the poetic system, called the TSINNORIT (a diminutive of *tsinor*), which occurs directly above its letter, and is (almost?) never on the last letter of its word. (More modern printing tends to put the *zarqa* right on top of its letter too, but that's just a printing preference). If you look closely at some old manuscripts, you can tell that *tsinnorit* has a slightly different shape than *zarqa/tsinor*.

As encoded in Unicode, there are ZARQA (U+0598) and ZINOR (U+05AE) [sic]. By the usual meanings of those names, those should properly be synonyms, the same accent, but they're not. While the word "zinor" would be mnemonic of "tsinnorit," it's the wrong way around in the character names: ZINOR has the combining class of above-postposed, and ZARQA is encoded to go directly above the letter. So, to encode a *zarqa* or a *tsinor*, you need to use ZINOR, and to encode a *tsinnorit*, you need to use ZARQA.

Here is my suggested version:

There are two cantillation systems in the Hebrew Bible. One system, the poetic system, is used for Psalms, Proverbs, and most of Job. The other system, the prose system, is used everywhere else. The two systems have structural similarities and share some but not all graphemes. The prose system has an accent called *zarqa*, which is postposed (on or to the left of the last letter). The poetic system has an accent called *tsinor*, which has the same shape and placement as *zarqa*, and even an analogous function in the structure of the cantillations. (*Tsinor* is sometimes called *zarqa* and vice-versa; each of these has many names).

There is a third accent, only in the poetic system, called *tsinnorit* (a diminutive of *tsinor*), which is always printed directly above its letter. Its letter is rarely if ever the last of its word. (More modern printing tends to place the *zarqa/tsinor* directly above its letter too, but that's just a printing preference). If you look closely at some manuscripts, you can tell that *tsinnorit* has a slightly different shape than *zarqa/tsinor*.

Unicode encodes all three of these accents using only two code points: U+0598 and U+05AE. Unfortunately, these code points are named ZARQA and ZINOR respectively. This is unfortunate because by the usual meanings of those names, those accents are, among the three accents, the two that do not need to be distinguished.

Fortunately, Unicode has extra information about ZINOR that allows us to infer that ZINOR is intended to encode *zarqa/tsinor*. This in turn allows us to infer that ZARQA is intended to encode *tsinnorit*. This extra information comes in the form of combining class 228,

prescribing ZINOR to have an above-left placement relative to its letter. While this is too prescriptive, at least it allows us to infer that ZINOR encodes *zarqa/tsinor*. We can infer this because while *zarqa/tsinor* need not have above-left placement, *tsinorit* **cannot** have above-left placement.

So, ZINOR encodes *zarqa/tsinor*, and ZARQA encodes *tsinorit*.

Accordingly, ZINOR rather than ZARQA should be used as the stress helper for ZINOR, just as PASHTA rather than QADMA is used as the stress helper for PASHTA, and just as other prepositives and postpositives are used as the stress helpers for themselves: TELISHA GEDOLA, TELISHA QETANA, and ACCENT SEGOL.

Though it is not appropriate to get into it in TN 27, I can't resist making a final note here regarding the use of combining class 228 for ZINOR. No other Hebrew accent, including, notably, none of the other postpositives, have this above-left combining class (228). The use of combining class 228 for ZINOR is yet another error among many errors surrounding the ZINOR and ZARQA code points.

Suggested edits to *Core Spec* Chapter 9

In Section 1.1 (the section on Hebrew), I suggest that some reference be made to the Hebrew-relevant parts of TN 27. Or, if it is undesirable to make such a reference, I suggest that some or all of the Hebrew-relevant parts of TN 27 be replicated in Section 1.1.

Standards history

The problems with (and confusion surrounding) ZARQA and ZINOR pre-date Unicode's support for them. Like most of the accents, ZARQA and ZINOR entered Unicode in version 2.0 (July 1996), via a fairly literal adoption of SI 1489 (1992).⁵ Already in SI 1489 we see the problems that Unicode inherited. The problems in Table 9 of SI 1489 are summarized in the following table:

⁵ SI = Israeli Standard (issued by SII (מת"י), the Standards Institution of Israel (מכון התקנים הישראלי)).

	listed with POA (poetic-only accents)?	position code (קוד מיקום)
ZARQA	No, not with POA, as if it were <i>zarqa/tsinor</i>	above[-center] (ע), as if it were <i>tsinorit</i>
ZINOR	Yes, with POA, as if it were <i>tsinorit</i>	above-left (ש/ע), ⁶ as if it were <i>zarqa/tsinor</i>

Here are some images⁷ from a scan of SI 1489, giving some historical flavor, showing the parts of Table 9 in which ZARQA and ZINOR are listed:

קוד מיקום	שם מקוצר	ייצוג גרפי	שם אנגלי	שם עברי
--------------	----------	---------------	----------	---------

[...]

ח	Tipeha	◌	HEBREW ACCENT TIPEHA	טעם עברי טפחה
ע	Revia	◌	HEBREW ACCENT REVIA	טעם עברי רביע
ע	Zarqa	◌	HEBREW ACCENT ZARQA	טעם עברי זרקא

[...]

ע	Iluy	◌	HEBREW ACCENT ILUY	טעם עברי עילוי
י/ח	Dehi	◌	HEBREW ACCENT DEHI	טעם עברי דחי
ש/ע	Zinor	◌	HEBREW ACCENT ZINOR	טעם עברי צינור

⁶ Too prescriptive, but informative.

⁷ I apologize that these images are of such low resolution but this is what SII provides.

The problems of SI 1489 are not limited to its Table 9. Here is another table from SI 1489, this one showing supposed equivalents⁸ between poetic and prose accents:

טעמי מקרא לפי טבלה 9	טעמי אמ"ח
טפחה	טרחה
זרקא	צינורית
קדמה	אזלא
ירח בן-י ומו	גלגל
מרכה	יורד

The names of accents vary across time and place, but these equivalences are fine except for the *zarqa/tsinorit* one.⁹ This table shows explicitly what Table 9's position codes ע and ש/ע showed only implicitly: that SI 1489 mistook *tsinorit* to be the equivalent of *zarqa*.

It is perhaps of interest to note these two related errors in SI 1489:

Error 1. In Table 9, the accent *segol* is given the above-center position code (ע), an error of no redeeming value, in contrast to the analogous error for ZARQA, which was fortuitous since it allowed ZARQA to be used for *tsinorit*.

Error 2. In Table 9, the poetic-only accent *geresh muqdam*¹⁰ is listed with the accents that are **not** poetic-only. Perhaps the temptation to list all types of *geresh* together was too great (*geresh* proper, *geresh muqdam*, and *gershayim*). SI 1489 reinforces the absence of *geresh muqdam* from the poetic-only accents when it says:

עולה, עילוי, דחי וצינור ייחודיים לספרי אמ"ח.
Oleh, iluy, dehi and *tsinor* are unique to the *Sifrei Emet*.¹¹

⁸ SI 1489 proposes these as equivalents only for the (superficial) purposes of sharing an encoding. I.e. no claim is made that they are equivalent any deeper, semantic sense.

⁹ The spelling מרכה rather than מרצה is nonstandard and is amended to מרצה in 1996 in SI 1311.2 (1311 part 2, also sometimes written 1311-2).

¹⁰ There are five cases in *Tanakh* in which either *geresh* or *gershayim* are combined with *telisha gedolah*. In these cases, the prepositive nature of the *telisha gedolah* seems to "rub off" on the *geresh* or *gershayim*. In the two cases that involve *geresh*, I suppose it is possible to interpret that *geresh* to be a *geresh muqdam*. I have never seen such an interpretation, and I consider such an interpretation to be far-fetched, but I thought it would be worth mentioning the possibility, if only to dismiss it.

¹¹ The *Sifrei Emet* are the poetic books of *Tanakh*. Note that this statement not only neglects *geresh muqdam* but also includes *tsinor* rather than *tsinorit*, a far more serious error, and one that is at the root of all the past and present confusion regarding Unicode ZARQA and ZINOR.

Yet, in the section on sorting order (collation), SI 1489 is at pains to specify that:

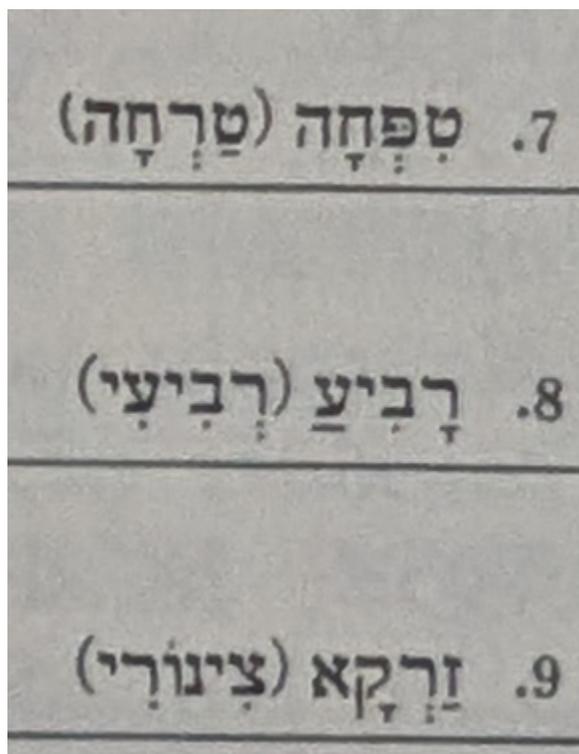
גרש מוקדם (טעמי אמ"ת) ימוין לפני טפחה.

Geresh muqdam ([one of the] *Emet* accents)¹² should be sorted before *tipeḥa*.

Some of these problems in SI 1489 may have come from the limitations of its main source for the names of the accents: האנציקלופדיה העברית.¹³ The relevant entry in that encyclopedia does not cover the poetic system in as much detail as the prose system. Notably, its table of accents has the following limitations:

- It lists *zarqa/tsinor* but not *tsinorit* or any other poetic-only accent.
- It lists *zarqa/tsinor* as זַרְקָא (צִינוּרִי), i.e. as “*zarqa (tsinori)*”. *Tsinori* is a name I have never seen. It may simply be a typo. It sits in a confusing middle ground between the standard names *tsinor* and *tsinorit*, although much closer to *tsinorit*.

For some historic flavor, here is an image of the part of that table leading up to *zarqa/tsinor[i]*:



¹² I.e. one of the poetic accents.

¹³ Sometimes called the *Encyclopedia Hebraica*. The entry on accents is in volume 18, on pages (columns) 866–872. SI 1489 mentions its use of this entry as follows:

לצורך תקן זה נלקחו השמות מהאנציקלופדיה העברית, כרך י"ח, עמוד 866, ערך "טעמי המקרא".

For the purposes of this standard, the names [of the accents] were taken from the האנציקלופדיה העברית, volume 18, page 866, entry "טעמי המקרא".