

Masoretic Graphemics and Graphetics: Challenges and Solutions

Yannis Haralambous

Abstract. Written Hebrew uses two systems of diacritics: vowel marks (*niqqud*) and cantillation marks (*te'amim*). In the Hebrew Bible these marks carry syntactic, accentual, and musical information, and a single accent can decide the meaning of a verse, as the long-disputed reading of Isaiah 40:3 shows. Setting them in type is hard: dozens of sub- and supralinear marks must combine on one line while the base letters keep their normal spacing, each mark stays centered on its letter's vertical axes, and nothing collides with descenders or ascenders. In this paper we first describe the Masoretic writing system as a stack of graphemic layers, define the notion of a graphemic cluster and the graphotactic rules that govern it, and locate the *otiyot meshunot* ("altered letters") in a twilight zone between graphemes and allographs. Tracing one common word across twenty-six printed Bibles from 1517 to 2021, we show that even the finest typography has only partially solved the alignment problem. We then present Tiqwah v2, a corpus-driven system that generates an optimal Masoretic font from a given corpus: diacritic positions are computed by linear-constraint minimization, and the glyph set is kept minimal by a graph-coloring procedure that merges clusters indistinguishable below the perceptual threshold. The method is demonstrated on three fonts of very different character—SBL Hebrew, Monotype Peninim, and the deliberately informal Yarden.

Written Hebrew employs two types of diacritic marks: vowel (or absence of vowel) marks, called *niqqud*, and cantillation marks, called *te'amim*. The former are still used in Modern Hebrew for foreign words or disambiguation, in a way similar to Arabic *tashkeel*, while the latter are specific to the text of the Hebrew Bible (as well as a 19th-century translation of the New Testament into Hebrew) and their use is taught only in religious schools.

Yannis Haralambous  0000-0003-1443-6115

Data Science Department, IMT Atlantique & UMR CNRS 6285 Lab-STICC, Technopôle Brest-Iroise, CS 83818, 29238 Brest Cedex 3, France
E-mail: yannis.haralambous@imt-atlantique.fr

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2024. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 11.
Fluxus Editions, Brest, 2026, pp. 297–326. <https://doi.org/10.36824/2024-graf-hara>
ISBN: 978-2-487055-10-0, e-ISBN: 978-2-487055-11-7

1. Introduction

1.1. History of the Masorah

The word *masorah* means “tradition” or “transmission” in Hebrew. The term *Masoretic Text* refers to the corpus of 24 books of the Hebrew Bible as annotated and standardized by the Masoretes, Jewish scribes and scholars active between the 6th and 10th centuries CE. Historically, three competing notation systems have emerged: the Babylonian and Palestinian schools, which used supralinear marks, and the Tiberian system, which was mostly sublinear and combined with a refined vocalization. The Tiberian system eventually prevailed, and the Tiberian Masoretes, especially the Ben Asher and Ben Naphtali families in the 9th and 10th centuries, fixed both vowels and accents on the inherited consonantal text. The Aleppo Codex (c. 930), pointed¹ by Aaron ben Asher, became the touchstone manuscript, and the Leningrad Codex B19a (1008/1009) (see Fig. 1) remains the oldest complete Tiberian witness.

It is widely accepted (Futato, 2020; Park, 2020; Wickes, 1881; 1887; Хаздан, 2020) that cantillation marks serve three primary functions: they segment the text (analogous to punctuation) (Wu and Lowery, 2006), they indicate phonetic stress (Hedánek, 2017), and they prescribe melody (according to various musical traditions) (Kelley, Mynatt, and Crawford, 1998; Levine, 1989). Nevertheless, some scholars, notably Hoop (2008), have contested this view, arguing that cantillation marks constitute predominantly a system of musical notation influenced by other Mediterranean musical notation systems of the period. Mitchell (2017) draws parallels between these marks and the Greek cantillation marks found in a third-century BCE papyrus of a Euripidean tragedy, advancing sixteen arguments for why they could not have been invented by the Masoretes but must rather have been transmitted to them from sources potentially predating the Masoretic period by more than ten centuries.

1.2. Importance of the Masoretic Layer

The following example is paradigmatic of the importance of *te‘amim* as syntactic markers (Futato, 2020, pp. 84–85). In Isa 40:3, there is verse:

1. By “pointed” we mean that the consonantal text was written by a scribe (in this case Shlomo ben Buya’a, around 930 CE), and Ben Asher then added the vowels and cantillation marks on top of that already-written consonantal layer. It was a distinct, highly specialized task—the *naqdan* (pointer) was a master of the oral reading tradition, while the *sofer* (scribe) was responsible for the letters themselves. The division of labor reflects the theological status of the consonantal text, which was fixed and sacred, versus the vocalization, which was understood as a learned tradition laid over it.

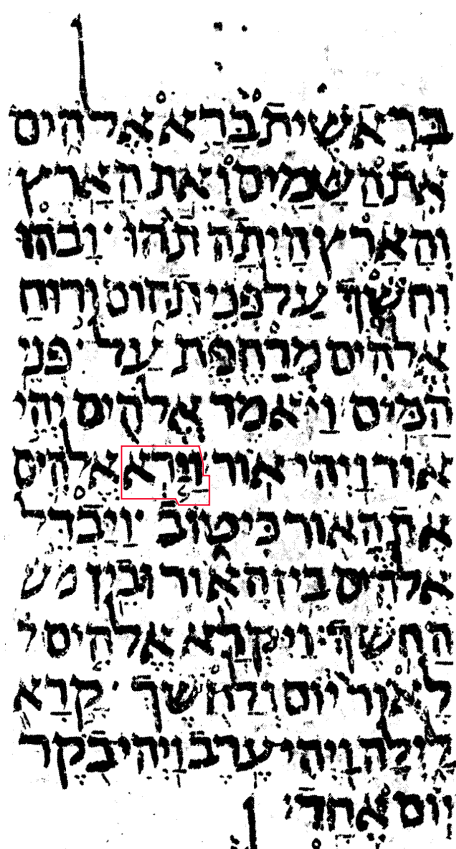


FIG. 1. The first five verses of Genesis from the Leningrad Codex B19a (1008/1009). The word marked is the one used in Fig. 7 to compare 26 printed Bibles from 1517 to 2026.

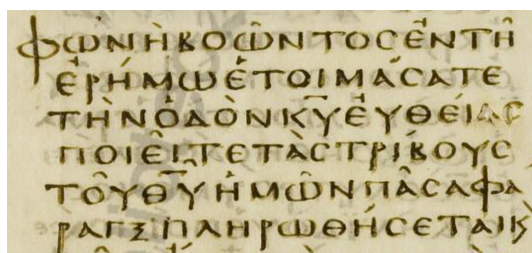
zaqef gadol → וְ
zaqef katan → וְ

קוֹל קוֹרֵא בַּמִּדְבָּר פָּנּוּ דֶרֶךְ יְהוָה

that can be roughly translated as (we purposely do not include punctuation):

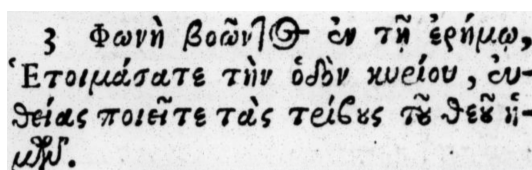
a voice is calling in the wilderness prepare the way for the Lord

This has been translated into Greek and appears in the Septuagint Bible (3rd c. BCE) as shown in this excerpt of the *Codex Vaticanus Vat.gr. 1209* (300-350 CE):



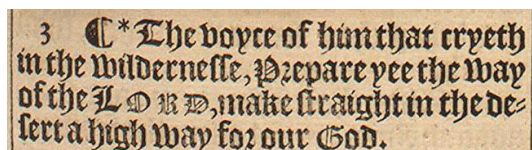
φωνή βοῶντος ἐν τῇ ἐρήμῳ ἐτοιμάσατε τὴν ὁδὸν κυ...

Following 4th-century CE conventions, there are no interword spaces or punctuation. In modern times, punctuation has been added, and we find this passage, in Roger Daniels' *Old Testament* (London, 1653), written as



Φωνὴ βοῶντος ἐν τῇ ἐρήμῳ, Ἐτοιμάσατε τὴν ὁδὸν κυρίου...

where both the comma and the capital epsilon indicate that “in the wilderness” (ἐν τῇ ἐρήμῳ) is syntactically bound to “the voice” (φωνή) rather than to “prepare the way for the Lord” (ἐτοιμάσατε τὴν ὁδὸν κυρίου...). Similarly, in the 1611 King James Bible:



“The voyce of him that cryeth in the wilderness,
Prepare yee the way of the LORD...”

there is a comma after “wilderness,” and “Prepare” is capitalized, as in the Greek text, so that the interpretation is the same as in Roger Daniels.

We could stop the interpretation at this point, if it weren't for the *zaqef katan* cantillation mark on the second word of the Hebrew text, which denotes a primary syntactic division after “calling,” and therefore binds “in the wilderness” not to “the voice” but to “prepare,” leading to the interpretation

A voice rings out: Clear in the desert a road for the LORD,

as it appears in the NJPS Bible edition of the Jewish Publication Society of America (Philadelphia, 1985), which invalidates the “voice calling in the wilderness,” recalled as a prophecy in the New Testament, where

Saint John the Baptist is presented as the personification of this voice (in all four gospels: Matthew 3:3, Mark 1:3, Luke 3:4, and John 1:23). Futato (2020) suggests that this is an error in the Masoretic text, while Hoop (2008, pp. 113–115) is more nuanced.

This example shows that the presence of a specific Masoretic mark suffices to give an entirely different interpretation to a verse that is immensely popular since it is the origin of the idiomatic expression “a voice in the wilderness”.

1.3. The Printed Masora

When Hebrew printing began in late 15th-century Italy, the *niqqud* and *te'amim* posed a serious technical challenge: dozens of marks had to combine cleanly with consonants and vowel points on a single line of metal type. The Soncino family produced the first complete printed Hebrew Bible in 1488, but the decisive achievement was Daniel Bomberg's Rabbinic Bible (*Miqra'ot Gedolot*) in Venice, especially the 1524–25 edition edited by Jacob ben Hayyim, whose pointed and accented text became the printed textus receptus for nearly four centuries. The great Christian polyglots—Complutensian, Antwerp, Paris, London—carried this typography into European scholarship. In the 19th and 20th centuries, critical editions refined the marks: Baer-Delitzsch, Ginsburg, then Kittel's *Biblia Hebraica* (1906; its 3rd edition of 1937 was the first to follow the Leningrad Codex). The *Stuttgartensia* (BHS, 1967–77) and the ongoing *Biblia Hebraica Quinta* (since 2004) reproduce the Leningrad accents diplomatically, while the *Hebrew University Bible* follows the Aleppo Codex.

1.4. The New Testament with Cantillation Marks

Besides the Tanakh, the New Testament has also been translated into Hebrew with Masoretic signs prepared by an English scholar, Ezekiel Margoliouth, and published by the London Society for Promoting Christianity Amongst the Jews, in 1865. The London Society had been producing pointed Hebrew New Testaments since 1821, and a committee revised that text between 1836 and 1838. Margoliouth's was the third revision, and its distinctive contribution was the addition of the full Tiberian apparatus of *niqqud* and *te'amim*, with all the editorial decisions that implies—chiefly, how to map syntactic and rhetorical structures of Koine Greek (parallelism, direct speech, parenthetical clauses) onto the disjunctive/conjunctive accent system designed for biblical Hebrew prose (McClintock and Strong, 1870).

The project's purpose was twofold—missionary outreach to Jewish readers, and a deliberate “naturalizing” of the Greek NT into the lin-

guistic register of the Tanakh. Cantillating the text was the typographic climax of that program: it gave the New Testament the visual and aural signature of Scripture.

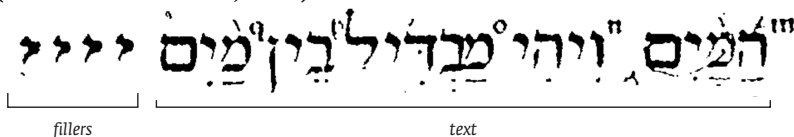
2. Layers of the Masoretic Writing System



FIG. 2. The phrase “[I give] unto him My covenant of peace” (Num 25:12) with colors for the following layers: (1) in black (grey for punctuation), (1') in deep green (broken *vav*), (2) in orange, (3) in violet, (4) in blue, (5) in light green, and (6) in red.

We can define *layers* of elements of a writing system (be they graphemes or non-linguistic) in the following way: two elements belong to different layers if they are not mutually exclusive. Here are the layers we observe (see Fig. 2), inspired by the layers of Arabic (Osborn, 2017):

1. basic letter forms and punctuation;
- 1' special letter forms for the Bible (inverted *nun*, closed *qof* and broken *vav*), and graphe[mt]ic hapaxes (§2.3);
2. a supralinear dot placed on the left or on the right of basic form ש, to distinguish between *sin* and *shin*;
3. the internal dot *dagesh/mappiq* which represents a plosive feature or gemination, or the consonant nature of letter *heh* at word end, and the upper impositive *rafe* that is the binary opposite of *dagesh* and represents the fricative feature;
4. the supralinear short vowel *holam* and sublinear short vowels *niqqud*, half vowels *bataf*, and absence of vowel *sheva*;
5. the sublinear impositive sign *meteg*;
6. *te'amim* cantillation marks;
7. *line fillers* as in the following excerpt of the *Complutensian Polyglot Bible* (Alcalá de Henares, 1517)



2.1. Upper and lower systems

Masoretic text has the unique feature that it can convey not only one but *two* systems of *nikkud* and *te'amim*: the Upper System (*ta'amey ba-elyon*, also called “high cantillation”) and the Lower System (*ta'amey ba-tachton*, also called “low cantillation”). This occurs in texts such as the Decalogue in Exo 20:2–14 and Deu 5:6–18 (Jacobson, 1995), Gen 35:22, and other places. As primary graphemes are the same, and to avoid writing the text twice, secondary graphemes of the two systems are written simultaneously so that a given consonant can have two vowels (the first one always belongs to the high system and the second to the low system) and/or two cantillation marks (see Fig. 3).



FIG. 3. Upper/lower systems

Ashkenazi custom is to use the Lower System for private study and ordinary reading, and the Upper System for public reading of the Decalogue, including the festival of Shavu'ot.

2.2. Graphemic Clusters

Nikkud and *te'amim* can be supralinear/internal/supralinear, prepositive/centered/postpositive (from right to left in the following figure):



These features are graphemic in the sense that a change of positional category can affect the semantics of the text.

More generally, we can subdivide the graphemes used in the Masoretic text into primary and secondary graphemes, the former being stand-alone graphemes and the latter being necessarily “carried” by primary graphemes. A primary grapheme (called “base character” in

Unicode) together with the secondary graphemes it carries is called a *graphemic cluster*. In a cluster, primary graphemes (layers 1 and 1') represent consonants, long vowels, and punctuation, while layers 2-6 are represented by secondary graphemes.

2.3. Graphe[mt]ic Hapaxes

When studying the phonology of a language, we consider that two phones belong to distinct phonemes if we can find two words, differing only by these phones, and having different semantics. We can define graphemes in a similar way and conclude, e.g., that the graphs “i” and “ı” (as in “fi”) represent the same English grapheme ⟨i⟩ since there is no pair of English words differing only by these two graphs and having distinct semantics, while, e.g., they represent distinct graphemes in Turkish. The problem with this statement is that “distinct semantics” is rather vague.

In this section, we describe cases where primary graphemes are transformed in four distinct ways: they can be broken, have their strokes joined, be larger than their surroundings, be smaller than their surroundings, or be suspended. These are called *otiyyot meshunot* (Ginsburg, 1897; Tov, 1999) and belong to several functional categories at once. They are *indexical* (in Peirce’s sense): they do not represent a sound or a morpheme but point at the word, saying “look here, this is special”. They are *critical*: the puncta extraordinaria in particular are now widely understood as the Masoretes’ way of flagging words their tradition regarded as textually doubtful—a kind of pre-modern critical apparatus encoded inside the text rather than alongside it. They are *hermeneutic*: they triggered, and are retroactively explained (in a way similar to the *ḥurūf al-muqaṭṭa’a* in the Qur’an, Bauer, 1921; Nguyen, 2012), by a large body of midrashic interpretation, e.g., the small *alef* of Lev 1:1 read as Moses’ humility; the raised *nun* of Jdg 18:30 preserving the tradition that the original reading was Mosheh but had to be screened to protect the prophet’s name from association with idolatry, etc. And they are *redactional fossils*: in several cases they almost certainly preserve, in petrified graphic form, evidence of earlier textual states—alternative readings, scribal corrections, or the boundaries of redactional units (the inverted *nuns* around the “When the ark went forth” pericope in Num 10:35–36 are the textbook example).

With these explanations in mind, we may raise the question: are *otiyyot meshunot* graphemes or graphetic allographs? It is easy to find minimal pairs (words that differ only by having or not having the *meshunot* feature), but do they carry “distinct semantics”? Depending on the individual interpretations given, they are more or less semantic, so we consider that they belong to a twilight zone between graphemes and allographs. Here is a list, based on Haralambous (1994). In two cases (Est 9:9 and Dan 6:20), we have two *meshunot* in the same word.

2.3.1. *Transformed Letters*

These are: the joined *qof* (Ginsburg, 1897, p. 48) in Exo 32:25 and Num 7:2; the broken *vav* (*loc. cit.*) in Num 25:12. For the sake of completeness we provide an example of the use of the inverted *nun* (*nun hafukba*), which is the only graphe[mt]ic hapax that found its way into Unicode (HEBREW PUNCTUATION NUN HAFUKHA), and which is used as punctuation.

Exo 32:25	אֶהְרֹן לְשִׁמְצָה בְּקִמְיָהֶם
Num 7:2	הֵם הָעֹמְדִים עַל־הַפָּקָדִים
Num 25:12	לֹא אֶת־בְּרִיתִי שָׁלֹם
Psa 107:23	יִיְרְדֵי הַיָּם בְּאַנְיוֹת

2.3.2. *Large Letters*

Gen 1:1	בְּרֵאשִׁית בָּרָא אֱלֹהִים
Gen 5:1	זֶה סֵפֶר תּוֹלְדֹת
Gen 30:42	וּבַהֲעֵטֶרָה הִצָּאן לֹא יִשְׁיִים
Gen 34:31	וַיֹּאמְרוּ הַכְּזֹנָה יַעֲשֶׂה
Gen 50:23	בְּנֵי שְׁלֹשִׁים גַּם
Exo 34:7	נִצֵּר חֹסֶד לְאֻלָּפִים
Exo 34:14	לֹאֵל אֶתֶר כִּי
Lev 11:42	הוֹלִיף עַל־גִּחְוֹן וְכָל
Lev 13:33	וְהִתְגַּלַּח וְאֶת־הַנֶּתֶק לֹא
Num 14:17	וְעִתָּה יִגְדֵּל־נָא כָּח
Num 27:5	מַשֶּׁה אֶת־מִשְׁפָּטָן לִפְנֵי
Deu 6:4	שְׁמַע יִשְׂרָאֵל יְהוָה
Deu 6:4	אֱלֹהֵינוּ יְהוָה אֶחָד
Deu 11:21	כִּימִי הַשָּׁמַיִם הָאָרֶץ
Deu 22:6	יִקְרָא קֹן־צִפּוֹר לִפְנֵיהֶ

Deu 29:27	גָּדוֹל וַיִּשְׁלַחְכֶּם אֶל-אֶרֶץ
Deu 32:6	הַלִּיהוּהַ תִּגְמְלוּ-זֹאת
Deu 33:29	אֲשֶׁרִידֶיךָ יִשְׂרָאֵל מִי
Rth 3:13	לִינִי הַלַּיְלָה וְהָיָה
Jos 14:11	מִשֶּׁה כָּכָחִי אִזּוֹ
2Ki 17:31	עָשׂוּ נִבְחָז וְאֶת-תַּרְתָּק
1Ch 1:1	אָדָם שֵׁשֶׁת שָׁנָה
Est 1:6	חֵינֹר כַּרְפָּס וּתְכֵלֶת
Est 9:9	וְאֶת אֶרְדֵּי וַיִּזְתָּא
Est 9:29	וְתִכְתֹּב אֶסְתֵּר הַמַּלְכָּה
Job 9:34	מֵעַלִי שְׁבֻטֹּי וְאִמָּתוֹ
Psa 80:16	וְכִנֹּה אֲשֶׁר-נִטְעָה יְמִינָהּ
Psa 84:4	וּדְרוֹר קָן לָהּ
Pro 1:1	מִשְׁלֵי שְׁלֹמֹה בֶן-דָּוִד
Ecc 7:1	טוֹב שֵׁם מִשָּׁמֶן
Ecc 12:13	סוּף דְּבַר הַכֹּל
Sng 1:1	שִׁיר הַשִּׁירִים אֲשֶׁר
Isa 56:10	צִפּוֹ עוֹרִים כָּלָם
Dan 6:20	בְּשִׁפְרָפְרָא יָקוּם בְּנִינְהָא
Mal 3:22	זָכְרוּ תּוֹרַת מֹשֶׁה

2.3.3. Small Letters

Gen 2:4	וְהָאָרֶץ בְּהִבְרָאָם בְּיוֹם
Gen 23:2	לְסָפֹד לְשָׂרָה וּלְבִכְתָּהּ
Gen 27:46	אֶל-יִצְחָק קָצַמְתִּי בְחַיִּי

Lev 1:1	וַיִּקְרָא אֶל־מֹשֶׁה וַיְדַבֵּר
Lev 6:2	עַל־מִזְבֵּחַ עַל־הַמִּזְבֵּחַ
Deu 9:24	מִמֵּרִים הֵייתָם עִם־יְהוָה
2Sa 21:19	אֶלְחָנָן בֶּן־יֵעָרִי אֲרָגִים
Est 9:9	וְאֵת אֲרָדִי וַיִּזְחָא
Job 7:5	רָמָה וְגִישׁ עָפָר
Job 16:14	יִפְרָצֵנִי פֶּרֶץ עַל־פְּנֵי־פָרֶץ
Job 33:9	בְּלִי־פֶשַׁע חָף אֲנֹכִי
Pro 16:28	מִדּוֹן וְנִרְגָּז מִפָּרִיד
Pro 28:17	אָדָם עֹשֶׂק בְּדָם־נַפֶּשׁ
Pro 30:15	בְּנוֹת־הַבִּבְבָּא
Isa 44:14	נָטַע אֲרָז וְגִשְׁם יִגְדֵּל
Jer 14:2	לְאָרֶץ וַיִּזְחַת יְרוּשָׁלַם
Jer 39:13	רַב־טַבָּחִים וְנִבְשׁוּבָו רַב־סָרִיס
Lam 1:12	לֹא אֵלֵיכֶם כָּל־עֲבָרִי
Lam 2:9	טָבְעוּ בָאָרֶץ שְׁעָרֶיהָ
Lam 3:36	לְעֵינֵי אָדָם בְּרִיבּוֹ
Dan 6:20	בִּשְׁפָרָא יָקוּם בְּנִגְהָא
Nah 1:3	יְהוָה בְּסוּפָה וּבְשַׁעֲרָה

2.3.4. *Suspended Letters*

Jdg 18:30	בֶּן־גִּרְשָׁם בֶּן־מְנַשֶּׁה הוּא
Job 38:13	וַיִּנָּעְרוּ רַשְׁעִים מִמָּנָה
Job 38:15	וַיִּמָּנַע מִרַשְׁעִים אוֹרָם
Psa 80:14	חֲזִיר מִיַּעַר וְזִז

2.4. Graphotactic Rules

Let us compare two musical notation systems (the Byzantine Greek system and the Armenian *kbaz* system) with the system of Masoretic cantillation marks.

Both the Greek and the Armenian systems place neumes on a separate line above the text.

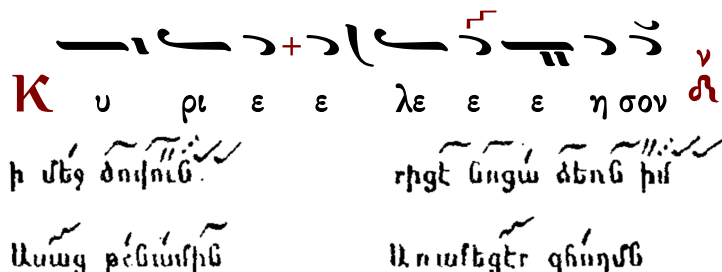


FIG. 4. Samples of Byzantine Greek (above) and Armenian (below) systems of musical notation

In the Byzantine Greek system (Troelsgard, 2011; Ἀθανασόπουλος, 1950) neumes are significantly larger than the corresponding text syllables so that the latter are necessarily spaced to be vertically aligned (and therefore temporally synchronized) with the neumes. In the Armenian *kbaz* neume system (At'ayan and Nersessian, 1999), neumes are sparser and smaller, so that words do not necessarily need to be spaced. Also, the precision of alignment between syllable and neume does not seem to be of paramount importance.

Not so for Masoretic text, where cantillation is represented by secondary graphemes that are “carried” by specific primary graphemes, and therefore their alignments have to be extremely precise.

Subscripts can be horizontally centered or prepositive (right-aligned), while superscripts can be horizontally centered, prepositive, or postpositive. Position is a graphemic feature since, in several cases, visually similar marks denote different signs when changing position (e.g., for example, a centered sublinear “<” denotes mark *mahapakh*; in contrast, in the prepositive position, the same graph denotes mark *yetiv*).

Primary graphemes have vertical accentuation axes for upper and lower diacritics (see Fig. 5). Sometimes, the same consonant can have two lower accentuation axes intended for different marks (see rule 8). For placing the various marks, the following rules have to be followed in the given order:

- Rule 1: supralinear *shin/sin* dots and internal *dagesh/mappiq* marks are placed at invariant positions (at least in printed texts);
- Rule 2: primary graphemes have (sometimes distinct) supralinear and sublinear *vertical axes* (Fig. 5), e.g., in the cases of א, ל, נ, the sublinear axis is their vertical stroke;
- Rule 3: sublinear *niqqud* and “centered” *te’amim* are horizontally centered with respect to these axes: notice the position of diacritics in א, ל, נ, vs. ב, ז, ד, etc.,;
- Rule 4: the *holam* vowel is a postpositive superscript—an exception occurs for letter *vav* that sometimes carries a prepositive *holam* mark (*holam baser*) and sometimes a postpositive one (*holam male*);
- Rule 5: the subscript *meteg* mark is either centered or follows an existing sublinear vowel—in some cases it may appear between the two components of a semi-vowel;
- Rule 6: in case there is both a sublinear *niqqud* and one or more sublinear *te’amim* of the centered type, the *niqqud* is centered with respect to the vertical axis (see Rule 2) and the *te’amims* are concatenated to its right;
- Rule 7: prepositive and postpositive cantillation marks are placed on their respective positions, and if there is a postpositive *holam* vowel and a postpositive cantillation mark, the latter will always follow the former;
- Rule 8: when the last syllable of a word has a א, נ or ז consonant with *patah* and is preceded by a long vowel, then the *patah* becomes prepositive and is called *patah furtivum*;
- Rule 9: secondary graphemes must be equally distanced from each other and must not cross or touch descenders (letters *qoph*, *ayin*, and final *kaf*, *nun*, *pe*, and *tsadi*) or ascenders (letter *lamed*);
- Rule 10: most importantly, *the graphs of primary graphemes are set at their regular distances, as if there were no secondary graphemes.*²

It should be noted that the properties of centrality or pre-/postpositivity of *te’amim* may depend on the context. Indeed, some cantillation marks appear twice in the same word: once at its beginning or end, in prepositive or postpositive form, and once on the stressed syllable, in *centered* form—these are called *anticipatory* marks.

2. This was the case in manuscript Bibles since, as already mentioned in note 1, a manuscript Masoretic text was produced in two steps: first the primary graphemes were written by the *sofer* and then the secondary ones were added by the *naqdan*. The technology of printing has reduced Bible production into a single-step process.

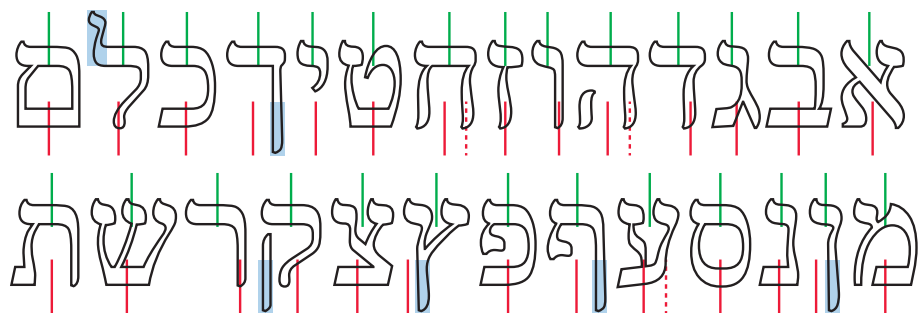


FIG. 5. Vertical axes of the graphs of the Monotype Peninim font. In green: upper vertical axes; in red: lower vertical axes; in dotted red: axes for placement of *patah furtivum*; in light blue: ascenders and descenders

3. Graphetic Challenges

The rules given in the previous section are extremely challenging on the graphetic level: it is often impossible to respect the perfect centering of diacritics with respect to vertical axes, and therefore, *niqqud* and *te'amim* may need to be decentered due to the graphetic context.

3.1. An Example

To illustrate the various Masoretic graphetic challenges, we use the example of word וַיֵּרָא ("and (he) saw" as in "And God saw the light, that it was good") from Gen 1:4.

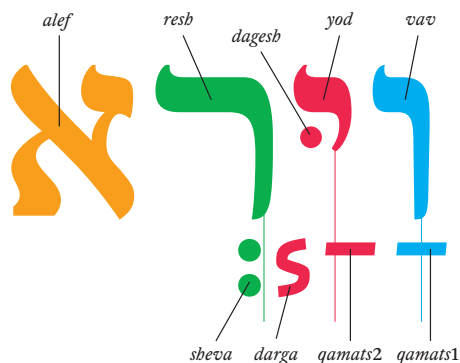


FIG. 6. The word וַיֵּרָא, analyzed

In Fig. 6 we illustrate the graphemic structure of this word: it has four letters: *vav*, *yod*, *resh*, and *alef*, and therefore four graphemic clusters. The *yod* carries a *dagesh*. As for sublinear secondary graphemes, the *vav* carries a *qamats* (we denote it by *qamats1*); the *yod* carries both a *qamats* (= *qamats2*) and the Masoretic diacritic *darga*; the *resh* carries a *sheva*.

The graphetic challenge lies in the fact that, on the one hand, the *yod* is a quite narrow letter so that the group of two secondary graphemes it carries (*qamats2* and *darga*) is necessarily wider than the letter itself, while, on the other hand, both letters surrounding the *yod* also carry sublinear secondary graphemes. The accumulated width of the four marks will fatally exceed the space assigned to them, which is delimited by the sublinear vertical axes of the *vav* and the *resh*.

In Fig. 6, typeset in the Monotype Peninim font by our system (Tiqwah v2), we have drawn the sublinear vertical axes of the three letters. We see that *qamats1* is perfectly centered on the vertical axis of the *vav* and that the *sheva* is slightly on the left of the axis of letter *resh*. The other noticeable fact of Fig. 6 is that the letters have kept their standard spacing, i.e., they are located at exactly the same distance as they would if they did not carry secondary graphemes.

Let us now take an overview, in Fig. 7, of how this word appears in 26 printed Bibles dating from 1517 to 2021 (#27 is our Tiqwah v2 output). The bibliographical data of each of the 26 Bibles is given in Appendix A (p. 320). 23% of these examples leave extra space on the left of the *yod*, 8% on its right, and 19% on both sides of the *yod*, i.e., half of these examples introduce parasitic spacing between letters and therefore break Rule #10. In many examples, the *darga*, which belongs to the *yod*, is placed on the axis of the *resh* and sometimes even to its left, introducing ambiguity into the visual recognition of graphemic clusters by raising the suspicion that the *darga* belongs to the *resh* since it is placed on the left of its vertical axis.

Example #1 contains only *niqqud* (it comes from the *Complutensian Polyglot Bible* (1517), in which the Hebrew text has only a very limited set of *te'amim*). Nevertheless, despite the fact that there is no *darga* in #1, the *sheva* has been moved to the left.

Example #2 is very important because it stems from the first printed Bible with full Masoretic marks. It was printed in 1525, in Venice, by the Antwerpian printer Daniel Bomberg. It is astounding that the word is typeset perfectly in #2, without any extra spacing between letters, and with the *sheva* perfectly centered on the *resh*'s axis. Among our examples, we have to wait until #22 (1999) to obtain a version of similar quality.

Example #3 moves the *qamats1* definitely too far to the right, and not at the same height as *qamats2*, but it otherwise manages to keep the *sheva* almost centered under the *resh*.

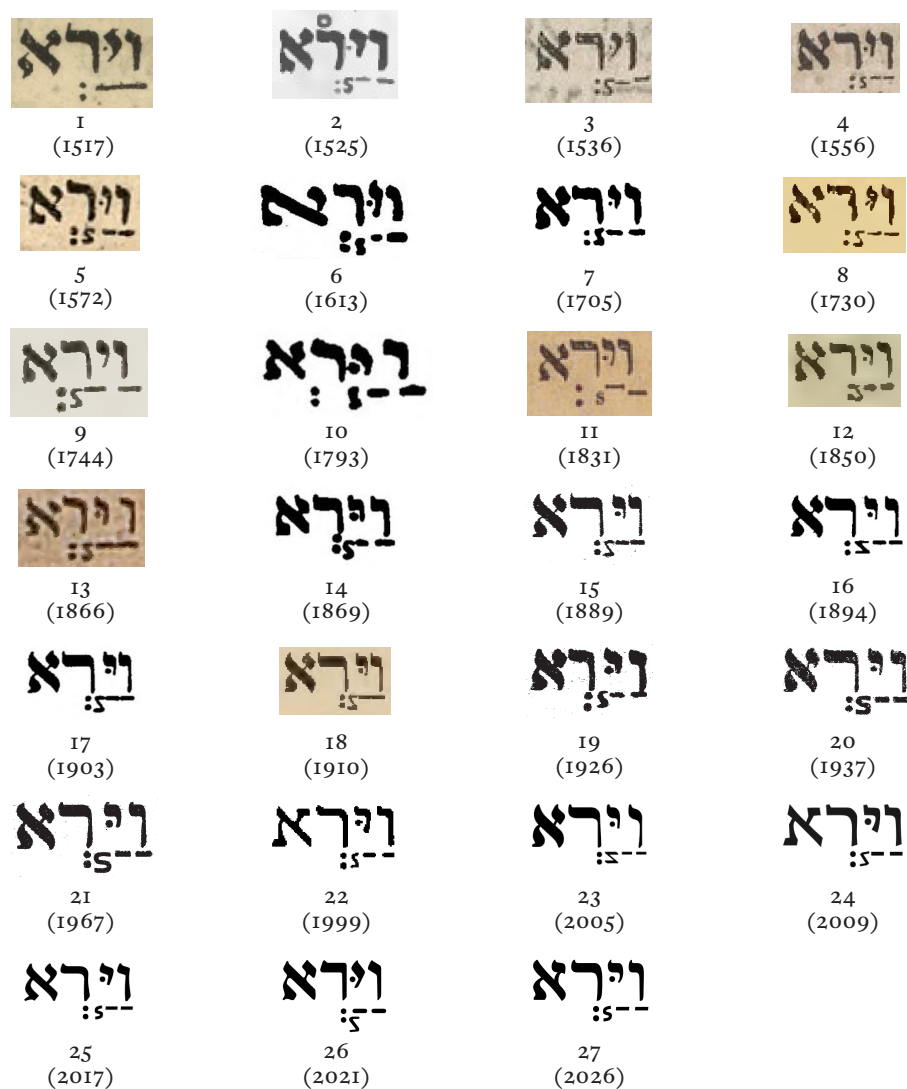


FIG. 7. The word **וירא** in 26 different printed Bibles, from 1517 to 2021 (#1–#26) and typeset by the Tiqwah v2 system in the Monotype Peninim font (#27)

Example #4 is quite good, but the *sheva* is slightly to the left, and it is not visually clear whether the *darga* belongs to the *resh* or to the *yod*. We encounter the same problem in #8.

Examples #5, #6, and #7 are of very bad quality and definitely ambiguous: the *darga* is centered on the axis of the *resh*, and the *sheva* is pushed to the left.

Example #9 is, by far, the worst of all cases: the *qamats2* of the *yod* has moved under the *resh*, so that the *darga* is pushed to the left and the *sheva* almost touches the *alef*.

Example #10 is also very bad: while the two *qamats* and the *darga* could be considered as correct, the *sheva* is pushed to the left for no reason, and appears at the boundary between the (quite large) *resh* and the *alef*.

Example #11 is from a 19th-century French edition of the Hebrew Bible. The *qamats1* is placed before the *vav*, the *qamats2* is correctly centered under the *yod*, but on a different height, probably to distinguish it from *qamats1*. The *darga* looks like an “s” letter and is placed on the *resh*’s axis, but a bit lower than the *qamats2*, and the *sheva*, once again, is pushed to the extreme left of the *resh*.

Example #12 has no parasitic spaces and the niqqud are well centered, but the *darga* is so narrow that it becomes unreadable, and touches the *sheva*.

Example #13 is reminiscent of Examples #9–#11, except that it has a lot of parasitic space between the *vav* and the *yod*.

Example #14 is quite interesting: the marks are quite legible, the letters are sufficiently spaced so as not to give the impression of parasitic space, but the most exciting fact is that the two *qamats* have different widths: *qamats2* is almost 50% the width of *qamats1*, and therefore leaves sufficient space to the *darga*. The *sheva* is only slightly moved to the left of the vertical axis.

Examples between 1889 and 1967 (#15–#21), as well as #23 (typeset using the Tiqwah v1 system, cf. Haralambous, 1994) and #25, use extra parasitic space between letters to solve the problem, the most extreme being #21, which is the only one that strictly respects the alignment of the *sheva* with the *resh*’s axis, even though it has a huge *darga* graph.

Examples #22 and #24 are set in the Hadassah font, which is very well suited for Masoretic text: it has large letters and narrow diacritics, so that no extra spacing is needed. The alignment is perfect.

Example #26 is a very special example: it comes from the *Koren Tanakh* and uses the Koren typeface. This typeface has been especially designed for Masoretic text, and we see that the *darga* is placed *under* the *qamats2*. This transgresses five centuries of printed Masoretic tradition, but the result is very pleasing and would be impossible to achieve otherwise, since the Koren graphs are very narrow and close together. Somehow we have the feeling that in 1831 the French printer of #11 also wanted to place the *darga* under the *qamats2* but didn’t manage to do so—#11 and #26 are the only examples where the *darga* is lower than the *qamats2*.

It is interesting to note that the technique of vertically stacking secondary graphemes has already been used in manuscript Bibles. Here is our example word in the Leningrad Codex (cf. Fig. 1 to see it in context):



We observe that (1) primary graphemes are indeed placed at their regular distances, (2) taking advantage of the fact that the *yod* descends only to half-height, the *qamats2* is raised to leave space for the *darga*. Nevertheless, the *naqdan* did not manage to center the *sheva* under the *resh*'s vertical axis, even though there was room enough to do so.

3.2. Algorithms

Algorithms solve this problem (cf. Haralambous (1994)) only partially by using letter spacing in extreme situations. In modern typefaces, OpenType features (Haralambous, 2007) address the problem to some extent. Nevertheless, OpenType features are limited because they process glyphs in linear order: in a situation where we have Cv_1Cv_2 , and we need to move v_1 and/or v_2 , a kerning rule involving them can hardly be applied since the glyph C lies between them.

To adapt a legacy-format font to the many interactions arising between clusters, one would have to include entire clusters as glyphs in the font. But this again presupposes that all difficult configurations are known to the font designer and can be anticipated. And that is not possible since, even though the Masoretic corpus (Hebrew Bible and New Testament) is relatively well-delimited, one cannot anticipate all textual variations in scholarly editions.

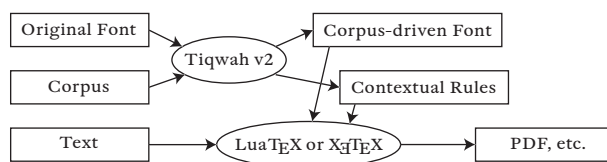
These considerations lead us to the conclusion that an optimal Masoretic font must be *built out of a given corpus*, and updated after every update or extension of the corpus, since every change may potentially introduce new problematic cluster interactions.

We call such a font a *corpus-driven* font. In the following section, we present our solution as well as the technologies involved.

4. The Tiqwah v2 System

4.1. System Architecture

The Tiqwah v2 system is an entirely rewritten version of the Tiqwah v1 system developed more than thirty years ago (Haralambous, 1994). Its architecture is as follows:



where the “Original Font” is any font containing standard Hebrew glyphs and *niqqud*, but not necessarily *te’amim*; “Corpus” is a corpus of texts using *te’amim* such as the Hebrew Bible, Margoliouth’s New Testament, etc.; “Text” is the text we wish to typeset (ideally the Hebrew text contained in “Text” should also be part of the “Corpus”); and “PDF, etc.” is the output. This system takes advantage of the separation between input and output in professional typesetting systems such as XeTeX or LuaTeX.

As for “Contextual Rules,” these are of two kinds: (1) reordering rules that ensure that glyphs are processed in a standard order that fits glyph names (see below) and (2) rewrite rules, with left and right context, which will map Unicode strings with lookahead- and lookbehind-endowed regular expressions, written either as a list of TECKit rules for XeTeX (Kew, 2021) or as a Lua program for LuaTeX (Hagen, Henkel, Hoekwater, and Scarso, 2025).

The output is searchable in any Acrobat-compliant PDF reader, since the glyphs of clusters have names that include the glyph names of all members of the cluster, in the logical order. So, for example, the cluster װ (which actually appears 282 times in the Bible, including in the first word of Genesis 1:1), the Unicode representation of which consists of characters U+FB2A (*shin* with *shin* dot), U+05B4 (*hiriq*), and U+0596 (*tipeba*), in this order, will have a glyph named uniFB2A05B40596, so that any search or copy action in the PDF file will use the underlying Unicode character string.

4.2. Volumetry

We adapted three fonts to Masoretic text: SBL Hebrew (designed by John Hudson in 2007), in which we kept the Masoretic glyphs but removed all OpenType features; Monotype Peninim (see Turkenich (2013) for the history of this emblematic font); Yarden (designed by Yarden Barzilai³ in 2023), a “playful, informal look [that] is great for light-hearted or humorous designs, comics, food packaging, and more”. We chose the Yarden font as an extreme Proof-of-Concept, because of its very free design. The reader can see in Figs. 9–11, on p. 322–324, the

3. <https://alefalefalef.co.il/font-yarden/>

beginning of the Book of Genesis typeset in Monotype Peninim, SBL Hebrew, and Yarden, as adapted by Tiqwah v2.

Let us give some elements of volumetry of the three fonts: SBL Hebrew, Monotype Peninim, and Yarden, all three built using the same corpus (Sefaria Tanakh⁴, Margoliouth’s New Testament⁵, and some other texts) of 160,099 distinct words containing at least one *niqqud* or *te’amim*:

	SBL Hebrew	Monotype Peninim	Yarden
# of added glyphs	9,625	12,877	18,664
# of .alt glyphs	2,108	5,111	10,898
avg. # of .alt	2.59	3.37	5.16
max # of .alt	101 (ך)	151 (ך)	123 (ך)
# of contextual rules	3,505	11,815	21,802
size of original *TF	61 kB	184 kB	30 kB
size of adapted OTF	449 kB	555 kB	755 kB
# of lines of .fea file	138	442	164
After font reduction with $\delta = 10$ glyph units (see §4.4):			
# of added glyphs	8,875	11,142	13,451
# of .alt glyphs	1,796	4,063	6,372
avg. # of .alt	1.72	2.32	2.82
max # of .alt	18 (ך)	27 (ך)	25 (ך)

where by “avg. # of .alt” we mean the average number of alternate glyphs for those that have at least one alternate glyph. We observe from the table that the number of glyphs of Yarden is almost double that of SBL Hebrew. This is due to irregularities in Yarden’s design, whereas SBL Hebrew has been tailored to be compatible with the Masoretic apparatus. The “.alt” glyphs are alternative versions of the same cluster, where one or more *niqqud* and *te’amim* have been moved for the cluster to be used in a specific context. The number of contextual rules depends on the number of alternative glyphs. As for the size of the OTF files obtained, it may seem small, e.g., when we compare them with CJK fonts, which also have tens of thousands of glyphs—this is because glyphs representing clusters are built using internal references rather than actual glyph code. The “.fea” file is the file of OpenType features, in Adobe syntax; its size represents the complexity of the font with respect to OpenType features.

In the following, we will describe the technology we use to build the cluster glyphs.

4. <https://www.sefaria.org/texts/Tanakh>

5. <https://www.jer-31.com/Margoliouth.html>

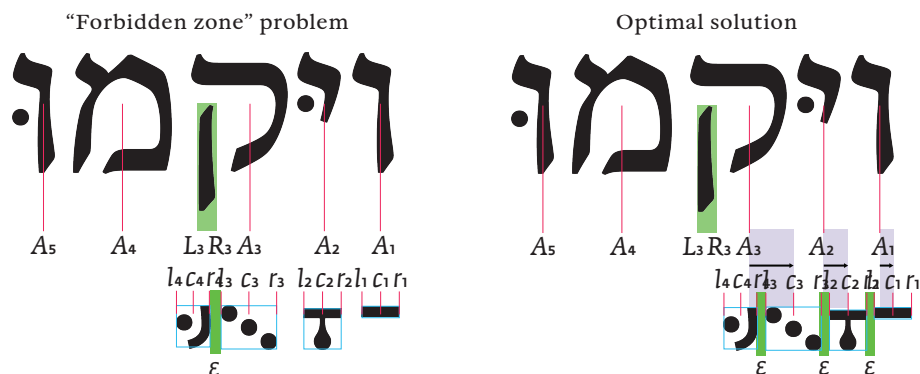


FIG. 8. The linear constraint minimization method applied to glyph positioning

4.3. Linear Constraint Minimization

Positioning of glyphs requires calculations in glyph space. In Fig. 8 we see the word **ויקמו** (Gen 22:19), in which we have drawn in red color the sublinear vertical axes, whose x-coordinates are called $A_1, \dots, A_5$. We have also drawn the four sublinear diacritics (three *niqqud* and one *te'amim*), each one carries three x-coordinates: the right boundary of its bounding box $r_1, \dots, r_4$, its stacking axis $c_1, \dots, c_4$ (which, depending on the diacritic's shape, are not necessarily the centers of the bounding boxes), and the left boundary of its bounding box $l_1, \dots, l_4$. We have drawn two zones in green color: one is the minimal distance ε between two sublinear (or supralinear) marks belonging to the same cluster, and the other, delimited by L_3 and R_3 , is a "forbidden zone," namely the bounding box of the descender of the *qof*.

On the left part of the figure we have placed the diacritics according to their natural position *in the absence of "forbidden zone"*, that is we have matched the vertical axes c_1, c_2, c_3 of the three *niqqud* with the corresponding vertical axes of letters A_1, A_2, A_3 , and we have concatenated the *te'amim* next to the *niqqud* of the third letter, so that the distance between r_4 and l_3 is equal to ε . The problem is that this contradicts graphotactic Rule 9 (§ 2.4), namely that marks should not cross letter descenders, which, in our case, can be expressed as the constraint $l_4 - R_3 \geq \varepsilon$.

On the right part of the figure, we see the optimal solution: all diacritics are moved to the right so that $l_4 - R_3, l_3 - r_4, l_2 - r_3$, and $l_1 - r_2$ are all equal to ε . In other words, the presence of the *qof* has influenced the position of the diacritics of both letters preceding it. We have drawn the shift vectors of each diacritic in violet.

Let us model this situation in mathematical terms, keeping the same notation.

A *linear constraint minimization problem* consists of:

1. a linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to be minimized by some value x ,
2. for which we have inequalities $g_i(x) \geq c_i$ for $i = 1, \dots, n$, where the g_i 's are linear functions $\mathbb{R} \rightarrow \mathbb{R}$.

The application of this mathematical problem to our situation is obvious: the linear function f is the sum of the absolute values of the differences between the c_i 's which respect graphotactic Rule 3 (left part of Figure 8) and those actually used; the g_i 's are the constraints given by graphotactic Rule 9, for example the fact that diacritics may not cross descenders (and even better: that they must stay at a distance of $\varepsilon + \kappa$, where κ is a potential kerning value. Other constraints occur at boundaries: a diacritic may not be “too far” on the right of the first letter of the word or on the left of the last letter of the word, etc.

This method provides us with optimal solutions in any situation. It will not work when the arrangement is physically impossible (this occurs if we have two descenders and the total width of the diacritics placed between them exceeds the width of the letter glyphs). These are extreme cases, and extreme cases require extreme solutions: we will either use narrower glyphs for the diacritics or stack them vertically. There is always a solution since these problems have been faced already and (mostly) solved by printers of the last five centuries...

4.4. Font Reduction

The problem with linear constraint minimization is that it provides optimal solutions up to a single glyph unit (which, for 10-point type, corresponds roughly to $3.53 \mu\text{m}$, far below the visibility threshold that, according to Westheimer (1981), is $\sim 10 \mu\text{m}$ at normal reading distance). Having distinct glyphs for clusters that differ by only a few glyph units is unnecessary and artificially conflates the font.

But how do we choose between two glyph clusters that differ by only a few glyph units, knowing that a cluster may have many diacritic components that may move in different directions (both horizontally and vertically)?

This is a typical graph theory problem, namely a *minimal clique partition in unit disk graphs*. Let S be a set of points in a Euclidean space $\mathbb{R}^d$, corresponding to the coordinates of the various alternate glyphs of a given cluster with d diacritics. Let us define a distance between members c, c' of S by $\text{dist}(c, c') := \sum |d_i - d'_i|$ where d_i, d'_i are the centers of the diacritics of c, c' .

Let us define the “ δ -near-graph” G as a graph the vertices of which are the elements of S and edges connecting vertices whenever their distance is less than δ . The “ δ -far-graph” G' is the complementary graph of G (same vertices but edges only where G has no edges). Covering S with

balls of diameter δ corresponds to finding cliques in G . But cliques in G correspond to discrete subgraphs in G' , and discrete subgraphs can be obtained by applying a coloring algorithm, and that is what we do.

Once we have found a minimal number of δ -balls covering S , we choose representatives for each, which are the only glyphs kept in the font.

As can be seen in the table of p. 316, font reduction allowed us to reduce the number of added glyphs by up to 28% (in the case of font Yarden). The impact of this change is most noticeable in the maximum number of .alts for a given glyph: it drops from 151 for Monotype Peninim to a mere 27.

5. Conclusion

We have argued that the Masoretic apparatus is best understood as a layered writing system in which a small set of primary graphemes carries a dense and tightly constrained set of secondary graphemes, and that the difficulty of typesetting it is not orthographic but graphetic. The marks themselves are well defined; the problem is realizing them on the page—centered on the correct axes, evenly spaced, clear of descenders and ascenders, and without disturbing the standard spacing of the consonants. This is a constraint-satisfaction problem that metal type, OpenType features, and even careful manuscript pointing solve only partially. The survey in Fig. 7 makes the point concretely: across five centuries, only a few editions place the marks of a single ordinary word without ambiguity or parasitic spacing.

Our response is the notion of a *corpus-driven font*. Because the difficult configurations arise from interactions between clusters, and because these cannot all be anticipated by a font designer, the font must be generated from the actual corpus to be typeset and regenerated whenever that corpus changes. Tiqwah v2 does this in two stages. Linear-constraint minimization computes, for every cluster, diacritic positions that respect the graphotactic rules and the “forbidden zones” of neighboring descenders, falling back on narrower marks or vertical stacking only in the rare physically impossible cases. A graph-coloring procedure then collapses clusters whose diacritics differ by less than the perceptual threshold into a single representative glyph, keeping the glyph count tractable even for a corpus of some 160,000 distinct pointed words. The output is a standard OpenType font whose cluster glyphs are named after their Unicode components, so the resulting PDF remains fully searchable.

That the same pipeline adapts fonts as different as SBL Hebrew, Monotype Peninim, and the playful Yarden—the last chosen precisely as an adversarial case—suggests the approach is general rather than tuned to one typeface, and it extends naturally beyond the Tanakh, as our

treatment of Margoliouth's cantillated New Testament shows. Several directions remain open: enlarging and diversifying the corpus (scholarly editions with non-standard pointing, the upper and lower systems of the Decalogue, Babylonian and Palestinian vocalization), refining the perceptual model that governs glyph merging, and applying the same corpus-driven, constraint-based methodology to other scripts whose diacritics interact as richly as the Masoretic ones—a portability the layered analysis of §2 was meant to support.

Acknowledgments

The author would like to thank Yosef Ginsberg and Gabriel Wasserman for their invaluable assistance in understanding Masoretic graphemics and graphetics and their guidance in improving the Tiqwah v2 system.

A. List of Bibles Used in Fig. 7

- #1 (Complutensian Polyglot Bible) *Vetus testamentu multiplici lingua nuc primo impressum. Et imprimis Pentateuchus Hebraico Greco atq[ue] Chaldaico idiomate, ediderunt: Antonius Nebrissensis, Demetrius Ducas, et al. Academia Complutensi Industria Arnaldi Guilielmi de Brocario: sumptibus F. Francisci Ximenez de Cisneros, Alcalá de Henares, 1514-1517*
- 2 (The Second Rabbinic Bible, first printed Hebrew Bible with the Massorah) *Great Scriptures with Onkelos' translation and commentary by Rasbi and Ibn Ezra, Venice, 1525* **מקרא גדולות עם תרגום אונקלוס ופירוש רש"י ואבן עזרא**
- 3 *Primus liber legis Mosaeicae, qui Genesis à Græcis inscribitur, ab Hebræis Beresith appellatur. Parisiis. Ex officina Christiani Wecheli sub scuto basiliensi, Paris, 1536*
- 4 *Liber Genesis. Parisiis. Ex officina Caroli Stephani (= Charles Étienne), Paris, 1556*
- 5 *Hebraicorum Bibliorum Veteris Testamenti Latina interpretatio. Christophorus Plantinus Regius Prototypographus Antuerpia excudebat, Antwerpen, 1572*
- 6 *Biblia Hebraica, cum interlineari interpretatione Latinæ Xantis Pagnini Lucensis, ex officina Plantiniana Raphelengii, Antwerpen, 1613*
- 7 *Biblia Hebraica, Secundum ultimam editionem Jos. Athias, a Jobanne Leusden denuo recognitam, Amstelædami et Ultrajecti, ediderunt Boom, Waesberge, etc., Amsterdam, 1705*
- 8 *Biblia Hebraica, ex aliquot manuscriptis et compluribus impressis codicibus, item massora tam edita, quam manuscripta, selectae variantes lectiones subiiciuntur cura ac studio Io. Heinr. Michaelis Halæ Magdeburgicæ, Typis & Jumptibus Orphanotrophei, Magdeburg, 1730*
- 9 *Book of Forty-Two, Main Section, Pentateuch, Mantua, 1744* **ספר ארבעה ועשרים החלק הראשון חמשה חומשי הורה כה מנטובה בשנת ויחכמ ידידיה שלמה לפ"ק**
- 10 *Biblia Hebraica olim a Christiano Reineccio edita et ad optimorum codicum et editionum fidem recensita et expressa nunc denuo ad fidem recensione masoreticæ ediderunt D. Io. Christoph. Doederlein et Ioannes Henricus Meisner, Lipsiæ, impensis*

- I. G. I. Breitkopfii et Soc.* (the same Breitkopf as in Breitkop & Härtel!), Leipzig, 1793.
- 11 *La Bible, traduction nouvelle, avec l'hébreu en regard, par Samuel Caben. Penteuque, La Genèse*, Paris, 1831
 - 12 *Biblia Hebraica, secundum ultimam editionem Jos. Athie, a Jobanne Leusden denuo recognitam. Ab Everardo van der Hooght, V.D.M. editio nova, recognita, et emendata, a Judab d'Allemand, Londini, typis excudebat A. Macintosh, impensis Jacobi Duncan*, London, 1850
 - 13 *A Torab scroll, Prophets, and Writings, accurately transcribed according to tradition, carefully studied by Yedidya ben Avram and the distinguished scholar Meir ba-Levi Letteris*, Berlin, 1866
ספר תורה נביאים ומכתובים מדויק היטיב יל פי המסורה
הוגה בעיון נמרץ על ידי ידידיה בן אברם ומאח החכם המובהק מהוי מאיר
הלוי לעטערס בערלין
 - 14 *Liber Genesis. Textum masoreticum accuratissime expressit, e fontibus masorae varie illustravit, notis criticis confirmavit S. Baer. Ex officina Bernhardi Tauchnitz*, Leipzig, 1869
 - 15 *Torab, Prophets, and Writings, accurate according to tradition, carefully studied by the distinguished scholar Meir ba-Levi Letteris*, Vienna, 1889
ספר תורה נביאים ומכתובים מדויק היטיב על פי המסורה הוגה בעיון נמרץ על ידי החכם המובהק מהוי מאיר הלוי לעטערס ווען שנת ה'תרמ"ט ליצירה
 - 16 *Torab, Prophets, Writings, Masoretic Text according to Jacob ben Chayim and C.D. Ginsburg*, London 1894
 - 17 *The Holy Scriptures of the Old Testament, Hebrew and English, printed for the British & Foreign Bible Society*, Berlin, 1903
 - 18 *A Torab scroll, Prophets, and Writings, accurately transcribed according to the Masorab, carefully studied by the distinguished scholar Mordechai Meir ba-Levi Letteris*, New York: Hebrew Publishing Company, 1910
ספר תורה נביאים ומכתובים מדויק היטב על פי המסורה הוגה בעיון נמרץ על ידי החכם המובהק מהוי מאיר הלוי לעטערס
 - 19 *Biblia Hebraica edidit Rud. Kittel*, Stuttgart, 1926
 - 20 *Biblia Hebraica edidit Rud. Kittel, textu masoreticum curavit P. Kable*, Stuttgart, 1937
 - 21 *Biblia Hebraica Stuttgartensia editio funditus renovata ediderunt K. Elliger et W. Rudolph, textum masoreticum curavit H. P. Rüger*, Stuttgart, 1967
 - 22 *JPS Hebrew-English Tanakh*, 2nd ed., The Jewish Publication Society, Philadelphia, 1999
 - 23 *Torab, Prophets, Writings, Masoretic Text according to Jacob ben Chayim and C.D. Ginsburg*, London 2005 (using the Tiqwah v1 system)
 - 24 *Hebrew-English Tanakh*, Varda Books, Skokie, Illinois, 2009
 - 25 Schmidt, Muhammad Wolfgang G. A., *The Complete Original Hebrew Text of the Old Testament or Tanakh. According to the Leningrad Codex*, Hamburg, disserta Verlag, 2017
 - 26 *The Magerman Edition. The Koren Tanakh, Hebrew-English*, Koren Publishers Jerusalem, 2021

1 בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת
 הָאָרֶץ: 2 וְהָאָרֶץ הִיְתָה תֹהוּ וָבֹהוּ וְחָשֶׁךְ
 עַל־פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי
 הַמַּיִם: 3 וַיֹּאמֶר אֱלֹהִים יְהי אֹר וַיְהי־אֹר:
 4 וַיֵּרָא אֱלֹהִים אֶת־הָאֹר כִּי־טוֹב וַיְבַדֵּל
 אֱלֹהִים בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ: 5 וַיִּקְרָא אֱלֹהִים
 | לְאֹר יוֹם וְלַחֹשֶׁךְ קָרָא לַיְלָה וַיְהי־עֶרֶב
 וַיְהי־בֹקֶר יוֹם אֶחָד: 6 וַיֹּאמֶר אֱלֹהִים יְהי
 רָקִיעַ בְּתוֹךְ הַמַּיִם וַיְהי מַבְדִּיל בֵּין מַיִם לַמַּיִם:
 7 וַיַּעַשׂ אֱלֹהִים אֶת־הָרָקִיעַ וַיְבַדֵּל בֵּין הַמַּיִם
 אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר מֵעַל
 לָרָקִיעַ וַיְהי־כֵן: 8 וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמַיִם
 וַיְהי־עֶרֶב וַיְהי־בֹקֶר יוֹם שֵׁנִי: 9 וַיֹּאמֶר
 אֱלֹהִים יִקְווּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם
 אֶחָד וְתִרְאֶה הַיַּבְשָׁה וַיְהי־כֵן: 10 וַיִּקְרָא
 אֱלֹהִים | לַיַּבְשָׁה אָרֶץ וּלַמְקוֹהַּ הַמַּיִם קָרָא
 יַמִּים וַיֵּרָא אֱלֹהִים כִּי־טוֹב: 11 וַיֹּאמֶר אֱלֹהִים
 תִּדְשָׂא הָאָרֶץ דָּשָׂא עֵשָׂב מִזְרִיעַ זֶרַע עֵץ פְּרִי
 עֹשֶׂה פְרִי לְמִינּוֹ אֲשֶׁר זֶרַע־בּוֹ עַל־הָאָרֶץ
 וַיְהי־כֵן: 12 וַתּוּצֵא הָאָרֶץ דָּשָׂא עֵשָׂב מִזְרִיעַ
 זֶרַע לְמִינֵהוּ וְעֵץ עֹשֶׂה־פְרִי אֲשֶׁר זֶרַע־בּוֹ
 לְמִינֵהוּ וַיֵּרָא אֱלֹהִים כִּי־טוֹב: 13 וַיְהי־עֶרֶב
 וַיְהי־בֹקֶר יוֹם שְׁלִישִׁי: 14 וַיֹּאמֶר אֱלֹהִים

FIG. 9. The beginning of the Book of Genesis typeset in the Monotype Peninim font, adapted by Tiqawah v2

1 בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ:
 2 וְהָאָרֶץ הָיְתָה תֹהוֹ וָבֹהוּ וְחָשֶׁךְ עַל־פְּנֵי תְהוֹם וְרוּחַ
 אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמַּיִם: 3 וַיֹּאמֶר אֱלֹהִים יְהִי
 אֹר וַיְהִי־אֹר: 4 וַיֵּרָא אֱלֹהִים אֶת־הָאֹר כִּי־טוֹב וַיַּבְדֵּל
 אֱלֹהִים בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ: 5 וַיִּקְרָא אֱלֹהִים | לָאֹר
 יוֹם וְלַחֹשֶׁךְ קָרָא לַיְלָה וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם אֶחָד:
 פ 6 וַיֹּאמֶר אֱלֹהִים יְהִי רָקִיעַ בְּתוֹךְ הַמַּיִם וַיְהִי מַבְדִּיל
 בֵּין מַיִם לַמַּיִם: 7 וַיַּעַשׂ אֱלֹהִים אֶת־הָרָקִיעַ וַיַּבְדֵּל בֵּין
 הַמַּיִם אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר מֵעַל לָרָקִיעַ
 וַיְהִי־כֵן: 8 וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמַיִם וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר
 יוֹם שֵׁנִי: פ 9 וַיֹּאמֶר אֱלֹהִים יִקְווּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם
 אֶל־מָקוֹם אֶחָד וְתִרְאָה הַיַּבֶּשֶׁה וַיְהִי־כֵן: 10 וַיִּקְרָא אֱלֹהִים
 | לַיַּבֶּשֶׁה אָרֶץ וּלְמִקְוֵה הַמַּיִם קָרָא יַמִּים וַיֵּרָא אֱלֹהִים
 כִּי־טוֹב: 11 וַיֹּאמֶר אֱלֹהִים תְּדַשֵּׂא הָאָרֶץ דָּשָׂא עֵשֶׂב מִזֶּרַע
 זֶרַע עֵץ פָּרִי עֵשָׂה פְרִי לְמִינֹו אֲשֶׁר זֶרְעוּבוּ עַל־הָאָרֶץ
 וַיְהִי־כֵן: 12 וַתּוֹצֵא הָאָרֶץ דָּשָׂא עֵשֶׂב מִזֶּרַע זֶרַע לְמִינֹהוּ
 וַעֲץ עֵשָׂה־פְרִי אֲשֶׁר זֶרְעוּבוּ לְמִינֹהוּ וַיֵּרָא אֱלֹהִים כִּי־טוֹב:
 13 וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם שְׁלִישִׁי: פ 14 וַיֹּאמֶר אֱלֹהִים
 יְהִי מְאֹרֶת בְּרָקִיעַ הַשָּׁמַיִם לְהַבְדִּיל בֵּין הַיּוֹם וּבֵין הַלַּיְלָה
 וְהָיוּ לְאֹתוֹת וּלְמוֹעֲדִים וּלְיָמִים וְשָׁנִים: 15 וְהָיוּ לְמְאֹרֹת
 בְּרָקִיעַ הַשָּׁמַיִם לְהָאִיר עַל־הָאָרֶץ וַיְהִי־כֵן: 16 וַיַּעַשׂ אֱלֹהִים
 אֶת־שְׁנֵי הַמְּאֹרֹת הַגְּדֹלִים אֶת־הַמָּאֹר הַגָּדֹל לְמַמְשֶׁלֶת
 הַיּוֹם וְאֶת־הַמָּאֹר הַקָּטָן לְמַמְשֶׁלֶת הַלַּיְלָה וְאֵת הַכּוֹכָבִים:
 17 וַיִּתֵּן אֹתָם אֱלֹהִים בְּרָקִיעַ הַשָּׁמַיִם לְהָאִיר עַל־הָאָרֶץ:
 18 וּלְמַשֵּׁל בַּיּוֹם וּבַלַּיְלָה וּלְהַבְדִּיל בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ
 וַיֵּרָא אֱלֹהִים כִּי־טוֹב: 19 וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם רְבִיעִי:
 פ 20 וַיֹּאמֶר אֱלֹהִים יִשְׂרְצוּ הַמַּיִם שָׂרָץ נֶפֶשׁ חַיָּה וְעוֹף

FIG. 10. The beginning of the Book of Genesis typeset in the SBL Hebrew font, adapted by Tiqwah v2

1 בְּרֵאשִׁית בָּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ:
 2 וְהָאָרֶץ הִיְתָה תֵהוֹ וְבָהּ וְחָשֶׁךְ עַל־פְּנֵי תְהוֹם וְרוּחַ
 אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמַּיִם: 3 וַיֹּאמֶר אֱלֹהִים יְהִי
 אֹר וַיְהִי־אֹר: 4 וַיֵּרָא אֱלֹהִים אֶת־הָאֹר כִּי־טוֹב וַיַּבְדֵּל
 אֱלֹהִים בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ: 5 וַיִּקְרָא אֱלֹהִים א
 לְאֹר יוֹם וְלַחֹשֶׁךְ קָרָא לַיְלָה וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם
 אֶחָד: 6 וַיֹּאמֶר אֱלֹהִים יְהִי רָקִיעַ בְּתוֹךְ הַמַּיִם וַיְהִי
 מַבְדִּיל בֵּין מַיִם לַמַּיִם: 7 וַיַּעַשׂ אֱלֹהִים אֶת־הָרָקִיעַ
 וַיַּבְדֵּל בֵּין הַמַּיִם אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר
 מֵעַל לָרָקִיעַ וַיְהִי־כֵן: 8 וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמַיִם
 וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם שֵׁנִי: 9 וַיֹּאמֶר אֱלֹהִים יִקְוּ
 הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם אֶחָד וְתִרְאָה הַיַּבְשָׁה
 וַיְהִי־כֵן: 10 וַיִּקְרָא אֱלֹהִים א לַיַּבְשָׁה אֶרֶץ וּלְמִקְוֵה
 הַמַּיִם קָרָא יַמִּים וַיֵּרָא אֱלֹהִים כִּי־טוֹב: 11 וַיֹּאמֶר
 אֱלֹהִים תִּדְשֵׁא הָאָרֶץ דָּשָׁא עֵשֶׂב מִזְרִיעַ זֶרַע עֵץ פֶּרִי
 עֹשֶׂה פֶרִי לְמִינֹו אֲשֶׁר זֶרְעוֹ־בּוֹ עַל־הָאָרֶץ וַיְהִי־כֵן:
 12 וַתּוֹצֵא הָאָרֶץ דָּשָׁא עֵשֶׂב מִזְרִיעַ זֶרַע לְמִינֵהוּ וְעֵץ
 עֹשֶׂה־פֶרִי אֲשֶׁר זֶרְעוֹ־בּוֹ לְמִינֵהוּ וַיֵּרָא אֱלֹהִים כִּי־טוֹב:
 13 וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם שְׁלִישִׁי: 14 וַיֹּאמֶר אֱלֹהִים
 יְהִי מְאֹרֶת בְּרָקִיעַ הַשָּׁמַיִם לְהַבְדִּיל בֵּין הַיּוֹם וּבֵין
 הַלַּיְלָה וְהָיוּ לְאֹתוֹת וּלְמוֹעֲדִים וּלְיָמִים וְשָׁנִים: 15 וְהָיוּ
 לְמֹאֲרוֹת בְּרָקִיעַ הַשָּׁמַיִם לְהַאֲרִי עַל־הָאָרֶץ וַיְהִי־כֵן:
 16 וַיַּעַשׂ אֱלֹהִים אֶת־שֵׁנִי הַמְּאֹרֶת הַגְּדֹלִים אֶת־הַמָּאֹר
 הַגָּדֹל לְמַחְשֶׁלֶת הַיּוֹם וְאֶת־הַמָּאֹר הַקָּטָן לְמַחְשֶׁלֶת
 הַלַּיְלָה וְאֶת הַכּוֹכָבִים: 17 וַיִּתֵּן אֹתָם אֱלֹהִים בְּרָקִיעַ
 הַשָּׁמַיִם לְהַאֲרִי עַל־הָאָרֶץ: 18 וּלְמַשֵּׁל בַּיּוֹם וּבַלַּיְלָה
 וּלְהַבְדִּיל בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ וַיֵּרָא אֱלֹהִים כִּי־טוֹב:

FIG. 11. The beginning of the Book of Genesis typeset in the Yarden font, adapted by Tiqwah v2

References

- At'ayan, Robert and Vrej N. Nersessian (1999). *Armenian Neume System of Notation*. Routledge.
- Bauer, Hans (1921). "Über die Anordnung der Suren und über die geheimnisvollen Buchstaben im Qoran". In: *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 75, pp. 1–20.
- Futato, Mark D. (2020). *Basics of Hebrew Accents*. Zondervan Academic.
- Ginsburg, Christian D. (1897). *Introduction to the Massoretico-Critical Edition of the Hebrew Bible*. London: Trinitarian Bible Society.
- Hagen, Hans et al. (2025). "LuaTeX Reference Manual". <http://mirrors.ctan.org/systems/doc/luatex/luatex.pdf>.
- Haralambous, Yannis (1994). "Tiqwah, a Typesetting System for Biblical Hebrew, Based on TeX". In: *Actes du Quatrième Colloque International Bible et informatique, matériel et matière*. Amsterdam, pp. 445–470.
- (2007). *Fonts & Encodings. From Advanced Typography to Unicode and Everything in Between*. Sebastopol, CA: O'Reilly, pp. xx+1018.
- Hedánek, Jiří (2017). "Phonology of Masoretic Hebrew II. Accents as Prosody". PhD thesis. Charles University, Czech Republic.
- Hoop, Raymond de (2008). "Stress and Syntax; Music and Meaning: The Purpose and Function of the Masoretic Accentuation System". In: *Journal of Northwest Semitic Languages* 34, pp. 99–121.
- Jacobson, Joshua R. (1995). "The Cantillation of the Decalogue". In: *The Journal of Synagogue Music* 24.1, pp. 12–30.
- Kelley, Page H., Daniel S. Mynatt, and Timothy G. Crawford (1998). *The Masorah of the Biblia Hebraica Stuttgartensia*. Grand Rapids, MI & Cambridge, UK: William B. Eerdmans Publishing Company.
- Kew, Jonathan (2021). "The TECKit Language. Mapping byte encodings to Unicode". https://software.sil.org/downloads/r/teckit/TECKit_Language.pdf.
- Levine, Joseph A. (1989). *Synagogue Song in America*. Crown Point, IN: White Cliffs Media Company.
- McClintock, John and James Strong (1870). "Hebrew Version of the New Testament". In: *Cyclopedia of Biblical, Theological and Ecclesiastical Literature*. New York: Harper & Brothers.
- Mitchell, David C. (2017). "The Origins of the Masoretic Cantillation". A paper read at the Society of Old Testament Studies Centenary Conference, Nottingham University. URL: https://d1wqtxts1xzle7.cloudfront.net/51712224/The_Origins_of_the_Masoretic_teamim-libre.pdf.
- Nguyen, Martin (2012). "Exegesis of the ḥurūf al-muqaṭṭa'a: Polyvalency in Sunnī Traditions of Qur'ānic Interpretation". In: *Journal of Qur'anic Studies* 14/2, pp. 1–28.
- Osborn, J.R. (2017). *Letters of Light. Arabic Script in Calligraphy, Print, and Digital Design*. Princeton University Press.

- Park, Sung Jin (2020). *The Fundamentals of Hebrew Accents: Divisions and Exegetical Roles beyond Syntax*. Cambridge: Cambridge University Press.
- Tov, Emanuel (1999). "Paratextual Elements in the Masoretic Manuscripts of the Bible Compared with the Qumran Evidence". In: *Antikes Judentum und frühes Christentum: Festschrift für Hartmut Stegemann zum 65. Geburtstag*. Ed. by Bernd Kollmann, Wolfgang Reinbold, and Annette Steudel. Vol. 97. Beihefte zur Zeitschrift für die neutestamentliche Wissenschaft. De Gruyter, pp. 73–83.
- Troelsgard, Christian (2011). *Byzantine Neumes: A New Introduction to the Middle Byzantine Musical Notation*. Museum Tusculanum Press.
- Turkenich, Liron Lavi (2013). *From Frank Rübl to Peninim. A story of people, type and technology*. Master of Arts Thesis, University of Reading.
- Westheimer, Gerald (1981). "Visual hyperacuity". In: ed. by Hansjochem Autrum et al. Vol. 1. *Progress in Sensory Physiology*. Springer, pp. 1–30.
- Wickes, William (1881). *A Treatise on the Accentuation of the Three So-Called Poetical Books of the Old Testament*. Oxford: Clarendon Press.
- (1887). *A Treatise on the Accentuation of the Twenty-One So-Called Prose Books of the Old Testament*. Oxford: Clarendon Press.
- Wu, Andi and Kirk Lowery (2006). "A Hebrew Tree Bank Based on Cantillation Marks". In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Ed. by Nicoletta Calzolari et al. URL: http://www.lrec-conf.org/proceedings/lrec2006/pdf/6_pdf.pdf.
- Ἀθανασόπουλος, Γεώργιος Δ. [Athanasopoulos, Georgios D.] (1950). *Θεωρία τῆς βυζαντινῆς μουσικῆς [Theory of Byzantine Music]*. Πάτραι [Patras]: Τούλας-Μαυρίκιος [Toulas-Mavrikios].
- Хаздан, Евгения В. [Khazdan, Evgeniya V.] (2020). "От масоретских помет к знакам кантилляции: динамика восприятия [From Masoretic Notes to Cantillation Marks: The Dynamics of Perception]". In: *Богослужебные практики и культовые искусства в современном мире [Liturgical Practices and Sacred Arts in the Modern World]*. Vol. 4. Краснодар [Krasnodar]: Российский фонд фундаментальных исследований [Russian Foundation for Basic Research], pp. 288–310.