

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Proposal to add two Masoretic punctuation marks to the BMP of the UCS****Source: Mark Shoulson, Peter Kirk, John Hudson, Michael Everson, and Peter Constable****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2004-03-04**

This document requests an additional character to be added to the UCS and contains the proposal summary form.

A. Administrative**1. Title**

Proposal to add two Masoretic punctuation marks to the BMP of the UCS.

2. Requester's name

Mark Shoulson, Peter Kirk, John Hudson, Michael Everson, and Peter Constable

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

Individual contribution.

4. Submission date

2004-03-04

5. Requester's reference (if applicable)**6. Choose one of the following:****6a. This is a complete proposal**

Yes.

6b. More information will be provided later

No.

B. Technical – General**1. Choose one of the following:****1a. This proposal is for a new script (set of characters)**

No.

Proposed name of script**1b. The proposal is for addition of character(s) to an existing block**

Yes.

1b. Name of the existing block

Hebrew

2. Number of characters in proposal

2

3. Proposed category (see section II, Character Categories)

Category B.1

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Spacing characters.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson. TrueType.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson. Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes, see bibliography below.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes, see below.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Yes, see Unicode properties below.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Professor Aron Dotan, script expert

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Yes. Printing of materials for scholars and Bible translators.

4a. The context of use for the proposed characters (type of use; common or rare)

Scholarly and ritual use.

4b. Reference

See examples below.

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

See examples below.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Other Masoretic characters are also in the BMP.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference

D. Proposal

The traditional Masoretic text of the Hebrew Bible contains several anomalies of printing that have no parallel in other Hebrew usages. Many of these are already encoded in UCS, but there are a few that have thus far slipped through the cracks.

05C5  HEBREW MARK LOWER DOT
lower *punctum extraordinarium* (Psalms 27:13)

05C6  HEBREW PUNCTUATION NUN HAFUKHA (glyph variants , )
does not historically derive from the letter nun

1. HEBREW MARK LOWER DOT

In several places in the Hebrew Bible, manuscripts show dots above and below the letters, even in unpointed texts. These dots, called by the scholarly community *puncta extraordinaria*, are of obscure meaning, but are always included in ritual and scholarly texts, as well as the ritual scrolls used in Jewish ceremony. They also appear in various non-Tiberian documents (Tiberian orthography became the standard for Hebrew, but there had been others before it).

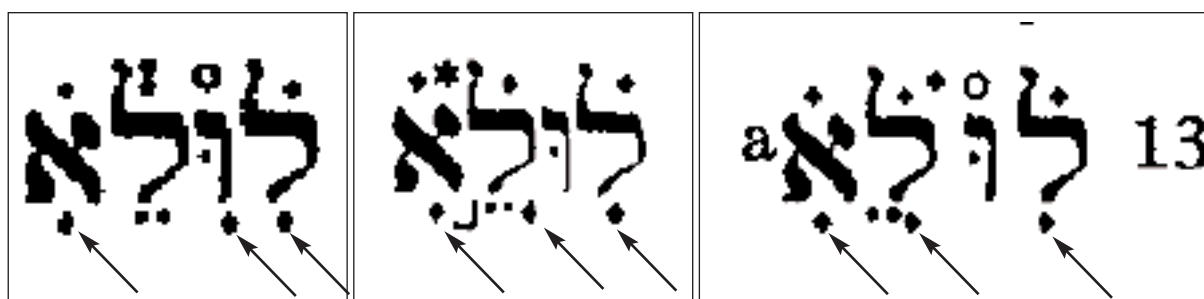
The upper puncta appear 53 times in 16 words (Genesis 16:5; 18:9; 19:33; 33:4; 37:12; Numbers 3:39; 9:10; 21:30; 29:15; Deuteronomy 29:28,28; 2 Samuel 19:20; Psalms 27:13; Isaiah 44:9; Ezekiel 41:20; 46:22). And in fact, the upper punctum is already encoded in UCS, as U+05C4 HEBREW MARK UPPER DOT. But the lower punctum, which occurs only in Psalms 27:13 (see scans below), due to its even greater rarity, is as yet uncoded. In order to represent the text of the Hebrew Bible properly, and for parity with the upper punctum, the lower one needs to be encoded also. Although there already is a codepoint for U+0323 COMBINING DOT BELOW, this is not acceptable for the lower punctum, because of the use of other dots above Hebrew letters when used as numbers in some contexts. The lower punctum is paired with the upper punctum, which is encoded in the Hebrew block: 05C4. It is probably acceptable to use 0307 and 0308 for Hebrew number dots, but 0323 would be confusing, especially if 0307 is used for hundreds and 05C4 is used for the upper punctum: it will make it more likely for people to get mixed up and think that

0307 is the punctum and 05C4 is for hundreds, especially if they're familiar with the single dot for hundreds but not the double dot for thousands. Also, many printers (for instance, see the BHS example below) make the puncta rather diamond-shaped instead of round like most other dots (Similarly, we suggest that the explanatory text “Upper *punctum extraordinarium*” be added to the listing for U+05C4).

In general, the puncta are different from most other Hebrew marks in a qualitative sense. Hebrew vowels are optional, and accents are so optional as to be hardly written at all outside of restricted contexts. However, the laws regarding the writing of a Torah scroll are more precise in their prescription. In a ritual Torah scroll, *only* the letters are written: not the accents, not the vowels, not the sof-pasuq punctuation, not verse-numbers, nothing else. Just the letters, spacing (as indicated by traditional divisions)... and the puncta (and the NUN HAFUKHA, see below). A ritually fit scroll must have them, and may not have anything else.

Examples

Since the lower punctum occurs only on one word in standard Hebrew Bibles, it suffices to show examples of that word from various sources.



Snaith Bible (1958)

Letteris Bible (1866)

Biblia Hebraica Stuttgartensia (1997)



IDF Tanakh (1982)

Leningrad Codex (1009)

Aleppo Codex (c. 920)

Here is an example of the upper punctum, taken from a modern Torah scroll in current use (note highlighted word). Since only the Pentateuch (and the Book of Esther) are ritually required to be written and read from a scroll, it is not common to find scrolls of Psalms (the only book which has the lower punctum), so the example given must be from the Pentateuch (Numbers 3:39), and thus shows an upper punctum. But the example is still valid, as it demonstrates the special nature of the puncta: that even when *no* other non-letter marks are written, in ritual contexts, the puncta remain, and this is true of the lower punctum as well.



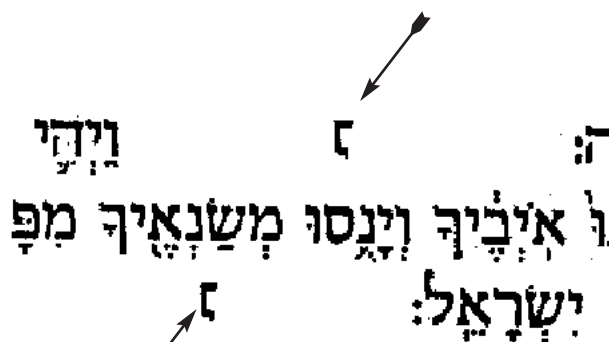
2. HEBREW PUNCTUATION NUN HAFUKHA

In two places in the Hebrew Bible, an unusual character appears, often in the form of a letter nun (U+05E0 HEBREW LETTER NUN) that is reversed or inverted or turned (depending on the particular scribe or printing). Although it is referred to in Judaic literature as *nun hafukha* (“reversed nun”), it plainly does not function as any sort of letter in the text. It is not part of a word, and it is not read aloud in any sense: it is just a mark that is written, and thus we classify it as punctuation and not a letter. In many manuscripts, it does not even resemble a transformed nun at all, and when it does, as mentioned, sometimes it appears reversed, sometimes inverted, sometimes turned 180°, etc. All these glyphic variants make any official name such as “REVERSED NUN” or “INVERTED NUN” potentially misleading, so we choose simply to give it the name by which Rabbis have referred to it for centuries: *nun hafukha*. It might be appropriate to gloss this in the Unicode standard as “not historically a *nun*, though conventionally named so” and “also called *nun m’zuneret*”.

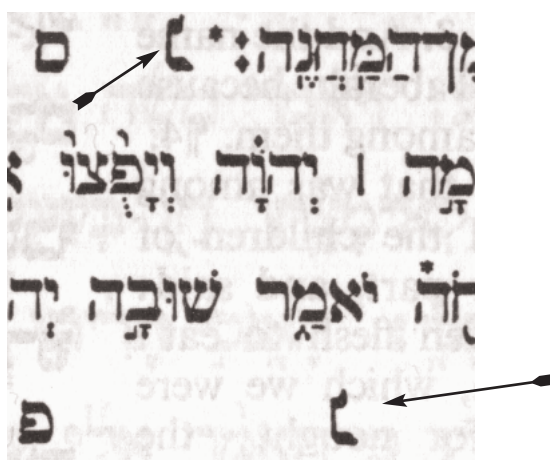
The NUN HAFUKHA appears in the book of Numbers, around verses 10:35-36, apparently setting them off from the surrounding text. According to the Talmud (Shabbat 115b-116a):

Our Rabbis teach: “And it came to pass, when the ark set forward, that Moses said...” — the Holy One, Blessed be He, made signs for this passage above and below, to show that this is not its place. Rabbi [Judah the Prince] says, not for this reason, but because it is an important book in its own right.... Rabbi Jonathan says, “Wisdom... hath hewn out her seven pillars...” (Proverbs 9:1) these are the seven books of the Torah [i.e. Genesis, Exodus, Leviticus, Numbers up to 10:35, Numbers 10:35-36, Numbers after 10:36, Deuteronomy] (translation mine (Mark Shoulson))

It also appears seven times in the book of Psalms, chapter 107, separating verses (or, in some versions, in the margin).



Stone Edition Tanakh (1996)



J. H. Hertz Pentateuch (1980)

לְתוֹר לָהֶם מְנוּחָה: 34^{ba} וְ

חֲנָה: וְ 35 וְהִי בְנִי

בֵּית וַיִּנָּסוּ מִשְׁנֵאִיד מִפְּנֵיךְ

וּלְפִי יִשְׂרָאֵל: וְ פ

Biblia Hebraica Stuttgartensia (1997)

וַיְחַנּוּ עַל יָדָם וּמִסְבָּנָם עִם
מִזְבֵּחַ חֲנָה:
וַיַּחֲבִיבֵם עַל הָאָרֶץ
וַיֹּאמְרוּ מִשָּׁח קוֹמַח וַיְחַנּוּ
וַיַּעַז אֶבְיָד וַיִּנָּסוּ מִשְׁנֵאִיד
מִפְּנֵיךְ וּבְנֵיךְ וַיֹּאמְרוּ שׁוּבוּ
וַיְחַנּוּ רַב־כּוֹת אֶלְפֵי יִשְׂרָאֵל
וַיַּחֲדִי חַ עֵס כְּמֵתָאנִם

Leningrad Codex (1009)

וַיַּחֲבִיבֵם
וַיֹּאמְרוּ
וַיִּנָּסוּ
וַיֹּאמְרוּ
וַיִּנָּסוּ
וַיֹּאמְרוּ

Aleppo Codex (ca. 920) – Psalm 107

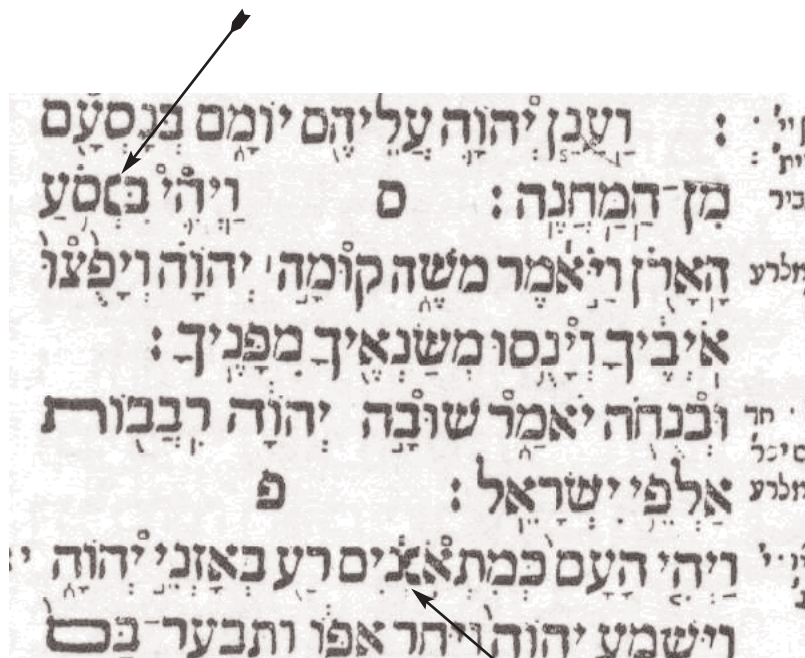
יְהוָה כְּתוּב ל כְּרִנָּה: יוֹרְדֵי הַיָּם בָּאֹ
 מְלֹאכָה כְּמִסְרָבִים: ל וְסִיף
 מַעֲשֵׂי יְהוָה וְנִפְלְאוֹתָיו בִּ
 וַיֹּאמֶר וַיַּעֲמֵד רוּחַ סְעָרָה ל 5 ר
 גְּלִיּוֹ: יְעֹלֹ שָׁמַיִם יְהוָה ל 5 ה
 נִפְשָׁם בְּרַעַה תִּתְמוּגָג: ל 5 ח
 בְּשִׁכּוֹר וְכָל־חֲכָמַתְסַתָּ: ל 5
 וַיַּעֲקוּ אֶל־יְהוָה ל ←

Biblia Rabbinica (1525) – Psalm 107

לָהֶם מַנוּחָה וְעֵצָן יְהוָה עָלֵיהֶם יוֹמָם בְּנִסְעָם בֵּין
 הַמּוֹחֲזָה ל וַיְהִי בְּנִסְעָה
 הָאָרֶץ וַיֹּאמֶר מִשֶּׁהָ קוֹמוֹת יְהוָה וַיַּפְעֵנוּ אִיבִיךְ וַיִּנָּסוּ
 מִשִּׁנְאִיךָ מִפִּנֵּיךְ וּבְנִחָה יֹאמֶר עֹבְדֵי יְהוָה
 רַבּוֹת אֲנִפֵּי יִשְׂרָאֵל ל

Modern Torah Scroll

It is true that there are some scrolls and even printed editions (e.g. the Biblia Rabbinica, in Numbers, see below) which indicate the NUN HAFUKHA with actual inverted NUNS used as part of the text. We would argue, however, that such cases qualify not as instances of NUN HAFUKHA, which is never read, never part of a word, but is only a punctuation, but rather as instances of an ordinary NUN written in an idiosyncratic way. Indeed there are many such “special letters” in the Hebrew Bible, including (but not limited to) the various large and small letters, the famous “broken” letter VAV in Numbers 25:12, and also many lesserknown (and less-frequently observed) reversed, odd-shaped, or oddly-decorated letters in various traditions. These are all glyph variants of their respective letters, as they are treated as parts of words and read just like any other letters, thus they should be handled by special glyphs and/or higher-level markup.



Unicode Character Properties

05C5;HEBREW MARK LOWER DOT;Mn;220;NSM;;;;N;;;;;

05C6;HEBREW PUNCTUATION NUN HAFUKHA;Po;0;R;;;;N;;;;;