

Title: Response to “Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS” (L2/04-310)

Source: Peter Kirk

Status: Individual Contribution

Action: For consideration by the UTC

Date: 2004-08-02

This document is a response to the “Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS” submitted by Michael Everson and Mark Shoulson, Unicode document L2/04-310 and ISO/IEC JTC1/SC2/WG2 N2840. This proposal offers an alternative solution to the same problem addressed in the “New proposal on the Hebrew vowel HOLAM” submitted by a group including myself, Unicode document L2/04-307 (also available as <http://qaya.org/academic/hebrew/Holam3.pdf>). These comments should be taken as an extension and more specific clarification of some of the comments made in the section “Justification” of the latter proposal.

I wish to present to the UTC the following comments in response to the HOLAM HASER FOR VAV proposal:

1. The HOLAM HASER FOR VAV proposal refers, in sections C 2b and D, to discussion of these issues with the user community on the hebrew@unicode.org list, but it does not summarise the drift of this discussion. In fact the users of Hebrew script involved in this discussion have been almost unanimously opposed both to the principle of solving the problem by encoding a new character, and specifically to the solution in the HOLAM HASER FOR VAV proposal. There is no general request or requirement for the proposed new character from the user community, but rather a general opposition to it. The only exceptions have been a few users who have argued reluctantly that such a solution might be preferable because of inadequacies of current rendering engine implementations; but standardisation should not be driven by accommodation to existing implementations. The raw archives of these discussions are accessible at <http://www.unicode.org/~ecartis/hebrew/>.
2. The most serious objection to encoding the new character HEBREW POINT HOLAM HASER FOR VAV is that this character is identical to the existing U+05B9 HEBREW POINT HOLAM both in its semantics and in its visual appearance. The answers in sections C 8a and C 10c of the HOLAM HASER FOR VAV proposal are seriously misleading: there is no general distinction in position, size or height between the proposed character and the existing one. The reference glyphs in the proposal are also misleading. There is a graphical distinction only when these combining characters are combined with the base character U+05D5 HEBREW LETTER VAV; in this case U+05B9 HEBREW POINT HOLAM is generally rendered further to the right, relative to the base character, than its usual position, whereas the proposal is for the new HEBREW POINT HOLAM HASER FOR VAV to retain its regular position. At the level of interpretation, the distinction here is in the VAV, which is a consonant in *Vav Haluma* (VAV with *Holam Haser*) but silent in *Holam Male*, rather than in the *Holam*, which has the identical function and pronunciation in both cases.

The image

Peter Kirk, Response to “Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS” Page 1 of 4

proposal is that two different characters should be used for the graphically identical combining mark when only the base characters are different, and when their interpretations are identical.

The HOLAM HASER FOR VAV proposal seems to treat the HOLAM in *Vav Haluma* as specially marked, whereas in fact the marked case, both graphically and semantically, is *Holam Male*. The logical implication of this is that it would be theoretically preferable to encode a new variant *Holam* character either for use only in the combination *Holam Male* or for all cases of *Holam Haser*. (The latter would be slightly more disruptive for the Hebrew Bible, in which of all occurrences of HOLAM about 47% are part of *Holam Male* and 53% are *Holam Haser*; however, *Holam Male* is relatively more frequent in pointed modern Hebrew, probably a little more frequent than *Holam Haser*.) This would correspond to the actual graphical distinction, and to a small semantic distinction between the functions of the *Holam* dots. However, such solutions have been unanimously rejected by both sets of proposers and on the hebrew@unicode.org list, because they would require an incompatible change to a large body of existing texts, in which both varieties of *Holam* dot are represented by the same character.

3. The HOLAM HASER FOR VAV proposal makes an inappropriate comparison between the case for the proposed character HEBREW POINT HOLAM HASER FOR VAV and the accepted proposal for HEBREW POINT QAMATS QATAN. These two cases are quite different, in several ways:
 - a. The graphical distinction between *Qamats Qatan* and *Qamats Gadol* is a novel one, used in only a very few recently published special purpose texts. There are no known texts represented in Unicode in which a distinction is attempted. But the graphical distinction between *Holam Male* and *Vav Haluma* is an ancient one, made consistently for more than 1000 years in exact editions of certain texts. The distinction continues to be made in a large minority of existing pointed Hebrew texts: not only liturgical texts as stated in the HOLAM HASER FOR VAV proposal, but also religious texts used for private and public study (including scholarly study by historians, linguists etc, and as the base text for translation), and modern poetic and educational texts. (See the examples given in the “New proposal on the Hebrew vowel HOLAM” L2/04-307.) And there are several existing Unicode texts in which *Holam Male* and *Vav Haluma* are distinguished by a variety of non-standard mechanisms.
 - b. The new character HEBREW POINT QAMATS QATAN has been provisionally accepted on the basis that it has both a distinct graphical form and a distinct interpretation as indicating a variant pronunciation. Neither of these applies to HEBREW POINT HOLAM HASER FOR VAV, which differs from U+05B9 HEBREW POINT HOLAM neither in form nor in interpretation.
 - c. HEBREW POINT QAMATS QATAN may be used with all Hebrew base characters. The proposed HEBREW POINT HOLAM HASER FOR VAV may be used only with a single base character.
 - d. My own preference would have been to encode HEBREW POINT QAMATS QATAN not as a separate character but as a variant of U+05B8 HEBREW POINT QAMATS. However, since *Qamats Qatan* can occur with any Hebrew base character and with other combining characters e.g. DAGESH, the only available mechanism for selecting a variant would be a Variation Selector, and use of Variation Selectors with combining characters is not permitted. But there is no such restriction with *Vav Haluma* because only a single base character is involved, and because the UTC has recently approved

an appropriate mechanism for selecting variant combinations (more connected and less connected renderings) of base characters with combining marks, by inserting ZWNJ or ZWJ after the base character.

4. The HOLAM HASER FOR VAV proposal states that “in texts which do not distinguish [*Holam Male* and *Vav Haluma*], HOLAM is used as the generic mark; reading rules only distinguish them. In texts which do make a distinction, the HOLAM HASER FOR VAV can be used.” But this is based on a clear misunderstanding of the nature of the distinction. In every *text*, there is a distinction between *Holam Male* and *Vav Haluma*; there are no ambiguous marks in the context of a text because the context disambiguates. Where the distinction is sometimes lost is in *rendering*. Now it is understood that in practice some texts will continue to be represented in Unicode with no distinction between *Holam Male* and *Vav Haluma*. But other texts will need to be represented, in the forms stored in databases and used for various processing, with distinct Unicode representations for these two. However, if the distinction is made in a particular text, that does not imply that it should always be made in rendering that text; it should rather be the decision of the rendering engine, based on the end user's choice of font etc, whether to render *Holam Male* and *Vav Haluma* distinctly. Fonts for general use for modern Hebrew would probably not want to make a distinction. If the HOLAM HASER FOR VAV proposal is accepted, this would imply that with the most commonly used fonts the new character HOLAM HASER FOR VAV would have to be rendered identically to the existing HOLAM in *all* contexts.
5. The HOLAM HASER FOR VAV proposal presupposes that the distinction between *Holam Male* and *Vav Haluma* is a semantic one. But this is actually debatable. A distinction which is only optionally marked cannot indicate a distinctive interpretation. Consider the pointed Hebrew words מִצְוֹת *mitzvot* “commandments”, with *Vav Haluma* pronounced [vo], and מַצּוֹת *matzot* “wafers”, with *Holam Male* pronounced [o]. (I have been forced to represent *Vav Haluma* and *Holam Male* in the same way in the Hebrew script forms of these two words.) Readers of pointed Hebrew are able to determine the correct pronunciations of these words, but not so much from the unreliable distinction between dot positions as from the rule that *Vav Haluma* must follow a vowel (including SHEVA as here) but *Holam Male* must follow a consonant. This shows that the semantic distinction is derived from the context within the whole word, and not from the position of the *Holam* dot. Therefore the graphical distinction between *Holam Male* and *Vav Haluma*, although important for the correct exact rendering of the text, is not actually semantically significant. This conclusion makes irrelevant the argument against the the “New proposal on the Hebrew vowel HOLAM”, made on the hebrew@unicode.org list, that solutions using ZWNJ and ZWJ are unsuitable for making semantically significant distinctions. (Anyway, ZWNJ is already used to indicate non-standard cursive joining behaviour which is semantically significant in Persian and other Arabic script languages.) On the other hand, if the distinction is not semantic, the Unicode character/glyph model implies that a new character should not be encoded and so that the HOLAM HASER FOR VAV proposal should be rejected.
6. It is argued in the HOLAM HASER FOR VAV proposal that a representation using ZWNJ is inappropriate for use in a script which is “not a ligating script”. But this ignores the fact that ZWNJ and ZWJ are explicitly defined, in *TUS* section 15.2, for use not only in regularly ligating scripts like Arabic and Indic scripts, but also for control of the ligatures in Latin script. The Latin and Hebrew scripts are alike in being not generally ligating scripts but in including occasional ligatures, such as U+FB01 LATIN SMALL LIGATURE FI and U+FB4F HEBREW LIGATURE ALEF LAMED. If solutions using ZWNJ are appropriate for Latin script, as explicitly stated in *TUS*, they should not be rejected as inappropriate for Hebrew script.
7. It is stated in the HOLAM HASER FOR VAV proposal that *Holam Male* “is no more a ligature

than ö is.” I accept that neither of these is a ligature in the conceptual sense which is probably intended by the Unicode definition of “ligature”, in <http://www.unicode.org/versions/Unicode4.0.0/b1.pdf>. On the other hand, both of the sequences <VAV, HOLAM> and <LATIN SMALL LETTER O, COMBINING DIAERESIS> are commonly implemented within rendering systems as ligatures, i.e. with “A glyph representing a combination of two or more characters.”

I accept that there is some confusion of terminology in the “New proposal on the Hebrew vowel HOLAM” L2/04-307, in the first paragraph after the heading “Justification”. This is because when drafting that proposal I was thinking in terms of ligatures as used in rendering engines. It would have been more in accordance with Unicode definitions to write in more generic terms of more and less connected renderings. That proposal states: “Because *Holam Male* is much more common than *Vav Haluma*, this ligature is taken as the default. The function of ZWNJ in the proposed representation of *Vav Haluma*, ... is to inhibit this ligature formation or equivalently to select the less connected rendering of VAV with HOLAM, ...”. This would have been better worded: “Because *Holam Male* is much more common than *Vav Haluma*, this more connected rendering is taken as the default. The function of ZWNJ in the proposed representation of *Vav Haluma*, ... is to select the less connected rendering of VAV with HOLAM, ...”

I accept also that the “New proposal on the Hebrew vowel HOLAM” is not theoretically perfectly neat. It does involve some slight stretching of definitions, but only in a benign way: while *Holam Male* may not strictly be a ligature in the conceptual sense, it is perfectly possible within the Unicode model to treat it as if it were a ligature, and the results of doing so seem to be acceptable. It is unfortunately impossible to find a theoretically perfect solution which does not also undermine the stability of large amounts of data. Nevertheless, the theoretical imperfections of the “New proposal on the Hebrew vowel HOLAM” are far less than those of the HOLAM HASER FOR VAV proposal, which ignores basic Unicode principles such as the character/glyph model.

8. The HOLAM HASER FOR VAV proposal is essentially a resubmission of an early Unicode proposal for a LEFT HOLAM character, which was provisionally assigned to U+05BA but never formally accepted (see <http://www.unicode.org/mail-arch/unicode-ml/Archives-Old/UML019/0559.html>). The UTC should review its reasons for previously rejecting this proposal before accepting it now. Nevertheless, this provisionally accepted character seems to have been widely implemented, including in the Microsoft 2003 distributions of the fonts Arial and Times New Roman (version 3.00 of each) as well as in the Fontographer program. One implication of this is that if the new character HEBREW POINT HOLAM HASER FOR VAV is accepted, it would be sensible to allocate it to U+05BA, and to reallocate the proposed and provisionally allocated HEBREW POINT QAMATS QATAN.

I conclude that the HOLAM HASER FOR VAV proposal L2/04-310 is badly thought out. It is based on serious misunderstandings of the Hebrew script and a false analogy. The grounds given for rejecting the alternative proposal do not stand up to close scrutiny. The HOLAM HASER FOR VAV proposal is also apparently not acceptable to the user community for whose benefit it is supposedly being proposed. It should therefore be rejected in favour of the “New proposal on the Hebrew vowel HOLAM” L2/04-307, which has widespread support among the community of users of both ancient and modern Hebrew.