Title:	On the Hebrew mark *METEG*	
Source:	Peter Kirk	**L2/04-194**
Status:	Individual Contribution	
Action:	For consideration by the UTC	
Date:	2004-06-05	

Summary

The Hebrew combining mark METEG is in origin part of the Hebrew accent system, although it is sometimes used in otherwise unaccented texts. It is unique among the Hebrew accents in that its positioning relative to certain Hebrew vowel points is not fixed or determined by the context. When it occurs in the same combining sequence (i.e. associated with the same base character) as a vowel point below the base character, the METEG is usually positioned to the left of the vowel point. However, in some printed texts and manuscripts it is sometimes positioned to the right of the vowel point. Furthermore, when the vowel point is one of the three *Hataf* vowels, each of which consists of two separate elements, in some printed texts and manuscripts the METEG is generally, but not always, positioned between the two elements of the *Hataf* vowel.

Thus, in some printed texts and manuscripts METEG appears in two different positions relative to many vowel points, and in three different positions relative to some vowel points. The origin of this distinction dates back at least 1000 years, and it is still made in some printed texts of the Hebrew Bible. Its semantic significance is uncertain. However, other modern Hebrew Bible editions make no such distinction.

The current proposal is for a defined mechanism for making the distinction between these positions of METEG in plain text. According to the proposal, METEG to the right of a vowel point should be indicated by placing the METEG character before the vowel point in the character stream, and inserting CGJ between these two to avoid canonical equivalence to the regular order and to inhibit canonical reordering. Positioning of METEG relative to *Hataf* vowels, to the left and medially, should be specified with ZWJ and ZWNJ, with the default position when neither of these characters is used to be determined by the rendering system and font.

As this proposal involves specific use of ZWJ, ZWNJ and CGJ, it is being presented to the UTC for approval.

Background

The Hebrew mark METEG, i.e. U+05BD HEBREW POINT METEG, consists of a short vertical line written underneath the centre of the base character. It is in origin part of the Hebrew accent system, which is used in full primarily for Hebrew Bible texts. However, it is sometimes used in otherwise unaccented texts, either to indicate stress or secondary stress, or to disambiguate certain otherwise ambiguous forms. For this reason METEG is included in the Israeli national standard for vowel points and miscellaneous marks (SI 1311.1), and not only in the standard covering accents (SI 1311.2) (see http://qsm.co.il/Hebrew/stdisr.htm for further details); and because of this in Unicode METEG is called a POINT rather than an ACCENT, and is included among the miscellaneous points and marks rather than among the accents in the Hebrew block.

In fact two major functions within the accent system have been unified, for Unicode purposes, in this one character METEG, because they share the same glyph. The first function is *Silluq*, which marks the stressed syllable in the last word of a verse; the second is *Meteg* proper, also known as *Ga'ya*, which has several functions including marking secondary stress. *Silluq*, which is never found with a *Hataf* vowel, is apparently always positioned to the left of any single vowel, and so almost all of the positioning variations described in this proposal refer to *Meteg* proper. However, the METEG positioned between two vowels in a form of the name of Jerusalem is in fact *Silluq*.

In a fully accented Hebrew text, most words carry one or, less commonly, more accents. In most cases there is also a vowel point combined with the same base character as the accent, and quite commonly the vowel point and the accent have to compete for the same space. The general rule is that the vowel point remains in its normal position, and the accent is moved to avoid a collision. In most cases the movement is entirely

predictable, according to the following simplified summary: when the vowel is centred below the base character, an accent which would otherwise collide with it is shifted to the left; when the vowel is HOLAM, positioned above the top left of the base character, the otherwise colliding accent is shifted to the right, except at the end of a word when it is shifted to the left.

METEG is unique among the accents in that its movement to avoid collision with a vowel point is not predictable. The general rule is the same as for other accents centred under the base character: it is shifted to the left; and indeed in some editions of the Hebrew Bible text, apparently as far back as the Rabbinic Bible of 1524-25, it is always in this position. But in many biblical manuscripts including the earliest fully pointed ones (the Aleppo and Leningrad codexes), and also in modern printed Bible editions based on these manuscripts, the position of METEG varies. In a small minority of cases, mostly at the beginning of a word, METEG is written to the right of a vowel point. Furthermore, in some texts, on the rather few occasions when METEG is combined with one of the *Hataf* vowels (HATAF SEGOL, HATAF PATAH and HATAF QAMATS), each of which consists of two separate elements, METEG is commonly written medially, i.e. between the two elements of the Hataf vowel; however, even more rarely METEG may be written either to the left or to the right of a *Hataf* vowel. In the Aleppo codex, however, METEG seems to be most commonly written not medially but to the right of *Hataf* vowels.

The significance of these positioning distinctions is uncertain. They may have originally arisen because scribes wrote METEG in the most convenient available space. However, these distinctions have been copied from manuscript to manuscript and thence into printed editions for more than 1000 years. In addition, at least one 20th century Bible edition used METEG positioning to make a rather different distinction: apparently METEGs added by the editors for consistency were positioned to the right. Thus the METEG positioning distinctions should be considered part of the plain text of the Hebrew Bible. This is a plain text issue, not one to be handled by markup, because it relates to character positioning and plain text legibility; it is also outside the scope of known markup schemes.

In the electronic text of the Hebrew Bible based on the Leningrad Codex, METEG occurs 41,311 times, of which 905 are to the right of a vowel point and 78 are medial within a *Hataf* vowel. With *Hataf* vowels, METEG occurs six times to the left and twice to the right, as well as 78 times medially (in fact none of these are with HATAF QAMATS). The statistics for other Bible editions may vary considerably.

Samples

Codex Leningradensis (1006-7) (note the ambiguous positioning of some of the METEGs, marked in green)

Lisbon Bible (1492) (verses 7-8 only, note that right METEG is used in some places where left METEG is used in *BHS*)

וַיְהִי מַבְדִּיל בֵּין מַיִם לַמָּיִם : וַיַּעַשׂ אֱלֹהִים אֶת־הָרָקִיעַ וַיַּבְדֵּל בֵּין הַמַּיִם אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר מֵעַל לָרָקִיעַ וַיְהִי־כֵן : וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמָיִם וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם שֵׁנִי : פ וַיֹּאמֶר אֱלֹהִים יִקָּווּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם אֶחָד וְתֵרָאֶה הַיַּבָּשָׁה וַיְהִי־כֵן : וַיִּקְרָא

Rabbinic Bible (1524-5) (note that there are extra METEGs, and that all METEGs are left METEG)

7 בֵּין מַיִם לַמָּיִם : 7 וַיַּעַשׂ אֱלֹהִים אֶת־הָרָקִיעַ וַיַּבְדֵּל בֵּין הַמַּיִם אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר מֵעַל לָרָקִיעַ וַיְהִי־כֵן : 8 וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמָיִם וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם שֵׁנִי : פ 9 וַיֹּאמֶר אֱלֹהִים יִקָּווּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם אֶחָד וְתֵרָאֶה הַיַּבָּשָׁה וַיְהִי־כֵן : 10 וַיִּקְרָא אֱלֹהִים לַיַּבָּשָׁה אֶרֶץ וּלְמִקְוֵה הַמַּיִם

Biblia Hebraica Stuttgartensia (1976)

Figure 1: Right METEG (marked in red) contrasted with regular METEG (marked in blue), sometimes in otherwise identical words, in ancient and modern editions of the Hebrew Bible - these words are from Genesis 1:7-9.

Codex Leningradensis (1006-7)	Lisbon Bible (1492)	*Biblia Hebraica Stuttgartensia* (1976)

Figure 2: Medial METEG in ancient and modern editions of the Hebrew Bible - this word is from Leviticus 21:10.

Aleppo Codex (10th century CE) (note that right METEG is used where *BHS* has medial METEG, marked in green)

Codex Leningradensis (1006-7)

Biblia Hebraica Stuttgartensia (1976)

Figure 3: Left METEG (marked in blue) contrasted with medial METEG (marked in red), both with a *Hataf* vowel - these words are from Job 39:10-11

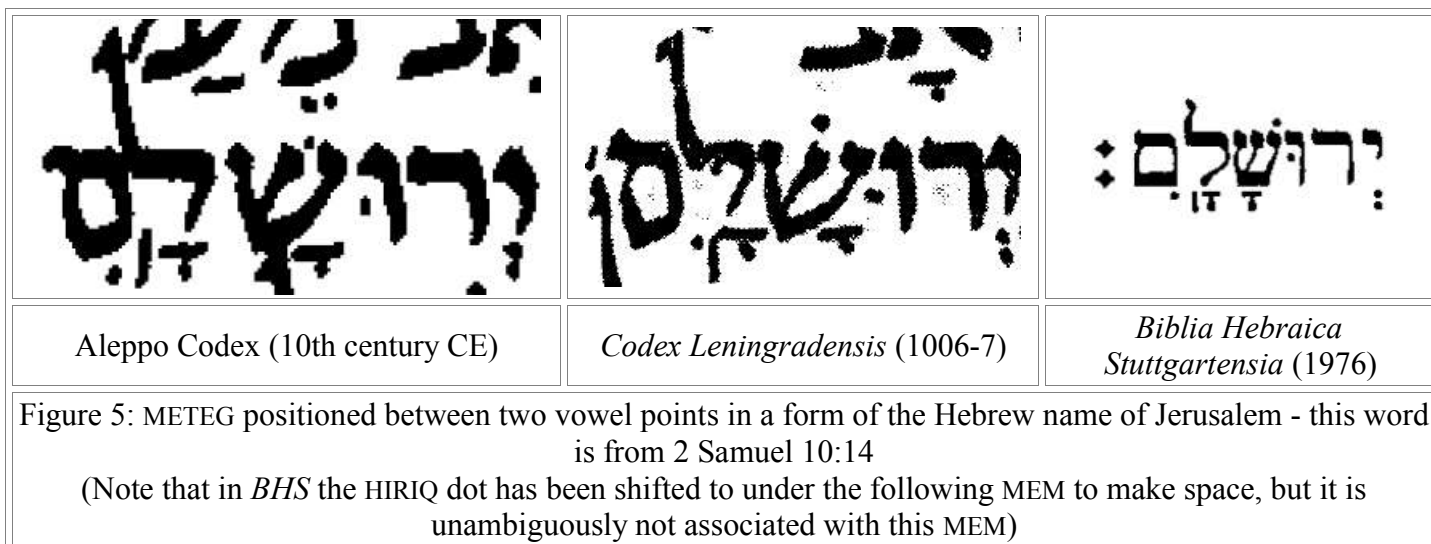|		
Aleppo Codex (10th century CE)	*Codex Leningradensis* (1006-7)	*Biblia Hebraica Stuttgartensia* (1976)

Figure 4: Right METEG with a *Hataf* vowel - this word is from Psalm 85:7

Aleppo Codex (10th century CE)	*Codex Leningradensis* (1006-7)	*Biblia Hebraica Stuttgartensia* (1976)

Figure 5: METEG positioned between two vowel points in a form of the Hebrew name of Jerusalem - this word is from 2 Samuel 10:14
(Note that in *BHS* the HIRIQ dot has been shifted to under the following MEM to make space, but it is unambiguously not associated with this MEM)

METEG Positioning and Unicode

The Unicode combining class mechanism was designed to allow a graphical and semantic distinction to be made between different orderings of combining marks which interact typographically, by assigning such marks to the same combining class, but to disallow any such distinctions between different orderings of other combining marks by defining the orderings to be canonically equivalent. This mechanism works well when the combining classes are correctly assigned. However, it breaks down when marks which actually do interact typographically are assigned to different combining classes, as unfortunately the Unicode stability policy forbids changes to existing combining class allocations even when these demonstrably contradict the definitions in the standard.

A large number of Hebrew points, marks and accents are by default centred under the base character and so interact typographically when combined with the same base character. According to the normal rules these should all be assigned to combining class 220. However, although these accents, apart from METEG, are in class 220, the vowel points centred below and METEG are all assigned to different individual combining classes. This implies that for each vowel point <METEG, vowel point> is canonically equivalent to <vowel point, METEG>. In fact the latter is the normalised canonical order as METEG is in a higher combining class than any of the vowel points. So it is impossible to distinguish positions of METEG simply by the order of the characters.

Furthermore, even if the canonical classes could be adjusted, this would allow only two distinct orderings of METEG and a vowel point, not the three orderings required to include medial METEG with *Hataf* vowels.

It is therefore necessary to propose alternative mechanisms for distinguishing positions of METEG. In the light of discussions on the Unicode Hebrew list, it seems best to propose two rather distinct mechanisms, one for METEG positioned to the right and the other to distinguish between medial and left positioned METEG with *Hataf* vowels.

Proposed Encoding for Right *METEG*

The proposed encoding for right METEG, i.e. METEG positioned to the right of a vowel (including a *Hataf* vowel), uses CGJ, i.e. U+034F COMBINING GRAPHEME JOINER, to block canonical equivalence and reordering, according to a principle accepted by the UTC at its August 2003 meeting. The principle was accepted primarily to support the rare cases of two Hebrew vowel points under a single base character, but is equally applicable to the slightly less rare cases of right METEG.

According to this proposal, a base character with a vowel point and right METEG should be encoded <base character, METEG, CGJ, vowel point>. The commoner case of METEG to the left of the vowel point should be encoded <base character, vowel point, METEG> (or its not normalised canonical equivalent <base character, METEG, vowel point>), except when the vowel is a *Hataf* vowel, in which case the following section applies.

A particular font or rendering system may choose to render all METEGs as left METEG, according to the practice in some Hebrew Bible editions. In this case the sequence with CGJ should be treated as equivalent to the same sequence without CGJ. At this point there is a potentially complex interaction with any procedures to reorder for normalisation the character stream presented to the rendering system, because CGJ should block such reordering.

Proposed Encodings for *METEG* with *Hataf* Vowels

The rare case of right METEG with *Hataf* vowels is covered above. This combination should be encoded <base character, METEG, CGJ, *Hataf* vowel point>.

The proposed encodings for the other two positions of METEG with *Hataf* vowels are based on treating the *Hataf* vowel with medial METEG as a ligature of the vowel with a regular left METEG. For this purpose the regular mechanism should be used, in which formation of the ligature is promoted by insertion of ZWJ, i.e. U+200D ZERO WIDTH JOINER, and inhibited by insertion of ZWNJ, i.e. U+200C ZERO WIDTH NON-JOINER. Until recently this mechanism could be used only with base characters. However, a decision of the UTC in February 2004 has made it permissible to use ZWJ and ZWNJ between combining characters, apparently subject to the UTC's approval of each such case, which is being sought with this proposal.

The following encodings are proposed, according to this mechanism:

<base character, *Hataf* vowel point, ZWJ, METEG>	METEG should be positioned medially within the *Hataf* vowel, if possible
<base character, *Hataf* vowel point, ZWNJ, METEG>	METEG should be positioned to the left of the vowel point, if possible
<base character, *Hataf* vowel point, METEG>	METEG should be positioned according to the default for the font and rendering system

As before, a particular font or rendering system may choose to render all METEGs, including those with *Hataf* vowels, as left METEG. It may also choose to render all METEGs with *Hataf* vowels as medial METEG. In either case such implementations should ignore both ZWJ and ZWNJ, as well as CGJ, in these combinations.

Normalisation Issues

There is a potential problem with all of the above encodings, that they need to be stable under normalisation. As there are no precomposed Hebrew characters, apart from some presentation forms which are composition exclusions (and should never be used in conjunction with any of the encodings proposed here), the only relevant issue is canonical ordering of combining marks.

There are no relevant combining marks in non-zero combining classes less than those of the various Hebrew vowel points and of METEG. Any such marks are from specific different writing systems, and as such the rendering of combinations of them with Hebrew combining marks need not be defined precisely.

There is only one combining mark in a combining class between those of the Hebrew vowel points and that of METEG: DAGESH, i.e. U+05BC HEBREW POINT DAGESH OR MAPIQ. This is likely to occur commonly combined with the same base character as a vowel point and METEG (in fact it occurs about 75 times with right METEG in the electronic text based on the Leningrad codex, but never with medial METEG; for example, TAV with DAGESH, right METEG and PATAH in Genesis 6:14), and if positioned after METEG in any of the defined encodings it will be moved to before METEG by normalisation. Therefore each of encodings defined here needs to be extended by allowing for an optional DAGESH before METEG. This results in two possible positions for DAGESH in each combination including CGJ, ZWJ or ZWNJ, which are not canonically equivalent. DAGESH is logically more closely associated with the base character than is any vowel point or METEG, although this is not reflected in the incorrectly assigned combining classes. Because of this, it is proposed here that DAGESH should always be positioned before any CGJ, ZWJ or ZWNJ, and that all combining character sequences with DAGESH after any of these control characters should be treated as spelling errors. For example, in Genesis 6:14 the following sequence should occur: <TAV, DAGESH, METEG, CGJ, PATAH>; this should not be spelled <TAV,

METEG, CGJ, PATAH, DAGESH>.

There are many combining marks in combining classes greater than those of the various Hebrew vowel points and of METEG. These marks include Hebrew accents, RAFE, SIN DOT and SHIN DOT. Hebrew accents should always be ordered after any of the encodings specified here, and will remain separate from them during normalisation. In fact there are no examples of right or medial METEG combined with any other accent in the electronic text (in which MASORA CIRCLE is not coded), but in principle these may occur in other texts.

The situation with RAFE, SIN DOT and SHIN DOT is more complex, because these marks, like DAGESH, are logically more closely associated with the base character than is any vowel point or METEG, although again this is not reflected in the incorrectly assigned combining classes. It is therefore proposed here that these three marks should, like DAGESH, always be positioned before any CGJ, ZWJ or ZWNJ, and that all combining character sequences with any of these marks after any of these control characters should be treated as spelling errors. In fact the only such combination which occurs in the electronic text is right METEG with SHIN DOT. This occurs four times in combinations like SHIN with SHIN DOT, right METEG and QAMATS in Genesis 8:17; this should be spelled <SHIN, METEG, SHIN DOT, CGJ, QAMATS>, not <SHIN, METEG, CGJ, QAMATS, SHIN DOT>.

The remaining case is when METEG is combined with two or more vowel points in the same combining sequence. This occurs approximately 144 times in the Hebrew Bible, all in a form of the name Jerusalem as spelled in biblical Hebrew, and all the attested sequences in the electronic text of the Leningrad codex consist of LAMED combined with QAMATS, METEG and HIRIQ (but there may be other sequences in other texts). In all of these cases METEG is in fact *Silluq*, marking the major stress in the word, which is on the syllable with the QAMATS vowel, and in this case alone any accent may be substituted for METEG. In every such case the METEG, and indeed any other accent below the base character, should be positioned between the two vowel points. In order to represent QAMATS to the right of HIRIQ as required, it is necessary in this case to use an encoding with CGJ. Because of the position of the METEG, and because METEG is associated with the syllable with QAMATS and so with the actual base character whereas HIRIQ is in fact logically associated with an omitted base character, the proposed encoding for this combination is <base character, vowel point, METEG, CGJ, vowel point>, i.e. <LAMED, QAMATS, METEG, CGJ, HIRIQ>. In principle (although not in the attested text) DAGESH may be added before METEG, and RAFE, SHIN DOT or SIN DOT after it. But none of these should be positioned after CGJ; this should be treated as a spelling error.

The spelling rules proposed here should be considered as guidelines for users and implementers. It is not intended that they should be normative parts of the Unicode standard.

Summary of Proposed Encodings

Encoding	Example(s)	Recommended rendering
<base character, [DAGESH,] METEG, [RAFE+,] CGJ, vowel point>	<TAV, DAGESH, METEG, CGJ, PATAH> <SHIN, METEG, SHIN DOT, CGJ, QAMATS>	METEG should be positioned to the right of the vowel point, if possible
<base character, METEG, CGJ, vowel point, DAGESH+>	<TAV, METEG, CGJ, PATAH, DAGESH> <SHIN, METEG, CGJ, QAMATS, SHIN DOT>	This should be treated as a spelling error
<base character, *Hataf* vowel point, [DAGESH+,] ZWJ, METEG>	<ALEF, HATAF PATAH, ZWJ, METEG>	METEG should be positioned medially within the *Hataf* vowel, if possible
<base character, *Hataf* vowel point, ZWJ, DAGESH, METEG>	(Not found in the Hebrew Bible)	This should be treated as a spelling error
<base character, *Hataf* vowel point, ZWJ, [DAGESH,] METEG, RAFE+>	(Not found in the Hebrew Bible)	This should be treated as a spelling error

<base character, *Hataf* vowel point, [DAGESH+,] ZWNJ, METEG>	<HE, HATAF PATAH, ZWNJ, METEG>	METEG should be positioned to the left of the vowel point, if possible
<base character, *Hataf* vowel point, ZWNJ, DAGESH, METEG>	(Not found in the Hebrew Bible)	This should be treated as a spelling error
<base character, *Hataf* vowel point, ZWNJ, [DAGESH,] METEG, RAFE+>	(Not found in the Hebrew Bible)	This should be treated as a spelling error
<base character, *Hataf* vowel point, [DAGESH,] METEG, [RAFE+]>	<ALEF, HATAF PATAH, METEG>	METEG should be positioned according to the default for the font and rendering system
<base character, non-*Hataf* vowel point, [DAGESH,] METEG, [RAFE+]>	<SHIN, QAMATS, DAGESH, METEG, SHIN DOT>	METEG should be positioned to the left of the vowel point
<base character, vowel point, [DAGESH,] METEG, [RAFE+,] CGJ, vowel point>	<LAMED, QAMATS, METEG, CGJ, HIRIQ>	METEG should be positioned between the two vowel points
<base character, vowel point, METEG, CGJ, vowel point, DAGESH+>	(Not found in the Hebrew Bible)	This should be treated as a spelling error

Elements in [...] are optional. DAGESH+ means any one or more of DAGESH, RAFE, SHIN DOT and SIN DOT; RAFE+ means any one or more of RAFE, SHIN DOT and SIN DOT. These encodings are all normalised; all canonical equivalents should be rendered identically. Note that none of these encodings apply when the vowel point is HOLAM, because this is positioned above the base character and so does not interact with METEG.