Title: **Proposal to Reclassify Ethiopic Wordspace as a Space Separator (Zs) Symbol**
Source: **Daniel Yacob (Ge'ez Frontier Foundation)**
Status: **Individual Contribution**
Action: **For consideration by the Unicode Technical Committee**
Date: **2015-05-02**

Issue

The classification of the Ethiopic Wordspace (U+1361) as a punctuation symbol (Po) has had unintended and undesirable consequences on text layout. Layout engines, as found in word processing and web browser software, will apply horizontal spacing rules based upon a symbol's categorical identity. The "Po" classification of Ethiopic Wordspace leads to the treatment of a paragraph of text as a single contiguous unit rather than comprising individual words. Consequentially, software applications will not perform line breaking correctly allowing the wordspace to appear at the first symbol on a new line. Applications that can perform line breaking correctly (wrapping to the right of the symbol) are unable to justify text with wordspace as expected. This deficiency is most apparent with text under "full justification". These issues would be avoided if the Ethiopic Wordspace were treated as a regular Space during line layout calculations. Equivalency with Space is behaviorally appropriate for the symbol from a layout perspective.

A semantic argument for the reclassification of Ethiopic Wordspace as a space separator is grounded in the classic role of the symbol in literature. As the symbol name "Wordspace" implies, the sign is simply a word separator ("Z" category). Unlike general punctuation, the symbol plays no part in segmenting a sentence to disambiguate meaning as does, for example, Ethiopic and Latin comma, semicolon and colon. Nor does the Ethiopic Wordspace in any way modify or impact the words adjacent to it as does a hyphen in joining words, or like a period in abbreviating them. Given this application of the symbol, Ethiopic Wordspace should not be classified under a punctuation, "P", category.

Proposal

To help support the Ethiopic Wordspace's layout behavior meet with user expectations, and to align the functional role of the symbol with its character classification property, it is proposed that the U+1361 be assigned the "Zs" character property in place of the initial "Po" assignment. A best practices note is offered to assist vendors in supporting both the full justification requirements and the contextual difference of the symbol when not employed in the wordspace context.

Ethiopic Justification Discussion with Samples

Since the arrival of the printing press to Ethiopia in 1863 (Pankhurst, 1998), full justification of Ethiopic has been a common typesetting practice in Ethiopian, and later Eritrean, publishing houses. Earlier, Ethiopic justification rules are a feature of Hiob Ludolf's *Historia Æthiopica* which is noted as the first use of movable type for Ethiopic script (Ludolf, 1681). Prior to letterpress typography, calligraphic manuscripts rendered on parchment also featured full, or approximately full, justification. Though the latter likely reflects the scribe's desire not to waste a millimeter of available lateral writing space.

Two predominant forms of Ethiopic Justification developed and would be of value to the present day user community. Each form in turn has two sub-modes that govern spacing following an Ethiopic Full Stop. Both forms and sub-modes are discussed in this technical note.
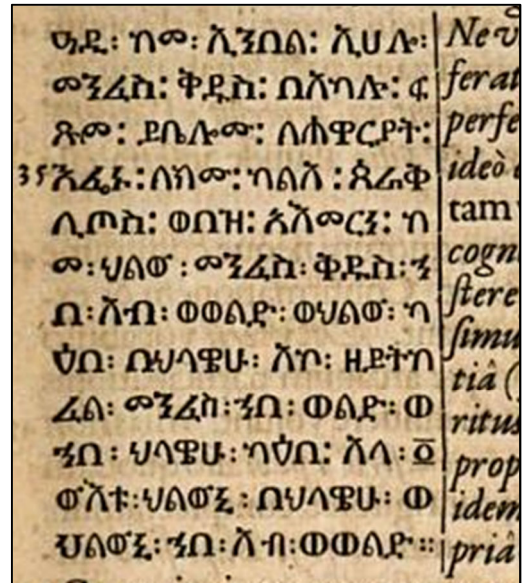

Figure 1: Ethiopic Justification in *Historia Æthiopica* (Ludolf, 1681)

Justification with Left Bounded Wordspace and Punctuation

In keeping with line justification for Latin script, the non-printed or "white space" between words is treated as stretchable. The width of the space symbol itself will be elongated to some aesthetic width value that may vary from space symbol to space symbol across a printed line. In Ethiopic Justification, the white space between the Ethiopic word separator and the words it separates is likewise allowed to stretch. This stretching of white space may be either symmetrical ("centered") or asymmetrical but in the latter case space stretching is always between the right side of the separator and the following word –referred to here as "left bounded".

In "left bounded" justification the word separator, which may be either a punctuation symbol or Ethiopic Wordspace, appears to adhere to the word to the left as if it were its final character.
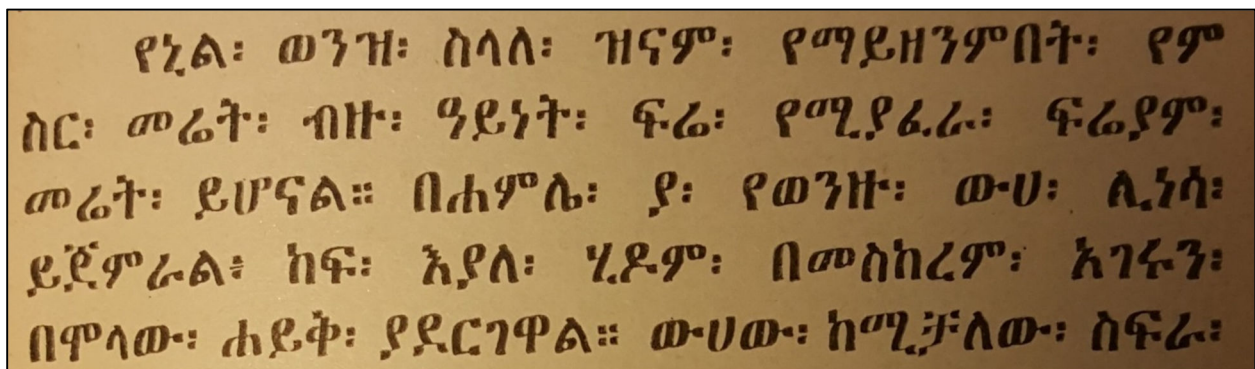

Figure 2: Ethiopic Justification in Left Bounded Style (Erikson, 1921 (1913 EC))

Justification with Centered Wordspace and Punctuation

In the second major form of Ethiopic Justification the white space around word separators is stretched equally on both the left and right sides; giving the appearance of the separator being centered between the words it divides. Figure 4 depicts the white space stretching in both forms of justification using Ethiopic Wordspace as an example, though the stretching rules apply equally to Ethiopic punctuation as well.
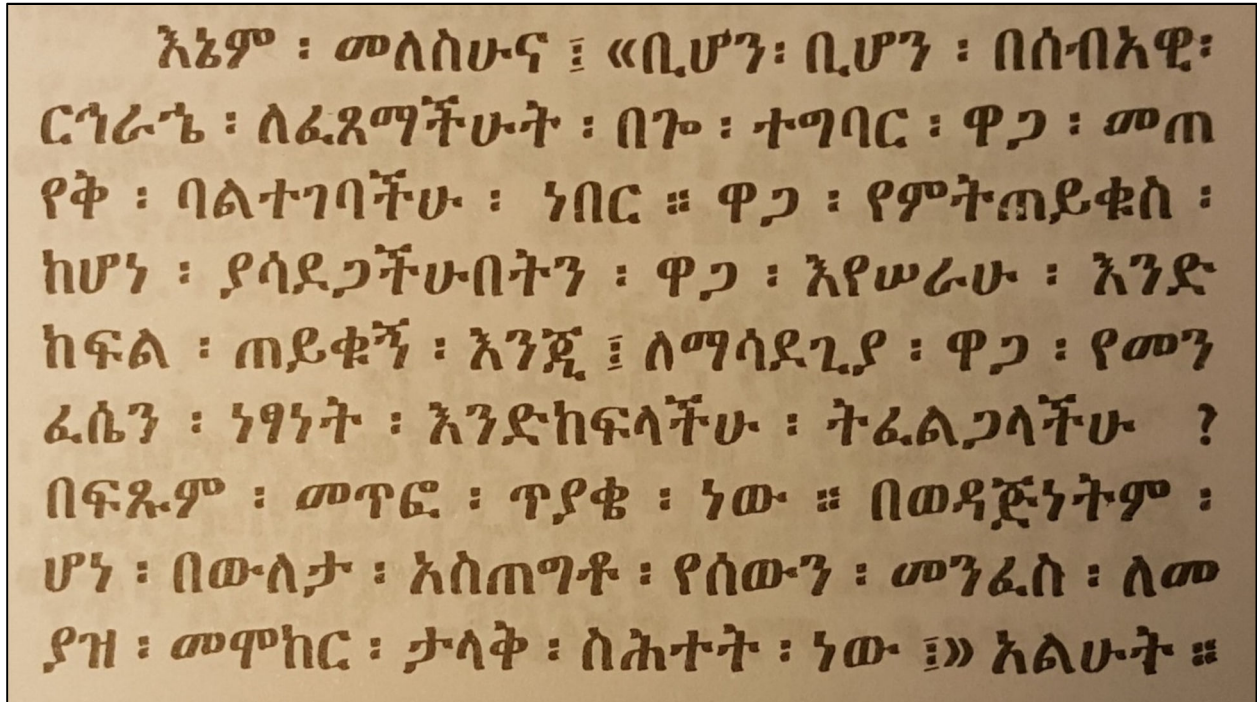

Figure 3: Ethiopic Justification in Centered Style (Gubenya, 1973 (1966 EC)).

To further illustrate the justification spacing applied to both Ethiopic Punctuation and Wordspace, Figure 4 presents white space stretching from the point of view of the symbol's typographic bounding box. Here the "design white space", the space between the visible symbol and the box border, is itself stretched as need to meet line justification:
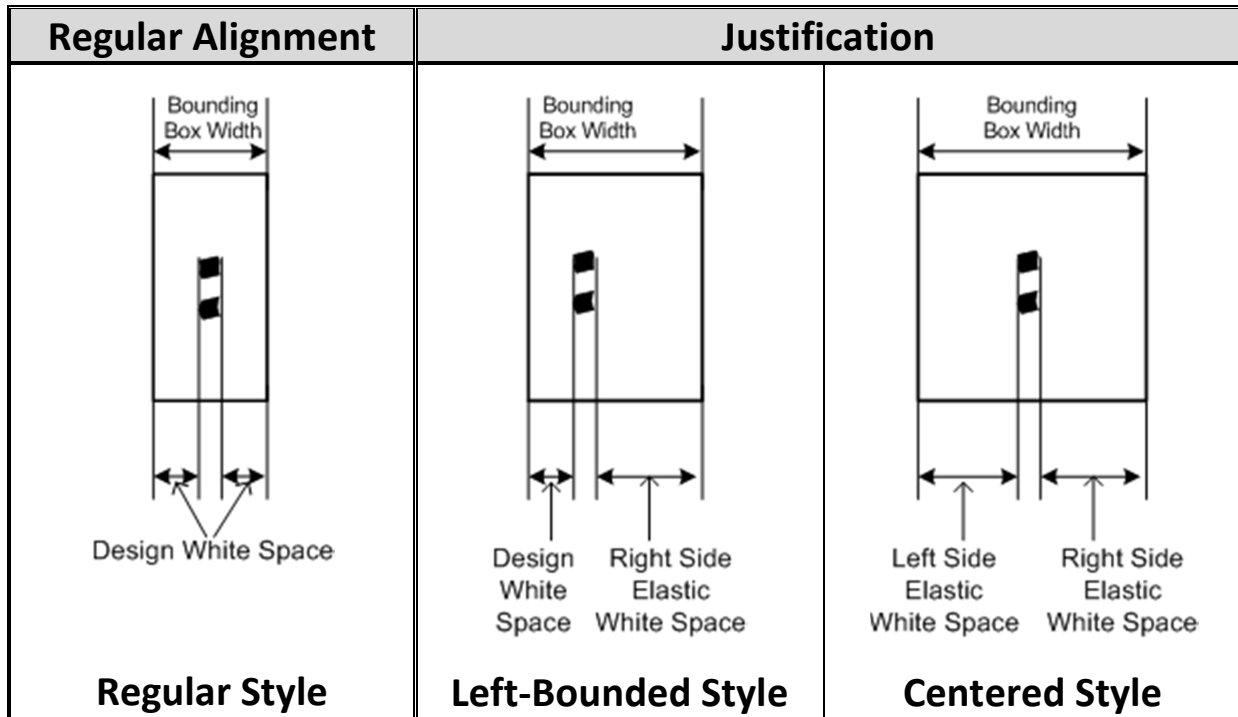
Regular Alignment	Justification	
Regular Style	Left-Bounded Style	Centered Style

Figure 4: Depiction of white space around Ethiopic Wordspace for three modes of text justification.

Spacing Threshold Following Full Stop

In the regular mode of Ethiopic Justification (both forms) the Ethiopic Full Stop will be treated equally with all other punctuation symbols. In a second mode, Ethiopic Full Stop will have special spacing rules applied to it whereby more separation space is allowed following the symbol and the start of the next word. In a sense, the right side space of the Full Stop is "more elastic" than in the regular mode. The elasticity rule and the visual effect are similar to that of the final line of a fully justified paragraph in Western text. When the final line of a paragraph of Latin script crosses a certain horizontal threshold, the line will become fully justified. Below that threshold the line will appear left aligned. The same rule appears to be applied to the Ethiopic Full Stop but on any line of the paragraph. An illustration of this sub-mode is depicted in the following:
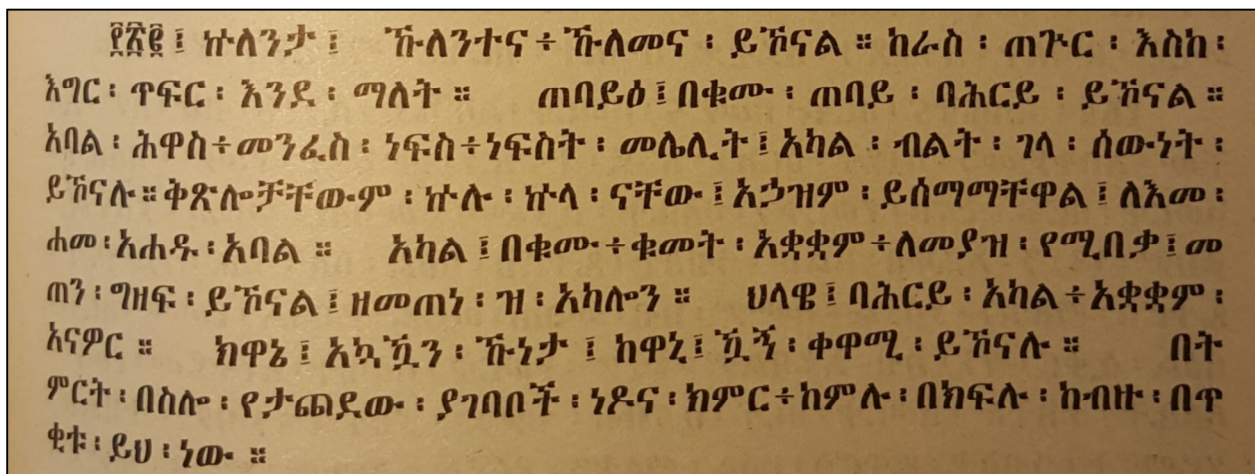

Figure 5: Justification Sample from *Maṣḥafa Sawāsew*, Page 159 (Kifle, 1955 (1948 EC)).

Short Comings Found in Electronic Typesetting Systems

To date, computer software that typesets text has applied justification rules for white space stretching that were designed to meet publishing requirements in the Western world. When the same rules are applied to Ethiopic text, the results are unsatisfactory as they do not meet user expectations. Largely responsible for the formatting dissonance when Western justification is applied to Ethiopic text, is the absence of a white space symbol in the writing system. There is no explicit white space symbol (in classic Ethiopic writing) to be "stretched".

Formatting algorithms will then process Ethiopic Wordspace as a punctuation symbol where word binding rules, rather than word spacing rules, will be applied. While still stretchable, "white space" in Ethiopic Wordspace is implicit rather than explicit. For a complete solution, software will ultimately need to be enhanced to stretch implicit space as required. Reclassifying Ethiopic Wordspace as a "Zs" symbol is expected to help alleviate justification issues and clears the way for software firms to implement comprehensive support for Ethiopic Justification. Since Ethiopic Wordspace interferes with justification in present day software, users may opt not to use it or may "pad" Wordspace and Ethiopic punctuation with explicit white space to obtain the formatting desired. The following samples depict formatting of Kidane Wolde Kifle's seminal work *Maṣḥafa Sawāsew* with a popular word processor (Kifle, 1955 (1948 EC)) under the limitations of Western spacing rules justification.
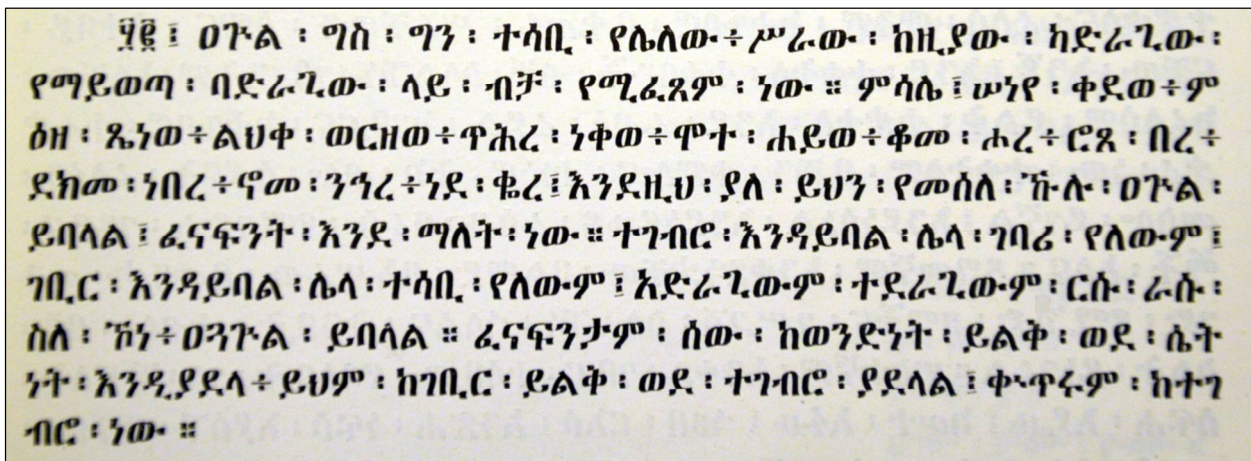

Figure 6: Full Justification Sample from *Maṣḥafa Sawāsew*, Page 65 (Kifle, 1955 (1948 EC)).
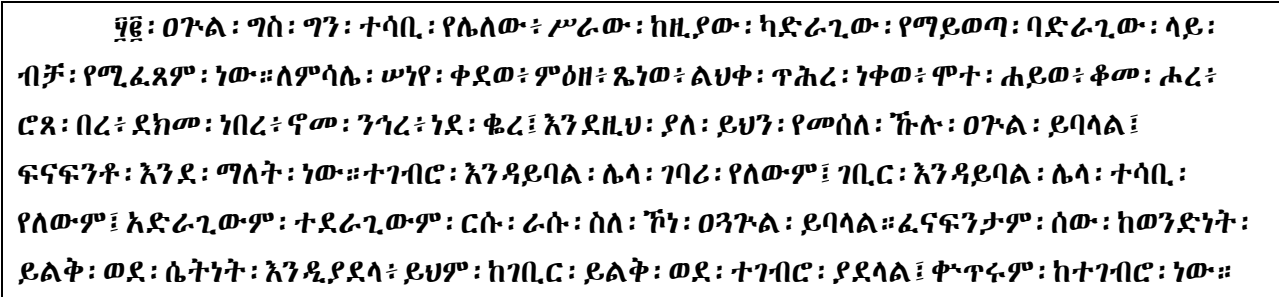

Figure 7: The sample from Figure 6 with a 12 point font and fully justified by Microsoft Word 2010 within a 6.5 inch margin with line-breaking rules applied. No spaces (U+0020) in sample.

፶፪ ዐጉል ግስ ግን ተሳቢ የሌለው፥ ሥራው ከዚያው ካድራጊው የማይወጣ ባድራጊው ላይ ብቻ የሚፈጸም ነው። ለምሳሌ ሠነየ ቀደወ፥ ምዕዘ፥ ጌነወ፥ ልህቀ ጥሕረ ነቀወ፥ ሞተ ሐይወ፥ ቆመ ሐረ፥ ሮጸ በረ፥ ደክመ ነበረ፥ ኖመ ንጎረ፥ ነደ ቄረ፤ እንደዚህ ያለ ይህን የመሰለ ኹሉ ዐጉል ይባላል፤ ፍናፍንቶ እንደ ማለት ነው። ተገብሮ እንዳይባል ሌላ ገባሪ የለውም፤ ገቢር እንዳይባል ሌላ ተሳቢ የለውም፤ አድራጊውም ተደራጊውም ርሱ ራሱ ስለ ኾነ ዐንጉል ይባላል። ፈናፍንታም ሰው ከወንድነት ይልቅ ወደ ሴትነት እንዲያደላ፥ ይህም ከገቢር ይልቅ ወደ ተገብሮ ያደላል፤ ቍጥሩም ከተገብሮ ነው።

Figure 8: The sample from Figure 7 with space (U+0020) replacing Ethiopic Wordspace. Space characters have additionally been added following Ethiopic punctuation.

In digital documents such as in web pages and eBooks, it is recommended that the appearance of either Space or Ethiopic Wordspace be configurable as a user preference. An easy to access "space" toggle button would enhance a viewing application's usability.

Ethiopic Semi-Justification

"Semi-Justification" is mentioned here as a historical note and to provide an interpretation of formatting observed in hand written manuscripts. In "Semi-Justification" the line spacing rules of either mode of Ethiopic Justification has been used, however, the text does not necessarily align perfectly flush with the right side margin. This practice is depicted in the following sample (Brita, 2012):
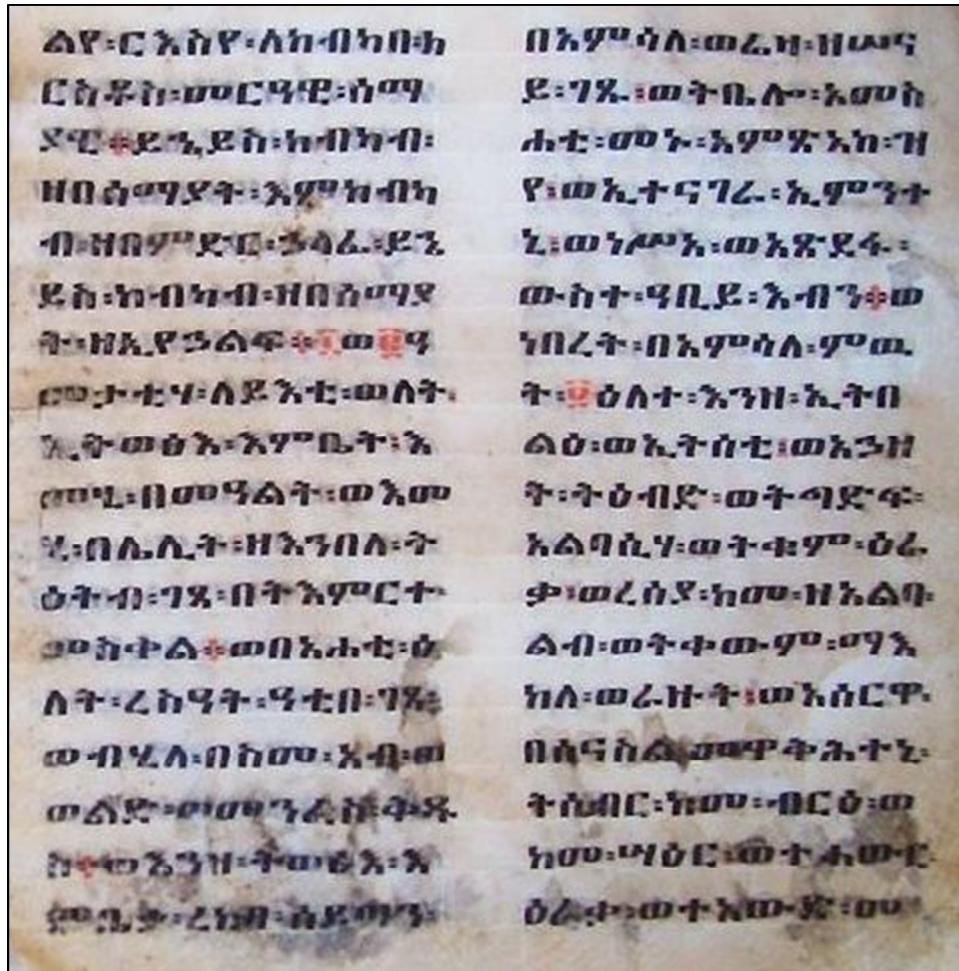

Figure 9: 'Adwā (Tegrāy), Dabra Madarā, Endā Abbā Garimā, ms. C2 IV 92, fol. 42r, Cropped (Brita, 2012).

Ethiopic Wordspace in the Context of Comma

In recent decades some communities have adopted a practice of employing the wordspace symbol as a comma when Space is used as the word separator. The interpretation of the symbol is then dependent on the context of the writing convention in use by the author. Accordingly, an application user setting could be offered to set the symbol context.

An alternative view point on this practice is that the Ethiopic Comma (፣ U+1363) is in fact in use by these user communities; however its *glyph* has decayed whereby the line segment is lost and so it visually coincides with Ethiopic Wordspace. Under this perspective, a simple solution would be modify an Ethiopic font for these users (perhaps adding an alternative glyph in an OpenType stylistic set) where the Ethiopic Comma character address and semantics remain intact though the visual form has been tailored to meet aesthetic needs.
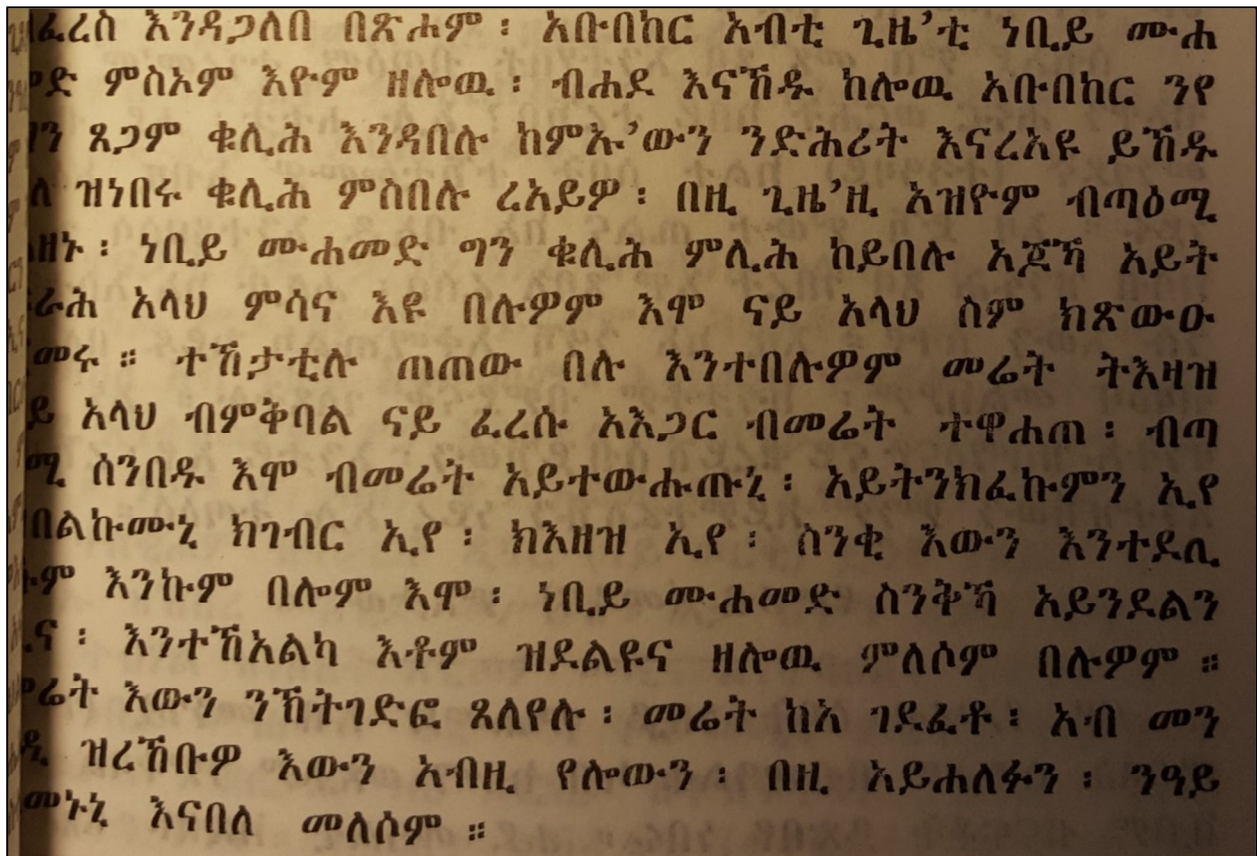

Figure 10: Ethiopic Wordspace in the context of comma (Abubeker, 1987 (1979 EC)).

Works Cited

Abubeker, J. H. (1987 (1979 EC)). *Ḥaṣir Tarix Nebiy Muḥamed.* Asmara: Selam Publishing House.

Brita, A. (2012, 11). *With the Saints Came the Manuscripts...* Retrieved 05 02, 2015, from Centre for the Study of Manuscript Cultures (CSMC): http://www.manuscript-cultures.uni-hamburg.de/mom/2012_11_mom_e.html

Erikson, O. (1921 (1913 EC)). *Yeografi LeEthiopia Lijoch Tiqim.* Asmara: Swedish Mission.

Gubenya, A. (1973 (1966 EC)). *Alweledem.* Addis Ababa: Brana Publisher.

Kifle, K.-W. (1955 (1948 EC)). *Maṣḥafa Sawāsew Wages Wamazgaba Qālāt Hadis.* Addis Ababa: Artistic Printers.

Ludolf, H. (1681). *Historia Aethiopica, sive brevis et succincta descriptio regni Habessinorum.* Frankfurt: prostat apud Joh. David Zunne.

Pankhurst, R. (1998, August 28). Ge'ez Literature, Church Libraries, and the Coming, from Europe, of the Printed Word. *Addis Tribute*.