

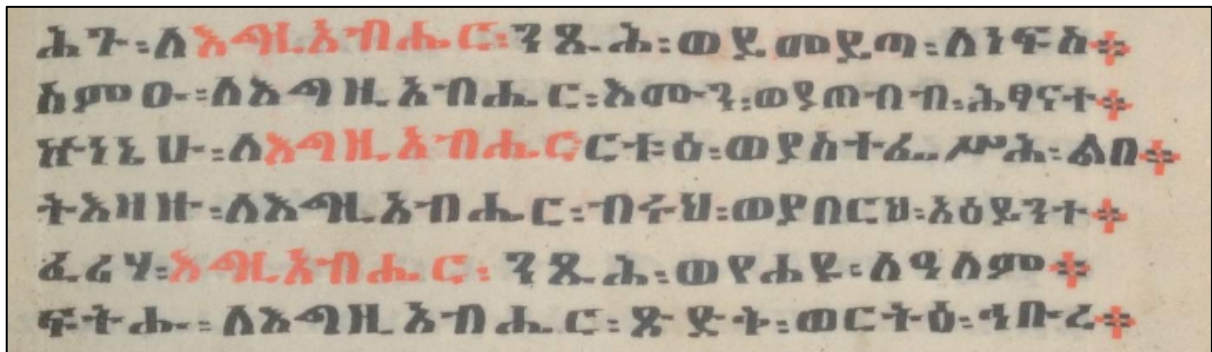
Application of the Zero Width Joiner Mark for Selective Ethiopic Ligation

Introduction

This note presents a convention for the elective use and preservation of ligatures in Ethiopic documents.

Ligatures are a regular occurrence in handwritten manuscripts in the Ge'ez language and are also found in manuscripts in Amharic and other languages prior to the 20th century. The most frequent of which is the well-known “ግዚ” ligature observed in the Ge'ez word for “God”, “እግዚአብሔር”. For people already familiar with the “ግዚ” ligature, it will be just as well known that its use is entirely elective and left to the scribe’s discretion. Within a single paragraph of text, the scribe may choose to use the ligature in one occurrence of “እግዚአብሔር” and not the next, and then use it again as the scribe sees fit. The same rule of elective ligation applies for all other ligatures and as a consequence makes their occurrence unpredictable.

Accordingly, a modern typist and digital preservation system will need control over where ligation occurs in a run of text. A typical sample is shown below from the Gunda Gunde collection, showing six occurrences of the word “እግዚአብሔር”; three with the “ግዚ” ligature and three without.



GG00053 fol. 20r, <https://w3id.org/vhmm1/readingRoom/view/789564>

Applying the Zero-Width Joiner (ZWJ)

The Zero Width Joiner (ZWJ, U+200D) offers a simple way to mark sequences of characters for ligation as well as to preserve the marking in plain text for transmission and machine analysis. The ZWJ symbol can, for example, easily be emitted via keyboard input and output in an OCR style digitization system. Using our sample word, “እግዚአብሔር”, the following demonstrates the byte sequence and visual appearance to the user in two examples:

እግ<ZWJ>ዚአብሔር will appear as እግዚአብሔር

also:

እግ<ZWJ>ዚአብ<ZWJ>ሔር will appear as እግዚአብሔር

and:

እግ<ZWJ>ዚአብ<ZWJ>ሔ<ZWJ>ር will appear as እግዚአብሔር

A Non-Exhaustive List of Sample Sequences and Their Results

Sequence	Appearance
[1203 ሃ] [ZWJ] [120C ሌ]	ኸ
[1203 ሃ] [ZWJ] [120C ሌ] [ZWJ] [1209 ሱ]	ኸፍ
[120C ሌ] [ZWJ] [1209 ሱ]	ሉፍ
[120C ሌ] [ZWJ] [1209 ሱ] [ZWJ] [12EB ያ]	፳፭
[120C ሌ] [ZWJ] [120A ሊ]	ሊሊ
[120C ሌ] [ZWJ] [120C ሌ]	ሌሌ
[120D ል] [ZWJ] [120C ሌ]	ልሌ
[121B ማ] [ZWJ] [122D ር]	ማሮ
[121D ም] [ZWJ] [12CD ው]	ምው
[121E ሞ] [ZWJ] [1219 ሙ]	ሞሙ
[1260 ቦ] [ZWJ] [136A ጀ]	ቦጀ
[1260 ቦ] [ZWJ] [120D ል]	ቦል
[1264 ብ] [ZWJ] [120D ል]	ብል
[1264 ብ] [ZWJ] [1214 ሔ]	ብከኔ
[1264 ብ] [ZWJ] [1214 ሔ] [ZWJ] [122D ር]	ብከኔሮ
[1264 ብ] [ZWJ] [122D ር]	ብሮ
[1286 ተ] [ZWJ] [1288 ጉ]	ተግ
[12A4 ኤ] [ZWJ] [1215 ል]	ኤል
[12A8 ከ] [ZWJ] [1215 ል]	ከል
[12AB ካ] [ZWJ] [1206 ሆ]	ካሆ
[12AD ክ] [ZWJ] [1219 ሙ]	ክሙ
[12AD ክ] [ZWJ] [122D ር]	ክሮ
[12AD ክ] [ZWJ] [122E ሮ]	ክሮፎ
[12CE ያ] [ZWJ] [12CD ው]	ያው
[12D8 ዘ] [ZWJ] [122D ር]	ዘሮ
[12DD ዝ] [ZWJ] [12E8 የ]	ዝየ
[12EB ያ] [ZWJ] [12E9 ዩ]	ያዩ
[1308 ገ] [ZWJ] [122D ር]	ገሮ
[130C ጊ] [ZWJ] [1215 ል]	ጊል
[130D ግ] [ZWJ] [1225 ሥ]	ግሥ
[130D ግ] [ZWJ] [122E ሮ]	ግሮፎ
[130D ግ] [ZWJ] [1261 ቡ]	ግቡ
[130D ግ] [ZWJ] [12DA ዘ]	ግዘ
[130D ግ] [ZWJ] [12DD ዝ]	ግዝ
[130D ግ] [ZWJ] [1309 ጉ]	ግጉ
[1313 ኧ] [ZWJ] [1310 ጉ]	ኧጉ
[133D ጸ] [ZWJ] [133A ጸ]	ጸጸ
[1345 ፅ] [ZWJ] [1341 ፀ]	ፅፀ
[1246 ቆ] [ZWJ] [1218 መ]	ቆመ
[124D ቆ] [ZWJ] [1218 መ]	ቆመ