Considering Ethiopic Character Classes

by Daniel Yacob

Abstract

This paper offers a tour of properties and features of the Ethiopic syllabary with discussion of their relevance to text and information processing, localization, and special purpose computing. Character entity names are developed from the traditional names of the letters and then used for the remainder of the paper. Ethiopic Gematria is covered as well as the more broadly practical issue of collation, which is shown to be language specific. Localized use of the syllabary finds its greatest impact on orthography practices where new POSIX style character classes have the potential to make text pattern matching manageable for the average developer, not only for the Ethiopic syllabary but for syllabaries in general.

Introduction: Discretizing a Cloud

The author feels compelled at this time to qualify the content of this paper before allowing the reader to read further. Perusing the material that follows, dear reader, you are embarking on a journey beyond the outer veil of Ethiopic that you know as a table in a Unicode handbook. Behind that veil lies, for lack of a better word, a gas. Gases can be fun but they can also be very confusing. It doesn't help that the gas in this case is both sentient and often times facetious...

The author can best describe his experience with Ethiopic script as having attempted to study and detail the features of a cloud. Consider looking at a cloud from a single vantage point and how that cloud appears to you. Someone else from a different vantage point, north, south, east or west from where you are would produce a description that you probably could not recognize as being of the same cloud.

From the same vantage point two people would likely come up with two different descriptions simply because we might perceive color and shading differently or we might focus our attention on different features. By the time you could even communicate your description to someone else, the cloud itself would have changed shape. It is a whole other story if you happen to be above or even inside the cloud itself.

So the cloud definitely changes with time and space and you have invisible influences like temperature and pressure acting upon it all along. All descriptions you would be given of the cloud, however different, are fully accurate and valid to the observers and no one description is necessarily more right or wrong than the other.

Ethiopic script and seemingly every aspect of the cultures that nurture it are highly resistant to discretization. This makes writing locales quite a challenge, which is great if you like a challenge, but it's not for the faint of heart. Unicode, for better or worse, mostly better, has come a long way to help standardize the cloud that the writing system is.

This paper attempts to go a step further and detail features of the Ethiopic syllabary from information collected across both space and time (literally) in the framework of a standardized cloud. Journey onward dear reader, look behind the veil. Be aware that the cloud knows your eyes are now cast upon it and has already changed shape.

Ethiopic Entity Names

Entity names are nothing new for Ethiopic letters and symbols but they are not often utilized. The letter names are in all likelihood as old as the letters themselves and in many cases even older. Ge'ez language, the first Abyssinian language to adopt writing, with its original 26 member alphabet is often compared with other Semitic languages and scripts from the same period. In particular Hebrew notably shares the largest number of terms with classic Ge'ez as well as present day languages. Hebrew script

with its 22 members also provides us with insight into the origins of the Ge'ez entity names, which we will see are often identical, or have a derivational relationship.

The Unicode definition for Ethiopic includes character names for Ethiopic elements based on transcriptions of the elements' phonetic value, or translations in the case of punctuation and numerals. These names are indeed useful identifiers and in many instances easier to work with than character codes. The Unicode names however were never intended for use as entity names as we see today in markup and other ASCII based document languages.

The motivation for deriving alternative entity names here is to produce a workable collection of names based on conventions from the character set's region of origin. The approach used is to construct a logical name in Ge'ez language first and to then transcribe the result into English. It is believed a naming scheme derived in this way will be more intuitive to software developers working with them on a day-to-day basis. It should also be noted before proceeding that the Ge'ez language plays an identical role in Eritrean and Ethiopian society, as does Latin for the Romance languages of Europe. Ge'ez is utilized in the same way by the sciences for the derivation of new terms in a way that is neutral to national and cultural boundaries. For these reasons Ge'ez is working language for the terminologies developed here.

Derivation of Names

For syllographic elements only the first syllable (or "form") of a series (which made up the original Ge'ez alphabet) are given proper names. As with other writing systems the names of the members evolved over time and experienced regional variations as well. In the present day there remains no consensus agreement on the element names, so our first task will be to come up with an average of some sort.

The approach used here is to come up with a weighted average where the terms recorded by Aba Gorgorios, a noted seventeenth century monk and savant, are selected first[10]. The Aba Gorgorios term is then abandoned in favor of another if the alternative term appears in four or more other references. Aba Gorgorios is given the most credibility for the combination of his stature in Ethiopian society and the age of his work being placed closer to the origin of the names themselves. While the Dillman reference cites even older names found in a translation of the New Testament in Rome, it is not given a greater weight as little else is known about the material by the present author[6]. Hebrew is shown in the following table only for comparison value and does not influence the outcome selection process.

Hebrew	1548 Rome (Dillman 1899)	አባ ግርጋርስ 1661	Praetorius 1879	Dillman/ Criton 1907	መዝገበ ፊደል 1934	Chaine 1938	መጽሐፍ፡ ጥበብ 1948	ኪዳነ፡ ወልድ፡ ክፍሌ 1948	ተስፋ ሁበተ ማኅከል 1958	መረጌታ ግርጋይን 1984	Weighted Average	Entity
ה	ሀወይ	ሆይ	ሆይ	ሆይ (ኋይ)	ሆይ (ሆወይ)	ሆይ	ሆይ	ሆይ	ሆይ	ሆይ	ሆይ	Hoy
למד	ለወ	ላዊ	ላዊ	ለወ	ለዊ (ለወ፡ይ)	ላዊ	ላዊ	ላዊ	ላዊ (ለው·)	ላዊ	ላዊ	Lawi
חת	ሐወት	ሐወት	ሐወት	ሐወት	ሐው (ይ) ት	ሐወት	ሐወት (ሐወት)	ሐወት	ሐወት	ሐወት	ሐወት	Hawt
מים	መይ	መይ	ማይ	ማይ	ማይ (መይም)	ማይ	ማይ	ማይ	ማይ	ማይ	ማይ	May
שמכת	ሠወት	ሠወት	ሠወት	ሠወት	ሠወት	ሠወት	ሣወት (ሠወት)	ሠወት	ሠወት	ሠወት	ሠውት	Sewt
רס	ርኣስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	ርእስ	Ris
סת	ሳት	ሳት	ሳት	ሳት	ሳት (በዓተ)	ሳት	ሳት	ሳት (በዓተ)	ሳት	ሳት	ሳት	Sat
קף	ቃፍ	ቃፍ	ቃፍ	ቃፍ	ቃፍ	ቃፍ	ቃፍ (ቃፍ)	ቃፍ	ቃፍ	ቃፍ	ቃፍ	Qaf
בת	በት	በት	በት	በት	በት	በት	በት (በት)	በት	በት	በት	በት	Bet
תו	ተወ	ተወ	ተው	ተው	ታው (ታዌ)	ታፀ	ታው (ታው·)	ታው	ታው	ታው	ታው	Taw
n/a	ሃርም	ሃርም	ሃርም	ሃርም	ሃርም	ሃርም	ሃርም	omitted	ሃርም	ሃርም	ሃርም	Harm
הר	ነሐስ	ነሐስ	ነሐስ	ነሐስ	ነሐስ (ነስ)	ነሐስ	ነሐስ	ነሐስ	ነሐስ	ነሐስ	ነሐስ	Nehas
אלף	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ	አለፍ/አለፍ	አለፍ	Alef
כף	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	ከፍ	Kaf
וו	ወዉ	ወዉ	ወዉ	ወዉ (ወዉ)	ወዉ (ወዉ)	ወዉ	ወዉ (ወዉ)	ወዉ	ወዉ	ወዉ	ወዉ	Wewe
עין	ዐይን	ዐይን	ዐይን	ዐይን	ዐይን (ታይ)	ዐይን	ዐይን	ዐይን	ዐይን	ዐይን	ዐይን	Ayin
זין	ዘይ	ዘይ	ዘይ	ዘይ	ዘይ (ዛይን)	ዘይ	ዘይ	ዛይ	ዘይ	ዘይ	ዘይ	Zey
ימן	የመን	የመን	የመን	የመን (የመን)	የመን	የመን	የመን	የማን	የመን	የመን	የመን	Yemen
דלת	ድንት	ድንት	ድንት	ድንት (ድንት)	ድ (ን) ለት	ድንት	ድንት	ድለት	ድንት	ድለት	ድንት	Dint
גמל	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	ገመል	Gemel
טת	ጠይት	ጠይት	ጠይት	ባይት	ጠይት	ጠይት	ጠይት	ጠይት	ጠይት	ጠይት	ጠይት	Teyt
n/a	ጰይት	ጰይት	ጰይት	ጰይት	ጰይት (ጰት)	ጰይት	ጰይት	omitted	omitted	omitted	ጰይት	Peyt
צד	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	ጸደይ	Sedey
n/a	ፀጳ	ፀጳ	ፀጳ	ፀጳ (ፀጳ)	ፀጳ (ፀጳ)	ፀጳ	ፀጳ	omitted	ፀሕ	omitted	ፀጳ	Sepa
פ	አፍ	አፍ	አፍ	አፍ	ፈ፡ (ፈ·)	አፍ	አፍ (ፈ·ፈ·)	ፈፍ	ፈፍ	ፈ·/ፈፍ	አፍ	Af
n/a	ፐሳ	ፐሳ	ፐሳ	ፐ (ፐስ)	ፐ (ፐ)	ፐ	omitted	omitted	omitted	omitted	ፐሳ	Psa

Note that Ethiopian references are shown with 7 years added to the year of publication as an approximation for the year under the Gregorian calendar, which is ahead of the Ethiopian by 7 years 8 months and 11 days (12 days during Ethiopian leap years). Guidi[9] offers a name list in keeping with Dillman[7] until ጠት which is given as ጠት.

Basic Ge'ez Letters

With our base names now available as a handy starting point we may simply derive entity names for the basic syllabary. In doing so we can borrow from the kindergarten practice of reciting the syllabary as per the character sound and form name: ለ - ግዕዝ፣ ሉ - ካዕብ፣ ሊ - ሣልስ (le-ge'ez, lu-ka'ib, li-salis ...). The form names translated have the meaning of "first", "second", "third" thru "seventh". In place of the characters' spoken value we can simply make the substitution for our base names.

	ግዕዝ Ge'ez	ካዕብ Kaib	ሣልስ Salis	ራብዕ Rabi	ኃምስ Hamis	ሳድስ Sadis	ሳብዕ Sabi
ሆይ Hoy	ህ ሆይ - ግዕዝ hoy-geez	ሁ ሆይ - ካዕብ hoy-kaib	ሂ ሆይ - ሣልስ hoy-salis	ሃ ሆይ - ራብዕ hoy-rabi	ሄ ሆይ - ኃምስ hoy-hamis	ህ ሆይ - ሳድስ hoy-sadis	ሆ ሆይ - ሳብዕ hoy-sabi
ላዊ Lawi	ለ ላዊ - ግዕዝ lawi-geez	ሉ ላዊ - ካዕብ lawi-kaib	ሊ ላዊ - ሣልስ lawi-salis	ላ ላዊ - ራብዕ lawi-rabi	ሌ ላዊ - ኃምስ lawi-hamis	ል ላዊ - ሳድስ lawi-sadis	ሎ ላዊ - ሳብዕ lawi-sabi
ሐውት Hawt	ሐ ሐውት - ግዕዝ hawt-geez	ሑ ሐውት - ካዕብ hawt-kaib	ሒ ሐውት - ሣልስ hawt-salis	ሓ ሐውት - ራብዕ hawt-rabi	ሔ ሐውት - ኃምስ hawt-hamis	ሕ ሐውት - ሳድስ hawt-sadis	ሖ ሐውት - ሳብዕ hawt-sabi
ማይ May	መ ማይ - ግዕዝ may-geez	ሙ ማይ - ካዕብ may-kaib	ሚ ማይ - ሣልስ may-salis	ማ ማይ - ራብዕ may-rabi	ሜ ማይ - ኃምስ may-hamis	ም ማይ - ሳድስ may-sadis	ሞ ማይ - ሳብዕ may-sabi
ሠውት Sewt	ሠ ሠውት - ግዕዝ sewt-geez	ሡ ሠውት - ካዕብ sewt-kaib	ሢ ሠውት - ሣልስ sewt-salis	ሣ ሠውት - ራብዕ sewt-rabi	ሤ ሠውት - ኃምስ sewt-hamis	ሥ ሠውት - ሳድስ sewt-sadis	ሦ ሠውት - ሳብዕ sewt-sabi
ርእስ Ris	ረ ርእስ - ግዕዝ ris-geez	ሩ ርእስ - ካዕብ ris-kaib	ሪ ርእስ - ሣልስ ris-salis	ራ ርእስ - ራብዕ ris-rabi	ሬ ርእስ - ኃምስ ris-hamis	ር ርእስ - ሳድስ ris-sadis	ሮ ርእስ - ሳብዕ ris-sabi
ሳት Sat	ሰ ሳት - ግዕዝ sat-geez	ሱ ሳት - ካዕብ sat-kaib	ሲ ሳት - ሣልስ sat-salis	ሳ ሳት - ራብዕ sat-rabi	ሴ ሳት - ኃምስ sat-hamis	ስ ሳት - ሳድስ sat-sadis	ሶ ሳት - ሳብዕ sat-sabi
ቃፍ Qaf	ቀ ቃፍ - ግዕዝ qaf-geez	ቁ ቃፍ - ካዕብ qaf-kaib	ቂ ቃፍ - ሣልስ qaf-salis	ቃ ቃፍ - ራብዕ qaf-rabi	ቄ ቃፍ - ኃምስ qaf-hamis	ቅ ቃፍ - ሳድስ qaf-sadis	ቆ ቃፍ - ሳብዕ qaf-sabi
ቤት Bet	በ ቤት - ግዕዝ bet-geez	ቡ ቤት - ካዕብ bet-kaib	ቢ ቤት - ሣልስ bet-salis	ባ ቤት - ራብዕ bet-rabi	ቤ ቤት - ኃምስ bet-hamis	ብ ቤት - ሳድስ bet-sadis	ቦ ቤት - ሳብዕ bet-sabi
ታው Taw	ተ ታው - ግዕዝ taw-geez	ቱ ታው - ካዕብ taw-kaib	ቲ ታው - ሣልስ taw-salis	ታ ታው - ራብዕ taw-rabi	ቴ ታው - ኃምስ taw-hamis	ት ታው - ሳድስ taw-sadis	ቶ ታው - ሳብዕ taw-sabi
ኀርም Harm	ኀ ኀርም - ግዕዝ harm-geez	ኁ ኀርም - ካዕብ harm-kaib	ኂ ኀርም - ሣልስ harm-salis	ኃ ኀርም - ራብዕ harm-rabi	ኄ ኀርም - ኃምስ harm-hamis	ኅ ኀርም - ሳድስ harm-sadis	ኆ ኀርም - ሳብዕ harm-sabi
ነሐስ Nehas	ነ ነሐስ - ግዕዝ nehas-geez	ኑ ነሐስ - ካዕብ nehas-kaib	ኒ ነሐስ - ሣልስ nehas-salis	ና ነሐስ - ራብዕ nehas-rabi	ኔ ነሐስ - ኃምስ nehas-hamis	ን ነሐስ - ሳድስ nehas-sadis	ኖ ነሐስ - ሳብዕ nehas-sabi
አልፍ Alef	አ አልፍ - ግዕዝ alef-geez	ኡ አልፍ - ካዕብ alef-kaib	ኢ አልፍ - ሣልስ alef-salis	ኣ አልፍ - ራብዕ alef-rabi	ኤ አልፍ - ኃምስ alef-hamis	እ አልፍ - ሳድስ alef-sadis	ኦ አልፍ - ሳብዕ alef-sabi
ካፍ Kaf	ከ ካፍ - ግዕዝ kaf-geez	ኩ ካፍ - ካዕብ kaf-kaib	ኪ ካፍ - ሣልስ kaf-salis	ካ ካፍ - ራብዕ kaf-rabi	ኬ ካፍ - ኃምስ kaf-hamis	ክ ካፍ - ሳድስ kaf-sadis	ኮ ካፍ - ሳብዕ kaf-sabi
ወዌ Wewe	ወ ወዌ - ግዕዝ wewe-geez	ዉ ወዌ - ካዕብ wewe-kaib	ዊ ወዌ - ሣልስ wewe-salis	ዋ ወዌ - ራብዕ wewe-rabi	ዌ ወዌ - ኃምስ wewe-hamis	ው ወዌ - ሳድስ wewe-sadis	ዎ ወዌ - ሳብዕ wewe-sabi
ዐይን Ayin	ዐ ዐይን - ግዕዝ ayin-geez	ዑ ዐይን - ካዕብ ayin-kaib	ዒ ዐይን - ሣልስ ayin-salis	ዓ ዐይን - ራብዕ ayin-rabi	ዔ ዐይን - ኃምስ ayin-hamis	ዕ ዐይን - ሳድስ ayin-sadis	ዖ ዐይን - ሳብዕ ayin-sabi
ዘይ Zey	ዘ ዘይ - ግዕዝ zey-geez	ዙ ዘይ - ካዕብ zey-kaib	ዚ ዘይ - ሣልስ zey-salis	ዛ ዘይ - ራብዕ zey-rabi	ዜ ዘይ - ኃምስ zey-hamis	ዝ ዘይ - ሳድስ zey-sadis	ዞ ዘይ - ሳብዕ zey-sabi

	ግዕዝ Ge'ez	ካዕብ Kaib	ሣልስ Salis	ራብዕ Rabi	ኃምስ Hamis	ሳድስ Sadis	ሳብዕ Sabi
የመን **Yemen**	የ የመን - ግዕዝ yemen-geez	ዩ የመን - ካዕብ yemen-kaib	ዪ የመን - ሣልስ yemen-salis	ያ የመን - ራብዕ yemen-rabi	ዬ የመን - ኃምስ yemen-hamis	ይ የመን - ሳድስ yemen-sadis	ዮ የመን - ሳብዕ yemen-sabi
ድንት **Dint**	ደ ድንት - ግዕዝ dint-geez	ዱ ድንት - ካዕብ dint-kaib	ዲ ድንት - ሣልስ dint-salis	ዳ ድንት - ራብዕ dint-rabi	ዴ ድንት - ኃምስ dint-hamis	ድ ድንት - ሳድስ dint-sadis	ዶ ድንት - ሳብዕ dint-sabi
ገምል **Gemel**	ገ ገምል - ግዕዝ gemel-geez	ጉ ገምል - ካዕብ gemel-kaib	ጊ ገምል - ሣልስ gemel-salis	ጋ ገምል - ራብዕ gemel-rabi	ጌ ገምል - ኃምስ gemel-hamis	ግ ገምል - ሳድስ gemel-sadis	ጎ ገምል - ሳብዕ gemel-sabi
ጠይት **Teyt**	ጠ ጠይት - ግዕዝ teyt-geez	ጡ ጠይት - ካዕብ teyt-kaib	ጢ ጠይት - ሣልስ teyt-salis	ጣ ጠይት - ራብዕ teyt-rabi	ጤ ጠይት - ኃምስ teyt-hamis	ጥ ጠይት - ሳድስ teyt-sadis	ጦ ጠይት - ሳብዕ teyt-sabi
ጰይት **Peyt**	ጰ ጰይት - ግዕዝ peyt-geez	ጱ ጰይት - ካዕብ peyt-kaib	ጲ ጰይት - ሣልስ peyt-salis	ጳ ጰይት - ራብዕ peyt-rabi	ጴ ጰይት - ኃምስ peyt-hamis	ጵ ጰይት - ሳድስ peyt-sadis	ጶ ጰይት - ሳብዕ peyt-sabi
ጸደይ **Sedey**	ጸ ጸደይ - ግዕዝ sedey-geez	ጹ ጸደይ - ካዕብ sedey-kaib	ጺ ጸደይ - ሣልስ sedey-salis	ጻ ጸደይ - ራብዕ sedey-rabi	ጼ ጸደይ - ኃምስ sedey-hamis	ጽ ጸደይ - ሳድስ sedey-sadis	ጾ ጸደይ - ሳብዕ sedey-sabi
ፀጸ **Sepa**	ፀ ፀጸ - ግዕዝ sepa-geez	ፁ ፀጸ - ካዕብ sepa-kaib	ፂ ፀጸ - ሣልስ sepa-salis	ፃ ፀጸ - ራብዕ sepa-rabi	ፄ ፀጸ - ኃምስ sepa-hamis	ፅ ፀጸ - ሳድስ sepa-sadis	ፆ ፀጸ - ሳብዕ sepa-sabi
አፍ **Af**	ፈ አፍ - ግዕዝ af-geez	ፉ አፍ - ካዕብ af-kaib	ፊ አፍ - ሣልስ af-salis	ፋ አፍ - ራብዕ af-rabi	ፌ አፍ - ኃምስ af-hamis	ፍ አፍ - ሳድስ af-sadis	ፎ አፍ - ሳብዕ af-sabi
ፕሳ **Psa**	ፐ ፕሳ - ግዕዝ psa-geez	ፑ ፕሳ - ካዕብ psa-kaib	ፒ ፕሳ - ሣልስ psa-salis	ፓ ፕሳ - ራብዕ psa-rabi	ፔ ፕሳ - ኃምስ psa-hamis	ፕ ፕሳ - ሳድስ psa-sadis	ፖ ፕሳ - ሳብዕ psa-sabi

Post Ge'ez Letters

Building upon the basic Ge'ez syllabary are additions made for the languages that followed as they adopted the writing system for their own use. Tesema Habte Michael offers ካፍ as an entity name for the ከ family and offers an approach for forming names for newer letters[17]. Accordingly, ቤት would become the name for the ቢ family by substituting ቢ for ቤ in ቤት from which the ቢ glyphs have been derived. Likewise ነመል for the ነ family, etc.

As the original entity names did themselves have literal meaning (ቤት is "house" and ገምል "camel" in our example[†]), this approach is not preferred as it implies the derived forms would also be meaningful. Nor was selecting a representative term from the primary language of the extended letters as doing so would require the availability of a fair number of experts in the given languages to come to a consensus opinion.

The solution applied here was to keep the original names and prefix the corresponding syllable of the first letter to the base name. Hence ቢ-ቤት, ነ-ገምል, etc. In this way the relationship between the new and derived form is also preserved.

[†] ገምል is an archaic Ge'ez spelling that survives as the syllable family name only. ግመል and ገመል are Amharic and Tigrigna spellings for "camel"[12].

	ግዕዝ Ge'ez	ካዕብ Kaib	ሣልስ Salis	ራብዕ Rabi	ኀምስ Hamis	ሳድስ Sadis	ሳብዕ Sabi
ሻ - ሳት Sha-Sat	ሻ - ሳት - ግዕዝ sha-sat-geez	ሹ - ሳት - ካዕብ sha-sat-kaib	ሺ - ሳት - ሣልስ sha-sat-salis	ሻ - ሳት - ራብዕ sha-sat-rabi	ሼ - ሳት - ኀምስ sha-sat-hamis	ሽ - ሳት - ሳድስ sha-sat-sadis	ሾ - ሳት - ሳብዕ sha-sat-sabi
ቻ - ቃፍ Qha-Qaf	ቻ - ቃፍ - ግዕዝ qha-qaf-geez	ቹ - ቃፍ - ካዕብ qha-qaf-kaib	ቺ - ቃፍ - ሣልስ qha-qaf-salis	ቻ - ቃፍ - ራብዕ qha-qaf-rabi	ቼ - ቃፍ - ኀምስ qha-qaf-hamis	ች - ቃፍ - ሳድስ qha-qaf-sadis	ቾ - ቃፍ - ሳብዕ qha-qaf-sabi
ቬ - ቤት Ve-Bet	ቬ - ቤት - ግዕዝ ve-bet-geez	ቩ - ቤት - ካዕብ ve-bet-kaib	ቪ - ቤት - ሣልስ ve-bet-salis	ቫ - ቤት - ራብዕ ve-bet-rabi	ቬ - ቤት - ኀምስ ve-bet-hamis	ቭ - ቤት - ሳድስ ve-bet-sadis	ቮ - ቤት - ሳብዕ ve-bet-sabi
ቻ - ታው Cha-Taw	ቻ - ታው - ግዕዝ cha-taw-geez	ቹ - ታው - ካዕብ cha-taw-kaib	ቺ - ታው - ሣልስ cha-taw-salis	ቻ - ታው - ራብዕ cha-taw-rabi	ቼ - ታው - ኀምስ cha-taw-hamis	ች - ታው - ሳድስ cha-taw-sadis	ቾ - ታው - ሳብዕ cha-taw-sabi
ኝ - ነሐስ Nye-Nehas	ኝ - ነሐስ - ግዕዝ nye-nehas-geez	ኙ - ነሐስ - ካዕብ nye-nehas-kaib	ኚ - ነሐስ - ሣልስ nye-nehas-salis	ኛ - ነሐስ - ራብዕ nye-nehas-rabi	ኜ - ነሐስ - ኀምስ nye-nehas-hamis	ኝ - ነሐስ - ሳድስ nye-nehas-sadis	ኞ - ነሐስ - ሳብዕ nye-nehas-sabi
ኸ - ካፍ Kha-Kaf	ኸ - ካፍ - ግዕዝ kha-kaf-geez	ኹ - ካፍ - ካዕብ kha-kaf-kaib	ኺ - ካፍ - ሣልስ kha-kaf-salis	ኻ - ካፍ - ራብዕ kha-kaf-rabi	ኼ - ካፍ - ኀምስ kha-kaf-hamis	ኽ - ካፍ - ሳድስ kha-kaf-sadis	ኾ - ካፍ - ሳብዕ kha-kaf-sabi
ዠ - ዘይ Zhe-Zey	ዠ - ዘይ - ግዕዝ zhe-zey-geez	ዡ - ዘይ - ካዕብ zhe-zey-kaib	ዢ - ዘይ - ሣልስ zhe-zey-salis	ዣ - ዘይ - ራብዕ zhe-zey-rabi	ዤ - ዘይ - ኀምስ zhe-zey-hamis	ዥ - ዘይ - ሳድስ zhe-zey-sadis	ዦ - ዘይ - ሳብዕ zhe-zey-sabi
ዽ - ድንት Dhi-Dint	ዽ - ድንት - ግዕዝ dhi-dint-geez	ዹ - ድንት - ካዕብ dhi-dint-kaib	ዺ - ድንት - ሣልስ dhi-dint-salis	ዻ - ድንት - ራብዕ dhi-dint-rabi	ዼ - ድንት - ኀምስ dhi-dint-hamis	ዽ - ድንት - ሳድስ dhi-dint-sadis	ዾ - ድንት - ሳብዕ dhi-dint-sabi
ጀ - ድንት Ji-Dint	ጀ - ድንት - ግዕዝ ji-dint-geez	ጁ - ድንት - ካዕብ ji-dint-kaib	ጂ - ድንት - ሣልስ ji-dint-salis	ጃ - ድንት - ራብዕ ji-dint-rabi	ጄ - ድንት - ኀምስ ji-dint-hamis	ጅ - ድንት - ሳድስ ji-dint-sadis	ጆ - ድንት - ሳብዕ ji-dint-sabi
ኘ - ገምል Nge-Gemel	ኘ-ገምል - ግዕዝ nge-gemel-geez	ኙ-ገምል - ካዕብ nge-gemel-kaib	ኚ-ገምል - ሣልስ nge-gemel-salis	ኛ-ገምል - ራብዕ nge-gemel-rabi	ኜ - ገምል - ኀምስ nge-gemel-hamis	ኝ-ገምል - ሳድስ nge-gemel-sadis	ነ-ገምል - ሳብዕ nge-gemel-sabi
ጨ - ጠይት Che-Teyt	ጨ - ጠይት - ግዕዝ che-teyt-geez	ጩ-ጠይት-ካዕብ che-teyt-kaib	ጪ - ጠይት - ሣልስ che-teyt-salis	ጫ - ጠይት - ራብዕ che-teyt-rabi	ጬ - ጠይት - ኀምስ che-teyt-hamis	ጭ - ጠይት- ሳድስ che-teyt-sadis	ጮ - ጠይት- ሳብዕ che-teyt-sabi

Labiovelar

The labiovelar extensions to the syllabary require a little more consideration for the selection of suitable names. In Ge'ez language ድቃል has the meaning of "derived from" (or "something originating

from something else"). In Amharic, ዲቃላ gets applied to the extra labiovelar letters as a group but also means "bastard". ተውሳከ might be a good choice for the name as it has the meaning for "additional" in Ge'ez but unfortunately in Tigrigna and Amharic the term is used for "parasite" which again does not conjur up an appropriate relationship. Both ዲቃላ and ተውሳከ do invoke confusion when used in the intended context and imply an illegitimacy to the characters that is best avoided. ዘመደ (the adjective form of ዘመድ) is then applied from Ge'ez for the generic form of "related". Fortunately the same meaning has carried over into Amharic and Tigrigna. Versus alternatives that also communicate a relationship (brother, sister, aunt, uncle) "zemede" nicely avoids debate over the appropriate gender associations.

	ዘመደ - ግዕዝ Zemede-Ge'ez	ዘመደ - ካዕብ Zemede-Kaib	ዘመደ - ሣልስ Zemede-Salis	ዘመደ - ራብዕ Zemede-Rabi	ዘመደ - ኃምስ Zemede-Hamis
ቃፍ Qaf	ቄ ቃፍ - ዘመደ - ግዕዝ qaf-zemede-geez	ቁ ቃፍ - ዘመደ - ካዕብ qaf-zemede-kaib	ቀ ቃፍ - ዘመደ - ሣልስ qaf-zemede-salis	ቂ ቃፍ - ዘመደ - ራብዕ qaf-zemede-rabi	ቃ ቃፍ - ዘመደ - ኃምስ qaf-zemede-hamis
ቓ - ቃፍ Qha-Qaf	ቔ ቓ - ቃፍ - ዘመደ - ግዕዝ qha-qaf-zemede-geez	ቑ ቓ - ቃፍ - ዘመደ - ካዕብ qha-qaf-zemede-kaib	ቐ ቓ - ቃፍ - ዘመደ - ሣልስ qha-qaf-zemede-salis	ቒ ቓ - ቃፍ - ዘመደ - ራብዕ qha-qaf-zemede-rabi	ቓ ቓ - ቃፍ - ዘመደ - ኃምስ qha-qaf-zemede-hamis
ጎርም Harm	ጐ ጎርም - ዘመደ - ግዕዝ harm-zemede-geez	ጉ ጎርም - ዘመደ - ካዕብ harm-zemede-kaib	ጐ ጎርም - ዘመደ - ሣልስ harm-zemede-salis	ጒ ጎርም - ዘመደ - ራብዕ harm-zemede-rabi	ጓ ጎርም - ዘመደ - ኃምስ harm-zemede-hamis
ከፍ Kaf	ኰ ከፍ - ዘመደ - ግዕዝ kaf-zemede-geez	ኩ ከፍ - ዘመደ - ካዕብ kaf-zemede-kaib	ኰ ከፍ - ዘመደ - ሣልስ kaf-zemede-salis	ኲ ከፍ - ዘመደ - ራብዕ kaf-zemede-rabi	ኳ ከፍ - ዘመደ - ኃምስ kaf-zemede-hamis
ኻ - ከፍ Kha-Kaf	ዀ ኻ - ከፍ - ዘመደ - ግዕዝ kha-kaf-zemede-geez	ዅ ኻ - ከፍ - ዘመደ - ካዕብ kha-kaf-zemede-kaib	ዀ ኻ - ከፍ - ዘመደ - ሣልስ kha-kaf-zemede-salis	ዂ ኻ - ከፍ - ዘመደ - ራብዕ kha-kaf-zemede-rabi	ዃ ኻ - ከፍ - ዘመደ - ኃምስ kha-kaf-zemede-hamis
ገምል Gemel	ጐ ገምል - ዘመደ - ግዕዝ gemel-zemede-geez	ጉ ገምል - ዘመደ - ካዕብ gemel-zemede-kaib	ጐ ገምል - ዘመደ - ሣልስ gemel-zemede-salis	ጒ ገምል - ዘመደ - ራብዕ gemel-zemede-rabi	ጓ ገምል - ዘመደ - ኃምስ gemel-zemede-hamis

The remaining labiovelar syllables appear only in the zemede-rabi form and would be named accordingly. An alternative naming scheme would have been one to apply extensions to the "first", "second", "third" name conventions for "eighth" thru "twelfth". The names of the forms after the seventh are less familiar to the average person and could even look contrived after the tenth form. Doing so would also meet with some resistance as the seven basic syllables are tied to the seven days of the week and one is consider as immutable as the other. Certainly the terms would not be as easy to remember and debate is likely over whether the remaining forms are rightfully an eleventh form or merely an eighth. In contrast, the "zemede" convention never has to confront these issues.

The ordering indicated in our last table is in the traditional system. The name convention unfortunately may invoke some confusion when compared against the Unicode sequence for these same elements. The Unicode definition orders the letters in a phonetic based sequence where the zemede-kaib form is shown in the zemede-sadis position. Linguistically these two forms are virtually indistinguishable and the same type of ordering is also found in Ethiopian references. In keeping with the traditional conventions, which underlie this work, the entity names here do offer a partial rectification to the non-traditional sequence used in Unicode.

Extended Zemede

A few relics of syllabary's evolution remain for us to contend with. The Alef series has what appears to be a non-labiovelar eighth form (ኧ) used primarily in Sebatbeit and Amharic. Rather than name it with an eighth form we adorn the character with the label "alef-zemede-geez" for consistency.

At the end of the Unicode Ethiopic syllabary we find the ligatures ፘ, ፙ and ፚ, which find no modern use (save for publishing Unicode tables). These letters could be treated as a thirteenth form (ala Dawkins[5]) but this would embrace the extended form confusion that we've made the effort already to avoid. As a resolution "zemede-ya" is applied to indicate the relationship to ይ from which the ligature is formed.

Character	Ge'ez Name	Transcribed
ኧ	አልፍ - ዘመደ - ግዕዝ	alef-zemede-geez
ፘ	ርዕስ - ዘመደ - ይ	re-zemede-ya
ፙ	ማይ - ዘመደ - ይ	me-zemede-ya
ፚ	ፈፍ - ዘመደ - ይ	fe-zemede-ya

Numerals

Ge'ez numeral names are the natural choice for naming the entities. The names of numbers will take a variety of forms depending on the context of the phrase they are invoked in. The entity names given below are those used when referring to the numerals themselves and are considered common knowledge.

Character	Ge'ez Name	Transcribed	Character	Ge'ez Name	Transcribed
፩	አሐዱ	ahadu	፳	ዕሥራ	isra
፪	ክልኤቱ	kileitu	፴	ሠላሳ	selasa
፫	ሠለስቱ	selestu	፵	አርብዓ	arba
፬	አርባዕቱ	arbaitu	፶	ኃምሳ	hamsa
፭	ኃምስቱ	hamistu	፷	ስሳ	sisa
፮	ስድስቱ	sidistu	፸	ሰብዓ	seba
፯	ሰብዐቱ	sebatu	፹	ሰማንያ	semanya
፰	ሰመንት	sementu	፺	ተስዓ	tesa
፱	ተስዐቱ	tesatu	፻	ምእት	meet
፲	ዐሥርቱ	asertu	፼	እልፍ	ilf

Punctuation

Ethiopic punctuation is unique in that unlike previous character elements most do have commonly used names in the present day. To remain language neutral however we seek out the Ge'ez names for the punctuation at the cost of the loss of familiarity.

Ge'ez names were not available for the punctuation marks ፥, ፦, ፡, ፨ and so had to be derived. For ፡ and ፨ names are constructed directly from the number of "dots" (ነጥብ or ነቁጥ in Ge'ez) in the symbol. ፦ is a modern symbol used primarily in a dialogue or in preface to a list. Accordingly the name አስተአምሮ is chosen from Ge'ez meaning "to explain". አቅርቦት was the next runner up here as it similarly means "to introduce" or "to present". አስተአምሮ was preferred however as it implies more, a greater amount of, information is being communicated. Finally we have ፥ which originated (and is still largely recognized as) another form of ፤. Fortunately two Ge'ez terms are available for both ነጠላ ሠረዝ and ንዑስ ሠረዝ. We apply ነጠላ ሠረዝ to ፤ as the name will already be familiar to Amharic speakers and ንዑስ ሠረዝ then remains for ፥.

Character	Ge'ez Name	Transcribed
፤	ንዑስ ነጥብ	neus-neteb
፨	ዐቢይ ነጥብ	ebiy-neteb
፣	ነጠላ ሠረዝ	netela-serez
፥	ንዑስ ሠረዝ	neus-serez
፦	አስተአምሮ	asteamro
፤	ዐቢይ ሠረዝ	ebiy-serez
፧	ሠለስተ ነጥብ	seleste-neteb
፠	ሰብዐተ ነጥብ	sebate-neteb

__Numerology__

Apart from the Ethiopic numerals (or ኍኃዝ - U+1369-U+137C), which were in fact derived from the Ethiopic letters[8], there is a fortune telling as well as mystical practice of associating numeric values to letters.

The simplest form of enumeration applies only to the ግዕዝ (first) form letters arranged in the Hebrew order. This arrangement appears to be an attempt to superimpose the Gematria discipline of Hebrew and Greek mysticism onto the Ge'ez character set. Gematria, derived from the Greek word "geometry" (also known as "Sacred Geometry" and "Isopsephia" in Greek[15]) is the practice of assigning numeric value to letters to reveal hidden meanings and relationships thru the summation of the letters in keyword found in sacred text. A phonetic based superposition would help keep the values of personal and place names equal across translated scriptures as the characters of each script are first phonetically aligned before the numeric assignments are made. Archaic letters were inserted into the Greek sequence (ϛ and ϙ) to stay in locked step with Hebrew (the character names also indicate the alignment). Or so it appears, there are convincing arguments that show the Hebrew enumerations were borrowed from an earlier Greek practice. Unquestionably the Gematria practice reached its most developed state in Judaism where it flourished and remains strongest in the present day.

Variations on the theme occur in Hebrew after the end of the Hebrew alephbet, which would otherwise end in 400. There appears to be two conventions that allow Gematria for Hebrew script to continue. Whereby final forms of certain letters may be employed or the primary group is used in combination to form values thru summation. It should be noted that enumeration of letters to serve as numbers for Semitic scripts was fairly common. It is not known to the present author if the enumeration had the same intended purpose. Resource readily available on the World Wide Web would indicate that enumeration after 400 is less common and may have become a practice lost to antiquity.

Gematria Enumeration

Ethiopic	Hebrew	Greek	Value	Ethiopic Numeral
ል Alef	א Alef	A Alpha	1	
በ Bet	ב Bet	B Beta	2	
ገ Gemel	ג Gimel	Γ Gamma	3	
ደ Dint	ד Dalet	Δ Delta	4	
ሀ Hoy	ה He	E Epsilon	5	
ወ Wewe	ו Vav	ϛ/Ϝ Vau	6	
		ϛ or Ϝ inserted into Greek sequence.		
ዘ Zey	ז Zayin	Z Zeta	7	
ሐ Hawt	ח Het	H Eta	8	
ጠ Teyt	ט Tet	Θ Theta	9	
የ Yemen	י Yod	I Iota	10	
ከ Kaf	כ Kaf	K Kappa	20	
ለ Lawi	ל Lamed	Λ Lambda	30	

Ethiopic	Hebrew	Greek	Value	Ethiopic Numeral
መ May	מ Mem	M Mu	40	
ነ Nehas	נ Nun	N Nu	50	
ሠ Sewt	ס Samekh	Ξ Xi	60	
ዐ Ayin	ע Ayin	O Omicron	70	
ፈ Af	פ Pe	Π Pi	80	
ጸ Seday	צ Tsadi	Ϙ Koppa	90	
		Ϙ inserted into Greek sequence.		
ቀ Qaf	ק Qof	P Rho	100	
ረ Ris	ר Resh	Σ Sigma	200	
ሰ Sat	ש Shin	T Tau	300	
ተ Taw	ת Tav	Y Upsilon	400	
		End of Hebrew Alephbet		

Note that the Greek letters Stigma (ϛ a Sigma-Tau ligature) and Digamma (Ϝ) could both be used for the number 6 and were known in numeric context as "Vau".

Continuing Named Orders of 10 for Completeness

Ethiopic	Name	Value	Ethiopic Numeral
የእ Yemen-Ge'ez Alef-Sadis	አእለፉ	100,000	፲፼
ቀእ Qaf-Ge'ez Alef-Sadis	አእላፉት	1,000,000	፲፼
ቄእ Qaf-Zemede-Ge'ez Alef-Sadis	ተእለፊት	10,000,000	፲፼፼
እእ Alef-Sadis Alef-Sadis	ተእለፊታት	100,000,000	፼፼
ቄእእ Qaf-Zemede-Ge'ez Alef-Sadis Alef-Sadis	ምእላፊት	100,000,000,000	፲፼፼፼
እእእ Alef-Sadis Alef-Sadis Alef-Sadis	ምእለፊታት	1,000,000,000,000	፼፼፼

Continued…

Ethiopic	Hebrew	Greek	Value	Ethiopic Numeral
ኀ Harm	ך Final Kaf	Φ Phi	500	፭፻
ሰ Sepa	ם Final Mem	Χ Chi	600	፮፻
ጰ Peyt	ן Final Nun	Ψ Psi	700	፯፻
ፐ Psa	ף Final Pe	Ω Omega	800	፰፻
ኈ Harm-Zemede-Ge'ez	ץ Final Tsadi	ϡ Sampi	900	፱፻
ቈ Qaf-Zemede-Ge'ez			1,000	፲፻
እ Alef-Sadis			10,000	፼
End of Ethiopic Numerals				

Debtera (Halehame) Enumeration

The applicability of the "Abegede" enumeration is not well understood; it may be a scholarly alternative to the preferred "Halehame" system of the debtere. The debtere are adjunct members of the Ethiopian Orthodox Church clergy. Though not ordained, debtere will have a wider range of learned skills than what is required of a priest. They perform the music and dance associated with church services and also function as astrologers, herbalists, scribes, wizards, and fortune-tellers and learn church lore. The system extends the enumeration for the original seven syllabic forms of the syllabary. The application of the numerals after the first form is not clear. Example usage of the system to find the numeric value of words invariably uses the letters of a word in the first form only. Seemingly as if the pre-syllabic Ge'ez was somehow more meaningful in the practice.

Proponents of the "halehame" enumeration testify to the better results it offers when applied as Gematria. A validation of the ordering comes from the numeric value of ክርስቶስ (Christos) which expands to: ክ 50 + ረ 6 + ስ 7 + ቶ 10 + ስ 7 = 80. Eighty in turn multiplied by ten is eight hundred, which is exactly the value of ፐ the last member of the first order of the syllabary[2].

A basic example use of the enumeration system applied to fortune telling is to take a person's mother's name, find its summation from the first form characters, take the modulo of twelve and compare it against the corresponding star for that month. The resulting star then indicates what fortune that person can expect in his or her lifetime. This use of numerology with astrology is in the much the same tradition as the monthly Zodiac system still popular in the west and the Chinese annual system where the animal for a given year indicates what fortune a person can expect. No doubt there are the available variations to offer a second opinion when needed.

	ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ
ሆይ Hoy	ህ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
ላዊ Lawi	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ሐውት Hawt	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
ማይ May	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
ሰውት Sewt	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
ርእስ Ris	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
ሳት Sat	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
ቃፍ Qaf	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
ቤት Bet	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
ታው Taw	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
ኅርም Harm	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
ነሐስ Nehas	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
አልፍ Alef	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
ካፍ Kaf	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
ወዌ Wewe	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ

ዐይን Ayin	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
ዘይ Zey	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
የመን Yemen	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
ድንት Dint	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
ገምል Gemel	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
ጠይት Teyt	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
ጸይት Peyt	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
ጸደይ Sedey	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
ፀጸ Sepa	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
አፍ Af	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
ፐሳ Psa	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

But is it Essential?

Ultimately no. This class of character information is not widely applicable to text and information processing. Ethiopic numerology while still in use plays a role in modern society analogous to a "Zodiac" type fortune telling system and would only be applied in specialized software for this purpose.

<u>Collation</u>

When we consider that Ethiopia and Eritrea together have upwards of 80 languages it should come as no surprise to learn that no single language uses all of the characters in the Unicode definition of Ethiopic. Ethiopic in Unicode is indeed a collection of the character sets of these languages integrated under a set of linguistic rules. As is the case in Europe and North America a person generally learns the letters they need for their particular language and might go their lifetime oblivious to additional letters, or even slightly different use of common letters, in neighboring languages.

Fortunately the variations in convention are not too unwieldy to manage in software and there is a good degree of tolerance in the area of sorting. Firstly, there is no real concern for the order of punctuation. Numerals should of course be sequential as per their value but there is no concern that their order be higher or lower than that of the syllographs or punctuation.

The syllabic series that go unused by a given language are generally just dropped from its syllabary. Accordingly, users are not concerned for the sorting order of jettisoned letters, as they were unlikely to have been used in their literature anyway. The traditional order is then suitable. For example the ቨ ("ve") family is not a phoneme in most Ethiopian languages but is important for transcribing foreign words such as "university" (ዩኒቨርስቲ). "ቨ" may then acceptably sort after "በ", from which the glyph is derived, though some conventions will place it after "ፐ" the last series of the syllabary.

Under Amharic practices the phonetically redundant (homophonic) characters are not dropped from the syllabary as they do retain their relevance for the canonical spelling of words. It is however a common convention in Amharic dictionaries to "fold" the series onto one another for brevity.

The convention of placing relatively newer characters at the end of the syllabary (including the zemede extensions) is required to maintain the Gematria encoding of the script and is also common in primary

education but is generally not a concern when applied to electronic collation. The Unicode approach of integrating the extension members into the syllabary proper will meet with little objection. The approach also has precedence set by Kidane Wold Kifle who applied the same technique in 1955 in an updated form of the "Abegede" syllabary[12].

The Abegede syllabary does indeed demand our attention and consideration when addressing collation issues. Proponents of the sequencing argue that it is the original order of the Ge'ez script as would be required by the Semitic alphabetic template. Many references on ancient writing and archaeology will use the Abegede order for sake of comparison when displaying the Mino-Sabaean script (the likely progenitor of Ge'ez), "Thamudic" style Ge'ez and later Ge'ez along side other Semitic scripts. The more thorough references will add the caveat to the comparison that the South Arabian convention (albeit sans the vowel demarcation) was indeed "Halehame".

The Ethiopian Orthodox Church did for a time embrace the "Abegede" order and until the present day materials are still published where the sequence is used for indexing[12]. The "Abegede" collation should remain an option in a Ge'ez language locale when dealing with this class of literature where it is expected. For every other case the "Halehame" collation convention is appropriate. Kidane Wold Kifle introduced a modernization of the Abegede order for the Amharic syllabary[11]. The extent to which it was adapted and impacted literature that followed is unknown to the author.

The Abugida syllabary is more familiar to the average person than is the Abegede syllabary that it is derived from. The two are often thought to be the same syllabary. This is not the case however as the Abugida syllabary is a simple rotation of the Abegede order intended to challenge the student who is has learned the syllabary thru rote memorization. The "Abugida" in the first column curiously follows the Gematria order. Each consecutive rightward column then shifts a single row upward as per its distance from the first column (shift = column number - 1). Elements shifted off the top then wrap around to the bottom.

A modern ordering of the syllabary is aimed at introducing the syllabary to kindergarteners following the glyph similarities of the elements. The kindergarten order is shown here for comparison value. We need only be aware of conventions as not in need of being addressed as a collation system.

Unicode	Amharic Dictionary	Ge'ez Heleheme	Ge'ez Abegede	KWK Abegede	Glyph Based
ሀ	ሀ/ሐ/ኀ/ኸ	ሀ	አ	አ	በ
ለ	ለ	ለ	በ	በ	ሰ
ሐ	መ	ሐ	ገ	ገ	ሸ
መ	w/ሠ	መ	ደ	ደ	ከ
ሠ	ረ	ሠ	ሀ	ዠ	ኸ
ረ	ሸ	ረ	ወ	ሀ	ለ
ሰ	ቀ	ሰ	ዘ	ወ	አ
ሸ	በ	ቀ	ሐ	ዘ	ዘ
ቀ	ተ	በ	ጠ	ዥ	ዥ
ቿ	ቸ	ተ	የ	ሐ	ሀ
በ	ነ	ነ	ከ	ኀ	ወ
ቨ	ኘ	ነ	ለ	ጠ	መ
ተ	አ/ዐ	አ	መ	ጬ	ገ
ቸ	ከ	ከ	ነ	የ	ተ
ኀ	ወ	ወ	ወ	ከ	ቸ
ኘ	ዘ	ዘ	ዐ	ኸ	ቿ
ኀ	ዠ	ዠ	ፈ	ለ	ቀ
አ	የ	የ	ጸ	መ	የ
ከ	ደ	ደ	ቀ	ነ	ደ
ኸ	ጀ	ገ	ረ	ኘ	ጀ
ወ	ገ	ጠ	ሰ	ወ	ጸ
ዐ	ጠ	ጸ	ተ	ዐ	ጸ
ዘ	ጨ	ጸ	ኀ	ፈ	ነ
ዠ	ጸ	ጰ	ዐ	ጸ	ኘ
የ	ጸ	ፈ	ጸ	ዐ	ኀ
ደ	ዐ	ፐ	ፐ	ቀ	ዐ
ጀ	ፈ			ረ	ዐ
ጀ	ፐ			ሰ	ወ
ገ				ሸ	ጠ

ሃ				ተ	ጨ
ጠ				ቸ	ሐ
ጨ				ጸ	ረ
ጸ				ፐ	ፈ
ጸ					
0					
ፈ					
ፐ					

It must be noted that for the six series (ቀ,ቸ,ጎ,ከ,ኸ and ገ) having the full complement of 12 forms, the Unicode ordering attempts a linguistic style of sequencing. Indeed the zemede-kaib and zemede-sadis can be shown to be separated by very little distance under linguistic metrics. In the present day this spoken difference is not stressed and we are likely seeing a phonemic decay of the zemede-kaib into a zemede-sadis form. While the Unicode sequencing can be found in Eritrean and Ethiopian reference, it is fairly common knowledge that the traditional sequence recognizes the zemede-kaib position as correct for the syllographs in questions. Our next table demonstrates the difference using the ከ syllable:

	ግዕዝ	ካዕብ	ሣልስ	ራብዕ	ኃምስ	ሳድስ
Traditional	ከወ	ኩ	ኩ	ኪ	ኬ	
Unicode	ከወ		ኩ	ኪ	ኬ	ኩ

The final three syllographic elements in Unicode, ቮ, ኟ, and ቿ, are in an order slightly different from that of how their base components appear in the syllabary. These last three elements could potentially sort as-is, or in order with the base syllabary as ኟ, ቮ and ቿ, or as the last member of their syllabic families following ሟ, ቯ and ፏ accordingly. The later option would likely be found to be the most intuitive of the three. Alas, since the 3 ligatures are virtually unknown archaic relics of the syllabary's history, the matter of their sort order could be no more than academic.

Collation is otherwise straightforward and no practices are known for Ethiopic script whereby two or more characters together sort in a higher or lower order than separately.

Legacy Character Classes

Character classes as defined in ANSI C and POSIX aim to provide a simple means to determine properties of an arbitrary character. The properties that we can analysis are the character's functional use in text (punctuation, alphabetic, digital, space, etc), relational ([=c=], case insensitive comparisons) which the locale mechanism lets us make specific to a language convention, orthographic semantics (uppercase, lowercase) and the character's address range (printable, ASCII).

A software developer working with Ethiopic text will want the same utilities for character analysis as he or she had for single byte text. Unicode aware resources will allow for generic character tests so that script specific functions can be avoided. For example a test for an Ethiopic punctuation might take the form:

```
if ( isrange ( ch, "ethiopic" ) && ispunct ( ch ) ) { ...
```

and avoids the need for an "isEthiopicPunct" equivalent.

While progress is being made towards Unicode based and multilingual aware pattern matching tools, by and large existing tools remain oriented towards the properties of western, Latin based, scripts. In this section we will look at extension of the character class paradigm to accommodate needs particular to Ethiopic script and languages while trying to maintain a perspective towards syllabic writing systems in general.

Syllabic Classes

The first simple extension to the standard character class lexicon would be the introduction of "[:syllable:]" as the logical analog to "[:alpha:]". "[:syllable:]" would match any character with a syllabic property, it is left to the user to restrict the syllabary of interest.

When we looked at Ethiopic character entity names we had our first introduction to the concept of Ethiopic character classes. In fact the naming convention proposed is built upon recognizing the "rows" and "columns" of the traditional syllabary with discretely named terms. It is a common problem to want to detect either the row or column property of an arbitrary Ethiopic character, as we will examine in more detail shortly. We can employ row and column names to define additional POSIX style character classes as follows:

Pattern	Expansion	Description
[:ላዊ:] (or [:lawi:])	[ለ-ᎀ]	Match any syllable in the "Lawi" family (ለ thru ᎀ).
[:ካፍ:] (or [:kaf:])	[ከ-ኮኰ-ኵ]	Match any syllable in the "Kaf" family (ከ thru ኵ but excluding undefined address in between, or all entity names beginning with "kaf-").
[:ግዕዝ:] (or [:geez:])	[ሀለሐመሠረሰሸቀበበተቸነኘአከ ኸወዐዘዠየደዸጀገጠጨጸፀፈፐ]	Match any ግዕዝ (first) form syllable.

This approach maintains a Ge'ez script perspective and requires the definition of a considerable number of pattern matching terms meaningful only in the applicable address range. The approach can be generalized for use with other syllabaries as follows:

Pattern	Expansion	Description
[:ለ:]	[ለ-ᎀ]	Match any character in the same family with ለ. Likewise for [:ሉ:], [:ሊ:], [:ላ:], etc.
[:ከ:]	[ከ-ኮኰ-ኵ]	Match any character in the same family with ከ. Likewise for [:ኩ:], [:ኪ:], [:ካ:], etc.
(isrange("ethiopic") && [:%1:])	[ሀለሐመሠረሰሸቀበበተቸነኘአከ ኸወዐዘዠየደዸጀገጠጨጸፀፈፐ]	Match any first form syllable in the Ethiopic context.

Note that in the final pattern for detecting a first form syllable some means was required to restrict the context to Ethiopic script so as not to match a first form syllable from *any* syllabary. The percentage sign, "%", is used here to specify the syllabic context by employing the "modulo" meaning of the symbol.

Locale Based Equivalence Classes

In the strictest sense no two members of the Ethiopic syllabary would have the same same phonemic value. The presence of ኸ (U+12A8) is the telltale indicator that this is not quite the case. As a consequence of the phonemic decay of many Ethiopic syllographs spelling correctness in Ethiopian and Eritrean languages works on the notion of proximity. While each language may recognize a canonical spelling for a given word, a rendering may still be regarded as "correct" or acceptable based on its orthographic distance from the canonical rendering. Further, as with American and British English spellings, the canonical spellings are also allowed to change for the same word across national or linguistic borders.

To handle these conventions, pattern matching software need be made aware of the localized rules. A demonstrative sampling is offered in our next table:

Equivalence Class	Locale	Expansion	Comment
[=ሃ=]	Amharic	[ሀሃሐሓጎጓ኱]	Also needed for [=ሁ=], [=ሂ=], [=ሃ=], etc.
	Tigrigna (Et)	[ሀሃጎጓ]	ሐ and ኸ series and different phonemes in Tigrigna.
	Tigrigna (Er)	[ሀሃ]	Redundant ጎ series is dropped.
[=ሰ=]	Amharic / Tigrigna (Et)	[ሰሠ]	Also needed for [=ሱ=], [=ሲ=], [=ሳ=], etc.
	Tigrigna (Er)	[ሰ]	Redundant ሠ series is dropped.
[=ቄ=]	All	[ቆቄ]	Example: ቄኖስ vs ቆኖስ.
[=ቊ=]	All	[ቁቊ]	Example: ቁጥር vs ቊጥር.
[=አ=]	Amharic	[አኣዐዓ]	Also needed for [=ኡ=], [=ኢ=], [=ኣ=], etc.
	Tigrigna (Et)	[አኣ]	አ and ዐ series and different phonemes in Tigrigna.
	Tigrigna (Er)	[አ]	አ is not interchangle with ኣ though phonetically equivalent.
[=ኮ=]	All	[ኮኰ]	Example: መኮንን vs መኰንን.
[=ጎ=]	All	[ጎጐ]	Example: ጎንደር vs ጐንደር.
[=ጸ=]	Amharic / Tigrigna (Et)	[ጸዐ]	Also needed for [=ጹ=], [=ጺ=], [=ጻ=], etc.
	Tigrigna (Er)	[ጸ]	Redundant ዐ series is dropped.
Eritrean conventions are based on the conventions taught in primary education since 1991. Ethiopian Tigrigna stresses the same conventions but is more forgiving in the use of the redundant syllabic series.			

The table demonstrates a sample of useful character classes and how they would vary with locale setting. With collectively over 80 languages in Eritrea and Ethiopia the table is not intended to be comprehensive but demonstrative with the most familiar classes. Notably a class to fold all ግዕዝ and ራዕስ forms would be desirable when working with southern language of Ethiopia - where many of the classes shown above would not be applicable. It is also worth noting that Ge'ez and Ari may share character classes while not sharing character phonemes. This helps highlight the separation of spoken language from orthography, in this case without consequence to pattern matching, and is certainly the exception and not the rule, as we will now see.

Demonstrating the classes we can consider the case of the Ethiopian Tafari Mekonnen who worked his way thru a military career up to the commander rank of "Ras". His life took a turn to the orthographically more complex when in 1930 Ras Tafari became Emperor of Ethiopia and assumed the coronation name Haile Selassie I. "Tafari Mekonnen" had only two possible spellings while "Haile Selassie" has numerous (not to imply for a moment though that HIM would have used anything but the canonical form). We'll look at these possibilities along with the female name ግለምዐሕይ and its many spellings.

Locale	[=አ=|ለም [=ጸ=]|[=ሃ=|ይ				[=ሃ=|ይለ [=ስ=|ለ[=ሴ=]			
Amharic	አለምጸሀይ አለምጸሃይ አለምጸሐይ አለምጸሐይ አለምጸጎይ አለምጸጋይ አለምጸኸይ	አለምዐሀይ አለምዐሃይ አለምዐሐይ አለምዐሐይ አለምዐጎይ አለምዐጋይ አለምዐኸይ	አለምጸሀይ አለምጸሃይ አለምጸሐይ አለምጸሐይ አለምጸጎይ አለምጸጋይ አለምጸኸይ	አለምዐሀይ አለምዐሃይ አለምዐሐይ አለምዐሐይ አለምዐጎይ አለምዐጋይ አለምዐኸይ	ሀይለ ሥላሜ ሀይለ ሥላሴ ሀይለ ሰላሜ ሀይለ ሰላሴ	ሃይለ ሥላሜ ሃይለ ሥላሴ ሃይለ ሰላሜ ሃይለ ሰላሴ	ሐይለ ሥላሜ ሐይለ ሥላሴ ሐይለ ሰላሜ ሐይለ ሰላሴ	ሐይለ ሥላሜ ሐይለ ሥላሴ ሐይለ ሰላሜ ሐይለ ሰላሴ
	ዐለምጸሀይ ዐለምጸሃይ ዐለምጸሐይ ዐለምጸሐይ ዐለምጸጎይ ዐለምጸጋይ ዐለምጸኸይ	ዐለምዐሀይ ዐለምዐሃይ ዐለምዐሐይ ዐለምዐሐይ ዐለምዐጎይ ዐለምዐጋይ ዐለምዐኸይ	ዓለምጸሀይ ዓለምጸሃይ ዓለምጸሐይ ዓለምጸሐይ ዓለምጸጎይ ዓለምጸጋይ ዓለምጸኸይ	ዓለምዐሀይ ዓለምዐሃይ ዓለምዐሐይ ዓለምዐሐይ ዓለምዐጎይ ዓለምዐጋይ ዓለምዐኸይ	ጎይለ ሥላሜ ጎይለ ሥላሴ ጎይለ ሰላሜ ጎይለ ሰላሴ	ጋይለ ሥላሜ ጋይለ ሥላሴ ጋይለ ሰላሜ ጋይለ ሰላሴ	ኸይለ ሥላሜ ኸይለ ሥላሴ ኸይለ ሰላሜ ኸይለ ሰላሴ	
Tigrigna (Eritrea)		አለምጸሀይ አለምጸሃይ			ሀይለ ሰላሴ	ሃይለ ሰላሴ		
Tigrigna (Ethiopia)	አለምጸሀይ አለምጸሃይ አለምጸጎይ አለምጸጋይ	አለምዐሀይ አለምዐሃይ አለምዐጎይ አለምዐጋይ	አለምጸሀይ አለምጸሃይ አለምጸጎይ አለምጸጋይ	አለምዐሀይ አለምዐሃይ አለምዐጎይ አለምዐጋይ	ሀይለ ሥላሜ ሀይለ ሥላሴ ሀይለ ሰላሜ ሀይለ ሰላሴ	ሃይለ ሥላሜ ሃይለ ሥላሴ ሃይለ ሰላሜ ሃይለ ሰላሴ	ጎይለ ሥላሜ ጎይለ ሥላሴ ጎይለ ሰላሜ ጎይለ ሰላሴ	ጋይለ ሥላሜ ጋይለ ሥላሴ ጋይለ ሰላሜ ጋይለ ሰላሴ
Ge'ez	አለምጸሀይ አለምጸሃይ		አለምጸሀይ አለምጸሃይ		ሀይለ ሰላሴ	ሃይለ ሰላሴ		

To be certain, while the renderings shown are logically possible they are not all necessarily probable, though the character classes used are entirely valid. It should be emphasized also that the renderings shown for languages following Amharic do not indicate the acceptable spellings in those languages but demonstrate how the pattern matching outcome would change with the corresponding locales. Indeed when searching for the same terms in a document known to be in the language indicated a matching pattern appropriate for the language would be applied.

The Syllabic Operator

Owing to the nature of language morphology where derivational rules are developed that rely on consonants and vowels as disassociated entities we expect to be able to apply these rules to written language. In an open syllabary the two (consonant and vowel) are fused together and so a developer is driven to seek out or create tools to isolate these character properties. Regular expressions resources are now indispensable in this field but are somewhat cumbersome to use when applied to syllabaries. The C-V property of a syllable corresponds directly to the rows and columns of the syllabary itself. The syllabic and form classes allow us to match a single character as a member of a group. This is analogous to folding cases ([:ለ:]) or specifying a specific case ([:%4:]). However, a limitation appears when we want to specify intersections of the two. An anticipated pitfall would be to attempt:

        [መበቀ][:%2,4-7:]

which matches two characters in sequence and not a member of the intersection:

        [ሙማ-ሞቡባ-ቦቁቃ-ቆ]

A convenient solution is to apply the same logic of the syllabic form matching class in operator notation.

        [ሙማ-ሞቡባ-ቦቁቃ-ቆ] → [መበቀ]%{2,4-7}

or in the negative expression:

        [መበቀ]%{^1,3,8-}

Applied in a small practical example we can develop an expression for the detection of the basic Amharic plural. We can start by defining a word stem as a sequence of syllables (assumed in the Ethiopic context):

```
$stem = "[:syllable:]+";
```

which without the utility of an operator becomes:

```
/^$stem(([:%4:][ቴታቴት])|([:%7:][ቾቿቾቾ]))/
```

and with the expressive power of a syllabic operator condenses nicely to:

```
/^$stem((%4ት)|(%7ት))%{2,4-6}/
```

This is of course a very simple example and it is intended that the power of the operator be evident from visual inspection by those only casually acquainted with regular expressions syntax. Applied to very large and real world text and natural language processing problems we can expect the operator to become an indispensable member of the regular expression toolbox.

Conclusion

The most widely applicable, as well as practical, property of Ethiopic characters is that of the implicit vowel of syllables. Syllabic properties are not tracked by the various character data files maintained by the Unicode Consortium, but could be.

Suggested amendments to the "Unicode Data" file then would begin with the introduction of the "General Category" character type "Ls" for "Letter, Syllable" to identify the code point as that of a syllable. With "Ls" to indicate the syllabic context the final three data fields for case types can be overloaded to indicate syllabic properties. The twelfth or "Uppercase Mapping" field would be applied to show the "family" association of a syllable where the character code of the base or "parent" syllable is recorded. The thirteenth field used for "Lowercase Mappings" would be used to record the vowel component of the syllable numerically, as demonstrated in our last section. The final fourteenth column for "Titlecase Mappings" would be applied to record in some way the final consonant in a closed syllabary (not an issue for Ethiopic which is an open syllabary).[†]

Accordingly the beginning of the Ethiopic region in the Unicode Data table would appear as:

```
1200;ETHIOPIC SYLLABLE HA;Ls;0;L;;;;;N;;;1200;1;
1201;ETHIOPIC SYLLABLE HU;Ls;0;L;;;;;N;;;1200;2;
1202;ETHIOPIC SYLLABLE HI;Ls;0;L;;;;;N;;;1200;3;
1203;ETHIOPIC SYLLABLE HAA;Ls;0;L;;;;;N;;;1200;4;
1204;ETHIOPIC SYLLABLE HEE;Ls;0;L;;;;;N;;;1200;5;
1205;ETHIOPIC SYLLABLE HE;Ls;0;L;;;;;N;;;1200;6;
1206;ETHIOPIC SYLLABLE HO;Ls;0;L;;;;;N;;;1200;7;
```

This approach is intended to be minimally intrusive on the current format by not adding additional columns to the table, which of course is also an option. It works so long as there would be no syllabaries having true case values. The family association is sufficient to indicate the consonant base of a syllable, which for the morphology of Abyssinian languages plays a less important role than do vowels.

Tracking the syllabic property of characters by an authoritative source such as the Unicode Consortium would undoubtedly encourage tool developers to support syllabic character classes. It is recommended that a technical committee be formed to address the matter for all syllabaries encompassed in the Unicode standard.

[†] Likewise, the eigth field for "Numeric value" could be overloaded for numerology.

Bibliography

1. Aklile Birhan Wold Kirkos, *Metsaheit Tibeb* መጽሐፈ፡ጥበብ, Neged Publishers, Addis Ababa, 1955 (1948 EC).

2. Asress Yenisew, *YeKam Metasebiya* የካም መታሰቢያ, Kokeb Printers, Asmera, 1958 (1951 EC).

3. Chaine, Marius, *Grammaire Ethiopienne Nouvelle Edition Beyrouth Imprimerie Catholique*, 1938.

4. Daniels, Peter T., <u>"Ha, La, Ha or Hōi, Lawe, Haut: The Ethiopic Letter Names"</u>, *Semitic Studies: in Honor of Wolf Leslau on Occassion of his 85th Birthday*, Otto Harrassowitz, Wiesbaden, 1991, ISBN 3-447-03168-9.

5. Dawkins, C. H. *The Fundamentals of Amharic*, Sudan Interior Mission, Addis Ababa, 1960.

6. Dillmann, August. *Grammatik der Aethiopischen Sprache*, Leipzig: C. H. Tauchnitz, 1899.

7. Dillmann, August, *Ethiopic Grammar* 2nd Edition Translated by James Crichton, London 1907.

8. Getatchew Haile (Editors: Daniels, Peter T., Bright, William), <u>"Section 51: Ethiopic Writing"</u>, *The World's Writing Systems,* Oxford University Press, 1996, ISBN 0-19-507993-0.

9. Guidi, Ignazio. *Storia Della Letteratura Etiopica*, Rome (Instituto per l'Oriente), 1932.

10. Ludolf, Hiob, *Lexicon Æthiopico-Latinum*, apud Thomam Roycroft, Londoni, 1661.

11. Kidane Wold Kifle, *Mezgebe Fidel* መዝገበ ፊደል, Artistic Printers, Dire Dawa, 1934 (1927 EC).

12. Kidane Wold Kifle, *Meshafe Sewasew Wegis Wemezgebe Kalat Hadis* መጽሐፈ፡ሰዋስው፡ወግስ፡ወመዝገበ፡ቃላት፡ሐዲስ, Artistic Printers, 1955 (1948 EC).

13. Mercer, Samuel Alfred Brown, *Ethiopic Grammar with Chrestomathy and Glossary*, Oxford: Clarendon Press, 1920.

14. Merigeta Girmstsion, *Lisane Agazi Ze'em Girma Nay Tigrigna Mezgebe Kalat* ልሳነ አገዓዚ ዘአም ግርማ ናይ ትግርኛ መዝገበ ቃላት, Indegena Ab Printers, Addis Ababa, 1984 (1977 EC).

15. Opsopaus, John, *Some Notes on the History of Isopsephia (Gematria),* http://www.cs.utk.edu/~mclennan/BA/SNHIG.html, 1995.

16. Praetorius, Franz, *Die Amharische Sprache*, Verlag der Buchhandlung des Waisenhauses, 1879.

17. Tesema Habte Mikael Gisew, *YeAmarigna Mezgebe Kalat* የአማርኛ መዝገበ ቃላት, Addis Ababa, 1958 (1951 EC).

18. Uncreditted, *Metsehafe Sewasew: Metseheit Aemro* መጽሐፈ፡ሰዋስው፦ መጽሐት፡አእምሮ፦, Tesfa Gebre Selassie Printers, Addis Ababa, 1989 (1982 EC).

19. Zena Markos Endalew, *Tintawi Ge'ez BeZemenawi Akerareb* ጥንታዊ ግዕዝ በዘመናዊ አቀራረብ, Artistic Printers, Addis Ababa, 1988 (1981 EC).