The System for Ethiopic Representation in ASCII

ይጥና ቢኒያም ፍርድይወቅ
Yitna Firdyiwek, *ybf2u@curry.edschool.virginia.edu*
ዳንኤል ያዕቆብ መኰንን
Daniel Yaqob, *yacob@padis.gn.apc.org*

1993 - 1997

Abstract

A convention for the transcription of Ethiopic script into the seven bit American Standard for Computer Information Interchange (ASCII) is presented. The convention provides a mechanism for computer interchange of Ethiopic text for single byte information systems. The system of transcription is readily human readable and encompasses all elements of Ethiopic text which includes; letters, numbers, and punctuation. The convention also provides a mechanism for the extension to multilingual text and may be employed for the keyboard entry of Ethiopic text elements.

Preface

In the time since the original publication of our paper in *The Journal of EthioSciences* [1] on the topic of representation of Fidel in 7-bit ASCII, the need became apparent to extend the system to encompass representation for Ethiopic numerals, punctuation, and mixed script notations. In the same period more was learned about the treatment of certain characters outside of Amharic that allowed for simplification of the ASCII representation. The following is a recapitulation of the original publication and an assessment of some of the more recent developments.

As we have indicated before, this system, though well developed, is still not in its final form. Further refinements will only come after many have had the chance to use it and test its strengths and weaknesses on their own. As the SERA system, joined now by a long awaited Unicode [7] domain assignement, work together to advance Ethiopic information processing we keep in mind the words of Abraham Demoz (1925 - 1994), to whom we have dedicated this work [2]:

> "...script reform calls not only for a competent professional assessment of the technical aspects of the script but also for a careful weighing of these against the psychological and socio-political factors that have a bearing on the written word and all that it stands for"

1 Ethiopic Letters

The Ethiopic writing system is a rich syllabary of at least 41 consonant classes each having generally 7 or 8 forms, and fewer having 12 or 13. The script has built in mechanisms for the extension to additional consonant classes -an occasional need as members of Ethiopia's and Eritrea's more than 70 language groups reach greater maturity in their own writing practices.

Cataloging of the complete writing system remains an on going task of the Academy of Ethiopian Languages (a division of Ethiopia's Ministry of Culture) in 1996 Ethiopia. Needless to say then that the Unicode and ISO-10646 introduction of Ethiopic in 1996 will not be comprehensive. Indeed, the character code assignment of nearly 50 Ethiopic letters will be left for future extensions to Unicode/ISO-10646.
The SERA system preserves the flexibility of the Ethiopic syllabary to adapt to changing needs.

SERA offers sound conventions suitable for any of the world's syllabic writing systems. The Ethiopic syllabary will be more readily know to its users by the names Ge'ez and Fidel; names that may be used through out this paper. We start now with the SERA table itself and follow with a discussion of its derivation.

The Ethiopic Script in ASCII

	1	2	3	4	5	6	7	8	(12)	9	10	11	12
	ግዕዝ	**ካዕብ**	**ሳልስ**	**ራብዕ**	**ሃምስ**	**ሳድስ**	**ሳብዕ**	**ዲቃላ** →					
ሀ	*he*	*hu*	*hi*	*ha*	*hE*	*h*	*ho*						
ለ	*le*	*lu*	*li*	*la*	*lE*	*l*	*lo*				*lWa*		
ሐ	*He*	*Hu*	*Hi*	*Ha*	*HE*	*H*	*Ho*				*HWa*		
መ	*me*	*mu*	*mi*	*ma*	*mE*	*m*	*mo*	*mWe*	(*mWu*)	*mWi*	*mWa*	*mWE*	*mW*
ሠ	*‘se*	*‘su*	*‘si*	*‘sa*	*‘sE*	*‘s*	*‘so*				*‘sWa*		
ረ	*re*	*ru*	*ri*	*ra*	*rE*	*r*	*ro*				*rWa*		
ሰ	*se*	*su*	*si*	*sa*	*sE*	*s*	*so*				*sWa*		
ሸ	*xe*	*xu*	*xi*	*xa*	*xE*	*x*	*xo*				*xWa*		
ቀ	*qe*	*qu*	*qi*	*qa*	*qE*	*q*	*qo*	*qWe*	(*qWu*)	*qWi*	*qWa*	*qWE*	*qW*
ቐ	*‘qe*	*‘qu*	*‘qi*	*‘qa*	*‘qE*	*‘q*	*‘qo*						
ቘ	*Qe*	*Qu*	*Qi*	*Qa*	*QE*	*Q*	*Qo*	*QWe*	(*QWu*)	*QWi*	*QWa*	*QWE*	*QW*
በ	*be*	*bu*	*bi*	*ba*	*bE*	*b*	*bo*	*bWe*	(*bWu*)	*bWi*	*bWa*	*bWE*	*bW*
ቨ	*ve*	*vu*	*vi*	*va*	*vE*	*v*	*vo*				*vWa*		
ተ	*te*	*tu*	*ti*	*ta*	*tE*	*t*	*to*				*tWa*		
ቸ	*ce*	*cu*	*ci*	*ca*	*cE*	*c*	*co*				*cWa*		
ኀ	*‘he*	*‘hu*	*‘hi*	*‘ha*	*‘hE*	*‘h*	*‘ho*	*hWe*	(*hWu*)	*hWi*	*hWa*	*hWE*	*hW*
ነ	*ne*	*nu*	*ni*	*na*	*nE*	*n*	*no*				*nWa*		
ኘ	*Ne*	*Nu*	*Ni*	*Na*	*NE*	*N*	*No*				*NWa*		
አ	*e/a**	*u/U*	*i*	*A/a*	*E*	*I*	*o/O*	*ea*					
ከ	*ke*	*ku*	*ki*	*ka*	*kE*	*k*	*ko*	*kWe*	(*kWu*)	*kWi*	*kWa*	*kWE*	*kW*
ኸ	*‘ke*	*‘ku*	*‘ki*	*‘ka*	*‘kE*	*‘k*	*‘ko*						
ዀ	*Ke*	*Ku*	*Ki*	*Ka*	*KE*	*K*	*Ko*	*KWe*	(*KWu*)	*KWi*	*KWa*	*KWE*	*KW*
ⷀ	*Xe*	*Xu*	*Xi*	*Xa*	*XE*	*X*	*Xo*						
ወ	*we*	*wu*	*wi*	*wa*	*wE*	*w*	*wo*						
ዐ	*‘e*	*‘u/‘U*	*‘i*	*‘A/‘a*	*‘E*	*‘I*	*‘o/‘O*						
ዘ	*ze*	*zu*	*zi*	*za*	*zE*	*z*	*zo*				*zWa*		
ዠ	*Ze*	*Zu*	*Zi*	*Za*	*ZE*	*Z*	*Zo*				*ZWa*		
የ	*ye*	*yu*	*yi*	*ya*	*yE*	*y*	*yo*				*yWa*		
ደ	*de*	*du*	*di*	*da*	*dE*	*d*	*do*				*dWa*		
ዸ	*De*	*Du*	*Di*	*Da*	*DE*	*D*	*Do*				*DWa*		
ጀ	*je*	*ju*	*ji*	*ja*	*jE*	*j*	*jo*				*jWa*		
ገ	*ge*	*gu*	*gi*	*ga*	*gE*	*g*	*go*	*gWe*	(*gWu*)	*gWi*	*gWa*	*gWE*	*gW*
ጘ	*‘ge*	*‘gu*	*‘gi*	*‘ga*	*‘gE*	*‘g*	*‘go*						
ⷘ	*Ge*	*Gu*	*Gi*	*Ga*	*GE*	*G*	*Go*	*GWe*	(*GWu*)	*GWi*	*GWa*	*GWE*	*GW*
ጠ	*Te*	*Tu*	*Ti*	*Ta*	*TE*	*T*	*To*				*TWa*		
ጨ	*Ce*	*Cu*	*Ci*	*Ca*	*CE*	*C*	*Co*				*CWa*		
ጰ	*Pe*	*Pu*	*Pi*	*Pa*	*PE*	*P*	*Po*				*PWa*		
ጸ	*Se*	*Su*	*Si*	*Sa*	*SE*	*S*	*So*				*SWa*		
ፀ	*‘Se*	*‘Su*	*‘Si*	*‘Sa*	*‘SE*	*‘S*	*‘So*						
ፈ	*fe*	*fu*	*fi*	*fa*	*fE*	*f*	*fo*	*fWe*	(*fWu*)	*fWi*	*fWa*	*fWE*	*fW*
ፐ	*pe*	*pu*	*pi*	*pa*	*pE*	*p*	*po*	*pWe*	(*pWu*)	*pWi*	*pWa*	*pWE*	*pW*

1.1 Considerations We Took in the Development of SERA

We have taken the following three considerations in coming up with our proposed standard

1. The system must be easy to type on a 101 keyboard. This entails:
   - finding the closest match between the Latin and Ethiopic phonetic systems (while being as systematic as possible with the inevitable exceptions),
   - limiting the number of keystrokes necessary for each Ethiopic character to a minimum, and
2. The system must be simple for humans to read without special decoders. This requires a flexibility of the phonetic mappings to accommodate differing writing practices of various language groups.
3. The system must also be easy for machine transcription. In this case, the systematicity of the mapping of Ethiopic to ASCII is exploited to make the machine transcription between ASCII and Ethiopic script (in word processors, for example) as fast as possible.

1.2 Development of the System

It may first occur to one when attempting to write Ethiopic script with Latin letters, to represent the first 7 forms with numbers as so:

Consonants:
h1 h2 h3 h4 h5 h6 h7
Independent Vowels:
a1 a2 a3 a4 a5 a6 a7

It is soon found in practice, however, that while this is a very simple system for representing the Ethiopic characters, it is not so pleasant to read or write with (e.g., "T5n1y6s6T6l6N6", "a1d5s6 a1b1b4"). This is true largely because our minds are not trained to associate the Latin script with Arabic numbers to form words. One will soon wonder why not use the Latin vowel letters to denote the 7 forms of the Ethiopic characters. This is where the trouble begins: How do you represent the standard 7 Ethiopic forms (plus the labiovelar "W" forms) with only 5 Latin vowels?

The first step we took was to assign a punctuation mark (the apostrophe ') and "I" for the two extra Ethiopic vowels (plus "W" for forms 8-12). So, following phonetic guide lines we came up with the following system:

Consonants:
h' hu hi ha he hI ho

Independent Vowels:
a' au ai aa ae aI ao

Again, after some trial use (e.g., "Ten'yIsITIlINI", "a'disI a'b'ba") we found that the writing can be made more readable if we used only one character for the pure vowel form. Then the system reduces to:

Consonants:
l' lu li la le lI lo

Independent Vowels:
' u i a e I o

and our sample text would look like: "TenayIsITIlINI", "'disI 'b'ba" which becomes a little easier to read as well as type.

After a short time a reader is likely to find that trying to "read a sound" from punctuation proves too difficult. Our minds have been conditioned for too long already to skip over apostrophes when reading possessive and contracted words. We introduce the principle now that whenever possible punctuation be avoided to represent spoken sounds and seek another alphabetic character to replace the apostrophe.

We find a suitable substitute in "E" but recognize right away the draw back of the extra "shift" required to type it. With only a small intuitive feeling one will come to realize that the 5^{th} form letters are used less often in writing than are 1st form. Hence a swap between the two forms makes the use of "E" a little easier and gives us the new table :

Consonants:
le lu li la lE lI lo

Independent Vowels:
e u i a E I o

and our sample text appears a little more naturally as: “TEnayIsITIlINI”, “edisI ebeba”
It is at this point that we began to notice two problems:

1. the 6th (or “sadis”) form of the Ethiopic characters occurs more often than any other form (about a third more often), and

2. the use of “e” for the first vowel makes the “look” of some familiar Amharic words peculiar, and the sound association is poor.

The quick solution:

1. stop using “I” for the sadis (sixth form) consonants, letting the consonants stand by themselves, and

2. allow the use of “a” for the first form independent vowel with “e”, and introduce “A” for the 4th form independent vowel.

Consonants:
le lu li la lE l lo
Independent Vowels:
e/a u i A E I o

Examples:

TEna ysTlN
adis abeba
Indemn kermachWal
zarE Tewat suq hEjE neber
manew smh? manew smx?

1.3 Ambiguity Problem With The Independent Vowel

This system is easier to read and type, but there is still a problem. If you have never before seen the word “TEna” how will you know if you are reading 2 Ethiopic characters or 4? I.E. “TE-na” or “T-E-n-a”? This problem of ambiguity usually occurs because it is not clear whether a consonant letter is a sadis (6th) form followed by an independent vowel form, or a syllable made up of the consonant and following vowel form. Of course, this is a problem only if the reader does not know the language. An Amharic speaker would not make such a mistake.

In another scenario, the name “Gabriel” can be read “ge-b-r-E-l” (correctly), or “ge-b-rE-l” (not quite correct, but okay when speaking fast). Though the ambiguity is there, whether you interpret the Latin as showing 5 (ge-b-r-E-l) characters or 4 (ge-b-rE-l) makes almost no difference.

These conditions may not always be true, however, and the difference does become a big problem for word processors and computer software for transcription. It is better then to insure that the characters are unmistakably represented; that we *not* delivered ሬ (rE) when ርኤ (r-E) is what we wanted. To accomplish this, our decision was to recycle the apostrophe ’ as a separator for independent vowels that appear after a sadis (6th form) consonant. Thus, we can rewrite Gabriel as “gabr’El” and modify our system, which now includes a third category, accordingly:

Consonants:
le lu li la lE l lo

Independent Vowels:
e/a u i A E I o

Independent Vowels Following a 6th Form Consonant:
l’a l’u l’i l’A l’E l’I l’o
l’e lU lO ←also

If we consider now an application for the remaining uppercase vowels; “U” and “O”, we find that in some instances, as shown in the 2nd row of the third category, the use of the apostrophe may be omitted without confusion.

Likewise, the leftover Latin consonants such as B, F, J, L, M, R, V, Y, are used as alternatives for their lowercase counterparts —this is often nice for personal names.

1.4 Lower Frequency Letters

Labiovelar series 8-12 may be mapped onto ASCII with a like logic using a two character syllable representation beginning with the upper case "W". A complication arises in representing 12$^{\text{th}}$ form syllables such as ኵ and ጕ which may be known to have both the sounds of "**k**^{w}**ï**" or "**k**^{w}**u**" and "**g**^{w}**ï**" or "**g**^{w}**u**". So in SERA both will be acceptable to indicate the same letter.

The extremely low frequency characters ፙ, ፘ, and ፚ may be given by "mYa", "rYa", and "fYa" respectively. Given the rarity of these letters "Y" is not deallocated as a secondary "y". It is acceptable to require the apostrophe to separate, for example ም and ያ in Roman script (as in m'Ya, or M'Ya) should the user find the uppercase "Y" construct necessary.

2 Punctuation And SERA Special Characters

SERA uses three special characters for writing Fidel in ASCII documents. Backslash, \ , is used to begin and end blocks of text written in Fidel or other writing systems such as Roman, Arabic, Hebrew, etc. Backslash can also separate blocks of text written in the same writing system but in different languages -this is useful to insure satisfactory transcription under different rules that languages can follow. Script and language separation is the primary purpose of backslash. We will see shortly that it can also be used for additional special purposes called "escapes".

Apostrophe, ' , also called single quote, serves the purpose of separation in SERA. Apostrophe appearing in an Ethiopic block of SERA text will not appear in the Ethiopic transcription. As discussed in Section 1.3 apostrophe is usually used after a 6$^{\text{th}}$ form consonant when a vowel follows. Another use of apostrophe is between two vowels to make the ASCII reading a little clearer such as ke'ityoPya instead of keityoPya. This use is left as a typist preference and makes no difference on the final outcome.

Backquote, ‘ , also known as "spacing grave", usually has the function of providing an "alternate" of the token that follows. Examples are the letters series ‘s, ‘h, ‘e, and ‘S for ሠ, ኀ, ዐ, and ፀ respectively. Also punctuations and numbers in Ethiopic writing that have alternatives in purpose such as ‘?, ‘: and numbers ‘1...‘10000.

Ethiopic Punctuation

,	፣	
;	፤	
:-	፦	
-:	፥	
‘:	፡	
:	፥	፡
?	?	፧
‘?	፧	?
‘!	¡	
::	።	
:|:	፨	
<<	«	
>>	»	
”	¨	
‘’	"Vocalized Sadis" control character for linguists	
‘ ’	Ignored If Alone	
‘1...‘10000	፩...፼	

The defaults for : and ‘? / ? use are presumed set by a user in the software using SERA. ፧ is the suggested default for ‘? use while no default is suggested for colon. Defaults may be reset to the alternatives at any point; \~‘| would set ? for ፧ which could be reset by \~? It is useful to have fixed ASCII definitions for ፡ and ፥ (namely ‘: and -:) while : is available for one or the other. Colon usage may be set and reset with \~-: \~‘: . To ensure the correct use of : and ? when exporting a document, the settings may be recorded at the start of the document.

2.1 Escapes

The core of SERA will always be its transliteration definition for the Fidel syllabary. SERA provides "escapes" or "switches" so that changes of language and scripts can be signaled to a reader without requiring special software to read the document. Special purpose escapes are also provided so that applications may communicate graphic elements and processing directives in an ASCII document.

The backslash character then is chosen for escapes in SERA as it is in agreement with the existing conventions of Unix, LaTeX, C, and other programming languages.

\	Script Toggle ... ← fidel → \ ← latin → \ ← fidel → etc. One or more punctuation characters following \ will not require a closing \ The script toggle terminates when the first nonpunctuation charcter is reached.
\\	\ The backslash escape is in following with normal rules for punctuation.
\~!	Escapes are ignored until closing \~!
\~xxx \~{do xxx}	Perform "xxx"

Whitespace is the required terminator for all escapes. When space, " ", is the following whitespace terminator it will be removed from the transcribed output.

2.2 Multilingualism

It is assumed that a document will be written primarily in two languages -which may be written in one or two scripts. The regular or bilingual script escape, "\" , always serves the two primary languages in the document. After switching to a third language, "\" will indicate a return to the first of the two major modes.

SERA applies the ISO 639 2 character and 3 character language names for multilingualism. The principle is identical to that adopted in HTML 3.0. The language name is then simply appended to the special purpose escape "\~".

Example Usage:

\~amh~eng	this is amharic	(Set Primary/Secondary)
\~tir	this is tigrigna	(New Third Language)
\	this is amharic	(Return To Primary)
\	this is english	(Secondary)
\~ar~gz	this is arabic	(Reset Primary/Secondary)
\	this is ge'ez	(Secondary)

Ethiopic and Arabic Numbers in SERA

The Arabic and Ethiopic numerals will both be given with the Arabic numbers found on Latin keyboards. The Arabic numbers, the more common, may be used in the usual way; Ethiopic numbers require the SERA alternate specifier, ' , before the number. An understanding of the Ethiopic number system will benefit the composer.

SERA uses a "Fidel For Numbers" of sorts where zeros play the role of vowels. This system allows for a simplification in the writing of Ethiopic numbers.

'10'9'100'80'7 = '109100807 = '10900807 = ፲፱፻፹፯

	ones	tens	hundreds	thousands	ten-thousands
	፩	፲	፻	፲፻	፻፻
		0	00	000	0000
፩	'1	'10	'100	'1000	'10000
፪	'2	'20	'200	'2000	'20000
፫	'3	'30	'200	'3000	'30000
፬	'4	'40	'400	'4000	'40000
⋮	⋮	⋮	⋮	⋮	⋮
፱	'9	'90	'900	'9000	'90000

3 Conclusion

We will conclude this paper with samples of applied SERA. Ethiopic text in Latin form will appear unnatural and even unsettling to the native reader reviewing SERA for the first time. The reader is reminded that SERA is *not* an effort to supplant Fidel with Latin script. SERA grew out of the needs of Email users to communicate with Fidel across mediums that would not support it. In the ultimate success of SERA, as a universal computer transport medium for Ethiopic writing, softwares alone are left to read and write the Latin script -providing the user with true Ethiopic.

As we have indicated before, this system, though well developed, may still not be in its final form. Further refinements will only come after many have had the chance to use it and test its strengths and weaknesses on their own.

References

[1] Firdyiwek, Y. and Yacob, D., "The Ethiopian Script in ASCII" *The Journal Of EthioSciences*, vo. 3 no. 1, 1993

[2] Demoz, Abraham, "Amharic Script Reform Efforts." Ethiopian Studies: Dedicated to Wolf Leslau on The Occasion of his Seventy-Fifth Birthday, November 14th 1981., S. Segert and A.J.E. Bodrogligeti, Eds., O. Harrassowitz, 1983.

[3] Daniel, Gabriel, Phrasebook English - Amharic Tigrina - Arabic, Washington, DC, (Self Published), 1990

[4] Diringer, David, Writing, New York, NY, Frederick A. Praeger Inc., 1962.

[5] Coulmas, Florian, The Writing Systems of The World, New York, NY, B. Blackwell, 1989.

[6] Committee for Standardization of Ethiopian Script, "View and Recommendation on The Unicode Technical Report ♯ 1 Draft Proposal on Ethiopian Script", unpublished, August, 1993.

[7] Becker, Dr. Joseph, et.al, "Conclusive Proposal for Encoding of Ethiopic Syllabary, USA (ANSI) & Unicode Consortium (UTC-95-055A)", XSoft/Xerox, December 9, 1995

Example Texts

Mixed Script Writing

Some Amharic (\amarNa)\ Expressions:

\1) TEnaysTlN!	\ Greetings!	(Lit: May He give health for me.) \
\2) Indemn : aderu?	\ Good Morning.	(Lit: How did you spend the night?) \
\3) amesegnalehu::	\ Thank You. \	
\4) mnm : aydelem::	\ You're Welcome.	(Lit: Nothing it is not.) \
\5) gra : gebeh?	\ Are you confused?	(Lit: Left understand?) \

Will be transcribed into:

Some Amharic (አማርኛ) Expressions:

1) **ጤናይስጥልኝ**!	Greetings!	(Lit: May He give health for me.)
2) **እንደምን፡አደሩ**?	Good Morning.	(Lit: How did you spend the night?)
3) **አመሰግናለሁ።**	Thank You.	
4) **ምንም፡አይደለም።**	You're Welcome.	(Lit: Nothing it is not.)
5) **ግራ፡ገባህ**?	Are you confused?	(Lit: Left understand?)

Classical Style

kexEkspir : befit : tEatr : yemiSfu : marlow-: jorj : pil : jon : lili :
tomas : kid : yemibalu : derasyan : neberu:: xEkspir : betenesa : gizE : gn : kesu :
befit : weym : besu : gizE : yeneberutn : hulu : belTo : asnaqacew::

be'10500907 : 'a : m : xEkspir : yetEatru : 'sra : Iyetesfafalet : hEdo : bzu :
genzeb : slageNe : wede : teweledebet : stratford : wedemibalew : ageru : hEdo :
wb : yehone : tlq : bEt : gezto : abatu : Inatu : mistuna : hulet : sEtoc :
ljocu : bezihu : bEt : wsT : IndiqemeTu : aderege::

Will be transcribed into:

**ከሼክስፒር፡በፊት፡ቴአትር፡የሚጽፉ፡ማርሎው፥ጆርጅ፡ፒል፡ጆን፡ሊሊ፡
ቶማስ፡ኪድ፡የሚባሉ፡ደራስያን፡ነበሩ።ሼክስፒር፡በተነሳ፡ጊዜ፡ግን፡ከሱ፡
በፊት፡ወይም፡በሱ፡ጊዜ፡የነበሩትን፡ሁሉ፡በልጦ፡አስናቃቸው።**

**በ፲፭፻፺፯፡ዓ፡ም፡ሼክስፒር፡የቴአትሩ፡ሥራ፡እየተስፋፋለት፡ሄዶ፡ብዙ፡
ገንዘብ፡ስላገኘ፡ወደ፡ተወለደበት፡ስትራትፎርድ፡ወደሚባለው፡አገሩ፡ሄዶ፡
ውብ፡የሆነ፡ትልቅ፡ቤት፡ገዝቶ፡አባቱ፡እናቱ፡ሚስቱና፡ሁለት፡ሴቶች፡
ልጆቹ፡በዚሁ፡ቤት፡ውስጥ፡እንዲቀመጡ፡አደረገ።**