**System for Ethiopic Representation in ASCII (SERA)**

*Daniel Yacob and Yitna Firdyiwek,*

*Indiana University, IN., University of Virginia, VA*

*yacob@apollo.aoe.vt.edu, ybf2u@curry.edschool.virginia.edu*

*Dedicated to the work and memory of Abraham Demoz (1935-1994)*

In the time since the original publication of our paper in "The Journalof Ethio Science & Technology" Volume 3, Number 1 on the topic of representation ofFidel in 7-bit ASCII, the need became apparent to extend the system toencompass representation for Ethiopic numerals, punctuation, andmixed script notations. In the same period more was learned about thetreatment of certain characters outside of Amharic that allowed forsimplification of the ASCII representation.

The following is a recapitulation of the original publication and anassessment of some of the more recent developments. A completediscussion of many of these changes (and a revised copy of the original publication) are available at the Rensselar Polytechnic Institute ftparchive under the file names ETHIOPIC and ADDENDUM.

As we have indicated before, this system, though well developed, isstill not in its final form. Further refinements will only come aftermany have had the chance to use it and test its strengths and weaknesseson their own. As Abraham Demoz, to whom we have dedicated this work,noted:

*"...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" (Demoz, "Amharic**Script Reform Efforts". ETHIOPIAN STUDIES. S. Segert and A.J.E.**Bodrogligeti, Eds. 1983).*

Any and all feedback will be greatly appreciated.

**PART I**

**1. The System for Ethiopic Representation of ASCII (SERA) Table**

Although some questions still remain to be answered regarding the number of "forms" to use for the ASCII/ETHIOPIC table, we have retained the original arrangement of twelve (12) for SERA pending decisions relating to the Unicode/ISO standards currently under discussion. (See Abass' paper in this conference.) We do not believe a change in the matrix of the table will affect the work discussed in this paper.

**The Ethiopic Script in ASCII**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |  |
| g`Iz | ka`Ib | sals | rab`I | hams | sads | sab`I | diquala |  |  |  |  |  |
| 1 | he | hu | hi | ha | hE | h | ho |  |  |  |  |  |
| 2 | le | lu | li | la | lE | l | lo | lWa |  |  |  |  |
| 3 | H2e | Hu | Hi | Ha | HE | H | Ho |  |  |  |  |  |
| 4 | me | mu | mi | ma | mE | m | mo | mWa |  |  |  |  |
| 5 | `se | `su | `si | `sa | `sE | `s | `so |  |  |  |  |  |
| 6 | re | ru | ri | ra | rE | r | ro | rWa |  |  |  |  |
| 7 | se | su | si | sa | sE | s | so | sWa |  |  |  |  |
| 8 | xe | xu | xi | xa | xE | x | xo | xWa |  |  |  |  |
| 9 | qe | qu | qi | qa | qE | q | qo | qWe | qWu | qWi | qWa | qWE |
| 10 | Qe | Qu | Qi | Qa | QE | Q | Qo | QWe | QWu | QWi | QWa | QWE |
| 11 | be | bu | bi | ba | bE | b | bo | bWa | (Q | is | Tigrignia) |  |
| 12 | ve | vu | vi | va | vE | v | vo | vWa |  |  |  |  |
| 13 | te | tu | ti | ta | tE | t | to | tWa |  |  |  |  |
| 14 | ce | cu | ci | ca | cE | c | co | cWa |  |  |  |  |
| 15 | `he | `hu | `hi | `ha | `hE | `h | `ho | hWe | hWu | hWi | hWa | hWE |
| 16 | ne | nu | ni | na | nE | n | no | nWa |  |  |  |  |
| 17 | Ne | Nu | Ni | Na | NE | N | No | NWa |  |  |  |  |
| 18 | e\a | u\U | i | A | E | I | o\O | e3(e3 | as | in | "e3re!") |  |
| 19 | ke | ku | ki | ka | kE | k | ko | kWe | kWu | kWi | kWa | kWE |
| 20 | `ke | `ku | `ki | `ka | `kE | `k | `ko | (`k | is | Chaha) |  |  |
| 21 | Ke | Ku | Ki | Ka | KE | K | Ko | KWe | KW | KWi | KWa | KWE |
| 22 | Xe | Xu | Xi | Xa | XE | X | Xo | (X | is | Chaha | ) |  |
| 23 | we | wu | wi | wa | wE | w | wo |  |  |  |  |  |
| 24 | `e | `u | `2 | `a | `E | `I | `o |  |  |  |  |  |
| 25 | ze | zu | zi | za | zE | z | zo | zWa |  |  |  |  |
| 26 | Ze | Zu | Zi | Za | ZE | Z | Zo | ZWa |  |  |  |  |
| 27 | ye | yu | yi | ya | yE | y | yo |  |  |  |  |  |
| 28 | de | du | di | da | dE | d | do | dWa |  |  |  |  |
| 29 | De | Du | Di | Da | DE | D | Do | (D | is | Oromiffa) |  |  |
| 30 | je | ju | ji | ja | jE | j | jo |  |  |  |  |  |
| 31 | ge | gu | gi | ga | gE | g | go | gWe | gWu | gWi | gWa | gWE |
| 32 | Ge | Gu | Gi | Ga | GE | G | Go | (G | is | Chaha) |  |  |
| 33 | Te | Tu | Ti | Ta | TE | T | To | TWa |  |  |  |  |
| 34 | Ce | Cu | Ci | Ca | CE | C | Co | CWa |  |  |  |  |
| 35 | Pe | Pu | Pi | Pa | PE | P | Po |  |  |  |  |  |
| 36 | Se | Su | Si | Sa | SE | S | So | SWa |  |  |  |  |
| 37 | `Se | `Su | `Si | `Sa | `SE | `S | `So |  |  |  |  |  |
| 38 | fe | fu | fi | fa | fE | f | fo | fWa |  |  |  |  |
| 39 | pe | pu | pi | pa | pE | p | po |  |  |  |  |  |

(Letters will be referred to both by their ASCII spelling and by their position on the above number matrix (e.g. "he" or 1/1). The columns arealso known as "forms" (e.g., first form, second form, .etc.) or by their Ethiopic names: e.g. g`Iz, ka`Ib, sals, . . .etc.)

**2. Considerations We Took in the Development of SERA**

We have taken the following two considerations in coming up with ourproposed standard

1. The system must be easy to type on a 101 keyboard. This entails:
   * finding the closest match between the Latin and Ethiopic phoneticsystem (while being as systematic as possible with the inevitable exceptions),
   * limiting the number of keystrokes necessary for each Ethiopic character to a minimum, and
   * placing the most frequently used keys as close as possible to the "home keys" row of the 101 keyboard
2. The system must also be easy for machine translation. Inthis case, the systematicity of the mapping of Ethiopic toASCII is exploited to make the machine translation between ASCII and Ethiopic script (in word processors, for example)as fast as possible.

**3. Development of the System**

It may first occur to one when attempting to write Ethiopic script with Latin letters, to represent the 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 simplesystem for representing the Ethiopic characters, it is not so simple toread or write in it (e.g., "T5n1y6s6T6l6N6", "a1d5s6 a1b1b4"). This istrue largely because our minds are not trained to associate the Latinscript with Arabic numbers to form words. One will soon wonder whynot use the Latin vowel letters to denote the 7 forms of the Ethiopiccharacters. This is where the trouble begins: How do you represent thestandard 7 Ethiopic forms (plus the "W" forms) with only 5 Latinvowels?

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

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

We find a suitable substitute in "E" but recognize right away the drawback of the extra "shift" required to type it. With only a smallintuitive feeling one will come to realize that the 5th form letters are used less often in writing than are 1st form. Hence a swap between thetwo forms makes the use of "E" a little easier and gives us the newtable:

**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, lettingthe 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?

**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 -- "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 translation. It is better then to insure that the characters are unmistakably represented. To accomplish this, our decision was to recycle the apostrophe ' as a separator for independent vowels that appear after a sadis (sixth form) consonant. Thus, we can rewrite Gabriel as "gabrEl" 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 rowof the third category, the use of the apostrophe may be omitted withoutconfusion.

**4. Some Commonly Asked Questions**

**1. Why not use "sh" for "x" and "ie" or "y" for "E"?**

These would make logical choices for readers familiar with rules in English but may not make sense in non-Englishspeaking nations where a form of the Latin script is used.It is desirable also to keep the keystrokes to a minimum forhumans, the parsing requirements of computers as simple aspossible, and media and transfer sizes to a minimum byavoiding multiple character representations when possible.

Further, the reader is left to infer the meaning "sh" as oneor two Fidel characters. The separator ' presents a solutionhere but again complicates parsing and introduces specialcase rules vs generalized ones. The acception to the generalrules also lends towards greater occurrences of spelling errors. "ie" may be an easier keystroke than "E" but againintroduces inference and parsing complexity. The choice isnot always logical as a phonetic model for the "ay" soundwith Latin letters when considering such examples as "die","vie", "pie", "lie", and "tie".

"y" occurs more commonly in speech and written text as aconsonant than as a syllabic form. Hence the lowercase Latincharacter is better reserved for the consonant to save onkeystrokes.

When an Ethiopic interface is available, these kinds ofquestions become input method issues and not file IO andtransfer which SERA was primarily designed for.

**2. What if I wish to show more sound for a sadis consonant?**

\* It is not always accurate to say that the vowel component of the sadis consonant is not spoken. For many words the vowelin the 6th form consonant is clearly enounced. If you wish to write in a more phonetic manner without loss of clarity;this may be accomplished by writing the 2-characterrepresentation form of the sadis consonant when it is needed.As you will recall we have redefined the 2-character form ofthe 6th consonant as "l|". We can mix the two character andone-character forms together in the same word to show whenthe vowel portion is voiced:

ysTlN = y|sT|l|N

tgrNa = tg|r|Na

alfelgm = alfel|g|m

TrE = T|rE

Writing with both the one- and two-character representationsof the 6th form consonant together may be more laborious tothe typist but has the advantage of giving the reader abetter demonstration of the word's sound when spoken. Themixed representation is not ambiguous and does not pose anyproblem for machine translation when going from Latin toEthiopic. If it would become a common practice to mix thetwo systems, we may wish to try alternate characters in place of the pipe ( "|" ).

**3. I see the ' used in other ways, what are the complete rules?**

The apostrophe was introduced as a separator to indicated that a vowel after a 6th form consonant does not modify theform of the consonant, ie "nE" is one Fidel letter and "n'E"is two. The principle of the separator may be appliedelsewhere when it enhances clarity. For instance, betweenvowels as in "beadis" vs "be'adis" or "keityoPya" vs"ke'ityoPya". Here, the ' helps prevent the reader fromslipping back into rules of English where the vowels would becombined into a single sound. Also ' following a consonantas in "t'" may be treated as another definition for the 6thform representation when convenient.

**4. Why Are Numbers Used with Letters?**

A problem that occurs when trying to represent Ethiopicscript phonetically in Latin is the presenceof Ethiopicletters that are phonetic equivalents. These cases areencountered with the two Ethiopic characters for "s" and "S"and the 4 characters for "h". Representing one of the 2ndforms with an unused Latin character, say F, R, or V, wouldbe a digression from phonetic norms and adds a level ofcomplication to the reading. In the case of what wouldbe h4 the uppercase "H" is chosen for representation. Thischoice models the husky "kh" sound that the character has inTigrinia and other languages.

For the more common type of email exchanges omitting the number 2 or 3 does not result in a loss of interpretation.The use of the ordinals becomes more important later if thetext is to be read and translated into Ethiopic script bycomputer.

**5. Why Does "s2" Come Before "s" ?**

The "2" is only needed to distinguish the difference between the two "s"s in Ethiopic script. In modern writing it isthe newer "s" (the 2nd "s" appearing in the fidel) that findsthe most frequent use in the spelling of words. The first "s"is represented as "s2" because it occurs less frequently inwriting. Were the 2nd "s" labeled as "s2" it would give thetypist considerably more finger work to perform.

**6. How was "e3" arrived at for the 8th vowel?**

The choice of "e3" is thought to be the best model for thesound of the character. The choice of a numeral to follow "e"will detract from the reading quality of the character, whichshould come at a small cost when its infrequent use isconsidered.

**7. Why is The Capital "W" Used For Diqala Forms?**

The uppercase "W" is used to remain phonetically consistent with the sound of the diqala forms (forms 8 - 12). The lower case "w" is reserved exclusively for consonant 21 with the "w" sound.

Thus, confusion and ambiguity is avoided with use of the uppercase "W".

**8. Why is "Wu" Used For the Letter I learned was "W"?**

Actually, both are valid under SERA. In different geographicregions, and at different times within the same region, people have been taught two different sounds for the 2nd form labiovelar (which one may have learned as a 6th form).Phonetic representation as "kWu" "kW" and "kW'", in example, is permitted for both ways a person may have been taught.

**9. Why is "hWa" used in place of "`hWa" or "h2Wa"?**

This is a break in consistency from how forms 1 through 7 of "h2" were represented. However, as "h" does not have forms after the sabI (the 7th form) there is no opportunity for confusion to arise from the omitted "2" of "h2W". Hence "hW"will be uniquely identifiable as representing diqala forms ofthe h2 consonant. The advantage of dropping the "2" in the diqala range, will be the keystroke saved for typists.

**5. A Full Sample Text with Statistics**

**WORD COUNT** : 170

**CONSONANT COUNT**

**Form 1:** 161 **Form 2:** 21 **Form 3:** 35 **Form 4:** 106

**Form 5:** 14 **Form 6:** 216 **Form 7:** 25 **Form 8:** 3

**VOWEL COUNT**

**Form 1:** 25 **Form 2:** 0 **Form 3:** 5 **Form 4:** 2

**Form 5:** 1 **Form 6:** 13 **Form 7:** 1

***From the Ethiopian Examiner January 1994***

**yeselamna yeIrqu konferans gizEyawiw mengst keslTan Indiwerd Teyeke**

bekefateNa gugut siTebeq yeneberew yeselamna ye`Irq gubae, ketahsas 9-13 1986 `a.m. beadis abeba ketema baderegew yeamst qen sbseba, beih`adEg yemimeraw gizEyawi mengst slTanun Indiyasrekb Teyeqe.

qedem blo paris lay sbsebaw Indidereg keTeyequt sebat teqawamibudnoc wsT, yesostun abalat wede ageracew sigebu awroplan Tabiya laybepolis bemasyazna bmaser mengst bzihu sbseba lay IndaysatefuadrgWacewal. yetasrut abalat, ato abera yemaneab, we/rit genet grma,ato mesfn tefera, ato alemayehu dErEsana ato genenew asefa (keidE`haq):ato seyum zenebe (kemed`hn) Ina ato ibsa gutema (keoneg) nacew.

mengst Inezihu sewoc lay yewesedew yeIsrat Irmja sewocnbeselamawi menged beageracew yepoletika hidet wesT Indaysatefaslemiyaderg bzu sewocn aseqoTtWal. beadis abeba yemigeNu diplomatocmyKEw yemengst Irmja yesbsebawn tesatafiwoc farhat lay bmeTal sbsebaw mnmbego wTEt IndayameTa yaderg yhonal bemalet hesabacewn gel`Sewal.

yeityoPya gizEyawi mengst (iH'adEg) besbsebaw lay saysatefqertWal. lezihum begizEyawi prEzidEntu beato meles zEnawi yeteseTewmknyat sbsebaw lepropaganda `alama bca yemidereg kentu sbseba new yemilnew.

**The following is an annotated list of appendixes that address various other aspects of SERA. The complete texts are**

**6. Other Appendixes** Not Included in this Version (as noted at the beginning of this discussion complete and up to date copies of all ofthese texts can be found at the Rensselaer Polytechnic Institute ftparchive under the file names ETHIOPIC and ADDENDUM.)

* 1. Questionnaire and Sample Text
  2. On The Character Specific Representation of Numbers:
  3. Generalized Pseudo-Code for Latin --> Ethiopic Translation
  4. Ethiopic in Emacs