

Localize or be Localized

An Assessment of Localization Frameworks

Daniel Yacob, yacob@geez.org

Ge'ez Frontier Foundation

Introduction

Localization in the broader sense is the transfer of cultural consciousness into a computer system, making the computer a natural extension of the society it serves. When computers are introduced into a society for the first time it has entered to displace an existing technology. When computers are integrated into the culture in such a way that it facilitates an aspect of the culture (production of literature, art, education, etc), or perhaps even helps revive a fading tradition, we may consider this adaptation as additive to the culture. If however the society must abandon some part of its culture that it values, then this adaptation is erosive upon the culture.

As computer scientists and localization engineers we have the responsibility to adapt computers and software to the needs of a society. When society is able to utilize new technologies in a way that is additive, we have succeeded. If society must lose some valuable part of itself simply to make use of a new technology, we have failed.

Localization infrastructures are the means by which traditions of a culture are supported in a computer environment. However “localization” for Ethiopian languages has thus far not extended beyond translation of the terminologies of foreign conventions. The challenge we have then is to go a step further and extend localization infrastructure to support Ethiopian requirements. Said another way, we must localize localization.

Elements of Cultural Subconsciousness

Another way of looking at localization is to consider what computers would be like today had they originally been created in Ethiopia. What would be different in the way we interact with and use software? The promise of localization is that we just might be able to achieve this hypothetical state. Hardware, being linguistically and culturally agnostic we need not consider. A possible exception is the computer keyboard which presents an alphabet for text entry. Amharic keyboards have been produced and marketed in the past and are still feasible today. With this one exception then, hardware has as much interest in language while powered off as it does when on –practically none. Language and culture do not enter into a computer system until the final level, the level of software. Software, like a sculptor’s block of clay, we can shape to the limits of our imagination.

Lets consider the basis of our imaginations. To a degree that psychologists oft debate and can not precisely specify, the workings of the mind are guided by (some would say “limited by”) the culture the mind was nourished in. Human language as elements of thought, composed of vocabulary and grammar, has been shown in tests to influence how a mind is able to perceive a problem and arrive at a solution. Cultures develop themes, motifs, and biases towards perceptions of visual and aural beauty, morality, justice and other values that define a sense of self. Learned at an early age and common within a culture, these preferences shape our collective psyche. Every culture develops these biases, all arbitrary, and none any better than the other.

Themes and motifs in Ethiopian culture would be found in the unique sound of Ethiopian music with its underlying pentave scale, played upon instruments found nowhere else. How about Habesha humour? Circles reside in our psyche as a visual motif, consider the shape of churches, a gojo beit, injera, a mesob, jebina, even crucifixes become round in Ethiopia. Spirituality is also so woven into the fabric of society that it is in our greetings and expressions that we use so frequently we forget the spirit within the words. The assumption here is that if computers reflected more of the frames of reference from our cultural subconsciousness, versus that of western society, computers would be more natural to us and thus easier to use.

How then would a window system created in Ethiopia, by software engineers with no prior exposure to other window systems, be different? Would the windows be circular instead of

rectangular? Would menus still exist in the same way? Would the window paradigm even be used at all? These questions are harder to answer, yet the door is always open to exploring new possibilities. We can assume where features are a matter of practicality (e.g. rectangular windows, following the shape of the computer monitor and books) that an Ethiopian window system would have arrived at the same place. The smaller details such as icon choices and audio warnings are the more likely to have changed.

Since we are ignoring hardware matters, assuming a degree of convergent evolution at the graphical level, what remains then? Whether it is for writing reports, exchanging email, reading news, accounting, or storing customer data, computers in the present day remain tools of text processing. Even graphical interfaces are dependent on text to explain the interface to the user. In the case of a computer server or a device like a mobile phone, the user environment may be completely text based.

So we are left with exploring textual expressions of culture. This may sound limited but we will soon see that quite a bit of what we understand about the world around us is expressed as text.

Discretizing Culture

It was inescapable that software would have a bias towards *some* language. Just as with literature software is written by people, and an author can not write a book in every language. If a book becomes popular and the publisher sees a big enough market, the publisher may invest the capital to have the work translated. But only then when the publisher is sure of a profit in return. Commercial software follows the same strategy; the larger, more profitable languages will always receive the highest priority.

In the OpenSource arena the rules are a little different. Software will be composed in the initial language of the author, the author will most often also be the publisher. *When* the software is available in a given language is no longer driven by commercial interests of the publisher. Rather, it is dependent on available volunteers with interest enough in the software to provide the translation. It may still be the case that we see translations appear first for the larger languages, but this is now just a consequence of the higher population of speakers of those languages and thus a higher potential pool of volunteers.

In both approaches to software development the technique of applying cultural conventions to the software is the same. The cultural aspects are first separated from the software itself. Lets consider soft drinks as an example. CocaCola has been sold world wide since before computers were invented. The soft drink container will have the CocaCola name on it in the local language. The container itself may be a glass bottle, aluminum can, tin can, or plastic. The size, shape and decoration of the bottle can be different in every regional market such that it is attractive and appealing to the local sensibilities. The soft drink within the container remains the same.

In software localization we are changing the Coke bottle only. The logic and computations that go on within remain unchanged. This approach relieves the author of the considerable labor and economic burden of having to change the software recipe for every market in the world. In this way only the external packaging need change. The notion of localization as “packaging” is of course an over simplification, but it upon this external aspect of an application that cultural conventions will be applied.

Software relies on other software. To run, software depends on other software components such as the operating system, libraries (DLLs), window system, registry, and so on. Internationalized software relies on an infrastructure for localization. The localization framework has evolved over time and is intended to support the conventions of all cultures and does so for the best known cultures. The localization infrastructure is again software, written by individuals, who can not know everything about all languages and cultures.

Most or all localization systems will be built upon or have some defined relationship to the ISO 14652 standard¹⁰. The Locale Data Markup Language, an open standard, may be considered more modern and reaches a little further in its scope, covering some additional conventions. What these systems provide is a computer’s ontology for culture, they are a means to identify, store, and provide to other software minute and discrete elements from cultural practices. The cultural aspects that will be

cataloged in such a way are those that are common, via direct or abstract correlations, with other cultures. We can review these conventions now and assess if Ethiopian practices are accommodated.

Perceptions of Time

Perhaps the clearest example of the different ways cultures view the world around them is demonstrated by our perceptions of the non-tangible entity of time. The seven day week is nearly universal but we disagree on what we consider the first day of the week to be. This is the case even under the same calendar system. Localization systems are prepared to handle at least this much difference in our perceptions of time. Though not yet as widely available, the more modern systems can go further and additionally support non-Gregorian (Western) calendars. This is of course important for Ethiopia which uses its own calendar system. Since this distinction from outside societies in how the year is understood is already familiar to most people, and can be supported in localization frameworks already, we need not explore it here.

More interestingly, Ethiopia views the passing of a day in a way this is not compatible with what localization systems can presently support. The difference in the day perception is also familiar to most; Ethiopia shares the 24 hour day convention with the rest of the world but differs on when the day begins. The two part day division around mid-day (AM for anti-meridian and PM for post-meridian) is also a foreign notion taken for universal in localization systems. In fact, other societies in the world, such as in Iran and Thailand, also recognize additional divisions in the day.

Localization systems permit translations for the AM and PM terms, but do not allow for additional day divisions to be added. Modern systems will apply XML to specify and store translations for language specific properties. For example, in LDML¹⁴, the AM and PM generalization is given by:

```
<am>A.M.</am>
<pm>P.M.</pm>
```

where the “<am>” and “<pm>” identifiers enclose a translation for the day divisions that will later be applied in a presentation of the time in a given language. The Amharic translation presently is an approximation that is no more than serviceable¹³:

```
<am>ጠዋት</am>
<pm>ከሰዓት</pm>
```

While these translations could be understood under the context of the foreign conventions that they map, they are not ideal for Ethiopia. Naturally, Ethiopia will want to apply its own conventions that are already millennium old. To do so would require revision of localization systems to support arbitrary day divisions and alternate start hours. A hypothetical way of doing so would be via the following XML structure^{2,6}:

```
<dayContext type="format">
  <dayDivision start="1:00">ወደቅት</dayDivision>
  <dayDivision start="4:00">ነጋት</dayDivision>
  <dayDivision start="6:00" dayStart="true">ጠዋት</dayDivision>
  <dayDivision start="9:00">ረፋድ</dayDivision>
  <dayDivision start="12:00">እኩለ ቀን</dayDivision>
  <dayDivision start="13:00">ከሰዓት በኋላ</dayDivision>
  <dayDivision start="16:00">ወደማት</dayDivision>
  <dayDivision start="18:00">ስደግዝ</dayDivision>
  <dayDivision start="20:00">ምሽት</dayDivision>
  <dayDivision start="23:00">እኩለ ሌሊት</dayDivision>
</dayContext>
```

In this structure the name for each day division is contained in the “<dayDivision>” tags, a starting hour for the period is given by the required “start” attribute (where hours are given under the western clock), and the start of the day is given by the “dayStart” attribute. By requirement this schema will also support the definition of the initial AM and PM divisions. The terms and start hours

in this schema instance are given for the purpose of demonstration only. Each day division name will have some alternative term available that may be more familiar to some readers than to others. This is in fact expected and does not pose a problem. Differences in word choice preference are most strongly associated with geographic regions; for instance the word used for a part of the day can change from Shewa to Gondar or Wello, etc. In essence, this is what localization exists for. A different lexicon can be applied for each region (or other way that distinguishes preference) as per the requirements of the effected users so long as an intuitive label is applied such that the intended users will be able to identify and select their preferred terminologies. Precise terminologies are not critical in this presentation, the proposal of a schema with the right abstractions to define the additional day divisions is the focus here.

This discussion of names for day divisions is only relevant if Ethiopia would want to use them when formatting time. The utility of the AM and PM notation is only helpful when presenting a time of the day in the 12 hour clock where the part of the day can be ambiguous (unlike when the 24 hour clock is applied). Since the 12 hour clock is used when time is spoken, it is assumed that it would be applied in digital presentations of time. Ethiopia does not have a well established precedence for digital time formats, but will want to establish them in the course of further integration of computers into society. An example digital time format under United States English conventions appears as:

Tue 21 Oct 2004 04:15:44 AM EAT

The equivalent date and time under the Ethiopian Amharic convention as available on Linux systems today appears as:

ማክሰ፡ አክቶ 21 ቀን 04:15:44 ጠዋት EAT 2004 ዓ/ም

As discussed earlier, this represents a loose mapping of some Amharic conventions onto an external reckoning of time. This is merely translation and not localization in its truest sense. What we see in this time presentation are the use of abbreviated day and month names (3 letters each), the Gregorian calendar, the western clock and numerals, and a time zone identifier (“EAT” for “East African Time”). A hypothetical idealized Amharic date and time presentation might render as:

ማክሰኞ፡ ጥቅምት 11 ቀን 10:15:44 ነጋት ኢትጊ ፲፱፻፺፯ ዓ/ም

Here the full day and month names are applied as is more common in Ethiopia, the Ethiopian calendar is used, the hour of the day reflects Ethiopian conventions as does the day division marker. Ethiopic numerals are also applied for the year and the time zone is translated. Localization systems will define many formats for date and time presentation for use in different circumstances such as short dates, medium, and full. It will still be useful to define abbreviated day and month representations for the shorter contexts. Time zone identifiers are optional, but will be useful when the time zone itself would be ambiguous (such as with email). Translations for time zone names is only newly supported in localization frameworks and Ethiopia will want to consider the advantages of translation (“ኢትጊ” is given here for example, short for “የኢትዮጵያ ጊዜ”). Where the time zone is placed within a time zone presentation is another detail to be worked out.

While predefined date and time formats are provided, each element of a date format is available such that a user may also construct their own formats:

ማክሰኞ፡ ጥቅምት ፲፩ ቀን 10:15:44 ነጋት ኢትጊ ፲፱፻፺፯ ዓ/ም

or

ማክሰኞ፡ ጥቅምት 11 ቀን 10:15:44 ነጋት 1997 ዓ/ም EAT

and

ማክሰኞ፡ ጥቅምት 11 (ሐና፡ማርያም) ቀን 10:15:44 ነጋት ፲፱፻፺፯ (ማቴዎስ) ዓ/ም

Thus far we have looked at how Ethiopia conceptualizes temporal divisions and points of reference. Aside from these differences in measure, Ethiopia also associates abstract properties with periods of time that is unique. In this last case we have imbedded into the day format the Tabot (ሐና፡ማርያም) and era name (ማቴዎስ) to present all information applicable to the particular day. While this additional information is helpful, it is upon Ethiopia to determine if it is “important enough” under its own criteria to warrant extension of localization systems to support.

At and before the time of Menilik II years could be given with respect to the creation of the world (-5500 EC) hence the present year would be given as ፪፬፻፺፯. Should the Ethiopian Orthodox Church express the wish to format dates with this reckoning, Ethiopia should consider it seriously. The Orthodox Church is a large and special case institute that may have special requirements, for instance the church represents the modern day Ge'ez user community.

Likewise the Borana, Sebatbeit and Silte are known to have used their own calendar systems⁸. The Sidama also use a five day market calendar along with a month system where each day in the 30 day month is given its own name. While not recognized as official calendars of the national government, Ethiopia may still want to support these systems if local interest is expressed to utilize them. Doing so in the OpenSource arena becomes a matter of collecting sufficient information about the systematic nature of the calendar such that it can be expressed mathematically and thus in computer algorithms.

Perceptions of Orthography

Perhaps the most fascinating aspect of writing is how the mind perceives the relationship between symbols and sounds. This cognitive development starts at an early age and depends very much on pedagogy, graphical features of a writing system, and writing practices for a particular script and language that have matured over the centuries. The psycho-linguistic perceptions are captured in localization thru the definition of character properties, terminology spellings, transcription and transliteration conventions, and some formatting rules.

Every symbol in a writing system will be assigned a set of descriptive properties that will later be used for textual manipulations, such as document formatting. At the highest level the classifications are very broad such as “letter”, “number” and “punctuation”. For Ethiopic script this is fairly straight forward, in the detailed properties, however, there remains some uncertainty. For example, should the Ethiopic symbols of Unicode ሳ, ኘ and ኙ receive ligature properties or be treated no different from other aggregate (ዲቃላ) letters? Do cognitive understandings of Ethiopic word space, ‘ ’, perceive the symbol to be more of punctuation or a space? Are these perceptions in agreement with practices which indicate a contextual dependency? The answer to this question, and others about the relationship of empty space, ‘ ’, to Ethiopic punctuation (required, optional, how many), and the interaction of Western punctuation with Ethiopic script are critical to the support of good quality formatting in word processor software. To achieve this end a typography reference would have to be provided by a respected authority on literature –possibly the AAU Amharic School.

No localization system supports the notion of Ethiopic character properties and classifications. For example the syllabic forms ግዕዝ, ካዕብ, ሣልስ, etc. present classifications that illuminate the relationship between letters and are important for abstract text processing such as pattern formation and the notion of character “folding”. No syllabary is any better off however. The property of equivalences between characters (ሀ and ሂ, ሠ and ሰ) is also not identified. Experience of the author here has shown that extending localization systems to support Ethiopic properties in place of western paradigms provides a more natural, easier to maintain, and effective means of working in the area of text comparison³.

Localization systems may support transliteration conventions (“ኢትዮጵያ” ⇒ “Ityop’ya”), and to a lesser degree for transcription (“ኢትዮጵያ” ⇒ “Ethiopia”). To date the focus has been to support transcription from non-Roman script into Roman. Ethiopia could benefit from standardized conventions here, for instance a single transcription for “ብርሃኑ” into Roman script would help database searches and other forms of record keeping¹. Ethiopia may also want to consider adapting localization of Ethiopic into the writing systems of major trade partners such as for Arabic, Hebraic even Chinese. Likewise adapting localization for the reverse direction, from non-Ethiopic into Ethiopic, could prove an aid to translation and other specialized work.

Other types of formatting are important to localization. Every country has a very specific way of formatting quantities like numbers and currencies, in particular how delimiter punctuation such as ‘.’, ‘;’ etc, the currency marker ‘\$’ (or other), and a negative indicator like ‘-’ are placed around the digits. The number of digits used to specify a zero value also needs to be defined. Ethiopia does not

follow strict conventions here, but they would be required by automatic formatting services in software. Structural formats such as a postal address will also be defined. Hyphenation is critical for the presentation of writing in many languages. Ethiopic writing has relied on the word space, ‘ ’, for tens of centuries and thus has had no need to consider the matter. Since the word space has fallen into decline in the era of mass publication it may now be ambiguous to know without any uncertainty where one word ends and another begins when falling across a line of text. This is a problem not so much for humans but for software that would reinterpret OCR (optical character recognition) scanned text, or further process text where word boundaries are critical such as text-to-speech software. Ethiopia may want to consider applying hyphenation in modern orthography, or at the very least provide a rules based system and some resource to recover unhyphenated text.

Localization infrastructures require translation for terms in a collection of important categories; such as the names of time zones mentioned earlier, also names of languages, nations, scripts and currencies. In the long term, most localization work will overwhelmingly be the translation of software vocabulary. Every other area of localization can be addressed once and not revisited. Software translation however does not end, it is a perpetual activity and thus is also the most labor and capital intensive area of localization and should not be underestimated. The reason for this stems from the continual introduction of new software and the continual improvement of existing software –new features will be added in news revisions and along with them come new vocabulary.

Localization of a spelling checker to support Amharic shares the same underlying limitations with translation activities. In both cases there are no longer technological limitations that impede good quality Amharicization but rather it is the loose conventions for spelling found in the modern era. The under-40 generation that will likely end up conducting vocabulary related localization does not feel as obliged to apply strict spelling practices as did the generation before. As computers become more and more the tool of education and the tool of composing and reading text, work here will be crucial to the future of Ethiopian languages. If poor spelling conventions are applied, they will become the standard for Amharic. If however care is taken, there is a great opportunity here to enhance the quality of Amharic writing itself. A carefully crafted Amharic spelling checker, for instance, would make it easier to apply the canonical word spellings for present and future generations. At issue then is for Ethiopian society to determine if spelling should be strict as it was in the past, or should be allowed to continue the present trend of permissible proximity spellings. In either case a reference lexicon, recognized by the applicable language authority, will need to be provided along with allowable variant spellings when applicable.

Perceptions of Sequence

Also considered a property of letters is the order in which they are sorted. In the simplest cases the sort order will also be the education order of the letters, that is, the order in which children learn the letters. For Ethiopic the sequence will be the familiar **ሀለሐመ**... (more precisely **ሀሁሂሃሄሀሆሐሉ**...) numerals would naturally sort as per their order **፩፪፫**... but what of punctuation? Also, should punctuation come before numerals or after? Should numbers come before letters? The ES 781:2002 standard of the QSAE provides some guidance here¹². The collation system presented in the ES 781:2002 standard is language neutral, individual languages may require specialized collation rules following historic practices.

Desta Tekle Wold and Kidane Wolde Kifle produced the foremost respected dictionaries for the Amharic and Ge’ez languages^{5,11}. Both dictionaries applied the **አበገደ**... order usually associated with Ge’ez. Supporting this order as an alternative for Amharic, Ge’ez and related languages should be considered. There is no extra cost to have an additional order other than in the time invested to define it in collation rules table. In the end it is just another option available to a user, it is never imposed upon a user.

Related to collation but serving a different purpose are list styles⁹. The most familiar list style is the “alphabetic” style used to enumerate a set of items. Traditionally the Ge’ez character set will be used here, e.g. **ሀ/ ... ለ/ ... ሐ/ ...** and not include extra letters introduced for Amharic and later day languages. When lists become long this rule might be broken and letters from a language’s alphabet might be inserted. The punctuation following the letter is also defined in a list style definition. Software needs to understand these styles exactly, such as the set of letters, punctuation rules, cycle

rules (after “T” the list typically will start over as per a numeral system, e.g. “ሀሀ”, “ሀለ”, ...) and a name to associate with each list style.

The Ethiopic numeral system is not presently supported in localization systems which, almost without exception, assume the western base-10 system. Consider the date examples presented earlier, no localization system can presently support Ethiopic numerals though some can support formatting with non-western numbers.

Different from these other examples is the format that Ethiopia applies to western numerals in a phone number sequence. The sequence in which phone number digits are formatted in is also defined by localization frameworks and is important for a good user experience. Consider the difference between the formatting of the identical phone numbers: 184-632 and the easier to read 18-46-32.

Perceptions of Honor

Another universal across societies are the titles that we assign to members of the society. Titles such as Mr., Miss, Ms. and Mrs. allow us to identify a person’s gender and marital status, no special qualifications are required. These titles are common in many parts of the world and thus have a simple abstraction. For a number of languages of Ethiopia: አቶ, ወ/ት, ወ/ት and ወ/ሮ, are the instances of these abstractions but in many cases they also change.

Some societies of the world do employ additional honorific titles, notably the Japanese. The Afar will use “Saqal” as special title of respect for men (except in the Sahoo dialect where it is used for Mr). The titles: ሊቀ, መምህር and አለቃ are widely known across Ethiopia, are not known to change with language, and could be defined for the nation as a whole. Other titles that might be added would have to be investigated on a community by community basis. For instance, in a localization project conducted for the Qebena community a survey of titles used by the Qebena would be made. Titles that represent an instance of a known abstraction (such as “Mr”) would be cataloged accordingly. If no available abstraction is found for a given title, then defining a new honorific type would be appropriate.

Perceptions of Quantity

Since Ethiopia has applied the International System of Measure (SI –the “metric” system) for a considerable amount of time, there is likely nothing here for Ethiopia to be concerned for. The system itself need only be identified and in some cases transcriptions for abbreviated metric units (i.e. millimeter ⇒ mm, ሚሊሜትር ⇒ ሚ/ሜ). If there are cases where local systems of measure remain in use, Ethiopia may want to consider extension of localization infrastructures to support them. As with honorific titles, doing so becomes more critical when the names change between Ethiopian languages.

Currency, discussed already with respect to formatting, also falls under this category. Names of international currencies are localized and less often currency divisions, such as the names of coins, are localized.

Perceptions of Symbolology

Not to be forgotten are the visual and audio signals that software invokes to indicate available services and statuses. Iconic symbols tend to apply easily recognizable images from everyday life. Often the symbols will embody a basic technology such as a house, dust bin, or post box. While the notion of a “house” is universal, architecture and building materials, and thus the appearance of a house, is not. For common icons used in a desktop environment Ethiopia would benefit from providing a local symbol set in the form of a human interface guideline⁷.

Audio signals are less likely to represent sounds from daily life. If a sound from software does unintentionally resemble an audio cue from the real world (police siren, alarms, etc) localization of the sound cue should be considered.

What Does Ethiopia Want?

System wide Ethiopian language localization has only been feasible in the last few years as systems and software libraries have been updated for end to end support of the Unicode character set. Localization efforts in Ethiopian languages have been successful to the extent that they were able to utilize the available infrastructure, but have now reached the limits of what localization systems can

presently support. The localization work conducted under these limitations is truly no more than translation of western conventions into a local language. While a promising start, Ethiopian languages have thus been localized for use in a computer setting most suitable for an Ethiopian living in North America or Europe wishing for a translated interface. This arrangement is not ideally suited for a computer that would be a natural part of an office environment in Ethiopia.

Localization endeavors should thus go further and pursue the course of extending localization frameworks to support the conventions of Ethiopian culture. Thus adapting localization for the needs of Ethiopia. This is done thru participation in open processes that govern the development of the localization frameworks themselves.

The lack of support for Ethiopian conventions in localization frameworks by no means stems from disregard for Ethiopia itself, but simply from the lack of available information to fulfill local needs. Ethiopia on its part has relied on outside developers yet has never come forward to state the local requirements. Thus the central question remains before both parties – “What does Ethiopia want?”

Of course it is a question that only Ethiopia can answer. This must be the starting question if true localization is ever to occur for Ethiopian culture. A Localization Centre directed by the Addis Ababa University Computer Science Department would offer a formalized way to develop an answer to this question and at the same time serve as an authoritative point of reference for the outside localization community. Such a centre would do well to determine the requirements for Ethiopian localization, drawn from Ethiopian culture, and following the ES 781 example that invited the public at large to participate in the process which is very much their own future.

Bibliography

1. Berhanu Beyene, *Issues with regard to application of IT in Ethiopian Studies*, International Conference of Ethiopian Studies XV, Hamburg, 2003.
2. Berhanu Beyene, Email correspondence, Hamburg, 2004.
3. Daniel Yacob, Regexp::Ethiopic, <http://search.cpan.org/~dyacob/Regexp-Ethiopic/>, CPAN, Version 0.14, 2004.
4. Davis, Mark, Loomis, Steven R., *Common Locales Data: Process, Issues, and Challenges*, 26th Internationalization & Unicode Conference, San Jose, September, 2004.
5. Desta Tekle Wold, አዲስ ያሳቢ ማረጋገጫ መዝገብ ቃላት *Addis Yamarña Mäzgba Qalat*, Artistic Printers, Addis Ababa, 1963 EC.
6. Fissehatsion Kassa, መርጫ አዳምና ገዕዝ *Merho Adamna Gə'ez*, Washington, DC, 1991.
7. GNOME Foundation, *GNOME Human Interface Guidelines*, <http://developer.gnome.org/projects/gup/hig/2.0/>, Version 2.0, 2004.
8. Gutt, Ernst August, Gutt, Eeva H.M., Hussein Mohammed Mussa, *Silt'e – Amharic – English Dictionary (with a Concise Grammar of Silt'e by Dr Ernst August Gutt)*, Addis Ababa University Press, Addis Ababa, 1997.
9. Hickson, Ian, Çelik, Tantek, *CSS3 module: List*, <http://www.w3.org/TR/css3-lists/>, Draft 7, 2002.
10. International Standards Organization (ISO), *Specification method for conventions*, ISO/IEC JTC1/SC22/WG20 14651, Geneva, 1997-2001.
11. Kidane Wold Kifle, መጽሐፈ ስዋስውጥ መዝገብ ቃላት አዲስ *Mäshäfü Säwasəw Wägəs Wämäzäb Qalat hädis*, Artistic Printers, Addis Ababa, 1948 EC.
12. Quality and Standards Authority of Ethiopia, *Ethiopian Standard ES 781:2002 Ethiopic Character Set*, Addis Ababa, October, 2002.
13. Unicode Consortium, *Common Locale Data Repository (CLDR) Project*, <http://www.unicode.org/cldr/>, Version 1.1, 2004.
14. Unicode Consortium, *Locale Data Markup Language (LDML)*, <http://www.unicode.org/reports/tr35/>, Version 1.1, 2004.