MODULE VI: Computer Localization

Language localization

Language localization(from the English term *locale*, "a place where something happens or is set") is the second phase of a larger process of product translation and cultural adaptation (for specific countries, regions, or groups) to account for differences in distinct markets, a process known as internationalization and localization. Language localization is not merely a translation activity, because it involves a comprehensive study of the target culture in order to correctly adapt the product to local needs. Localization is sometimes referred to by the numeronym "L10N" (as in: "L", followed by ten more letters, and then "N").

The localization process is most generally related to the cultural adaptation and translation of software, video games, and websites, and less frequently to any written translation (which may also involve cultural adaptation processes). Localization can be done for regions or countries where people speak different languages, or where the same language is spoken: for instance, different dialects of Spanish, with different idioms, are spoken in Spain than are spoken in Latin America; likewise, word choices and idioms vary among countries where English is the official language (e.g., in the United States, the United Kingdom, and the Philippines).

Keyboards that are sourced from the Gulf incorporate Arabic letters above each key. Regional languages in India are also being introduced into keyboards as part of localization.

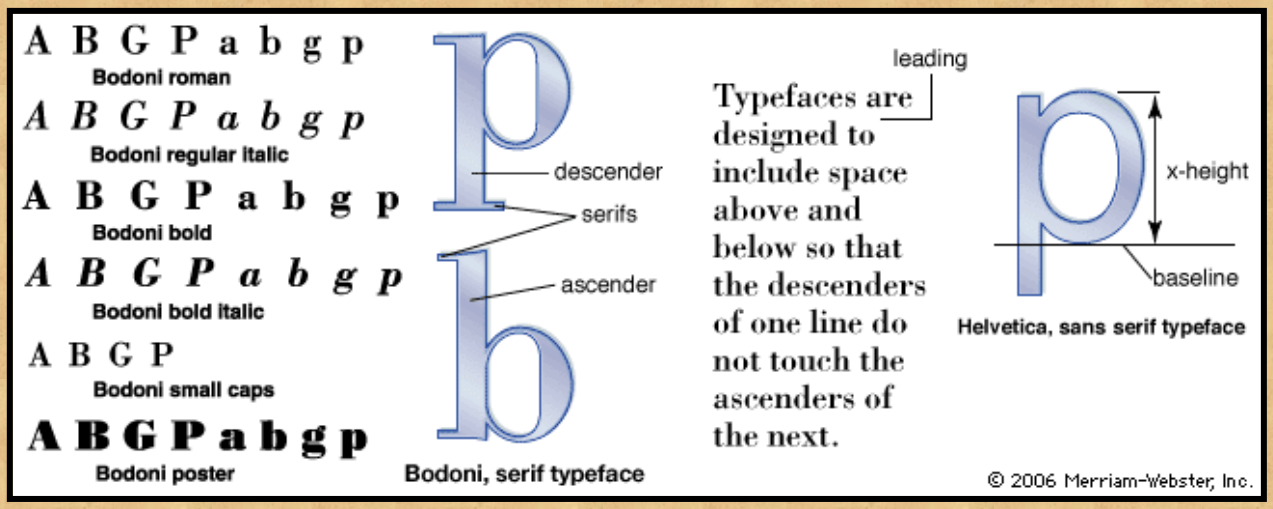
Using computers in the local language

Many organizations have been working to make the computers available in the different Indian languages. However, because of multiplicity of the languages (there are 18 languages recognized by the Constitution of India) the issue is quite complicated. There are only two essential components required to represent a language on computers - The language must have a Script, and it should be possible to represent the script on the computers. The computers understand English because they were developed by people who used English.

In the Bharatbhasha system, one can use computers in Indian languages without paying the extra cost for the hardware and software. Computer applications in Indian languages can be prepared by the people who know computer programming, therefore, this part of the work has to be taken up by those who know computers and also know Indian languages.

Nowadays there are virtual or "on-screen" keyboard that lets you type directly in your local language script in an easy and consistent manner, no matter where you are or what computer you're using. Malayalam is also included.

Fonts



The term *font* commonly refers to a type family such as Bodoni or Helvetica, which includes the entire alphabet in various weights (regular, bold, extra bold, etc.) and styles (roman, italics, or display type such as Bodoni poster). Type can be set in capitals ("caps"), lowercase, or small caps. The x-height of a font (the height of a lowercase letter that has no ascender or descender) will vary from typeface to typeface. The space between lines of type is referred to as "leading"—a term that dates back to a time when spacing was added with strips of lead. The specification of the example above is indicated as 10/11, or 10-point type with 11 points from baseline to baseline.

Unicode

International character-encoding system designed to support the electronic interchange, processing, and display of the written texts of the diverse languages of the modern and classical world.

The Unicode Worldwide Character Standard includes letters, digits, diacritics, punctuation marks, and technical symbols for all the world's principal written languages, using a uniform encoding scheme. The first version of Unicode was introduced in 1991; the most recent version contains almost 50,000 characters. Numerous encoding systems (including ASCII) predate Unicode. With Unicode (unlike earlier systems), the unique number provided for each character remains the same on any system that supports Unicode.

ASCII (American Standard Code for Information Interchange.

Data-transmission code used to represent both text (letters, numbers, and punctuation marks) and noninput device commands (control characters) for electronic exchange and storage.

Standard ASCII uses a string of 7 bits (binary digits) for each symbol and can thus represent $2^7 = 128$ characters. Extended ASCII uses an 8-bit encoding system and can thus represent $2^8 = 256$ characters. While ASCII is still found in legacy data, Unicode, with 8-, 16-, and 32-bit versions, has become standard for modern operating systems and browsers. In particular, the 32-bit version now supports all of the characters in every major language.

SOFTWARE TOOLS FOR TYPING LANGUAGES

NiLa, Varamozhi, Lipikaar are examples for Malayalam typing software. Lipikaar is a simple method for typing in Malayalam on an ordinary keyboard. It requires no learning, and within a few seconds you will be able to type in Malayalam any word that you can imagine. It works on all Windows Applications, MS Office, All Websites, Chat and E-mail.

Typing software is different from Transliteration software that is found in Gmail. Transliteration is a method in which you spell the pronunciation of the Malayalam word in English. The algorithm then converts the word into Malayalam script. There are several problems with this approach:

- ## *Ambiguous*

Transliteration is suitable for common words that can be spelled easily. However for words that are not part of our everyday conversation, figuring out the correct English spelling may not be as simple. Typing words accurately may require a trial and error approach and thus making it unsuitable for professional use.

- ## *Fluency in English*

Transliteration requires users to have fluency in English so that they can spell the Malayalam word phonetically.

- ## *Silent Characters*

There are many silent characters in languages like Malayalam, Tamil and other Indic scripts which may have different spellings but they are phonetically quite different. For an intelligent transliteration algorithm, it becomes difficult to interpret these words.

Writing Names, Addresses and other non-dictionary words

Since, transliteration is based on a dictionary approach, typing names, addresses and other non-dictionary or hybrid words becomes difficult.

Transliteration is more suitable for users who think in English and is meant for typing common words and few sentences.

TDIL (Technology Development for Indian Languages)

The Department of Information Technology initiated the TDIL (Technology Development for Indian Languages) with the objective of developing Information Processing Tools and Techniques to facilitate human-machine interaction without language barrier; creating and accessing multilingual knowledge resources; and integrating them to develop innovative user products and services. There are many issues that have to be tackled because of the diverse and complex language system in India.

Key issues include keyboard layout, IT localization, Vedic code set, Unicode for Indian scripts…

Indian Language Keyboard is categorized into three parts namely, Inscript, Phonetic and Typewriter keyboards. The Indian language alphabet table is divided into Vowels (Swar) and Consonants (Vyanjan). The INSCRIPT (Indian Script) keyboard was standardized by Department of Information Technology (DIT) and was declared as National Standard by Bureau of Indian Standard (BIS). Very recently a single solution was found out for the incorporation of a new Rupee symbol