Training needs for the Localization Industry in India

Shruti Khanna

Universitat Rovira I Virgili

shrutikhanna86@gmail.com

April 2008

A report prepared for the TechLink project.

http://isg.urv.es/publicity/isg/projects/2007_techlink/

Objective: Ascertain training needs for the Localization Industry in India.

Methodology: Compilation of information from various websites, newspapers, books; spoke to people involved in localization.

Contents:

1. Introduction

"Frog at the bottom of a well" is an old Asian proverb which states that a frog at the bottom of a well steadfastly believe that the size of heaven is only a small circle. When one climbs out of the well, only then, can the true vastness and magnitude of heaven be comprehended. Ageless as this proverb may be, it is also an appropriate description of the current commoditization challenge within the localization industry.

India is a vast country and plurality is its hallmark. Considering its linguistic, ethnic, social, cultural and geographical diversities, localization should have been the salient feature of the IT sector here. However, for a very long time, the major market for the Indian IT industry was clients based in developed countries. Even the boom in the Business Process Outsourcing (BPO) sector in recent years has produced localization of a different kind. The employees of the call centers located in India learn how to adapt to the ambience of foreign countries; they learn about the nuances of American accents, French slang and Australian names.

Hence, localization was not an issue. Even the government did not pay much attention to it. Most government websites continued to be predominantly in English, as against the official language policy. Although, the British variant of English is popular here; most word processors in government offices have American English as the default input language. Ideally, all the software sold in India must have Indic language support by default and the British variant of English should be the default.

2. Lack of localization

Considering the benefits of localization, we still don't see any major deployments, at least not in India. There is a lack of excitement about localization.

Some of the key challenges confronting the market at this point of time are:

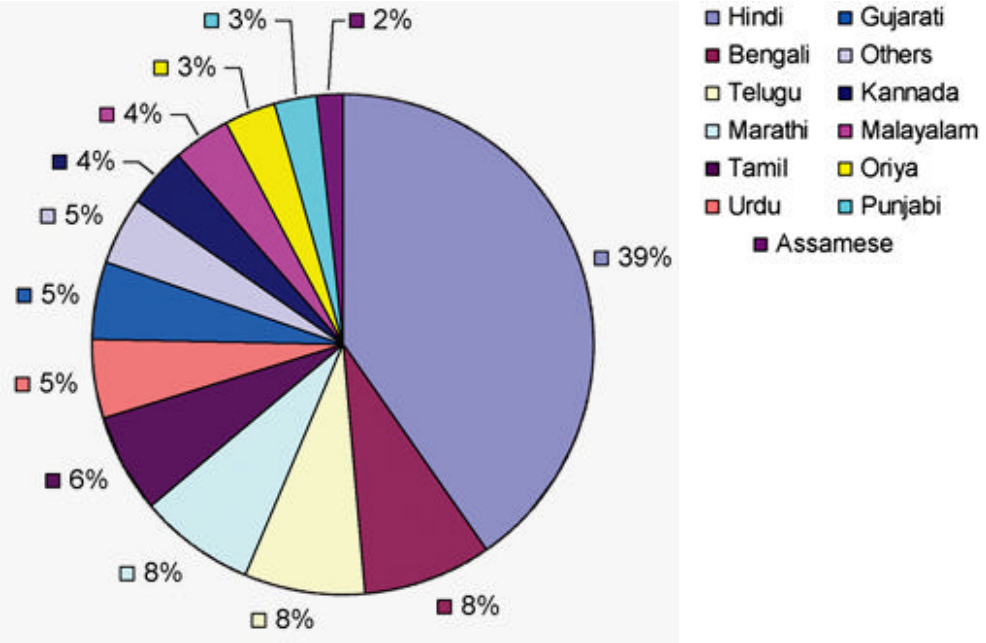
- Lack of universal standards for scripts and fonts, input devices and transliteration
- Limited availability of software and fonts
- Low availability of local language content

One plausible reason for the lack of localization might be the insufficient support for Indian languages at the operating system (OS) level. Interestingly, even though most foreign organizations use Unicode fonts for Indian languages, many Indian organizations still use the non-standard fonts. Popularity of non-standard fonts has hampered interoperability, collaboration and development of software tools in Indian languages. Indian computing researchers concentrated on high-tech problems like machine translation, speech recognition and optical character recognition, since these are intellectually more challenging and can help them get their papers published in research journals. For instance, the non-standard font encoding for Indian languages is often used, which is an unethical practice, as each developer puts fonts at his or her own free will. As a result, a glyph (character shape) is positioned at a location on the glyph table, by different vendors, at different places.

According to the information on portal "India Rising", India has a diverse list

of spoken languages amongst different groups of people. At least 800 different languages and around 2,000 dialects have been identified. Official languages include Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu and Urdu. Each of these has different dialects or variations. English is the 23rd official language. It is said in India that the dialect and food changes at every 200kms.

Distribution of Indian languages by native speakers



(Figure I)

On one hand IT is improving the quality of life in India; on the other hand the use of technology is still out of reach for many, the so called 'Digital Divide'. So, a person literate in Indian languages but not well versed in English is deprived of access to a vast store of information. Today 80% of the content on the Web is in English, which is spoken by only 8% of the World population and only 5% of Indian population. In a multilingual

country like India, with 22 official languages and 10 scripts, it is essential that tools for information processing in local languages are developed and made available at a low cost for wider proliferation of ICT, to benefit the people at large and thus pave the way towards 'Digital Unite and Knowledge for all' and arrest the sprawling Digital Divide.

To bridge this Digital Language Divide, one of the key technologies required is Cross Lingual Information Retrieval (CLIR). The proposed CLIR system aims at enabling a person to query the web for documents related to health issues and obtain the results, in Hindi. Still, the medium of communication in higher education, the judiciary, the bureaucracy and the corporate sector is English. Since English is the medium of interaction in IT systems too, structurally, such a situation aggravates the divide between segments of population that have access to computing and the ones that don't. India's English literacy rate is close to 65 percent, but most of these people cannot use the available computers because the User Interface (UI) isn't in their mother tongue The majority of the population is English-illiterate, computing has to speak a language the locals understand. This is where **user-interface localization** steps in. It enables non-English-speakers to access computers. With localization, developers can offer computer environments for education, as well as tools that give local and useful information -- such as water resource maps -- without requiring the knowledge of English.

To arrest this situation, an important step has come from the Ministry of Communication and Information Technology in the form of The Technology Development for Indian Languages (TDIL). TDIL has been mandated to bridge the digital divide by developing IT tools in local languages in India.

Since 1991, TDIL has sponsored research in developing Indian language computing resources, processing systems, tools and translation support systems and localization of software for Indian languages.

The other key initiatives have come in from development of Human-Machine Interface Systems and development of web centric applications. TDIL operates on a distributed innovation model through collaborations with 13 resource centers across India. Some of the notable milestones have come through CDAC, a collaborative partner of TDIL in form of GIST (Graphics and Intelligence-based Script) that has brought diverse users to employ local language IT tools. Applications have ranged from desktop publishing to subtitles in TV broadcast in various Indian languages. A Local Language word processor, 'LEAP' has brought desktop publishing to a large segment of population in a language they can communicate in naturally. Technical issues are to be solved; processes and procedures are to be put in place to attain the objective of localization.

(Figure II, Source WHP)

3. Where to start?

Computing involves a lot more than just the operating system. So, where should the developers working on localizing the GUI start? **Applications used by everyone -- browsers, web sites, email clients, office applications, and file managers -- should be considered first while localizing.**

The extent and usefulness of localization, especially in India, is immense. It is also a path to bring the experiences and opportunities of the world to India, thereby giving the local talent a scope to express at the global platform.

Localization can promote educational content in local languages. Localized education content on cheap localized Linux machines could promote the spread of IT in schools across the country, in no time. For proper localization, we need all the elements of a writing system (input method, editor, fonts, dictionaries, spell checkers etc.). More importantly, we need applications that give us the advantages of computer and communication. To get to many such resources on the Web, one has to use an operating system and applications which require basic English know-how but which a majority of the Indian population don't know.

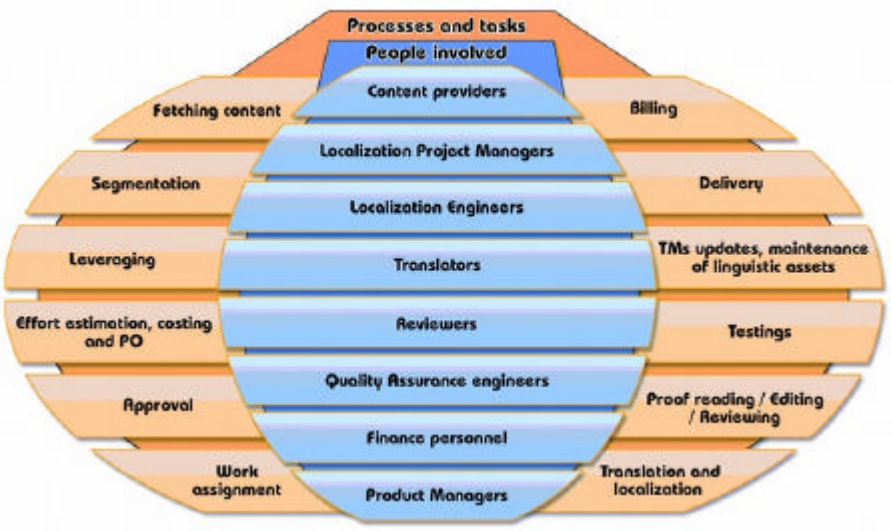
Where to start is a complex question, and the answer depends on whom you ask.

Mr. Mehta, whose primary focus is the Gujarati-speaking non-IT city population, believes a complete localized experience should empower

individuals to carry on personal and business activities in their own language. It should be able to position computers as simple tools to get the job done. The targets should be ease of use and productivity. Users should be able to create and dispatch documents, work with data and calculations, communicate with peers and associates, gain insight and knowledge about a field of interest, and satisfy their personal and business interests.

As a general progression, teams focus on getting the basic requirements taken care of and then move on to other things. Mehta and his project have the locale, fonts, keyboard layout and core translations for GNOME 2.6 for Gujarati completed. OpenOffice 1.1 translations are complete too and now the team is building an install set. People have started asking for applications specific to their industries and needs -- accounts, stock market, education and even software for opticians!

On the other hand, Mr. Shah is building localized solutions that cater to the general population. It involves enabling voter lists to be searchable in his native language, Marathi. Sankarshan's Ankur team is working with the aim of assisting the delivery of education.



(Figure III, Source WHP)

Building on popular software

While all the groups might be localizing for a different kind of audience, there is one similarity. All these projects are localizing a couple of open source software projects in their respective languages. Popular software is also tested over time, which means the chances of bugs halting the localization work are less.

Mozilla is a favorite application for localization teams because these are the core applications that a user would need.

4. Lack of awareness about localization

There is also lack of awareness regarding localization for the Indian ambience already available under various operating systems and application software. Localization for popular software is available in the form of language interface packs (LIPs). These packs are more than mere text substitution; since they take care of the localization issues. Language interface packs for Microsoft Windows and Office have been available for quite some time now and these are available for all major Indian languages. User interface in Indian languages are available for Free/Libre Open Source Software (FLOSS) such as Linux, OpenOffice etc. However, there is not much awareness about these tools. Support is also available for calendar based on Saka Era, but there appears hardly any popular demand for Saka Era calendar, even by the Indian government organization for which the Saka Era is the official calendar.

Localization and the government

India is a member of Unicode Consortium and World Wide Web Consortium (W3C). It became a member with the objective of proper representation of Indian languages in the Web Technology Standards and related standards. India has put forth several recommendations before that Unicode Consortium for better encoding of Indian languages. All modern and ancient Indian scripts have been or are being included in the Unicode standards. Proposals have been put forth for including special notations

and characters used in Vedic hymns, which have not been encoded so far. Also, there is need for further refinement in the formation of letters and ligature or conjuncts in Indian languages. C-DAC, IIT and many other organizations are contributing to the localization efforts, including machine-translation from one language to another.

The Constitution of India recognizes twenty-two languages,—apart from English which has the status of associate official language. The constitutional recognition of these languages is meaningless, unless these languages get their due share in the cyber space in this information age. The front-end of all government website and service delivery mechanisms should be simultaneously available in all the twenty-three languages, including English.

The central government and the state governments have embarked on an ambitious National e-Governance Plan (NeGP) and several Mission Mode Projects (MMPs) have been identified under the NeGP. E-governance is not so much about technology, as it is about efficient delivery of services. Good governance means citizens being able to avail of services from the government speedily, on a 24x7 basis, at a convenient location and in a cost-effective manner. Technology is used to improve the quality of service, and the delivery of service.

With many states adopting ICT (Information and Communication Technologies) as key instrument in improving the public service delivery especially in rural areas, rural India is all set for IT revolution like never before. Government organizations are creating huge IT infrastructure in terms of setting up of networks, financing rural cyber cafés and information centers. Government investment in e-governance is truly phenomenal.

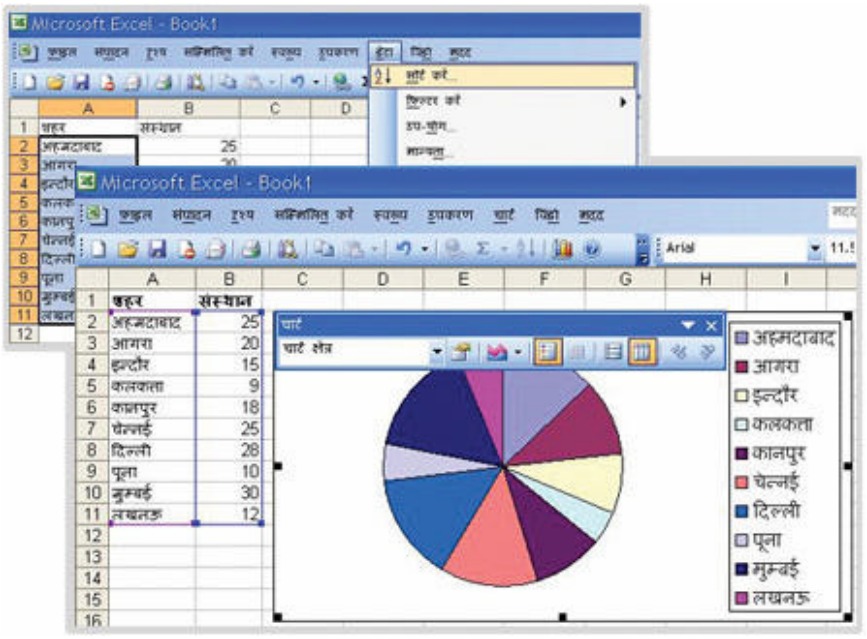
Several projects are under public-private partnership (PPP) mode. Indeed, considering the demand for services in rural areas, the spread of telecommunication network across the country and the computing power of today's inexpensive devices, the target of only 100,000 common service centers may appear too conservative.

Governments have set up several cyber kiosks to deliver government services to the people. There have been several initiatives across the country. **Gyandoot, Community Information Centers (CICs), e-Seva, CARD, LokMitra, BangaloreOne, e-Chaupal** etc are just a few examples. Experience shows that while the hardware was ready, the development of content for these projects lagged behind. The e-readiness of various government services offered to the people lagged behind. Consequently, instead of being self-sufficient, some such initiatives became dependent on government support leading to in-efficiencies. Even if all services provided by the government were e-ready, still the revenue from government services would not be enough to make a large number of centers a viable proposition. This is especially because governments have been withdrawing from several sectors and allowing private sector to provide these services. The private services will have to be the major source of revenue for the cyber kiosks. Thus, private sector will have to develop localized contents to cater to the local needs. Brief status of the achievements made so far:

- **Launch of Indian Language Fonts and Software Tools for free use:** The development of tools and software such as Fonts, Key- Board Drivers, TextEditors, Spell Checkers, Morph Analyzers, Dictionaries and Messaging Systems for 10 Official languages out of total 22 official languages have been completed and have been released to the public domain for free

use by the masses. The development of such tools for the remaining 12 languages is at advanced stage and will be completed shortly.

- **Operating System in Indian Languages:** The open source operating system has been localized and is available for Hindi & other 11 Indian languages namely Assamese, Bengali, Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Sanskrit, Tamil and Telugu. This will provide Indian languages text processing, web page designing facility and also Internet access in these languages.



(Figure IV, Source BhashaIndia)

- **e-Content:** UNICODE compliant e- content of approximately 16000 HTML & Dynamic pages in the domains of health, education, tourism and agri-business have been developed at various centers and made available.

- **Standardization:** Unicode Standards are widely being used by the Industry for the development of Multilingual Software. Department of Information Technology is the voting member of the Unicode Consortium to ensure the adequate representation of Indic scripts in the Unicode Standards. DIT finalized the changes in the Unicode Standard and majority of changes have been accepted and incorporated in Unicode Standards version. Initiatives have been taken to incorporate additional languages/ scripts such as Lepcha and additional characters and symbols of Vedic Sanskrit.

- **Web Internationalization Initiative:** Project "Web Internationalization Initiative" has been initiated with the objective of adequate representation of Indian languages in the Web Technology Standards being evolved by World Wide Web Consortium (W3C).

- **Human Resource Development in Language Technology:** There is shortage of trained manpower in the area of multilingual computing. A project for introducing 'Master' level and 'Post Graduate Diploma' level courses in the domains of Knowledge Engineering, Computational Linguistics and Software Localization has been initiated at eight institutions in India. The project aims at developing trained manpower in the field of Language Technology to overcome the present shortage of manpower.

- **Development of Linguistic Resources:** Centers for Indian Language Technology and Resources (CILTR) would be established to generate linguistic resources in all official Indian Languages. The proposed CILTR centers would co-ordinate with State IT department and Language

Departments for providing the inputs & feedback on technological issues such as Localization etc. The Linguistic Resources developed at these centers would be used synergistically for the development, roll out and subsequent improvement of the language technology products.

- **National Localization Research Centers (NLRC):** To promote localization in the country, it is proposed that National Localization Research Centers (NLRC) will be first set up as an autonomous organization under DIT. The outcome of the TDIL program in terms of standards, resources and technologies will be showcased and NLRC will focus on following:
  - Setting up standards wherever gaps are there
  - Providing training, consultancy in selection and application of tools
  - Provide test and certification facility
  - Maintaining a portal: for making available basic localization tools & Linguistic resources, eLearning, best localization practices, Market opportunities.

- **Basic Research in Language Technology:** Basic research would be undertaken to develop prototype technologies in the frontier areas of Language Technology such as Speech Synthesis, Semantic Web and Information retrieval etc.

However, some of the projects initiated by the government have failed primarily due to:

- Lack of commercialization of technology and lax timelines for projects.

- Moreover, the majority of the players in the sector are mid-sized

companies or educational institutions with limited financial muscle; hence they often tend to be restrained in terms of their research and development (R&D) spending on new technologies.

5. What is required?

- The key to success lies in reducing redundancies and enabling positive amalgamation of ideas and sharing of knowledge among government institutions, academia and vendors.
- A collective and combined approach is required to generate adequate content.
- Machine translation and creation of lexware, dictionaries and WORDNET also need a collaborative approach that can lead to a faster development and intelligent computer learning of the language.
- Both the central and the state governments need to encourage the use of local language applications in their departments. It is of equal importance to ensure that most of the software for workflow process and documentation systems is enabled in local languages. The government needs to ensure that all real-life applications step out of planning stage and get implemented at the respective departments, thereby providing relevant and real-time information in local languages to citizens of India.

E-governance is citizen-centric, rather than officer-centric. The services are necessarily to be interactive and transparent, rather than being merely presenting information. E-governance means taking government to the doorstep of the people. This implies that localization shall play a crucial role in e-governance.

Scope of localization

The recent dubbing of the popular film Spiderman-III in Bhojpuri language, a language that is not even recognized by the Constitution of India, points to the immense potential of localization in India. There is a vast and largely untapped potential in localization. Indeed, localization is not an India-specific requirement. Cyber space today is increasingly global and internationalized, reflecting the diverse linguistic, ethnic, social and cultural predilections of the world community.

Unlike in the past when the IT industry was primarily geared to cater to the demands in the developed economies, as the Indian economy develops and the service sector becomes more-and-more prominent, there is a growing market in India too. Thanks to the expansion of telecom networks with huge bandwidth throughout the country and the ever-declining cost of computers and networking equipments, IT has penetrated into even rural and tribal areas. Even geographically remote locations are enjoying excellent connectivity, and the remoteness, as far as IT is concerned, is not a handicap.

Usually a locale identifier for localization consists of a language identifier and a country identifier. Thus, only one variant of Hindi is accepted i.e. Hindi-India. However, Hindi as spoken and written in different Indian states are different. Hindi terminologies used in government offices in Bihar are quite different from those used in Rajasthan. It is not only true for Hindi but for most Indian language. Urdu of Hyderabad is quite different from Urdu of Meerut, so Urdu-India is an insufficient locale identifier. Bengali of Rajshahi is

quite different from Bengali of Chittagong, so Bengali-Bangladesh is an insufficient locale identifier to capture the localization requirements. Malayalam is written in two styles: traditional and reformed, which should be properly identified; localization should take care of this too. In terms of number of speakers as mother tongue, Hindi is the second most popular language in the world. Even languages like Bengali, Marathi, Punjabi, Tamil etc. are quite high in the popularity chart. Nevertheless, these languages are quite deficient in terms of localization in these languages. **A lot of work still needs to be done in machine-translation, voice recognition, optical character recognition, handwriting recognition etc. for these languages.**

G2C, G2G and G2E services offer a lot of scope for localization; even B2C, B2B services require localization in the globalized economy. As e-governance is transforming itself to m-governance, there is need for localization there too.

Localization and marketing

If the foregoing arguments appear to insinuate that localization is all about government, then it would be entirely incorrect. Internationalization and localization are primarily driven by the private sector in the globalized economy and can be viewed as part of the marketing processes of segmentation and targeting. The demand for localization is so huge that there are many companies catering to only localization of software, product design, packaging, promotion etc.

If you open the official website of some MNCs, it is likely that what you will

see in the browser will be the localized content as per locale identified by the website. So, do not be surprised if the official website of Microsoft opens in Hindi in your browser without your asking for the content to be in Hindi. As the penetration of PCs and other electronic devices increases and connectivity becomes widespread through broadband and WiMAX, there will be increasing demand for local content. A little more than a decade ago when the cable TV was being opened to the private sector, financial viability of regional cable TV channels was questioned. Today, there is a plethora of regional cable TV channels catering to whole lot of local and special interests. Indeed, it would be incorrect to view localization as addressing only to the needs of the rural market. Information technology has facilitated and accelerated the pace of globalization. It has bridged the gap between the people, and made distance irrelevant. Interestingly, the same information technology is also facilitating localization.

IT tools are increasingly being used in e-learning. E-learning offers immense scope for localization. Many ATMs offers options in Indian languages depending on the geographical locations. The landline and mobile telephony services also offer assistance in local languages.

Illiteracy and localization

An important aspect of localization in India is consideration for the illiterate and the semi-literate population. We just cannot afford to keep this vulnerable section of the population on the other side of the digital divide. In order to reach to the illiterate and semi-literate population, it is essential that the software have audio content and use symbols and jingles to communicate the message. Of course, this underlines the importance of

voice recognition and text-to-speech tools for Indian languages.

Localization and FLOSS

Unless there is a paradigm shift in the business model and proprietary software vendors bring about a sea change in their marketing strategy, Free/Libre Open Source Software (FLOSS) is likely to play an important role in the mass-based localization. **Proprietary software costs huge money and use of open source software may reduce cost to affordable levels.** Moreover, detection of pirated software will be easy in a wired world and it will be vulnerable to malicious attacks. Pirated software may be too risky as much as computers and other devices will be used not just to type text or browse static web pages, but also to make financial and contractual transactions and generate outputs of evidentiary value. Moreover, **volunteers can develop open source software even though it is financially unviable.** In this context, it is worthwhile to note that support for Tibetan script (used in Tibet/Bhutan/parts of India) was available under Linux, though not under Microsoft Windows, probably because Microsoft did not see enough revenue from support to Tibetan script.

6. Perspectives

The local language IT market in India is in a nascent stage, and the market is projected to grow at a healthy rate of 30% to 60%, from $40 million in 2007 to $100 million in 2010. The key drivers in this market at present are:

- E-governance initiatives that will have local language interfaces so as to reach all the citizens and get their feedback as well.
- Bundling together of multilingual websites and e-commerce sites so that citizens from all regions can search; get informed and educated; play, shop, buy and sell online all kinds of merchandise and services; make reservations for cinema, theatre, travel, hotels and so on at their convenience from the comfort of their homes.
- Localization of all audio-visual material including media campaigns, films, television serials, broadcasts and so on.
- Distance learning initiatives of educational institutes.
- Widespread use of mobile telephony and their auxiliary services.

The future of the mobile interface is the internet. For now, Indian youth are picking up mobile browsing rather fancily, but a few years down the line most users will connect to the internet through a mobile browser. So, India has to get ready with localization tools and content in local languages. “In the era of rising connectivity, localized content will be the king. Infinitely reproducible systems and applications in Indian language content may well be the kingmaker,” says Jitendra Shah in “Localization for e-governance.”

The Indian government is pushing forward all possible economic reforms to

take power and connectivity to remote areas. Hardware prices are going down every day. The growth rate of cell phones is explosive. Mobile users grow every month by about 2.5 million — a total of 30 million, nearly the population of Canada, in the course of a year. With some of the cheapest calling rates in the world, an investment of $25 and a monthly expenditure of $5 make it feasible for every Indian to possess a cell phone. Consequently, the language diversity of India offers a great opportunity for the localization industry to provide cell-phone interfaces in various regional languages. At the same time it also throws up great challenges for them to create the right kind of localization tools and services for the Indian broadband and mobile connectivity services market.

Local Language Software Market-Vendor Analysis

The Local Language IT market is in a development stage and the market is expected to grow at a healthy rate of 80 percent (CAGR) from $ 11 Million in 2002 to $ 64 Million in 2005. The key drivers that will drive exponential growth for this market will be

- Newer areas of application for Local language IT
- Broad based e-Governance initiatives that will employ local language as a front end to disseminate Government services to citizens and
- Bundling of multi-lingual software with PC's and other access devices

The market for Local Language IT is also likely to face a number of restraints that could inhibit the pace of adoption. They are:

- Lack of formal language-based IT training
- Limited usage of available local language applications
- Lack of spending

- Low connectivity

The Local Language IT market constitutes predominantly of word processing. Word Processing applications revenues in 2002 constituted 48 percent of the total market, with Packages and DTP constituting 20 percent and 18 percent respectively. While word processing software will continue to occupy a lion's share of the total revenues by 2005, package applications and local language multimedia and video applications are likely to grow at a significant pace. Reflecting the diverse application areas that local language IT will be used across in the future, consulting services revenues are expected to see a big jump. Consulting services revenues were 47 percent in 2002; by 2005 the consulting services revenues are expected to grow to 67 percent of the total market. Investments by Governments on e-Governance will find a way to the Local Language IT market. The share of e-Governance will increase from 38 percent in 2002 to 58 percent in 2005. The Local Language IT market constitutes of about 12 to 14 vendors. Most of the domestic players are regional and have limited access to the market. They offer both off-the-shelf products and custom made applications in all the major Indian languages.

The other set of key player in the Local Language IT market are international players. International vendors are yet to take off in a big way in terms of the application offering across different languages. IBM offers a Hindi version of Lotus Notes in India. However, the participation of international vendors is expected to increase in the next three years. C-DAC, owing to its pioneering initiatives in the Local Language IT market has acquired the leadership place with 48 percent market share in the year 2002. C-DAC is the top leader in both the product and consulting services space. Modular follows up with 23 percent market share. There is an overall

consensus on the benefits of e-Governance in India. While a wide variance exists between states in terms of their e-Governance initiatives, it is expected that over the medium term, a greater number of states will provide services to citizens over the electronic medium. Deploying Local Language IT as a part of State and Central e-Governance implementations will serve the cause of improving the reach and quality of services offered across a wide section of the citizens.

E-Governance Initiatives and Potential for Local Language Market State Governments have deployed citizen services in local languages and the early benefits are clearly visible. Early Government-to-Citizen Portals such as eSeva have proved the feasibility of the model. Frost & Sullivan expects this trend to extend on both scale and scope: a wider bouquet of services will be available to a larger section of citizens. Andhra Pradesh is the state with the biggest spends on Local Language IT contributing 23.6 percent to the total market revenues for the Industry. Gujarat is the second highest spender followed closely by West Bengal.

7. Key initiatives

In 2002, only 12 to 20 vendors were in India's local-language IT market; according to a report by Microsoft. Most of the domestic players were regional and had limited access to the market. They offered off-the-shelf products as well as custom-made applications in all the major Indian languages. The other set of key players in the local language IT market consisted of international players, but they were yet to take off in a big way in terms of their offerings across different languages. IBM once had a Hindi version of Lotus Notes in India. Since then, the participation of international vendors has increased many-fold in the last five years.

In 2002 some Indian state governments deployed citizen-services portals in local languages like government-to-citizen portals. This trend has continued, and almost all the state governments have a bouquet of services now available on the internet in regional languages. Since 1991, under the Ministry of Communication and Information Technology, TDIL has been working to develop IT tools in local languages of India. TDIL has sponsored research in developing Indian language computing resources, processing systems, tools and translation support systems and localization of software for Indian languages. The other key initiatives have come in from the development of human-machine interface systems and the development of web-centric applications. Some of the notable milestones have also come through CDAC, a collaborative partner of TDIL in the form of graphics and intelligence-based

script (GIST) that has brought diverse users to employ local-language IT tools. Applications have ranged from desktop publishing to subtitles in television broadcast in various Indian languages.

GIST is associated with the development of standards for Indian language applications in IT. The strength of the GIST Group has been to develop the backbone for Indian language technology and thereby deliver some of the most renowned Indian language products such as Indian ASIC Chip, Language Independent Program Subtitles, spell checkers, thesauri, OCX controls and application development libraries to implement Indian languages. A Local Language Word Processor, LEAP, has brought desktop publishing to a large segment of the population in a language they can communicate in naturally.

Though the efforts and initiatives taken by TDIL and CDAC are commendable, much of course, remains to be achieved. A number of global corporate giants such as Microsoft, Google, Nokia and Yahoo! are taking a keen interest in localizing their IT products and content into local Indian languages for computing and for the web.

Some of the IT giants such as Microsoft and Google have already taken big initiatives and made substantial investments to ensure that their latest versions of operating systems, as well as, other packages are available in Indian languages as soon as possible. This will ensure that they continue to dominate and maintain market share, as the non-English-speaking Indian population moves forward to play a dominant role in the development and growth of the modern Indian economy.

Similar initiatives have been taken by mobile hardware suppliers such as Nokia and Motorola and service providers such as Airtel, so as to maintain

and grow their market share in the Indian market. As we all know, the mobile phone is going to be a single device that is going to be used as a computer, television, phone, radio and so on.

Nokia learned quickly from the debacle of certain multinational corporations entry into India due to their one-size-fits-all approach. It started connecting with the people with the release of "Saare Jahan se Accha," a patriotic anthem, in the form of a ring tone in 1999. Further on, realizing the necessity of localization among various states, Nokia released India-specific handsets supporting Indic languages. Consequently, Nokia, with a market share of 68%, now leads the multibillion Indian cell-phone markets. The market is growing at an astounding rate of six million handsets per month. From 150 million at present, the mobile customer base is expected to rise to 500 million in four to five years.

Leading service providers on the web for surfing, such as Google and Yahoo!, have already taken steps to localize their software and provide interfaces in some of the major Indian languages. At the same time, some of the leading language service providers (LSPs) such as Lionbridge have established offices in India, so as to address the localization needs of global as well as Indian players.

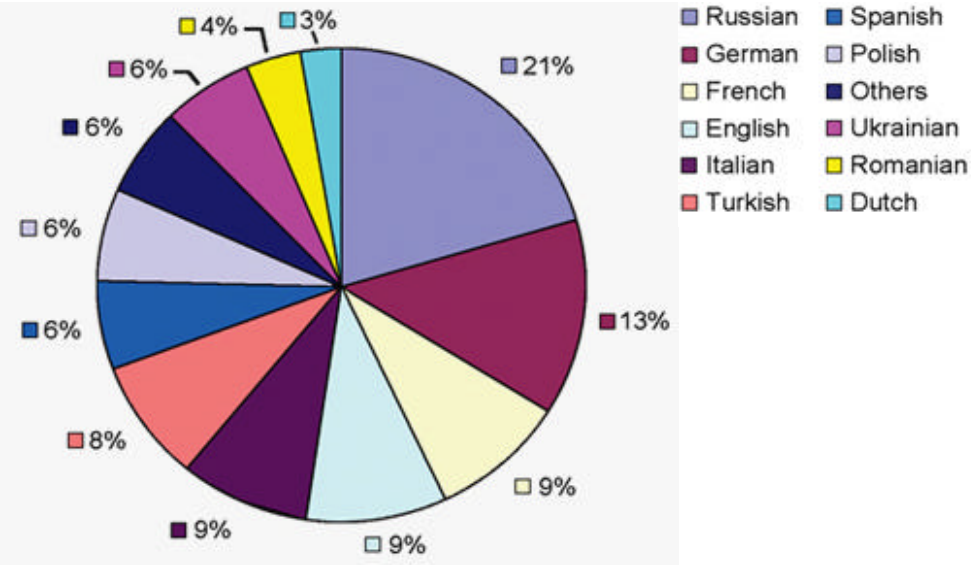
Others are contemplating the same course of action. Much of the localization business is being generated in India from sectors such as hospitality, travel, tourism, education, radio, films, television and so on. These sectors of the Indian economy have a massive scope for absorbing localized content. Commercial ventures — dubbing of English films such as Jurassic Park and Titanic into Hindi or even Tamil or Bengali, for example —

need a special mention for having set the biggest profit records that foreign films have ever made in India.

Comparison with Europe

Among the list of top 100 languages by population in the world, we observe that after Chinese, Spanish and English, the two most spoken languages are Bengali and Hindi. Together these two languages have more users in the world than English. In the same list we see more than 15 official Indian languages find a place in the top 40, whereas only ten European languages are on that list. India is, therefore, even more diverse than Europe as far as variety of languages and cultures is concerned.

Distribution of European languages by native speakers



(Figure V)

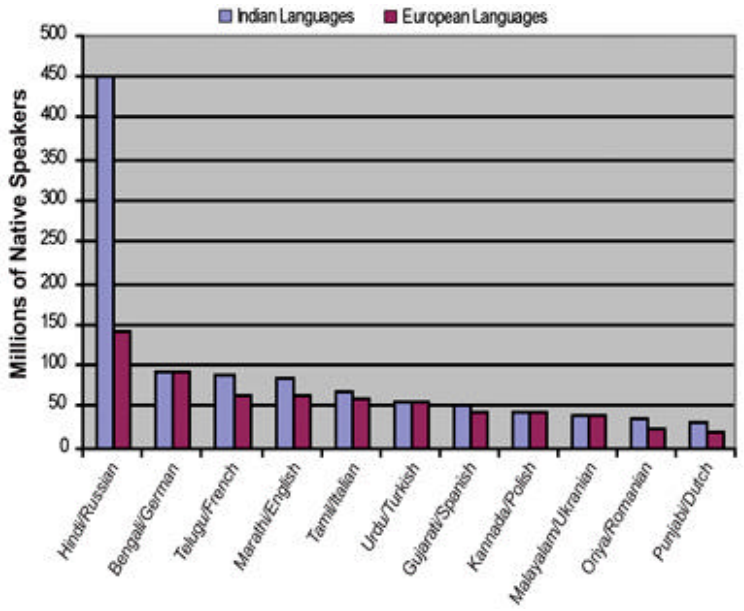
When we compare Indian languages with European languages sorted by the number of native speakers in India and Europe, we see that most

Indian languages have greater numbers of native speakers than their corresponding counterparts in decreasing number of native speakers from Europe.

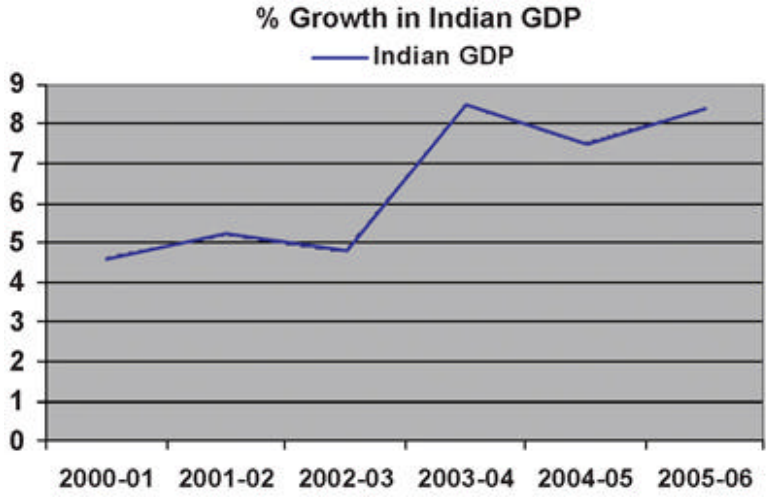
The population of India is more than one billion, and the country's economy has grown at a consistent annual economic growth rate in the range of 7% to 10% over the last few years. This economic growth is also reflected in the demand for content in local languages as is evident from the persistent growth in circulation of newspapers and magazines published in Indian languages.

With hardware, software and broadband connectivity costs continuously on the decline all over the world, the demand for converting content in English into regional Indian languages is certainly going to grow in the future to match the demand that exists today in Europe for conversion of content available in English into various European languages such as German, French, Spanish and so on - and vice versa.

Indian languages versus European languages



% GDP growth in India



(Figure VI)

8. Conclusion

Addressing the formal and informal education needs, as well as the information needs of more than one billion people speaking 23 official languages is a formidable task. In the Indian context, it becomes even more challenging when we take into consideration that 90% of the digital content available in higher education, judiciary and bureaucracy is in English and therefore, this content fails to address the educational and informational requirements of more than 95% of the non-English-speaking population.

On one hand, the non-English-speaking population is not motivated to use computers because digital content in their mother tongue is scarce. On the other hand, those who have the resources and knowledge to create digital content in regional Indian languages argue that only the English-speaking population is computer savvy. This is a vicious circle and therefore, needs to be broken by making content available on the internet in more and more regional languages of India. This will then trigger a positive growth loop when availability of more localized content in regional languages will attract more of the non-English-speaking Indian population to use computers.

The current state of the localization industry in India is far from perfect. It lags far behind the international level as far as the development and application of localization tools and technologies are concerned. We can say that it is still very much an improvised translation market. The various localization tools developed and available in English have to be first localized and tuned for use with Indian languages. Some of these will have

to be customized to suit the scripts of the Indian languages.

At the same time, the advancement that has taken place in the west while progressing from dictionary-armed translators to those using translation memories, linguistic verification tools, machine translation, statistical machine translators, translation quality assurance tools, translation project management tools and so on implies that the Indian LSPs have a long way to go before they can compete and secure projects directly from clients such as Microsoft, Google, Cisco, General Motors, Siemens, Nokia and Motorola. This, of course, is an opportunity for international LSPs and a challenge for the Indian LSPs. But there lies a great scope for co-operation amongst them. The vast translation resources available at competitive rates in India could be certainly used to outsource work to India. This is being done to an extent by some companies, while others are contemplating the merger-and-acquisition route to have a presence in India.

In the present era of human computer interaction, the educationally under privileged and the rural communities of India are being deprived of technologies that pervade the growing interconnected web of computers and communications. One good solution for this problem would be computers talking to the common man in the language he is comfortable to communicate in. Indian population has a significant percentage of people who are educationally under-privileged. There are still quite a large number of areas where people do not have the capabilities of 3R's. The digital divide under such circumstances is constantly on a rise, where on one hand we claim that India is leading in IT and on the other hand, the advances we make are totally inaccessible by a large number of countrymen. Under such circumstances, we cannot expect

rural/educationally under-privileged countrymen to use computers and IT products unless we remove the need of being literate, which exists as a barrier between them and computers.

The benefits of information technology will percolate to every Indian only when a computing interface is available in various local languages. This may happen sooner than expected. Computers, mobile phones and connectivity are fast spreading to every nook and corner of India. This is expected to result in a rapid growth in the demand for multilingual content, localization tools and technologies as well as translation and localization services. Keeping in view the various initiatives in progress in government and the public and private sectors, one could infer that India is poised to become a major supplier of localized content for the rest of the world in the near future.

India's economy also boasts an ever-growing middle class, attracting entrepreneurs from all over the world trying to capture a share of this growing market. The gross domestic product (GDP) has crossed $1 trillion, and by purchasing power parity one could claim it to be $4 trillion. India is gradually but steadily opening up sectors for foreign direct investment (FDI), this nearly tripled in the last fiscal year as more overseas investors flocked to the country. As per the International Herald Tribune, FDI into India rose in the last fiscal year to about $16 billion from just $5.5 billion a year before. The numbers do not include the billions of dollars that have been coming into the stock and bond markets.

Localization is the future in a globalized world and it cannot be any different for India.

References

Government of India (2007). Report by India to the UNESCO. Ministry of Communications and Information Technology, Department of Information Technology

DOHLER, Per N. (1997). *Facets of Software Localization*. A Translator's *View*. Translation Journal, nº1, Vol.1.

ESSELINK, Bert. (1998). *A Practical Guide to Software Localization*. John Benjamins Publishing.

WHP http://www.whp.fr:8080/whp/company/about-us

LISA http://www.lisa.org/

REDHAT http://www.redhat.com/

Acknowledgments: I wish to thank Dr. Anthony Pym, Director, Intercultural Studies Group, Universitat Rovira i Virgili, Tarragona, Spain for his generous support and feedback during the making of this report as well as for his insights and corrections to this report. I would also like to take this opportunity to thank Dr. Reinhard Schäler, Director, LRC and Dr. M. Sasikumar, KBCS/ETU/OSS Divisions, CDAC Mumbai, for giving me an insight to localization.