

## The SCALLA working conference “Crossing the Digital Divide”

This conference was the final conference of the SCALLA project<sup>1</sup>, funded by the European Union under the Asia IT&C programme. The SCALLA project aimed to increase collaborations between European and South Asian workers in human language technologies and software localisation. A first conference was held in Bangalore in 2001, followed by many visits to Europe by researchers from South Asia, culminating in this final conference in Kathmandu in January

At this final conference we aimed to consolidate what we had learnt over the previous two and a half years, and to look to the future. As well as revisiting computational linguistics for South Asian languages, we considered in greater detail the localisation of software to those languages and the reasons why this is important. As well as technical support for languages we also considered social and policy issues concerned with linguistic diversity. We aimed to identify gaps in current support, both technical and socio-political, contrasting the situation in South Asia with that in Europe. This led us to consider what advice and actions might be appropriate to help advance technical support for languages across South Asia, to help cross the digital divide. In this process we also expected to identify advice and actions which in turn should be taken back to help the European Union in its own efforts to remove its own internal digital divide.

How did this work out? The conference was organised into a number of sessions as set out below. Each session began with a challenge of key questions, with many presentations and much discussion. In the following we will start with those questions and discuss the presentations and how they answered, in part at least, those questions.

### 1. Setting the Scene

The conference was opened by welcoming addresses from

**The Honourable Mr Kamal Thapa**, Minister of Information and Communication, Nepal

**Mr Sharad Chandra Shah**, Vice Chairman of the High Level Commission for Information Technology, Nepal

**Mr Lochan Lal Amatya**, Computer Association of Nepal

**Mr Rudiger Wenk**, the representative of the European Union in Nepal

all of whom expressed their high level of commitment to information technology in the development of Nepal, and their appreciation at the conference being held in Kathmandu.

The keynote address for the conference was given by **Mr. Kanak Mani Dixit**, the member secretary of the Madan Puraskar Pustakalaya, the library of record of the Nepali speaking world. Mr. Dixit spoke of the need for full technological support for languages like Nepali, to enable citizens at large to benefit from the use of Information Technology. He illustrated his argument using Nepal, with its 23 million people most of whom are fluent in Nepali but have little English, who are barred access to computers not by cost since computers have become relatively cheap but by language. But he was optimistic, looking to the example set by the growing use of Radio, which shows such significant evidence of the creative application of technology within the Kingdom.

This keynote address set the scene for the conference which followed.

### 2. Language, culture, politics, and the digital divide.

We began by reviewing the state of languages in South Asia. Globally there are 6000 to 7000 languages, though the exact number cannot be determined due to the ambiguity concerning the difference between a language and a dialect. If two communities are unable to communicate with each other, is it due to a difference of dialect and accent, or a difference of language? This difficulty in identifying and cataloguing languages is very evident in South Asia, where the number of languages in the region has been estimated between the extremes of 500 and 2000 – the ‘true’ figure is probably around 1000.

The paper by **Dr. B. Mallikarjun** well illustrated this difficulty in counting languages. The 1991 census identified 1576 ‘mother tongues’ which were subsequently rationalised into 216 and then 114 languages from five language families. In this rationalisation, linguistic communities of less than 10,000 people were dropped,

---

<sup>1</sup> project ASI/B7-301/97/0126-05, partners Open University, Lancaster University, European Language Resources Association, National Centre for Software Technology (NCST, now part of CDAC) and the Indian Statistical Institute

and thus the number of languages is very much higher than that final 114. What is important to note is that there is a great variety of languages, written in a variety of writing systems. This variety and richness is repeated in other countries, and we heard from **Dr. Tariq Rahman** about the richness of languages in Pakistan with around 70 distinct languages, and from **Dr. Mark Turin** about the 100 or so languages of Nepal. Only in Sri Lanka, presented by **Vincent Halahokonege**, is the variety limited, with just 2 languages. In Bangladesh there is the single dominant language, Bangla, spoken by maybe 96% of the population, though it was contested that Sylheti seen by some as a dialect of Bangla is in fact a distinct language – but there are also a number of small ‘tribal’ languages.

However not all languages are equal. While in India there are no prohibitions about the use of languages in the media, 22 languages are given the special status as official languages named in the constitution, favoured with financial support for technology investment. In Nepal all languages have been viewed equally since 1991, but the previous two centuries of favouring Ghorkali/Nepali has left a legacy of Nepali dominance, with other languages still lagging far behind in the resources to support them. In Pakistan language policy has been to favour Urdu over all other languages even though Urdu is a minority language. In Bangladesh it is Bangla that is favoured over all other languages.

It is not simply a matter of treatment under the law that determines a language’s fate, there are significant social processes that are critical. The languages chosen by the powerful elite can bias people towards preferring those languages, and while this has usually meant favouring the official and national languages, it has often also meant a preference for English. One particularly surprising statistic from Pakistan is that Punjabi is by far the most used language, and yet it has low status and little support. While the people are drawn towards these languages of power, the elite themselves might use political and financial means to suppress other languages. This includes the level of technology provision to support these languages - the powerful languages get the support, the others may not.

This situation was compared with that in Europe, where there are also many minority languages, both indigenous and migrant, that are also inadequately supported with technology. Describing this situation **Prof. Jens Allwood** spoke of English as a ‘killer language’ (quoting Skutnabb-Kangas) and the fear that Swedish may itself be endangered as people move to English. Allwood spoke of the thousands of years of conceptual development stored in a language that would be lost if the language was lost, outlining many arguments in favour of preserving languages, distinguishing between museum preservation and sustaining a language in use, and proposing ‘language survival kits’.

We then turned to consider literacy and spoken versus written languages. Is literacy important, should all languages be given a written form? Spoken language is very different from written language in the way it works. Spoken language is interactive and multimodal with feedback and turn taking, the meaning being highly context dependent. It is very different from the written form of the language, and may even have a different syntax.

Contrasting the linguistic frameworks of South Asia and Europe, we noted that while there are not so many languages in Europe, both regions have a rich variety of languages with a relatively small number of dominant and more privileged languages whose dominant position is maintained a political, social, and economic forces.

### 3. Linguistics and Language Technology

We then turned to how much we know about these languages and what technology is available to support them.

**Prof. B.N. Patnaik** suggested that we should start by developing language technology to support tribal languages in order to overcome the ‘English divide’, the global dominance of English. Turning to models of language, he advocated the approaches of Panini and Tolkapiyam that had arisen in India, enriching them with western insights about language, but not to adopt western views of language. This moves away from universal views of language, a theme picked by **Prof. Harold Somers** in the context of Machine Translation which necessarily must use some common view of the two languages being translated - the conclusion from MT is that a universal interlingua is not possible, not even for closely related languages. As conjectured by Sapir and Whorf, languages reflect different conceptualisations of the world. However at the structural level there is much in common between languages, with the same grammatical phenomena occurring in many languages.

Before we can begin to develop models of languages and consider common models, we need basic data about those languages. **Prof. Rajeev Sangal** spoke about the development of language resources, corpora and lexicons, for South Asian languages and in particular for Indian languages. Over the past few years there have been significant advances here. with the development of dictionaries and the collection of many speech and text corpora which have been tagged for parts of speech. Tags derived from established western systems work adequately. This has led to a number of machine translation systems, but in the long run to cover the range of

languages needed, machine learning will be needed. **Prof. Yogendra Yadava** explained the situation in Nepal, with around 100 languages but where even the national language, Nepali, has very limited corpora and a dictionary that is in urgent need of updating and correcting and extending. **Prof. Pushpak Battacharya** then described a machine translation system based on the interlingual system UNL (Universal Network Language). A source sentence in one language is converted into UNL undergoing various normalisations in the process, and then the target sentence is generated by deconversion of the UNL. Studies are being made into English, Hindi, Marathi, and Bengali for application of this UNL-based approach to machine translation.

In discussions there was much concern that western and universal models of language might be inapplicable to South Asian languages, for both translation and tagging of corpora. BB Chaudhuri and BN Patnaik focused on examples of echo (eg guch-phuch) and reduplication (eg. the big big elephant) - in European languages reduplication denotes emphasis or de-emphasis, while in South Asian language the reduplication can take on new meanings.

We next moved on to statistical approaches to language engineering. **Dr. Ruvan Weerasinghe** reported an investigation into the use of statistical machine translation SMT) applied to the languages of Sri Lanka – English, Sinhala, and Tamil. SMT for European languages, such as English to French, works as well as the rule based systems like Systran. However when Weerasinghe applied SMT to English-Sinhala it did not work well, though it did do reasonably well for Sinhala-Tamil. Exactly why is not clear, perhaps because of the larger difference between Sinhala and English than between Sinhala and Tamil: Tamil is a Dravidian language while Sinhala is an Indo-Aryan language that has adopted many features from Dravidian languages. **Dr. Peter Juel-Henrichsen** then described his application of statistical methods to the description of unwritten unfamiliar languages: two different clustering methods were presented, which Battacharya conjectured resembled Wordnet similarity measures.

Henrichsen's talk had been about speech, and we continued now with speech technologies. **Dr. Roger Tucker** and **Dr. Ksenia Shalanova** presented their work using text-to-speech (TTS) software to give speech access to information sources, motivated by the need to overcome literacy barriers. They provide a toolset based on diphones and a speech database, give advice on the use of this toolset, and support knowledge sharing among a network of groups using these tools. Collaborators include Rajeev Sangal and Dipti Sharma for Hindi and Ramakrishan for Tamil, with other sites in South Africa and in Nigeria, though they hope in the long run for global outreach. **Dr. Hema Murthy** described work at IIT Madras on multi-modal interfaces, and in particular speech recognition for Indian languages. Speech recognition has started relatively recently with speech corpora being built up covering a range of speakers, native and non-native, dialects, speaking environments. There is still much to be discovered about the spoken forms of Indian languages before quality recognition systems can be created. Early work on speech had also been done on ISI Kolkata for Bangla, and Dr. B.B. Chaudhuri reported on the promising results that they had with syllables rather than diphones. He reported that the best speech recognition system he had seen was a statistically based Hindi system at IBM Delhi. Discussion then moved on to agreement that the best speech generation would mix data driven approaches with rule based approaches to cover unusual words.

This then led onto an extended discussion about literacy: its many degrees and types of literacy, the way information is structured can itself provide a lot of guidance to understanding (eg railway timetables), that technology can impose particular language practices (eg spelling and grammar checkers), that it could be possible to support spoken language directly without any written form, that communication is multimodal, that standardisation is about power, that video conferencing is the most popular software in village kiosks, that a pure speech word processor Castor had been produced by Dr. Raman at Stanford.

The development of language technologies in South Asia is progressing well, but is dominated by western approaches to language engineering and the models of language that underly these. There does seem to scope for building technology upon traditional South Asian views of language.

#### 4. Software localisation and local software.

Underlying all software that works in local languages are the computer encodings of the writing systems of those languages. In order to clarify the issues here, a short introduction into the history of the encoding of languages and writing system in the computer was given by **Prof Pat Hall**. There are three independent issues to be addressed in making computing technology work for a new writing system: the input methods including keyboard layout or other method; the internal code for the writing system, for example Unicode; and the output method, including fonts and rendering systems. It is the internal coding that is critical and this should capture the essence of the language regardless of the particularities of the writing systems. The ISCII encodings for the languages of India embody all the correct principles here. In discussion it was agreed that the situation in South Asia has been chaotic, with no acknowledgement of the need for standards, but that this is changing. Concern was also expressed about the lack of deep knowledge of Indian languages in Unicode, and thus the inclusion of inappropriate characters within Unicode, for example “leekar”.

The writing of Urdu using the Nastaleeq font was described by **Dr. Sarmad Hussain**. Urdu writing is cursive and multi-directional, with the 36 alphabetical characters changing into 15 to 20 different shapes as a function of context. Diacritical marks need to be moved around so as not to clash with each or the actual characters. So many open type font rules were required, well over a thousand, that the font designers ran out of space. While open type rendering works well for South Asian writing systems, technological limitations were encountered and need to be removed. Localised software needs to be used, and that requires both the marketing of the software and the training of people in its use.

The work on rendering Urdu in Pakistan is very recent, only possible now due to the development of Unicode and Open Type font rendering. Rendering for Indian writing systems came earlier, starting in the 1970s and 1980s, as described by **Dr. S.P. Mudur**. The need to separate input from output was known about then, but the limitations of computers at the time meant that Indic writing systems could not be rendered until the special GIST hardware had been produced. This GIST hardware and later software renderers were only possible because Indic writing follows very regular rules – however cursor positioning in composite characters (conjuncts) causes some problems. The renderers developed for Microsoft are also available for Linux.

In discussion two other OTF renderers – ICE and Pango – were described, together with issues around their use and the impact that these had on applications software. In India OTF fonts for Indian writings systems are being made freely available using public funds.

Keyboards were also raised in general discussion, and picked up by **Prof B.B. Chaudhuri** in the context of input devices in general. Keyboards should be reconfigurable so that a range of layouts can be used. OCR has been developed in Calcutta and is now being developed into a product at CDAC. Handwriting recognition and speech recognition are not well developed for Indian languages.

**Dr. Shailey Minocha** talked about human computer interaction and the way assumptions about users become embedded in the software and its interface. She illustrated this with banking websites in India and Taiwan, and explained the difficulty of getting evaluations of a site’s usability in non-western cultures, and the need for usability laboratories. Evaluating computer systems then becomes like evaluating any product in a new market. This led on to discussion of the ethical basis for localisation of software – whether it was driven by commercial interests in finding global markets, whether interface design was more like advertising.

The social reasons for localisation were then picked up by **Venky Hariharan**, who pointed out that only some 10% of India’s billion people were competent in English – which means that some 900 million people have no access to computers. A number of initiatives have distributed free fonts, and work is being done on localising Linux to 80 languages across Asia, which should be contrasted with MS Windows XP which has only been localised to 33 languages. These localised Linuxes add value to a billion dollar underlying open-source system, which are then distributed free. Harmonising terminologies is an important prerequisite, and the voluntary and distributed nature of open-source work makes this particularly challenging.

The contrasting situation in Nepal was described by **Amar Gurung** and **Dr. Rhoddy Chalmers**, with around 150,000 PCs mostly assembled locally and mostly without licensed software. Computer readable data has been converted to Unicode, keyboard drivers produced, a basic spell-checker and thesaurus produced, and Linux localised. The Linux localisation took around 18 person months of effort. All this had been achieved with much voluntary help and very little funding

In discussion Om Vikas emphasised the importance of standardisation, and the difficulties of multiple Linuxes – from IBM, Redhat, Mandrake and so on.

Localisation in Europe has become an industrialised process, with Ireland being the centre of this. **Prof. Reinhard Schaefer** explained that 95% of Ireland's output was localised US software and digital content, with exports of software from Ireland exceeding the exports of software from the US. The localisation is done in Ireland by outsourcing, with nearly 7000 companies ('vendors') dominated by 3 very large localisation companies Bowne Global Solutions, Lionbridge, and SDL International. These are supported by the industry association GALA, and standards making associations Unicode, W3C, LISA, and the new OASIS who developed XLIFF and the Translation Web Services Standard. There are also a number of publications supporting localisation, and several conferences. The industry is driven by short term financial considerations leading to bland interfaces. Recently the ELECTOnline website <http://www.electonline.org/index.php> has been produced to share knowledge within the industry. In discussion the potential of India and China as centres of localisation was indicated, particularly based on open source software. There are many difficulties such as the smallness of the local market, the smallness of the localisation groups, and the lack of standardisation of terminologies.

Computers had been developed in the West to handle languages written using the Roman writing system, and developments to handle the richer writing system of Asia only appeared very recently based on Unicode and Open Type rendering. This has enabled software in South Asia to be localised for local languages, based on open source systems like Linux using voluntary labour. This contrasts sharply with the commercial drive for localisation in the West where localisation only happens where the market is large enough to promise a return on investment.

## 5. Linguistics and language technology futures.

What is currently happening that could point the way forward in language engineering and linguistics for South Asia and South Asian languages? South Asian languages are important in Europe, and particularly in the United Kingdom, and two projects that point a way forward are the EMILLE and PLONES projects.

Language corpora are important for current linguistics, and **Prof Tony McEnery** has considerable experience in this, from gathering material in English in the UK for the British National Corpus and on the EMILLE project gathering material in South Asian languages in both the UK and in South Asia. Practices for gathering spoken corpora which were acceptable for English were just not acceptable to the UK South Asian communities. Some locations for data gathering were not acceptable, formal copyright assignment was an alien practice and less formal measures had to be used. **Vincent Halakonege** recounted similar experience in Sri Lanka where all contact had to be personal and obtaining the support of the top person in an organisation was critical. While the volume of corpus data for South Asian languages is very significantly less than that available for other world languages, the gathering of data has begun.

**Prof. Harold Somers** described his project to use language technology help health care patients with limited or no English (PLONES), where all too often untrained people act as interpreters during medical consultations. The current language of study is Somali, with plans to move to Punjabi or Bangla, using speech and icons to interact with the patient. Often the language does not have the vocabulary to handle the health situations, and often people speak a dialect different from that described by linguists. They may expect a formal interaction with their doctor, expected in their birth culture but no longer practiced in the UK. Literacy and familiarity with computers may be an issue, and the project has looked for help from research work on support for disabled people. The technology must be shaped to meet the people's needs. Discussion led to consideration of language technology having a pedagogical role, to teach language and even the doctor-patient relationships of the UK. We were reminded to use appropriate rather than high technology, that South Asia is multi-lingual rather than monolingual as in the UK or France, and that technical intervention should be accompanied by social and policy intervention.

In looking for ways forward, the PLONES project had looked outside linguistics to the assistive technologies developed for disabled people. Similarly translation theory might broaden our view of localisation and what language technologies might be appropriate. **Prof. Anthony Pym** gave us this translator's perspective. Localisation is a major employer of translators, but with the tools available to take care of equivalence, they become more concerned with controlling the source, with internationalisation. Human translators can do more than maintain equivalence, they can change the discourse using their knowledge of the cultural values of the target readers. However we must recognise that readers may wish to refer back to the original, particularly when that original is in an international language like English. We also need to consider translation versus language learning, and if the project is long enough it may be better to teach people the target language rather than translate material. Discussion confirmed this complex multilingual situation in localisation and internationalisation.

We see then that basic to all language engineering must be acquiring more knowledge about the languages to be enabled, and that this needs the gathering of raw data in the form of language corpora. Methods for gathering this data need to be different for South Asian languages and in South Asian contexts. While localisation can be viewed as translation, we need to look more widely to contextualise what localisation is aiming to achieve, from assistive technologies to translation theory.

## 6. Localisation and local software futures

While linguistics has been developing over millenia, and language technology has been with us since the advent of computers fifty years ago, software localisation is relatively new, arising only some 25 years ago with limited aims. Localisation as currently practiced is ad hoc, using techniques and facilities developed for other purposes. How should software localisation be taken forward?

The Language Observatory project of **Prof. Yoshiki Mikami** was inspired by his travels around Asia and the many forms of typewriter he discovered. While it took 3 to 4 hundred years for printing to become localised from West to East, it took only 50 years for typewriters and 15 years for computers. This has been accompanied by a simplification of the writing of the languages, but also by a reduction in the range of languages covered so that today Windows XP only covers 50 languages of which only 2 are Asian. Discussion covered the systematic layout of keyboards and where the non-spacing vowels were placed relative to the consonants – but even where such designs take place and are demonstrated as giving benefits, people are very resistant to change.

**Dr. S.P. Mudur** described the work that NCST did for Microsoft in enabling the input and output of Indic Languages at the code level, and the sad experience that this the only worked in Notepad because other Office packages did their own input and output. We need to recognise different levels of localisation as a function of return on investment. He identified six levels – 1 no translation, 2 just user instructions, 3 enabled for local language input/output, 4 translated menus and dialogues, 5 translated online help and tutorials and readmes., 6 convert all cultural elements. During discussion this was related to levels in international marketing, and to the ignorance of software producers about actual language diversity.

**Prof. Pat Hall** then described the Glossasoft project, in which software had a high-level knowledge model about the software, and used language engineering methods to generate error and help messages, with an API that enabled language specific components to be replaced as required. Discussion focused on more general issues of what the basic languages components are in order to enable language survival, with suggestions ranging from basic speech corpora and dictionaries, to translation aids, to enabling the computer platform for the writing system, distinguishing language documentation from language survival from localisation needs.

The various localisation issues were then brought together by **Dr. M. Sasikumar**, distinguishing between localisation which enabled content creation by encoding the script and providing fonts, and changing the interface and look and feel. He argued that the first was the critical facility to enable information sharing, with users learning to use foreign keywords. The key interface localisation should aim at enhancing usability locally: for example NCST is enabling the searching of web content in Hindi, where the Hindi search terms are translated, the web is searched, and the results gist translated back into Hindi. He highlighted a number of issues: different Linux localisation may be incompatible with each other, computer terminologies used in translation need to be intuitive and standardised, usage of localised software by the intended users needs to be studied, and the potential of multi-modal interfaces. The key lesson for localisation is that more collaboration and sharing will help progress.

## 7. Policy implications.

The final session was concerned with actions and policies to address the issues previously identified. We began by discussing the criteria that should be used in selecting languages to work on and support. Mark Turin pointing out the topicality of this in Nepal which had moved to an educational policy which is inclusive and supports education in all of the countries 100 or more languages. Others highlighted the importance of the languages used in broadcast television and radio, the role of national rivalries wanting their country to match the resourcing of national and local languages of other countries, that supporting a minority language could further marginalise its speakers, that unwritten languages were more endangered but among those that with the largest number of speakers should be supported, that some languages have precious written records even if no longer spoken, that China requires all software sold in China to have been localised resulting in far higher PC usage in China relative to India, that China's Mandarin-only policy has disadvantaged all other languages in China such as Tibetan and Cantonese.

We then contrasted the support for language technologies in South Asia and in Europe. **Dr. Om Vikas** has been leading TDIL in India, coordinating technology and language engineering developments. He noted the

increasing dominance of English as well as Spanish, French and German, measured by a range of criteria such as the proportion of Internet content or the volume of translation into those languages. He contrasted developed and developing countries in terms of: replacement of technology vs upgrades, interest in markets vs economic growth, IPR vs open source, affordable computers at \$500 vs \$40. India through TDIL has focused on adapting technology to meet core priorities, then developing its own technologies in collaborative programmes, and now focuses on creative technology within new sustainable structures. Areas of focus have been machine translation, OCR, text to speech generation and speech recognition, and many application areas like e-government and education. International collaborations are important. TDIL run an information and resource distribution service with a website and newsletter.

**Dr. Khalid Choukri** is the founding director of the European Language Resources Association (ELRA), which started in 1995. Europe is a multi-lingual multi-cultural region and faces many of the same problems as faced in South Asia. ELRA manages, evaluates, and distributes the full range of language resources of importance within the European Union. Before ELRA was founded in 1995 many language engineering projects had been funded, but their products had been lost. The languages covered include those of the enlarged European Union, plus many migrant languages now spoken within the union, as well as the languages of other regions like South America. The EU is now funding ambitious projects on speech-to-speech translation and multi-modal interfaces that will produce resources that will in due course be deposited with ELRA. There are still many gaps where no resources of a particular type are available for a language. Like TDIL, ELRA runs an information service with website and newsletter.

In discussion the idea of reusable generic application-neutral language resources was questioned, since theory-neutral language resources are not possible, questioning the whole idea of resource repositories like that managed for Europe by ELRA and being developed for India by TDIL, and suggested for South Asia. ELRA's resources are pre-competitive and not ready for commercial exploitation even though developed in part by industrial concerns using European Union funding. The irony is that the development of the 'knowledge society' has led to the loss of knowledge as languages are differentially advantaged and face extinction.

The final presentation to round off the conference was by **Prof Ken Keniston**, for has observed localisation developments in India for more than a decade. He had just spent a fruitful week at the WSIS meeting in Geneva. The tone of that meeting had been optimistic for the development benefits of ICTs with a focus upon local developments, but the absence of technology to support the local languages had not been recognised let alone seen as problematic. Investment in local languages is relatively small, even in India when the 5 million dollars invested so far is set beside the 7 billion dollars of revenue in 2003 from software services. Even though there have been significant developments in South Asia of software that handles South Asian languages, there is still little software that is standards compliant or robust enough for sustained use, and projects to remedy this make promises but have yet to deliver. There seem to be cultural pressures against working on or in Indian vernacular languages, with English viewed as more prestigious. And for those who do work on local language software there is always the danger that efforts fragment as developers take pride in their own solutions and reject the solutions of others. Where commercial collaboration is needed, IPR interests can cause difficulties.

The solution throughout is standards and collaborations, with resources devoted to support these.

## 8. Position Papers

available on the project website, <http://www.elda.fr/proj/scalla.html>

Allwood, Jens; *Linguistic Diversity and the Digital Divide*

Bhattacharyya, Pushpak; *Predicate Preserving Parsing* (with Jignashu Parikh, Yogadish Khot, and Shachi Dave)

Chalmers, Rhoderick and Amar Gurung; *Localising software: some experiences from Nepal*

Chaudhuri, B. B. *Input-Output Issues in Software Localisation of South Asian Language*. (abstract)

Choukri, Khalid *European Language Resources Association, History and Recent developments*

Dash, Niladri Sekhar; *Language Corpora: Present Indian Need*. (not able to attend)

Hall, Patrick A.V.; *Linguistic Components to save Endangered Languages*

Halihakonege, Don Joseph Vincent; Sinhala, *Language Engineering and an Attempt to take Internet to the Masses*

Hariharan, Venkatesh (Venky); *Linux Localization in India* (abstract)

Juel-Henrichsen, Peter; *Working in Darkness, corpus based tools for low-budget field linguistics.*

Hussain, Sarmad; *Complexity of Asian Writing Systems: A Case Study of Nafees Nasta'leeq for Urdu*

Keniston, Kenneth; *The Politics of the Digital Divide* (extended abstract)

Mallikarjun, B; *Indian Multilingualism, Language Policy and the Digital Divide*

McEnergy, Tony; *Constructing Corpora of South Asian Languages* (with Paul Baker, Andrew Hardie, and Sri B.D. Jayaram)

Mikami, Yoshiki; *The Language Observatory Project and its Experiment: Cyber Census Survey* (with Izumi Suzuki)

Minocha, Shailey; *Localisation Challenges in Usability and Customer Relationship Management of E Commerce Environments.*(with Pat Hall and Liisa Dawson)

Mudur, S. P.; *A Reference Model for Software Localisation* (with Rekha Sharma)

Murthy, Hema; *A Multimodal Indian Language Interface to the Computer.* (with C.S.Ramalingam, Srinivas Chakravarthy and C Chandra Sekhar).

Patnaik; B. N. *On the Choice of an Appropriate Framework for Computational Linguistic Research on Indian Languages.*

Pym, Anthony; *Localization from the Perspective of Translation Studies: Overlaps in the Digital Divide?*

Rahman, Tariq; *Language Policy and Localization in Pakistan: Proposal for a Paradigmatic Shift .*

Sangal, Rajeev and Dipti Sharma; *Creating Language Resources for NLP in Indian Languages* (Dipti Sharma unable to attend)

Sasikumar, M. and Jayprasad J Hegde; *Software Localisation: Some Issues and Challenges*

Schäler, Reinhard ; *A Framework for Localisation*

Singh, Udaya Narayana; *Language Technology: a Road Map for South Asian Languages.* (unable to attend)

Somers, Harold; *Language Technology and Patients with Limited English*

Tucker, Roger and Ksenia Shalnova; *The Local Language Speech Technology Initiative – localisation of TTS for voice access to information*

Turin, Mark; *Minority Language Politics in Nepal and the Himalayas*

Vikas, Om; *New Order of Knowledge-based Society: Rise, Raise & Race*

Weerasinghe, Ruvan; *A Statistical Machine Translation Approach to Sinhala-Tamil Language Translation*

Yogendra Yadava; *Use of Software Support in Nepal's Languages.*