

LangTech 2002 Conference

September 26-27, 2002, Berlin, Germany.

Visit by Professor BB Chaudhuri of ISI Kolkata and Durgesh Rao of NCST Mumbai, and Professor Pat Hall of the Open University.

Background

The LangTech 2002 conference on September 26 and 27 at Berlin was attended by 336 delegates from over 20 countries spanning 3 continents. LangTech has been newly established by ELRA (one of the SCALLA partners) and others, to be the main forum for European companies and institutes engaged in language technology to showcase their work, though participation is global. The unique aspect of this conference was the participation of industry, academia as well as government agencies, with emphasis on practical applications and business collaborations.

The conference followed a regular format. The morning session consisted of a keynote address on each day, followed by three parallel streams on Voice, Multilinguality and Knowledge on the first day, and a panel discussion on Industrial needs and outlook followed by a Venture Capital session on the second day. The afternoon session on both days included 'elevator pitch' presentations from Small and Medium Enterprises. There was also an exhibition area with about 20 exhibitors displaying various language technology applications, products and services. The second day concluded with a Future session.

Exposure to this conference was a very useful experience for both Professor Chaudhuri and Durgesh Rao, since many people and institutions have been trying to develop similar technology for Indian languages. The situation in India and more generally in South Asia is similar to that in Europe where quite a few languages are used in different regions. Perhaps the situation in India is somewhat more difficult because India has about a dozen scripts with about 350 basic and compound shapes in each script, giving a complexity that is much greater than largely Roman based European scripts. However, the linguistic complexity is of similar order in both situations and so is the phonetic complexity of speech. Given this background, the European experience has the potential of being immensely beneficial to the corresponding communities in South Asia.

Keynote addresses

The conference started with the opening address of Dr. Uwe Thomas, State secretary of the federal Ministry of Education and Research. He presented several figures about the thrust on language technology, research grants as well as the market potentials. Some of his experience is also applicable to the Indian and South Asian environment and appropriate motivation (though in a smaller scale because of economic constraints) needs to be taken by Government funding agencies. Professor Chaudhuri has already discussed these matters with TDIL program director in Indian Ministry of Information Technology and obtained some positive response.

The next talk was the keynote address of Bill Dolan, Head of the Natural Language Processing Unit of Microsoft Corporation. His talk was on language technology for consumer software. He gave a balanced perspective on the hype and the real possibilities of language technology, and pointed out the roles for 'deep' and 'shallow' NLP. He also highlighted the use of NL technology in Microsoft products such as the grammar checkers for several languages in Word, and the Smart Tags for English in Office XP. His main message was that current user interfaces are too clumsy, with users having to adapt to the computer, rather than the other way round, and since language is fundamental to human interaction, it must have a critical role to play in building natural user

interfaces. The talk appeared to be more biased to Microsoft's efforts to this end, although some interesting revelations were also there in his talk.

On the second day the keynote address was by Prof. Wolfgang Wahlster of DFKI. Though presented a bit fast, it was a brilliant mixture of methodology and application with some new and interesting solutions to the problem of mobile internet language technology. In discussion between ourselves after the talk we did note that the seminal example of finding your way around the streets of Heidelberg would not be relevant in South Asia, where you would just ask somebody.

Bente Maeggard of Euromap speaking at the Closing Session, acknowledged the institutions that made LangTech 2002 happen – Euromap, DFKI, ELRA and Elsnet, the funding agency Investitions Bank Berlin, and the organizing company, VDI/VDE-IT.

Technical Sessions

Most of the conference was taken up with three parallel sessions on Voice, Multilinguality and Knowledge Management. In general, the presentations were more industrial and application oriented than on theory and methodology (on a more narrow but focussed problem) than we notice in university and other academic organization conducted conferences.

The Knowledge Management session had interesting presentations. Geert Kobus of Knowledge Concepts, the Netherlands, spoke about "Selling Human Language Technology to the Knowledge Management Software Industry", and made the point that the Knowledge Management market for Human Language Technology is big, but the HLT people need to talk the language of the KM industry. Luca Dini of CELI, Italy, spoke about "Language based Web Discovery" with the help of two case studies – web driven customer acquisition and web based opinion survey. He also highlighted the need to have different technical and business models for big and small companies to deploy language technology. Donia Scott of ITRI, Brighton, UK, spoke about "KM in a Multilingual Environment" through an innovative combination of user interface, knowledge representation and natural language generation.

The Future Session had interesting presentations from XtraMind, a spin-off from DFKI, and Xerox Research Centre, Europe. They clearly brought out the trend towards pervasive computing and speech interfaces with language technologies playing a central role.

The industry and venture capital sessions were largely focused on how language technology companies are handling the current economic climate.

The exhibition featured several interesting exhibits. The interesting aspect of this for me was the presence of several associations devoted to promoting various aspects of HLT -- Euromap, ELRA, ElsNet and GALA.

Most of the presenters represented some industrial organization and they had brought their equipment to the exhibition. The delegates had the opportunity to test the efficacy of the technology directly on their systems. After visiting the stalls and trying out the systems Professor Chaudhuri developed the opinion that machine translation efforts have still a long way to go before being commercially usable by a large number of customers. However, both speech synthesis and recognition technology are more mature and shortly they can be used in various contexts including large scale biometrics. Moreover, multi-lingual and cross lingual information retrieval is quickly coming up and within the next year and so, multi-lingual and multi-media information retrieval engines will become commercially successful and popular.

There were some Asian components also in the conference. At least one organization informed us that they have developed advanced translation systems from Arabic to European languages. Also, Prof Tsujii of Tokyo University presented the activities on Machine Translation in Japan and Asia. His presentation was concentrated on efforts taken in Oriental Asia and focus on South Asia was somewhat lacking.

Post conference activities

During LangTech we had met to discuss the future plans for SCALLA. We decided to organize a special workshop on Computational Linguistics in South Asian languages, exploring synergies between Europe and South Asia, during EACL-03 in Budapest in April 2003. The idea for this was crystallized during LangTech, with encouragement from Prof Donia Scott. After the conference, Prof. BB Chaudhuri and Durgesh Rao stayed behind in Berlin 28 to work on the draft proposal, with further work being done between Durgesh Rao and Prof Hall in the UK.

After LangTech, Durgesh Rao took the opportunity to visit a few Universities in the UK. On October 1, he visited Hamish Cunningham and the GATE project at Sheffield University. GATE is an important project to create a General Architecture for Text Engineering. They spoke about integration of Indic languages into GATE, and in particular, code conversion of Indic language corpora collected under the Emille project, from font-based encodings to a standard character-based encoding, Unicode.

On October 2 and 3, Durgesh Rao visited Prof. Donia Scott and Adam Kilgarriff at the Information Technology Research Institute, Brighton University. October 2 was Open House day, where ITRI presented its work to other colleagues at the University. Rao attended this session, and met several of the staff at ITRI, and learnt about their many projects related to Language Technology. On October 3, Rao made a presentation about issues in Indic computing, and presented the new input method for phonetic scripts developed by him. They discussed how this might be of relevance in the UK, due to the large Indic language usage, and also how it might have wider applicability for European languages.

Conclusions

The main lessons of the conference were then summarized:

- The future lies in natural communication with concentration on the user.
- The technology is mature, in particular for voice and multilinguality, and it is now time to build applications.
- Language technology needs to get more visible in boardrooms.

For Rao and Chaudhuri LangTech was a real eye opener. It provided invaluable exposure to the state of the art in Language Technology in Europe. South Asia is similar to Europe in being linguistically rich and diverse. The language technology industry in South Asia is in its infancy, and there is a lot of synergy possible between Europe and South Asia in language technology research as well as industry. South Asia can learn from the European experience, and Europe can benefit from the rich linguistic environment as well as the large skilled manpower base in South Asia. This fits the objectives of the SCALLA project perfectly.

The next LangTech will be hosted by ELRA in Paris in 2003.