

should create and make accessible free of charge, a database on current experiences, standards and technical recommendations, software products and on-going innovations bearing on computerization of languages.

**Workshop on Lexical Resources for Natural Language Processing on 5-8 January 2001 at IIT, Hyderabad**

Lexical resources for Indian languages, for example, electronic dictionaries, thesauri, grammars, corpora etc. are far behind those for other languages of the world, even some smaller Asian languages. This in turn is affecting the development of applications. There is a need to make concerted and coordinated effort in building the lexical resources.

While developing a lexical resource, it is important to identify at least one application, which would benefit directly from the resource. Two important applications are: search engine for Indian languages, and machine translation. The following four LRs need to be built: (a) Transfer lexicon and grammar (between English & Indian languages), (b) Annotated corpus for each Indian language, (c) Bilingual dictionary of core meanings, (d) Word Net for Indian languages.

Details regarding how to build each of the above resources (except (d)) and a methodology was worked out. It would require a close monitoring of TIME and QUALITY. The distribution license for the LRs would be “free” software under GPL.

**Symposium on Translation Support System (STRANS) on February 15-17 2001 at IIT, Kanpur**

The symposium STRANS2001 was first of its kind after CPAL-2 and SMATAC-96 to provide a platform for presenting and discussing all the issues related to machine aided translation focussed to Indian languages including English. Prof. R.M.K. Sinha, Program Chair for the STRANS Symposium, presented a brief review of the on-going activities in the area. The first key-note address delivered at the inaugural session was that from Prof. Harold Somers of UMIST, UK. Prof. Somers group has been working on South Asian languages, which are minority languages in United Kingdom. Prof. Somers presented his thoughts on “A corpus based approach to resource development for low density languages”. Dr. Om Vikas, Ministry of Information Technology, Govt. of India, delivered another keynote address on issues related to Indian Language processing vis-a-vis the international scenario.

A special session was devoted to the presentation on TDIL VISION2010 document and Standardization issues. There was a general agreement on the Vision 2010 document. It was agreed that standardization of lexical format, various romanization codings, transliteration, OCR data and speech data as contained in the document be created in a phased manner and be forwarded to BIS for making them a national standard. Some people expressed apprehension about their execution in the stipulated time frame and wanted Govt. to take initiatives to coordinate the efforts.

Some of the major recommendations of the panel discussions are summarized below: (1) Software and technology be developed to cater to all Indian Languages and scripts. However, priority be given to developing interfaces between English and Hindi, English and regional languages, and among Indian languages in that order. Funding agencies at the Central and State Govt. levels must exchange information to avoid duplication of efforts or re-inventing of wheel. (2) Lexware is a basic infrastructure for all works on language processing. Lexware work is a highly time consuming, laborious and requires a skill for which training is required. It cannot be created at a single centre or by a single group. Therefore all researchers and developers for easy sharing and portability. (3) Follow a common standard. The TDIL web site maintained by MIT should be the gateway for disseminating and exchanging ideas. All major developments and policy matters should be posted there. Industries and users including state gov. agencies should also be encouraged to post their problems, experiences and exchange solutions through this site. (4) STRANS be made a regular annual affair to provide a platform for researchers, developers, industries and users to report their work and exchange their ideas.

**Universal Networking Language on April 10, 2001 at IIT, Bombay**

Universal Networking Language is a recent Interlingua proposed by a group of researchers from 18 countries of the world. The United Nations University, Tokyo, leads the UNL project. The overall goal of the project is to enable natural languages on the Internet so that the access of information on the Internet can be done by a large section of humanity crossing the language barrier.

In India, the Indian Institute of Technology, Bombay, is handling the UNL project. The responsibility of IIT Bombay is to link Hindi with the rest of the languages over the Internet. Towards that goal a team of researchers led by Prof. Pushpak Bhattacharyya of the Computer Science and Engineering Department, IIT Bombay is working on conversions of English and Hindi to UNL and also the reverse problem of language generation from UNL.