A one-day workshop on Morphology was organized by the Center for Indian Language Technology (CFILT), Computer Science and Engineering Department, IIT Mumbai on 31 March 2005. The event was sponsored by the Research and Training Center, Development Gateway Foundation at IIT Bombay. The aim of the workshop was to share experiences of researchers in morphology, and to evolve solutions and future directions for Indian language processing with emphasis on morphology. The importance of morphology is supreme in the context of Machine Translation, Information Retrieval, Information Extraction and many such applications, especially for Indian Languages.

The major topics covered in the workshop were (i) lexicology and morphology, (ii) morphology and syntax, (iii) semantics and morphology and (iv) morphology and its applications to language technology. Researchers from IIT Bombay, IIT Kharagpur, IIT Delhi, IIIT Hyderabad, CIIL Mysore, University of Hyderabad, Jadavpur University Kolkata, Banasthali Vidyapith Rajasthan, Pondicherry University, Goa University and Indian Statistical Institute, Kolkata presented their work in the workshop.

After a very melodious Bandana of Saraswati, the Goddess of Learning, the workshop was inaugurated by Prof. Kartic Khilar, Dean, Research and Development, IIT Mumbai in presence of Prof. G. Sivakumar, Head, CSE Department, IIT Bombay, Prof. Krithi Ramamritham, Head, Kanwal Rekhi School of IT, IIT Bombay and Prof. Pushpak Bhattacharyya, Head, CFILT, IIT Bombay. Prof. Sivakumar opened the workshop by underlining the importance of language processing research, and praising the good work being done at CFILT. Prof. Ramamritham welcomed all the participants; “The Language Technology research group here is vibrant, as departments like CSE, H & SS and KReSIT are involved in the research,” he said. He invited the participants to have a look at aAQUA- the question-answering forum in the domain of agriculture- where the research done at CFILT is being used.

Prof. Kartic C. Khilar said, “For me morphology is the chemistry of words. We Indians are proud of diversities like bio-diversity and language-diversity. Diversity implies scope for research and development. Our aim is to advance and employ the knowledge for industry and people, build environment for quality
education, promote research and build an R & D culture. We have a number of projects from industry. We interact with the government to enhance Science & Technology.” He further said, “Being involved in research activities as in CFILT gives us an opportunity for making a difference in people’s lives. We would like to see more of such workshops for interactions and sharing of knowledge between researchers.”

Prof. Pushpak Bhattacharyya echoed the sentiments of Dr. Om Vikas, Senior. Director and Head, Technology Development for Indian Languages (TDIL), Department of Information Technology, Government of India, who is the motive force behind Indian language technology research and who could not attend the workshop due to some unavoidable reasons. Prof. Bhattacharyya expressed the view, “Morphological Analyzer is a basic tool for NLP that is not yet available for Indian languages. We are developing important lexical resources like Wordnets for Hindi and Marathi. Another set of such tools are the Part-of-Speech taggers for Hindi and Marathi. The major difficulty in developing such tools is the unavailability of tagged data in Indian languages. But, since most of the Indian languages are morphologically rich, rules of morphology can help us develop such tools.”. “Our lifeline is our students and research staff. I am proud of a very young and dynamic group of researchers in NLP.” He thanked the Government of India for its generous support in funding the language technology research.

Prof. Vaijayanthi Sarma, Humanities and Social Sciences (H & SS) Dept., IIT Bombay, gave the keynote speech titled ‘Morph-logy’. She explained the key concepts in morphology like nature of morphemes, classification of morphemes, processes of word formation like inflections, derivations and compounding, vocabulary and creativity, mental lexicon and morphological typology. There was a discussion on questions like “How to coin new words in a local language?”, “Why some words don’t have plural forms?” and “Are bilinguals real bilinguals?”
There was an invited talk by Prof. Malhar Kulkarni, H & SS Dept. IIT Bombay on “Introduction to Sanskrit Morphology”. He discussed the fundamental concepts of padapaatha, padas, praatipadika, dhaatu, pratayaya, aagama, anga, ting and sup pratyaya, inflectional and derivational morphology, case endings, kaarakaartha, compounds and division of sentences.

Following papers were presented in the workshop:

1. The Affects of Inflectional and Derivational Morphology in Bengali Phonology and Its Implications in Speech Technology: Monojit Choudhury, Sudeshna Sarkar and Anupam Basu
2. Aspects of Concatenative And Non-concatenative Morphology Of Standard Hindi: Mrs. Rita Mathur
3. Ontology For Word-Formation Generation: Mr. Bira Chandra Singh
4. Morphological Processing of Kannada Verbs: B.Mallikarjun
5. Root Word Stemming and an Improved Porter's Algorithm: Sujit Kumar De, Sivaji Bandyopadhyay
6. Computational Account of Inflectional Morphology of Marathi Verbs: Veena Dixit, Sachin Burange, Prof. Pushpak Bhattacharyya
7. Morphological Analyzer for Hindi Text: Dr.(Mrs.) Madhavi Sinha & Prof.(Mrs.)Rekha Govil
8. Issues in Morphological Analysis of North - East Indian Languages: Vijayanand Kommaluri, R Subramanian, Anand Sagar K
9. Morphology Based Natural Language Processing Tools For Indian Languages: Manish Shrivastav, Nitin Agrawal, Bibhuti Mahapatra, Smriti Singh, Pushpak Bhattacharyya
10. Study of Hindi Noun Phrase Morphology for Developing a Link Grammar Based Parser: Shailly Goyal and Niladri Chatterjee
11. Towards Developing a Link Grammar Based Parser for Hindi: Shailly Goyal and Niladri Chatterjee
13. Exploring the Semantic Content of 'TO': Dr. Radhika Mamidi
A brief technical overview of the workshop follows:

1. For fixed word order languages like English, the position of a word within the sentence governs its semantics, while in the case of relatively free word ordered languages morphology plays an important role. Most of the Languages in the Northern part of India either originated from or are influenced by Sanskrit. These languages mostly have relatively free word order. Exhaustive morphological analysis is the strong platform for processing of such languages.

2. A presentation on basics of morphology and another on Sanskrit grammar set the tone for the workshop. The principles of Sanskrit Grammar can guide the research on NLP in the context of Indian Languages. The formalism of Sanskrit gives rise to a rule governed approach to the construction of NLP systems.

3. Hindi being the national language of India, a number of research centers are conducting research on Hindi language processing. Researchers presented their work on tools like morphological analyzer, stemmer, pos-tagger, parser and machine translation to and from Hindi during the workshop. At IIT Delhi, a parser for Hindi based on link grammar is under development. In the light of presentation on concatenative and non-concatenative morphology of Hindi, the criteria for identifying the morphemes and the extent to which they can be analyzed were discussed. Questions were raised to see whether suffixes like जा, की, क (chunk suffix) or क, आ, ई, ए (phoneme level suffixes) should be stored. Both of these approaches are being tried at IIT Mumbai. The Marathi pos-tagger utilizes the list of chunk suffixes while Hindi pos-tagger stores phoneme level suffixes. The Marathi pos-tagger attempts to handle word level analysis for all possible categories, while the Hindi pos-tagger follows two different tactics, such as word level analysis for nominal and group level analysis for verbal phrases. Both Marathi group and Hindi group construct noun paradigms based on phonological constraints. The Marathi pos-tagger uses the tabular representation of data and generates rules automatically, while the Hindi pos-tagger uses manually constructed rule base. It is evident that CFILT pursues multiple approaches for solving a problem.

4. At Hyderabad University, effort is on to establish the morphological rules considering the ontological properties of the words broadly and roots specifically. The aim of this effort is to generate the word forms. On the other hand, an improved Modified Porter’s Algorithm for Root Word Stemming for English was proposed by Jadavpur University, which consists of a modified rule base and an efficient table look up. This increases the accuracy of the algorithm from 95% to 97%.

5. IIT Kharagpur has investigated the problem of text to speech synthesis for Bengali at the interface of phonology and morphology. The group demonstrated the necessity to use Schwa Deletion Algorithm in case of grapheme to phoneme conversion.

6. A presentation on the languages of North East India from the Pondicherry university, essentially outlined the common morphological properties of the languages of this region and called for urgent action to prevent some of these languages becoming extinct because of the lack of script and the dwindling number of speakers.

At the end of the workshop, participants expressed their opinions and views on the workshop. The event-the participants said- was enriching to both linguists and computer scientists. The need to bridge the gap between linguists and computer scientists was emphasised. It was suggested that such workshops should be conducted frequently and for longer durations. An intensive interaction between the theoreticians and the practitioners would provide a comprehensive perspective of computational linguistics to both linguists and computer scientists.

The workshop ended with a detailed discussion on important issues in the areas of sharing of resources among various research centers, frequent interactions among researchers, working towards bringing India to the forefront of research in NLP and bridging the gap between linguists and computer scientists.
Prof. Pushpak Bhattacharyya closed the workshop with the promise, “Such activities will be continued. Resources and tools developed in the center will be shared. We will also hold a workshop on POS tagging in future. Such focused workshops are the key to make progress.”