Workshop on modern Insights from Indian Grammatical Tradition

Thursday, 05/11/2009, Hörsaal 3B (23.01)

13:45 Introduction by **Wiebke Petersen**

14:15 Amba Kulkarni (University of Hyderabad):

Anusaaraka – A Language Accessor cum Machine Translator

The conventional machine translation systems are fragile and do not provide a fallback mechanism. In this talk we look at the machine translation problem a fresh and observe that there is a need to a) share the load between man and machine, b) distinguish reliable knowledge from the heuristics, and c) to make use of existing resources instead of reinventing the wheel.

This calls for total restructuring of the machine translation system. We present an architecture for anusaaraka - an incremental robust machine translation system, which is influenced by the above observations.

The anusaaraka architecture differs from the conventional in three major ways a) the order of operations is reversed, b) a concept of interfaces to act as glues to improve the modularity of the system is introduced to help plugging in different components, and c) a GUI is developed to provide the right amount of information at right time.

This architecture of anusaaraka together with a user-friendly interface is convenient not only for users but also for a developer. It makes a clear-cut distinction between the resources that are in principle reliable and those that are in principle probabilistic.

Some of the aspects of Indian Grammatical Tradition that influenced this architecture will be briefly touched upon.

16:00 Amba Kulkarni (University of Hyderabad):

Pāṇini's Aṣṭādhyāyī: A Computational Linguist's perspective

Pāṇini's Aṣṭādhyāyī, though is around 25 centuries old, provides very good insights for Natural Language Processing. The importance Pāṇini has given to the information coding in a string of a language, is evident from the way he formulated the sūtras as well as the way he analysed Sanskrit language. We substantiate this claim with examples. We disccuss three sūtras from his work, which provide useful insight for understanding the information dynamics in a language. It answers some of the important questions such as: Where does a language code the information, how much information does it code and the manner in which it codes the information. Answers to these questions provide a clue to design the parsing strategy for a language, know precisely what is acheivable in NLP and what is not.

Amba Kulkarni – head of the department of Sanskrit Studies at the university of Hyderabad – is one of the best known computational linguists from India. Her main theoretical focus is on studying Pāṇini's Aṣṭādhyāyī from a computational and an NLP perspective and on showing the relevance of Indian traditional shastras in the modern context. She has developed various Sanskrit computational tools (e. g. machine translation systems, tagged Sanskrit corpora, ...).

Der Workshop wird veranstaltet von der Forschergruppe FOR 600