

# **Gujarati NLP workshop Report(Sardar Patel University,Gujarat),17th to 20th January, 2011**

The Workshop on Introduction to NLP for Gujarati started on 17th January, 2010. The workshop venue was the Post Graduate Department of Gujarati, Sardar Patel University, Vallabh Vidyanagar, Gujarat. There were 113 participants who attended the workshop, besides the staff members of the department.

The workshop started off with an inauguration function. Prof. Bhagirath Bhrambhatt, Head, Post Graduate Department of Gujarati, gave a warm welcome to the organizers and presenters from LDC-IL, CIIL, Mysore and felicitated them with flower bouquets. He also extended a warm welcome to the participants of the workshop. There was a short tea break after the inauguration function, after which Mona Parakh, gave an introduction to LDC-IL as well as gave a detailed presentation on the activities and goals of the LDC-IL. This presentation was followed by a presentation by Shahid Bhat, on a basic but detailed Introduction to Linguistics and NLP.

On the second day workshop a presentation on Corpus Linguistics and its relevance for the field of NLP was made by Mona Parakh. She was accompanied in this presentation by Mahesh Solanki who explained the method employed at LDC-IL for collecting Text Corpus. After tea session was devoted to giving demos of the tools that have been developed at LDC-IL. Mona Parakh gave demos of the Corpus related tools such as Transliteration tool, KWIC-KWOC, Frequency counter and the tool for inputting/editing the text corpus.

On the third day Shahid Bhat presented on Part of Speech Annotation and provided the general framework of the LDC-IL tag set. A more language specific discussion of the Gujarati Tag set was conducted by Purva Dholakia. During the explanation of the Gujarati tag set Purva Dholakia demonstrated the POS tagging tool and how to tag the corpus. After the tea break, there was a competition of 'Best Annotator' held by the LDC-IL Team. All participants took part in that competition, with enthusiasm.

The fourth day of the workshop started with the explanation of Speech Data collection for ASR as carried out at LDC-IL, which was presented by Hiren Gadhvi. Mona Parakh explained Speech Processing in general and the differences in data collection methods for ASR and TTS. This was followed by a demo of the speech segmentation tool (Wavesurfer) by Hiren Gadhvi, wherein he showed how recorded speech data can be segmented using the tool. After the tea break the workshop was concluded with a valedictory function, wherein prizes were given to winners of the 'Best Annotator' competition. Staff and students of the department gave their feedback on the workshop and two members of the staff gave a Vote of Thanks. The workshop ended with the vote of thanks by Purva Dholakia on behalf of LDC-IL.