

Assamese and Bodo NLP workshop Report(Dibrugarh University, Assam),

2nd to 5th of February, 2011

The workshop on introduction to NLP for Assamese and Bodo was organized in collaboration with the Department of Assamese, Dibrugarh University, Assam. The workshop was organized from the 2nd February to the 5th of February, 2011, at the conference hall of the Department of Assamese, DU. Each day had a four hour lecture from 10 a.m. to 2 p.m. with a tea break of 20 minutes in between. There were totally 80 participants excluding the department staff and helpers.

The first day started off with an inaugural function with an inaugural speech by the Head of the Department Prof. Karabi Deka Hazarika, followed by speeches by lecturers from the department Dr. Bhaskarjyoti Sarmah and Dr. Arpana Konwar. The chief guest, senior professor, Bhim Kanta Baruah in his speech stressed on the importance of NLP in today's world and offered his thanks to LDC-IL for arranging such workshops in the far north-east of India for the second time. This was followed by a speech by Atreyee Sharma on the essence of NLP workshops and LDC-IL. The academic session started off after tea. Atreyee Sharma spoke on LDC-IL and Arundhati Sengupta took over the next lecture on introduction to NLP.

The second day was devoted to an extensive lecture and discussion on Corpus linguistics and its relevance for the field of NLP by Atreyee Sharma. Jahnobi Kalita, Bridul Basumatary and Chuchen Dutta accompanied in this presentation by explaining real life corpus examples which are issues for people working on NLP. This triggered valuable inputs and suggestions from the participants. After tea session demos of the tools such as Transliteration tool, KWIC-KWOC, Frequency counter and the tool for inputting/editing the text corpus that have been developed at LDC-IL was given by Atreyee Sharma. This was followed by a lecture by Farson Daimari on Morphological analyser. The participants were given home assignments on corpus related issues.

The third day Samar Sinha talked on POS tagging, its importance and relevance on NLP work in general and LDC-IL in particular. He described the POS tagset of LDC-IL. This was followed by a competition on POS tagging for the participants. After this home assignments assigned to the participants were collected.

The fourth and the final day was fully devoted to speech and speech related work at LDC-IL. Arundhati Sengupta handled the session from 10 a.m. to 1p.m. Within this time frame she managed to explain basic phonetics(w.r.t. Assamese)to the participants and demonstrated speech data collection methodology under taken at LDC-IL. The valedictory session after tea was graced by the presence of the Vice Chancellor of Dibrugarh University Prof. Kandarpa Deka Hazarika, who was over whelmed by the response of the workshop and satisfaction of the participants. The VC distributed the certificates to all the participants and gave away the "best annotator" award to the three best annotators of the previous day.

The workshop ended with a feedback session and unanimous appeal by all the participants to have workshops of the like in the N-E of India by LDC-IL again in the future. Samar Sinha presented the vote of thanks on behalf of LDC-IL. The workshop was covered by the media in the local news papers.