Manipuri NLP Workshop Report
(Directorate of Language Planning and Implementation (DLPI), Imphal)
4th – 9th November, 2013

The Training Program on Natural Language Processing was organized by LDC-IL in collaboration with the Directorate of Language Planning and Implementation, Government of Manipur (DLPI), Imphal. The program was well thought-out from the 4th to 09th November, 2013, at the Lecture Hall of the DLPI, Government of Manipur, Imphal. There were 90 participants, excluding the Department staff and helpers, out of which 40 were trainees selected for the program through screening, 08, the Resource Persons engaged for presenting the talks and 50, novices in the field of the entire orientation program. There were tea break from 11:15-11:30 a.m. and lunch break from 1:30-2:30 p.m. in between the lectures each day.

The first day started off with an inaugural function. Kh. Raghumani Singh, IAS, Manipur Director, DLPI, Dr. L. Mahabir Singh, Joint-Director, DLPI, Imphal, kindly consented to grace the function as the Chief Guest and President respectively and Mr. A. Nandaraj Meetei, Co-ordinator, Lecturer/RP, LDC-IL, and Ms. Atreyee Sharma, Senior Lecturer/JRP, LDC-IL, took their respective chairs. Mr. Kh. Raghumani Singh, IAS, kindled the inaugural light and delivered his speech on the future perspectives of language studies in various domains, the NLP developments of Manipuri in particular. Dr. L. Mahabir Singh spoke about the status of Manipuri language and existing interests in incorporating technical aspects in the diachronic as well as synchronic stages of the language concerned. Mr. A. Nandaraj Meetei talked about the objectives of LDC-IL with respect to the repository of linguistic resources in all Indian languages, facilitation of database creation and training through workshops, seminars etc. as well as the goals of the workshop stating that the targets of training program are to disseminate the knowledge of NLP amongst student community, to equip students to work or pursue research in language technology and to promote technology development in Indian Languages. He also instructed the selected candidates as well as all the participants to attend all the days of the program and asked them to have interactive sessions. Dr. Th. Harimohan Singh, Language Officer, DLPI, Imphal, gave vote of thanks for the inaugural session by opening the fresh eyelids of every participant in the hall with his quick-to-remind delivery.

For the entire training program, after the Inaugural Function, Ms. Atreyee Sharma first gave a talk on *LDC-IL Activities & Workshop Goals*. She well illustrated the running activities in both academic and technical perspectives of LDC-IL in succession. Mr. A. Nandaraj Meetei gave a lecture on *Artificial Intelligence (AI)* and elaborated how machines are made intelligent artificially in processing natural languages through the modularizations of human linguistic assignment domains. He basically talked about how AI should be fitted in the Computer Science taxonomy. Dr. N. Pramodini Devi, Associate Professor, Linguistic Department, MU gave a special lecture on *Empowering Language through Technology*. Within the field of technical system, the language Manipuri can be upgraded upto the interface where lexical expansion as well as inclusion of new linguistic items can be inserted into the lexicon of the language concerned, she explained.

On the second day Mr. Kishorjit N., Asst. Prof. Dept. of Computer Sc. and Electronics, MIT, M.U. gave a special lecture on *Concept of NLP and Its Applications.* He elaborated the technical aspects involved in the area of NLP and some of the tools developed particularly in Manipuri were also exemplified thereafter. Ms. Atreyee Sharma, gave a lecture *on Corpus Linguistics: Corpus (Text, Speech & Sign), Concept, Balancing & Classification.* She explained the fact that the first NLP step is to develop a corpus and find the aftermath of any NLP module through systematic procedures where technical algorithms as well as linguistic ingredients are clubbed together. Dr. L. Anand Singh, Senior Research Assistance, LDC-IL, at the end of the first day, gave a presentation on *Bangla-to Meetei Mayek Auto Transliteration*. He successfully demonstrated how texts in Bangla Scripts could be transliterated into Meetei Mayek by using the tools developed in LDC-IL.

On the third day Mr. A. Nandaraj Meetei, gave a presentation *Part-of-Speech Tagging: Tagset and Annotation (Developing POS-Tagger & Demo).* He basically emphasized on the nature of agglutinative languages like Manipuri where the questions of splitting, pre-tokenization and emerging grammaticalization trends often arise. Mr. A. Lorendro, Technical Officer, DOEACC, Akampat, Imphal presented a special lecture on *Data storage and retrieval of Manipuri Language in UNICODE Standard*. The nature of the presentation was so innovative that participants paid attention to a great extent to the up-to-date information of UNICODE values and their slot occupation contained in it. Dr. H. Mamata Devi, Assoc. Professor, MU, gave a special lecture on *Manipuri Optical Character Recognition System (OCRS).* She particularized

Manipuri language both in printed and manual documents involving reading text from paper and translating the images into a form that the computer can manipulate.

On the fourth day Ms. Y. Premila Chanu, Junior Resource Person, LDC-IL, gave a lecture on *Speech Segmentation & Annotation (ASR: Concept, Applications, & Challenges)*. She systematically described how speech data got collected through field work and the data so collected could be stored in the server for the further NLP approaches. She further explained how speech data should be segmented and annotated for Automatic Speech Recognition (ASR), a technology allowing a computer to identify the words that a person speaks into a microphone or telephone and convert it to written text. There were Hands On Experience and Practice Sessions after the presentation. Ms. Atreyee Sharma also gave a lecture on *Sign Language and Annotation.* The presentation was so remarkable with the participants who had benefitted from the novel field and could open a new-fangled layer of linguistic means of insight over the signers' daily communicative perspectives.

On the fifth day, the last day of the program, Mr. A. Nandaraj Meetei, LDC-IL, CIIL, Mysore, gave a presentation on *Morphology and Morphological Analyzer*. The core idea of the paper was to demonstrate the agglutinative nature of Manipuri language and identification of the morpho-syntactic features. He explored the definition of atoms of the lexical categories as input and the analysis of these elements as output through the tool of morphological analyzer. After the presentation, there was a short program on a formal Valedictory Talk. The valedictory function was   chaired by Dr. L. Mahabir Singh, Joint Director, DLPI, Govt. of Manipur, Imphal, also the Local Co-ordinator of the program. He delivered speech encouraging the spirit of the organizing program committee. Mr. A. Nandaraj Meetei, LDC-IL, Co-ordinator of the program, extended his heart-felt congratulation for the complete training programme for the entire five days. For the comments and suggestions for more improving the programs to follow were from amongst the trainees and presenters. Dr. L. Mahabir Singh, the Local Co-ordinator of the programme, also gave vote of thanks for closing session by reminding everyone there of the fully academic vibrant past days in the conference hall. The chair person of the closing day declared the *Five Day Training Program on Natural Language Processing* closed.