

**Minutes of the Eighth Project Advisory Committee Meeting
of the Linguistic Data Consortium for Indian Languages (LDC-IL)
held on January 8 2023 at CIIL, Mysore (virtually and physically)**

The meeting was attended by the following members:

1	Prof. Shailendra Mohan	Director, CIIL & Chairperson
2	Prof. Girish Nath Jha	Chairperson, CSTT
3	Prof. Aadil Amin Kak	University of Kashmir
4	Prof. Niladri Shekhar Das	ISI Kolkata
5	Smt. Suman Dixit	Deputy Secretary, Language Division, MoE
6	Dr. Narayan Kumar Choudhary	Officer i/c, LDC-IL & Coordinator
7	Prof. TG Sitharam	Chairperson, AICTE
8	Dr. Buddha Chandrashekhar	AICTE
9	Dr. Kamal Kumar Choudhary	IIT Ropar, Punjab
10	Ms. Lenali Singh	C-DAC, Pune
11	Ms. Nimisha Sriavastava	Microsoft
12	Ms. Amrita Kamat	Google
13	Ms. Himani Detwani	Google
14	Mr. Ranjeet Kumar Singh	Keypoint Technologies, Hyderabad
15	Mr. Vivekananda Pani	Reverie Technologies, Bangalore
16	Mr. Pranjal Nayak	Reverie Technologies, Bangalore

Welcome

Prof. Shailendra Mohan, Director, CIIL & Chairperson welcomed the members to the 8th PAC meeting of Linguistic Data Consortium for Indian Languages (LDC-IL).

The members were briefed about the agenda items to be discussed on the day and the tasks at hand which needed a decision and consensus of the members.

Agenda Items and Recommendations

1. Release of 16 new datasets

LDC-IL Released the 16 new datasets covering 15 languages. This included the much awaited sentence aligned speech corpora in 14 Indian languages, including Indian English variants of Kannada and Bengali. Two new datasets of Chhattisgarhi were also released.

2. Extending the scope of LDC-IL in light of the NEP-2020 (inclusion of non-scheduled languages and mother tongues)

The committee agreed to extend the work of LDC-IL to additional mother 270 tongues of India in a staggered manner in alignment with the goals of the CIIL.

3. Development of new types of datasets (TTS and parallel corpora)

The Committee recommended to start the work developing TTS related resources in Indian languages and issue guidelines on the linguistic features of each of these languages. The committee also recommended to create parallel corpora in Indian languages in a curated manner such that it gets extended to minor languages and mother tongues of India.

It was also suggested that we need to develop a pan-Indian (elaborate guideline) for speech corpus generation as the standard norms and rules for the generation of written text corpora greatly differ with regard to the development of speech corpora.

One of the members also suggested that it is necessary to develop corpora of Indian (varieties of) English in both text and speech forms. This will be required for developing ELT materials for Indian learners.

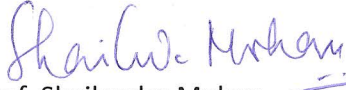
4. Updates in the license agreements based on the feedback received from the clients

The committee agreed to put a ceiling on the damage that can be claimed against the commercial users of the datasets as specified in the indemnity clause. This ceiling would now be capped at a maximum of USD 1,00,000/- per dataset.

5. Ratification of the Old PAC Recommendations

The committee sought action on the recommendations of the last PAC meeting and suggested that it should be acted upon ASAP. The recommendations of the 7th PAC are noted down once again below:

- i. Increase the size of the existing corpora to reach a uniform minimum critical data size in all the major languages, both text as well as speech data.
- ii. Involve 20 more non-scheduled languages.
- iii. Use automation to develop extensive corpora accessing various other resources.
- iv. A PG Diploma in Computational Linguistics can be started to train manpower to work in the area of language technology.
- v. To put pace on the daily workings of the scheme and to align with greater flexibility with the industry, this Scheme may attract funds from various governmental agencies such as Ministry of Education, TDIL, State Governments etc. as well as industry partners.
- vi. Filling of vacant positions, revision of salary and remunerations as per the current norms


Prof. Shailendra Mohan
Chairperson