Linguistic Data Consortium for
Indian Languages (LDC-IL)

CENTRAL INSTITUTE OF INDIAN LANGUAGES
DEPARTMENT OF HIGHER EDUCATION
Ministry of Human Resource Development, Government of India
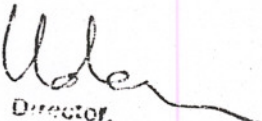**Manasagangotri, Hunsur Road, Mysore - 570 006**

June18, 2007

MINUTES OF THE FIRST PROJECT ADVISORY COMMITTEE MEETING OF LINGUISTIC DATA CONSORTIUM FOR INDIAN LANGUAGES (LDC-IL) HELD ON JUNE 5, 2007 FROM 12.00 NOON TO 4.00 P.M.

MEMBERS PRESENT

1. Prof. Udaya Narayana Singh
    Director
    Central Institute of Indian Languages
    Manasagangotri, Hunsur Road,
    Mysore - 570 006
    Chairperson

2. Director
    Indian Institute of Technology
    Bombay, P.O. IIT, Powai,
    MUMBAI - 400 076
    Member
    Represented by
    Prof. M.S.Malshe
    Dept. of H.S.

3. Director
    Indian Institute of Technology
    Madras, I.I.T Post Office
    Chennai - 600 036
    Member
    Represented by
    Prof. Hema Murthy
    Dept. of Computer Science

4. Prof. Rajeev Sangal
    Director,
    International Institute of Information
    Technology Hyderabad
    Gachibowli, Hyderabad - 500 019
    Member

5. Joint Secretary
    Human Centered Computing Division, Member
    TDIL, Deptt. Of Information Technology,
    Ministry of Communication & Information
    Technology, Room No. 3009,
    Electronics Niketan,
    6, CGO Complex, New Delhi - 3
    Member
    Represented by
    Shri Som Dutt Dadheech
    Scientist 'F' & HOD, TDIL

SPECIAL INSTITUTIONAL INVITEES

6. Director, C-DAC, Pune
    Special Invitee
    Represented by
    Dr. Hemant Darbari
    Programme Coordinator, AAI Group

Director,
निदेशक
Central Institute of Indian Languages
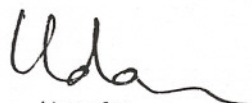भारतीय भाषा संस्थान
Mysore मैसूर.

1/6

7.   Prof. Vijayalakshmi Basavaraj   Special Invitee
     Director
     All India Institute of Speech & Hearing
     Manasagangotri, Mysore – 570 006

SPECIAL INDIVIDUAL INVITEES

8.   Prof. G. Umamaheshwar Rao
     Centre for Applied Linguistics & Translation Studies
     (CALTS), University of Hyderabad   Special Invitee
     P.O. Central University Campus
     Hyderabad - 500 046, A.P.

9.   Prof. C.N. Krishnan
     Director
     AU-KBC Research Centre for Internet & Telecom   Special Invitee
     Technology, Anna University MIT Campus
     Chromepet, Chennai - 600 044

10. Prof. Peri Bhaskararao   Special Invitee
     ILCAA, Tokyo University of Foreign Studies
     Asahi-cho 3-11-1, Fuchu-shi, TOKYO 183-8534, Japan

11. Ms. Rekha Sharma
     Centre for Speech Sciences
     Central Institute of Indian Languages   Special Invitee
     Manasagangotri, Hunsur Road,
     Mysore - 570 006

SPECIAL INDUSTRY INVITEES

12. MICROSOFT   Special Invitee
     DLF Cybergreens, 9th Floor   **Represented by**
     Tower A, Sector 25 A   Shri Pradeep &
     GURGAON - 122 002, HARYANA   Shri Meghashyam Karanam

13. MOTOROLA   Special Invitee
     Lake View Building   **Represented by**
     Bagmane Tech Park   Shri Ashwin Amarapur Chandramouly
     C.V. Raman Nagar,
     BANGALORE - 560 093

14. MAIT
     (Manufacturers' Association for   Special Invitee
     Information Technology)   **Represented by**
     PHD House, 4th Floor,   Shri Amit Chadda
     Opposite Asian Games Village,
     NEW DELHI - 110 016.

15. IBM   Special Invitee
     DLF Silokhera, NH-8   **Represented by**
     Sector - 30, GURGAON - 122 002,   Shri Abhijit Dutta
     Haryana

Director,
निदेशक
Central Institute of Indian Languages
भारतीय भाषा संस्थान
Mysore. मैसूर.

16. Dr. B. Mallikarjun   Member-Convener
    Reader cum Research Officer &
    Head, LDC-IL
    Central Institute of Indian Languages
    Manasagangotri,
    Mysore - 570 006

MEMBERS ABSENT

1. Director
   Indian Institute of Science Bangalore
   Bangalore - 560 012

2. Two Officials of the Language Bureau & Finance
   Division, Ministry of Human Resource Development
   Department of Higher Education
   Desk IV (L), Shastri Bhavan, 'C' Wing,
   New Delhi - 110 001

3. Vice Chancellor & Professor of Law
   National Law School University
   P.O. Box No. 7201, Nagarbhavi,
   Bangalore - 560 072

SPECIAL INSTITUTIONAL INVITEES

4. Director
   Indian Institute of Technology, Khargpur
   Kharagpur – 721 302

5. Additional Secretary
   Dept. of E-Governance of DIT
   Ministry of Communications & Information Technology
   Electronics Niketan, 6, C.G.O. Complex
   New Delhi – 110 003

SPECIAL INDIVIDUAL INVITEES

6. Prof. B.B. Chaudhuri
   Head, CVPR Unit
   Project Coordinator,
   MIT Resource Centre for Bangla,
   Indian Statistical Institute, Kolkata,
   203 Barrackpore Trunk Road,
   Kolkata 700 035, West Bengal

7. Prof. B. Yagnanarayana
   International Institute of Information Technology
   Hyderabad
   Gachibowli, Hyderabad - 500 019

SPECIAL INDUSTRY INVITEES

8. Hewlett-Packard Labs. India
   24, Salarpuria Arena
   Hosur Main Road, Adugodi
   Bangalore - 560 030

Director,
निदेशक
Central Institute of Indian Languages
भारतीय भाषा संस्थान
Mysore. मैसूर.

3/6

1. Welcome and Introduction about LDC-IL:

Prof. Udaya Narayana Singh, Director, Central Institute of Indian Languages, Mysore made a detailed presentation about the New Initiatives of the CIIL during the Eleventh Plan Period including LDC-IL and its relation with other projects and welcomed the members of the LDC-IL PAC for the first meeting.

2. Language Resources at CIIL

a. Speech: Ms. Rekha Sharma made a brief presentation about the work done by the Speech Sciences Group of the Institute.

b. Text: Dr. B. Mallikarjun made a presentation about Corpora Creation, maintenance in the area of Text and Speech at the Central Institute of Indian Languages.

3. Prescribed Measurable & Monitorable Outputs for the year 2007-08:

a. Before taking up this item, the members sought clarification regarding the structure of the Project Advisory Committee and its role in the functioning of the LDC-IL.

(i) It was clarified that the PAC is the Apex Body for guiding the activities of the LDC-IL including monitoring the work of the implementation agencies.

(ii) The Project Advisory Committee can form Sub-Committees to look into the different aspects of the work of the LDC-IL. Though it is proposed to have 2 meetings per annum, if need be, it can meet more than 2 times in a year. It was suggested that in the first year, the number of meetings could be more than two.

(iii) This being the first Project Advisory Committee meeting of the LDC-IL, it can set tone for the kind of its involvement in the guidance of the work on Indian Languages.

(iv) It can also take decisions relating to the fees structure (elucidated in the project proposal).

As desired by the members present, the copy each of the Project Proposal and Minutes of the SFC were given to all the members.

b. The difference between the LDC-IL of the Central Institute of Indian Languages and the Consortium Projects of the Ministry of Communication and Information Technology is that in case of the Central Institute of Indian Languages the deliverables are **Corpus** in different Indian Languages for Speech and Text and in case of the Ministry of Communication & Information Technology the deliverables are **Systems** or **Tools**.

c. It was agreed to form five groups to look into the aspects of standardization, implementation, plan for monitoring, quantum, languages, purchase of Corpus and other issues under each areas. (It was also agreed that each group would co-opt other specialists in the field).

Committee 1: Licensing Issues:

1. Dr. Hemant Darbari, C-DAC Pune
2. Shri Abhijit Dutta, IBM India, Haryana
3. Shri Pradeep Parappil, MICROSOFT, Haryana
4. Dr. B. Mallikarjun, LDC-IL, CIIL, Mysore.
   (Others to be Co-opted)

Director,
निदेशक
Central Institute of Indian Languages
भारतीय भाषा संस्थान
Mysore. मैसूर.

4/6

Committee 2: Natural Language Processing

1. Prof. Rajeev Sangal, IIIT, Hyderabad
2. Dr. Hemant Darbari, C-DAC Pune
3. Prof. Pushpak Bhattacharya
   (Others to be Co-opted)

Committee 3: Speech

1. Prof. Hema Murthy, IIT, Chennai
2. Prof. Peri Bhaskararao, Tokyo University of Foreign Studies Japan
3. Shri Ashwin Amarapur Chandramouly Motorola
4. Ms. Rekha Sharma, CIIL
   (Others to be Co-opted)

Committee 4: Character Recognition

1. Dr. B. B. Chaudhuri, ISI, Kolkata
2. Shri Shantanu Choudhuri
3. Shri Mahesh Kulkarni, C-DAC Pune
4. Shri A.G. Ramakrishna (Online Handwriting Recognition)
   (Others to be co-opted)

Committee 5: Speech Deficiencies

1. Prof. Vijayalakshmi Basavaraj, AIISH, Mysore
2. Ms. Smruthi, SNDT University, Pune
3. Ms. D. Vasanta, Osmania University, Hyderabad
4. Prof. Peri Bhaskararao, Tokyo University of Foreign Studies Japan.

These Working Groups will prepare the reports relating to languages to be covered, scope of the work, standardization, quality of data, financing model, open source/community model, different layers, end users , fees structures for R & D purpose, acknowledgement, buying and selling data and all other issues relating to the concerned area as the groups may deem fit.
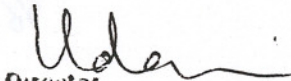
These Working Groups will submit their reports by mid of August 2007 and the PAC will meet during mid September 2007 to take stock of the situation and decide on the reports of the Working Groups.

d. The Institute will call for **Intent of Expression** from scholars and institutions in the field to work on different aspects of Linguistic Data Consortium for Indian Languages (LDC-IL) through newspapers, press release, web sites, State IT Departments, Circulars to Language Technology Institutions and through word of mouth including items/posters to be sent to different Indian Language Data Centre (ILDC).

e. The project proposals will be received up to August 31, 2007 and they will be placed before the PAC Meeting during mid September 2007. Receipt of proposals will be continuous and an on-going activity.

f. Dr. B. Mallikarjun gave details of proposed work relating to languages in the North-East being taken up by the LDC-IL for creating both Speech and Text Corpora out of NE funds of LDC-IL.

g. Indian English Data and Pronunciation Dictionary would also become part of the work of LDC-IL.

Director,
निदेशक
Central Institute of Indian Languages
भारतीय भाषा संस्थान
Mysore. मैसूर.

5/6

4. Proposed International Seminar, National Seminars/Events, Workshops, Training Programmes, Academic Meetings, Tasks, Projects etc.:

The e-mail sent by Prof. Pushpak Bhattacharya was considered.

It was agreed that the following programs would be conducted:

1. Workshop on Corpora Creation and Tagging, Prof. Uma Maheshwara Rao
2. Spoken Language Deficiency Workshop by AIISH, Mysore
3. One Regional Seminar by C-DAC, Pune
4. One International Seminar from January 7, 2008 by Prof. Rajeev Sangal, IIIT, Hyderabad.
5. Workshops by IIT, Bombay (Lexicon and WordNet, Morphology Analysis, Corpus Linguistics) by Prof. Pushpak Bhattacharya.
6. Workshop on Speech Recognition by IBM

7. The project proposal of Prof. Kavi Narayana Murthy regarding Tagging Telugu Corpora was considered and it was decided to ask the proposer to send the revised proposal so that the deliverables are clearly stated for the period of 6 to 8 months.

5. Any other Item with the permission of the Chair:

(1) The project proposals by different people could involve task driven fellowships. It was agreed that both individuals and institutions can apply for the projects.
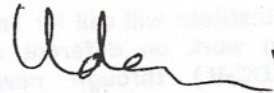
(2) **Infrastructure**: It was observed that the designations as given in the proposal (Scientists A to D) did not correspond to the Govt. of India designation for Scientists. It was, therefore, decided that all the positions listed under the project will be equated with the positions already existing in the CIIL and re-designated accordingly.

(3) The Director, CIIL has been authorized by the SFC to take necessary steps to recruit people and he has decided to have one Administration/Finance person and one Technical person to begin with so that the work could be initiated, the same were noted by the Committee.

(4) It was agreed that henceforth the CIIL will play a pivotal role in encoding language, yet to be encoded languages in the UNICODE.

(5) Prof. Rajeev Sangal communicated the opinion of Prof. Raj Reddy for providing low cost servers by TB size. The Chair-person agreed to consider this option.

The meeting ended with vote of thanks by Dr. B. Mallikarjun.

Prof. Udaya Narayana Singh
Chairperson
Project Advisory Committee, LDC-IL &
Director
Central Institute of Indian Languages, Mysore