

Abstract Booklet

5th International Endangered and Lesser-known Languages Conference (ELKL-5)

Organized by
Centre for Tribal Folklore, Language and Literature
Central University of Jharkhand

In collaboration with
Central Institute of Indian Languages, Mysuru



प्रधान मंत्री कार्यालय
नई दिल्ली- 110011
PRIME MINISTER'S OFFICE
NEW DELHI- 110011

MESSAGE

The Prime Minister is happy to learn that the Centre for Tribal Folklore, Language and Literature, Central University of Jharkhand, Ranchi in collaboration with Central Institute of Indian Languages, Ministry of Human Resource Development, is organizing 5th International Endangered and Lesser-known Languages Conference on the theme 'Scripts, Language Documentation, Policy and Cross Disciplinary Approaches for Endangered and Lesser-known Languages' from 24th - 26th February, 2017 and is publishing a souvenir to mark the occasion.

On this occasion, best wishes to the organizers and participants.

Chandresh Sona
Deputy Secretary

New Delhi
January 24, 2017

Upendra Kushwaha



Minister of State for
Human Resource Development
Government of India

I am delighted to associate with Central University of Jharkhand, Ranchi and the Centre for Tribal Folklore, Language and Literature through the timely 5th International Endangered and Lesser-known Languages Conference to be held from 24th-26th February, 2017.

As you rightly emphasised that this conference becomes more significant because of the fact that it is going to take place in the state of Jharkhand which is adorned with different indigenous groups speaking numerous languages coming from three different language families.

I believe that this conference may trigger a debate and sanitize the common mass towards the importance of language which defines and shapes one's identity. I wish all the academicians, researchers and scholars coming from across the globe to participate in the deliberation of the conference. I pray for the success of the conference.

Upendra Kushwaha



S.S. Sandhu
Joint Secretary (CU&L)
Ministry of Human Resource Development
Department of School Education and Literacy

It gives me immense pleasure to know that the Centre for Tribal Folklore, Language and Literature of Central University of Jharkhand in collaboration with Central Institute of Indian Languages (CIIL), Mysore is going to organize the 5th International Endangered and Lesser-known Languages Conference from 24th-26th February, 2017. It is indeed a step forward in the right direction as the university is located in a place which is repleat with various cultures, ethnic groups having rich folklore heritage and languages. I hope this conference would be able to trigger positive deliberation concerning the preservation and revitalization of languages.

I wish this conference all success and hope many such events will be organized in future.

Yours sincerely

S.S. Sandhu



Dr. Prof. Nand Kumar Yadav "Indu"
Vice-Chancellor
Central University of Jharkhand

It is a matter of happiness and contentment to know that the Centre for Tribal Folklore, Language and Literature, Central University of Jharkhand is going to organize the 5th International Endangered and Lesser-known Languages Conference from 24th-26th February, 2017. Academic pursuits like this is crucial for the fact that the university is located in a culturally and linguistically heterogenous place. Languages and folklores are the repository of human knowledge. Losing this treasure-trove would entail losing the vast source of indigenous knowledge and the very basic understanding of human evolution.

The theme of the conference includes cross-disciplinary approaches which enshrines the vision of this university towards cross-pollination of reseach methodologies and scientific techniques and temperatments.

I wish this conference a huge success and I hope that deliberations made during these three days are fruitful for one and all. My heartiest good wishes to the invited guest and the participants.

Prof. Nand Kumar Yadav "Indu"



Irina Bokova
Director-General
UNESCO

I wish to congratulate the Centre for Tribal Folklore, Language and Literacy at the Central University of Jharkhand, Ranchi, for organizing this important conference on international endangered and lesser-known languages.

This conference takes place during the same week as UNESCO celebrates International Mother Language Day to promote linguistic and cultural diversity and multilingualism. This year, we are shining the spotlight on multilingual education as an essential policy to promote equity, inclusion and successful learning, especially for children speaking minority or indigenous languages.

Every language embodies a culture, a system of unique knowledge and a wellspring of traditions. Every language is the carrier of values and a world-view that enriches all humanity. When a language is endangered, so are its people, and with it, part of humanity's heritage.

UNESCO's Atlas of the World's Endangered Languages estimates that there are some 3,000 endangered languages worldwide – this is about half of the languages spoken today. We also know that 40% of the global population does not have access to education in a language they speak or understand, paving the way for early exclusion for participation in society. This process is neither inevitable nor irreversible.

We must encourage stronger collaboration to raise awareness, share information about good practices, language teaching and education policies, because language matters for sustainable development, to build more inclusive and tolerant societies.

With its immense cultural and linguistic diversity, India plays a frontline role in safeguarding languages, through research, monitoring, documentation and policies that encourage their use, especially in education. Learning in mother tongue is a critical factor in improving student achievement, developing self-confidence and pride in one's culture – the foundation for being open to others.

We can and must safeguard and revive endangered languages and show their relevance to preserving our cultural and biological diversity. Technology, digital platforms and other creative mediums can give a new lease on life to languages; while advocacy and policies to encourage mother tongue languages in all areas, from education to the media and cyberspace, can champion their use.

I wish you every success in your invaluable work to safeguard the richness of India's multilingual landscape – a mirror of its wisdom and contribution to human civilization - because this will stand as an example to the world.

Irina Bokova



Prof. Dr. Nikolaus Himmelmann
Director
Institute of Linguistics
University of Cologne

Dear ELKL-5 organizers and participants

I am very sorry that I cannot make it to Jharkhand to your conference due to prior commitments in Australia. I enjoyed ELKL-4 in Agra last year very much and was impressed by the enthusiasm and dedication to the urgent documentation issues in your country.

Jharkhand is a particularly interesting place, linguistically speaking, for its mix of small languages of different genetic affiliations. Many of them highly endangered. I really would have liked to see this first hand.

But I am sure that this ELKL will be as successful and rewarding as the last one.

I wish you a productive time in Jharkhand and very much hope to be able to join in again.

Nikolaus Himmelmann

Dr. Tove Skutnabb Kangas
Retired Associate Professor
Åbo Akademi University Vasa, Finland

India is a country that, according to the Ethnologue, using EGIDS scale (<https://www.ethnologue.com/profile/IN>) has one of the highest percentages of languages on the endangered scale.

At the same time, much is done in India today to document these languages and to reduce them to writing. This is often seen as a prerequisite for revitalisation (even when this is not necessarily so). But many who document the languages (often outsiders, at least initially) do not have the time or competence or funds to revitalise them. This is work that requires constant presence of the revitalisers for years, even decades. Revitalisation also requires official support for the Linguistic Human Rights (LHRs) of the speakers/signers of the endangered languages.

This is especially important in formal school education. We have a good example of this, “Extracts from Odisha Guidelines on Multilingual Education, India, Government of Odisha” (mainly written by professor Ajit Mohanty), in Volume 4 of Skutnabb-Kangas, Tove & Phillipson, Robert (eds). Language Rights. London & New York: Routledge, a book of 1668 pages, four volumes. If the Odisha Guidelines were to spread to all India, and, especially, if they were implemented, so that

primary education with the mother tongues as the main teaching languages were to then.

With the speed than humankind is ruining the planet, it is not certain. We know that linguistic diversity is both correlationally and causally connected to biodiversity; much of the knowledge about how to maintain biodiversity is coded in small Indigenous/tribal and local languages, meaning exactly the languages that your conference is about. If these languages are killed off, this vital knowledge also disappears. Thus, the work you are doing is not only important for speakers and signers of the languages that you are documenting and revitalising. It is vital for all humankind.

We women are part of humankind too – why are 90% of your invited speakers men, I wonder? Have a fruitful conference!

Tove Skutnabb Kangas



Prof. Probal Dasgupta
Professor and Head
Linguistic Research Unit
Indian Statistical Institute, Kolkata

We are happy to hear that Centre for Tribal Folklore, Language and Literature at your university is hosting the fifth International ELKL (Endangered and Lesser-Known Languages) Conference in February 2017. It is important that you are holding the conference in Ranchi, which has the potential for becoming a hub for non-patronizing modes of inquiry in this domain, with serious participation by speakers of the endangered languages of these indigenous peoples – speakers not just playing games refereed by others but raising the stakes and changing the rules. No doubt your work at this conference will be part of the run-up the UN year of languages of indigenous peoples, 2019. I am associated with Universal Esperanto Association – an NGO that belongs to the UN family and actively pursues the linguistic rights of threatened communities, especially those whose languages are endangered. Both in that context and in terms of the language documentation work done at the department where I work, my colleagues are actively interested in ensuring that efforts such as yours help turn the tide. I speak for them when I emphasize that undermining the so-called mainstream has become a more urgent task than ever before. My colleagues in these circles join me in sending you our very best wishes for your conference and for other work you and your colleagues are doing in this domain.

Probal Dasgupta



Prof Anvita Abbi (Padma Shree)
Hon. Director: Centre for Oral and Tribal literature
Sahitya Akademi

I am happy to know that the Centre for Tribal Folklore, Language, and Literature, Central University of Jharkhand, Ranchi, India is going to organize the 5th International Endangered and Lesser-known Languages Conference (ELKL - 5).

Language documentation is the biggest challenge today as it has to ensure continuation of the existence of the indigenous knowledge-systems of the endangered and vanishing unwritten languages for the future existence of linguistic, cultural, and biological diversity. Documentation is necessary in understanding patterns in the structure and function of human language, human prehistory and the maintenance of the world's diverse ecosystems. Policies and strategies creating and safeguarding language archives should be undertaken by the linguists with the help of the local community members as key players.

I hope and wish that the Conference provides a healthy forum for scholars to discuss all aspects of language documentation and language archiving including safeguarding intellectual property rights of the speakers of languages that take part in documentation.

Wishing a grand success for the conference.

Padma Shri Prof. Anvita Abbi

Presentations

Segmental Phonology of Paite

Nianglianmoi

Jawaharlal Nehru University

Paite is a language of India having a total of approximately 64,100 speakers scattered across various states of North – East India such as Churachandpur district in Manipur, Champhai and Aizawl district in Mizoram, Tripura and Assam. Paite is also spoken by 13,000 speakers inhabiting Tiddim – Area of Chin – Hills of Myanmar.

This paper offers a description of the segmental phonology of Paite and discusses the relationship between vowel – lengths and tone.

National Unification and Endangered Languages:

Case study of China

Arpana Raj

Central University of Jharkhand

In China, among its 56 officially recognized ethnic nationalities, the Han ethnic group, comprising of 92% of the total 1.3 Billion Chinese people speak the official Chinese language, Mandarin; rest of the 55 ethnic minorities have more than 300 languages and dialects. But it is very obvious that, due to political and economic reasons many minority people like Man, Hui etc. willingly or unwillingly have accepted the ideology, language and culture of the majority

Han. Various policies of the government taken for national unification and strengthening has in turn forced numbers of Chinese lesser spoken languages to enter into the danger of extinct or endangered languages.

In this paper, I am going to focus on PRC's Language policies highlighting the policies towards its lesser spoken minority languages and its practice. I will try to analyze and highlight exact issues and its possible solutions and future prospects.

Classifiers in Taluitem

Chingrimung Lungleng

Dr. Bhim Rao Ambedkar University

This paper is a brief description of classifiers in Taluitem, a Tibeto-Burman language. It is spoken in Ukhrul district of Manipur by about 5000 people.

Taluitem has only nominal classifiers. The classifiers are not always exclusive in their use. As such, the same classifier may be used for both animate as well as inanimate objects as well as human or non human.

The context determines the use of classifiers. The classifiers mentioned can be used interchangeably with others that denote a similar object. The speakers have the choice of either using or omitting the schwa in the classifier. There is a lot of deletion in the language. A more comprehensive description will be given during the presentation.

A take on Adverbial Structures in Toto

Atreyee Sengupta

Jadavpur University

Toto is one of the critically endangered languages of India. Toto belongs to Sino-Tibetan language family and is spoken by Toto people in Totopara located in Indo-Bhutan border in Jalpaiguri district, West Bengal. Toto is a non-tonal language and uses SOV typology.

The principle objective of this paper is to investigate the Adverbial Clauses and Adverbial Quantification in Toto Language.

Given that the language documentation is a process through which a language can be preserved, especially ones which are endangered; there are many methods through which one can document a language. I used the questionnaire method. In Totopara I analysed the adverbial structures of the Toto Language by interviewing informants from a pre-set questionnaire.

Conjunctive Participle in Tinkar-lo and Hindi

Zoya Khalid

Jawaharlal Nehru University

Tinkar-lo is language spoken by Tinkar tribe. The '-lo' in Tinkar-lo means language. Tinkar originally is name of a

village in Byans valley in Nepal. Tinkari people now live in Dharchula district of Uttarakhand, India , and some Tinkari people have settled across Hahakali river Dharchula district of Nepal. Data of Tinkar-lo language, for this paper has been collected from native speakers of Tinkar-lo, living in Dharchula. It is an endangered language with only a few thousand speakers left. This is a work of language documentation and description of Tinkar-lo.

In this paper we see how there is a striking similarity between the use of conjunctive participle in both these languages; Tinkar-lo which is a Tibeto-Burman language and Hindi which falls under Indo-Aryan family of languages. But the use of conjunctive participle is freer, and more productive in Tinkar-lo than in Hindi.

Darma Language: A Phonological Description.

Ashish Kumar Pandey, K Srikumar

University of Lucknow

Darma belongs to the West Tibeto -Burman language family spoken by less than 3019 people of all villages of Darma valley (according to census 2011) in Dharchulla Tahseel, Pithauragarh District of Uttaranchal state. Darma speakers are found in the surroundings of “Dhaulti Ganges” river. Darma is also known by another name “Darmiya”, in Grierson’s Linguistic Survey of India (Varma, 1972). Annual migrations place them in two different language situations: Hindi and

Kumauni, dominating in the area of Dharchula.

In this paper, we shall present a sketch of the sound patterns of Darma language. The sounds of Darma language included here are presented in their phonemic and phonetic inventories, along with sound patterns of syllable, suprasegmental features and prosodic words. We have collected the information from people above 45 years, and got it verified by cross checking with other speakers. Approximately the same sounds were produced by every speaker. Information has been collected through questionnaire and personal interviews and direct observations in the field.

Revitalizing Mundari : Efforts, Effectiveness and Future Strategies

Gunjal Ikir Munda, Deep Lakshmi

Central University of Jharkhand

Mundari is a language belonging to the Austro-Asiatic Language family. The language is the L1 of the Munda Tribe which resides in the states of Jharkhand, Odisha and West Bengal. The population of the Munda tribe is 2.2 million (2011 census, Government of India), while the population of Mundari speakers is nearly 1.1 million (Ethnologue), nearly half the total population. The dwindling number of speakers is one of the indicators of vulnerable situation Mundari language is in.

No doubt, over the years measures have been taken by various leaders and organizations in countering the language shift. But even after all these efforts, the language shift among the Mundas still continues. Is it that language shift is such an enormous phenomenon that it will require a series of such efforts over a long period of time to reverse it? What is the status and the effectiveness of the ongoing revitalization efforts? What would be the best strategy to effectively further the language revitalization in Mundari? The present paper will try to answer these and other questions.

Noun Morphology in Bokar Language

Geyi Ete

Deccan College Post-Graduate & Research Institute

This paper attempts to study the Noun Morphology of Bokar, a Tani group of Tibeto-Burman language family, which is spoken in western Arunachal Pradesh, a state in the North-Eastern India. There are approx 5,500 speaker of this language in the State. The present paper discusses the inflectional categories of the noun such as gender, Number, Person and Case Marking in Bokar language.

In Bokar language, nouns are classified with classifier system involving shape, function arrangement etc. However there are some lexical nouns which can be differentiated either masculine or feminine.

Bokar classifiers are broadly divided into two types: sortal and menstrual. Sortal classifiers are used to denote nouns which are highly countable in nature such as human, animal and object e.g. table, book, utensil etc. and menstrual classifiers are used for measuring units of countable and mass nouns, nouns such as 'rice' 'Salt' 'Oil' etc are conceptualized as mass nouns which must take measure.

Reconstructing Language: Recapturing Oral Narratives the Koireng Example

T. Neishoning Koireng

Central University of Jharkhand

The Koirengs are a microscopic community found in Manipur. During the colonial rule, the Koirengs were referred to by different names by the British writers. Modern and present scholars usually refer to as Koireng but the community called themselves as Koren. Linguistically, the Koireng language belongs to the Kuki-Chin of the Tibeto-Burman family as researchers have pointed out. Till date, though the Koirengs have adopted roman script for writing, there is no well organised script or alphabet to follow in writing.

This paper would discuss about the land and people to understand the socio-cultural milieu of the Koireng. It would try to explain the present status of the Koireng language to keep in mind about its literary aspect and to discuss

necessary inputs to highlight activities from endangerment. Further, it would try to discuss how Koirang language could be reconstructed by recapturing oral narrative.

Developing a multimodal dictionary of Beda language

Ritesh Kumar, Girijesh Kumar Bharati, Yogesh Dawer

Dr. Bhim Rao Ambedkar University

Beda is a moribund language, currently being spoken by less than 50 people in a couple of villages in Leh district of Jammu & Kashmir. The language has remained completely undocumented and undescribed till now. As such its genetic affiliations remain unestablished. However the present fieldwork shows its close affinity with the neighbouring languages like Laddakhi and Balti and it might be safely posited that it belongs to the Tibetan group of Tibeto-Burman language family. Like most of the other languages of South Asia, it follows a canonical SOV word order.

In the present paper, we will present an effort to build a multimodal, trilingual Beda-English-Hindi dictionary. The data for dictionary was collected directly from the speakers using a questionnaire designed specifically for this purpose. The field work was primarily carried out in Chuchot Yokma village of Leh district. The dictionary is being developed using SIL FieldWorks Language Explorer.

**First language vulnerability to Attrition: Case of lexical
retrieval in Magahi Community of Patna**

Saloni Priya

Jawaharlal Nehru University

The Magahi Language community belongs to Bihar and spoken mostly in the southern and western parts of Bihar. The following paper examines the difficulties and the process of lexical retrieval of Magahi speakers living in Patna. The moment we think about the causes of language shift, the three main factors that come to our mind are ideology, attitude and need of the speakers. But these factors cannot be evident enough to prove that an individual is going through attrition in their linguistic knowledge because these three factors are totally psychological. In this type of study, the main focus would be on the linguistic features, which the attriters are losing gradually in their first language use. Therefore, this paper has been focused on the attriters ability of retrieving lexemes in their Mother tongue through a particular task given to them.

Tense and Aspect in Toto Language

Trishita Nandi

Jadavpur University

The paper discusses the tense and aspectual categories of Toto. It is often seen in languages that tense and aspect markers show special inflections. This work attempts to

investigate the tense and aspectual markers of Toto language and try to make generalized view of the categories. It is observed that different kind of same aspectual categories shows different markers with respect to verbs.

Negation in Toto

Asmita Gupta

Jadavpur University

This project explains how negatives work in Toto and the processes by which they get attached to the words and change their meanings. Toto Language belongs to the Himalayan subgroup of Tibeto-Burman language family, which is mainly spoken in Totopara located in the Alipurduar district of West Bengal. Researchers and even the members of this community admit that the language is under threat with the influence of Nepali and Bengali increasing day by day.

Major versus Minor regional languages: A Case study of Naxalbari

Bornini Lahiri

Jadavpur University

Dhimal is a Tibeto-Burman language spoken in both India and Nepal. In India, it is surrounded by Indo-Aryan languages like Bengali, Nepali and Rajbanshi. It does not have a script of its own though Bengali script is used to write Dhimal. There is no motivation to learn and speak Dhimal for the younger generation, as no one understands the language except

members of their community who are also slowly deserting the language. The young people have to go out of the community for studies and job where they cannot use Dhimal. This also makes the parents think that Dhimal is of no use for their children so the children should not learn Dhimal. This plays as a motivation for them to learn Bengali. There are various other reasons for the Dhimal community to abandon Dhimal and shift to Bengali like inter-speech community marriages and social gatherings. The paper elaborates the various reasons for this shift and the issues related to it. The present paper studies the trend of language shift of Dhimal. The paper, explores the reason behind the language shift and the reasons behind choosing one of the languages for the shift.

Challenges in Documenting an Endangered Language: A Case Study of Dhimal

Mayank, Bornini Lahiri

Jawaharlal Nehru University, Jadavpur University

The present paper discusses the challenges of documenting an endangered language of India, named Dhimal. The study is based on a fieldwork conducted in the two villages (Hatighisa and Mallabari) of Darjeeling district of West Bengal in November 2016. The Dhimal community is a small community with around 800 members as reported by the community head. This community is surrounded mainly by Bengali, Nepali and Rajbanshi communities. The interaction

with the other speech communities has resulted in language contact situation within Dhimal community. The Language contact situation has led to borrowings and complex structures. Language shift too can be witnessed in the Dhimal community as the younger generation is shifting towards major language of the area. As the language is already endangered, it is needed to be documented as soon as possible. We visited the field with the purpose of documenting the language. But in the field we had to face various challenges not only because the language is not fluently spoken by the members but also because of the certain set up of the community.

**“Remo” Folklore and Myths: Understanding Bonda
Identity**

Nancy Yadav

Jawaharlal Nehru University

Bonda tribe represents one of the small scale and un-differential indigenous societies. Bonda constitute to be one of the few tribal communities of India who are relatively less affected by the trends of modern civilization and come under Particularly Vulnerable Tribal Groups. Bonda tribe is part of Austro-Asiatic tribes and their language “Remo” comes under Munda group of Austro-Asiatic languages. The ‘Munda’ group inter-related to the ‘Remo’ genetically and structurally, with common source ancestry and cultural heritage.

In this paper, I have tried to examine the manifestation of Ramkatha in Bonda Folklore, its reception and reflection in ethnographic monograph and fiction. The study of Bonda folk tale in this paper is based on the hypothesis that epic, myths, legends and rituals comprise suggestive events. I will examine the reception, re-writing and re-evaluation of Bonda myths of creation and social institution of the indigenous tribe through ethnographic monographs by Verrier Elwin. Bonda Highlanders will provide us the emic perspective of the Bonda myth.

**Vowel phonemes in Muduga, Pal Kurumba and Irula
languages and their orthographic representation**

Binny Abraham

Central University of Kerala

Muduga, Pal Kurumba (Attappady Kurumba) and Irula are tribal languages spoken in Attappady area of Palakkad district in Kerala. In Kerala Muduga population is 4,668, Kurumbas 2,251 and Irula is 26,525 (Mahadevan. 2016). According to 1991 census the population of Irula in Tamilnadu is 1,55,606. All three languages are genetically classified as South Dravidian.

I have made multiple trips to Attappady region and gathered around 1000 words from Muduga and 750 words from Pal Kurumba languages, transcribed them using International Phonetic Alphabet (IPA), and audio recorded them in wave

format using a high quality audio recorder. I use the freeware Language Explorer to key in and organise the data, Phonology Assistant to facilitate phonological analysis, Audacity to edit audio files and Praat for spectrogram analysis.

I work with these communities to develop practical orthographies for them based on the methodology outlined by Easton and Wroge (2012).

Vitality and Endangerment of Contemporary Kurukh

Alisha Vandana Lakra

Indian Institute of Technology, Dhanbad

Kurukh is predominantly the language of the Oraon tribe of the Chotanagpur Plateau of East Central India and spoken by approximately two million people. For the present study the researcher considers Kurukh which is spoken in Ranchi and in its outskirts for describing Kurukh language in its social context. The study focuses on the linguistic variations of Kurukh at all the levels; people's attitude towards the language (insider and outsider); the endangerment of the language (degree and responsible factors); language vitality and code mixing and code switching found in contemporary Kurukh emphasizing on the social and linguistic factors that restrict the use of Kurukh language and also its restructuring in daily communication when it comes in contact with other neighbouring languages, like Hindi, Munda and English.

**Radio Bultoo – An Innovative Media Concept to Facilitate
the Preservation of Endangered Tribal Languages**

Shalini Saboo

Central University of Jharkhand

Radio Bultoo, a CG Net Swara initiative, invented earlier in 2016 is an innovative concept of using mobile phone and internet technology to save the dying local dialect of tribal people. The paper through its study, explores how tribals residing in remote villages can use mobile phones to record songs or voice their problems in their own indigenous dialects. This is one of the best ways for tribals to not only reach out to the mainstream or to voice their grievances to the concerned officials but to save their language which is dying a slow death due to lack of preservation and dissemination.

**A comparative phonological study of three varieties of
Tangkhul-Naga language**

Chakshang Siro

Deccan College Post-Graduate and Research Institute

In this paper, I present a phonological description of three varieties of Tangkhul language namely, Hunphun, Halang Aze and Laho (Leingangching). Tangkhul-Naga language is spoken predominantly in Ukhrul district of Manipur belonging to Tibeto-Burman language family.

Internal reconstruction and comparative methods are used to study the development of sound change. These methods are central to historical linguistics and they are the most important of the various methods and techniques we use to recover linguistic history. One of the commonly used principles is the majority principle. Since, Tangkhul languages didn't have recorded speech in writing until 1837 when the first vocabulary collections on Hunphun, Phadang and Khangoi varieties were taken, we have to depend more on synchronic variation to know the past in comparing the three varieties and diachronic variation is relevant only on the study of sound evolution in Hunphun variety.

Sounds and Syllables of Tai-Turung

Amalesh Gope, Manokh Hazarika, Prashanta Hazarika

Tezpur University

The present paper aims at providing a preliminary description of Tai-Turung phonemes. Tai-Turung belongs to Sino-Tibetan group of Tibeto-Burman language family and is widely spoken in the regions such as Jorhat, Golaghat and Karbi-Anglong districts of Assam and in some parts of Arunachal Pradesh.

According to the 2001 Census Reports of India, there are approximately 1300 people using Turung as their first language. Turung is considered as one of the endangered language by the experts. The native speakers prefer to use Turung at home, however they shift to Assamese while

communicating outside their home. The indigenous group considers the Pali script as their own, however the few documents that we came across representing this language are all written either in the Roman, Assamese or Tai script and some words of this language are gradually influenced by the neighboring Assamese language. The younger generation are prone to losing their native tongue way-faster and are comfortable of using Assamese in all the contexts of their daily life.

Generation of a Digital Lexical Database from a Multi-Domain Khortha Speech Corpus through Lexicographic Glossing and Annotation

Atul Aman, Niladri Sekhar Dash, Amalesh Gope
Indian Statistical Institute, Kolkata

The present work reports about the development of the first digital lexical database of Khortha from a multi-domain Khortha speech corpus. Each lexical item in the proposed lexical database is annotated with phonetic segmentation, part of speech, usage, and other lexicographic information within a Khortha sentence along with its English translation. The digital database constitutes a multi-domain lexical stock which is used commonly by the Khortha speech community in their day to day life. Khortha belongs to the Indo-Aryan language family and has been widely spoken in and around 15 districts of Jharkhand (Ohdar 2012). Officially, it is the second most populated variety in the state of Jharkhand with a population of 47, 25, 927 speakers (Census 2001). From

multiple surveys conducted, it has been observed that the speakers of Khortha are not able to maintain an independent linguistic identity as they are not sure about the linguistic uniqueness and characteristics of the language.

Anaphors in Toto

Atanu Saha

Jadhavpur University

In this paper, I have discussed reflexives and reciprocals of Toto. The main research questions are as the following:

- 1) How is reflexivity and reciprocity expressed in the language?
- 2) How forms and functions of those two types of constructions (reflexivity and reciprocity) are related as classified in Subbarao (2012) for the anaphors of South Asian Languages.

In order to elicit the data, I used the Berlin-Utrecht Reciprocals questionnaire, survey & outline edited by Dimitriadis (2006). A field survey (2016) was conducted at Totopara and the data was collected from two speakers (Santi Toto & Bhabesh Toto) and verified from a third Toto Speaker (Sanchita Toto).

Verbal Morphology in Mising

Barshapriya Dutta

The paper describes the verbal morphology in Mising, a Tibeto-Burman language variety spoken by a community who are also called Mising. The Misings are one of the major

ethnic groups of Assam .The Mising population of Assam spreads over the districts of Dhemaji, Lakhimpur and Sonitpur in the North Bank and Tinsukia, Dibrugarh, Sibsagar, Jorhat and Golaghat in the south bank of the Brahmaputra. The study is based on the data collected from various places all over Assam.

Mising language has extensive verbal morphology. Most of the morphemes are suffixes which are attached to the verbs resulting in inflectional or derivational morphology. This paper reflects on the use of different suffixes that are added in terms of tense, aspect, interrogative and the negative. This paper will also discuss certain possibilities where a native speaker of this language uses different suffixes to refer to the same aspectual category based on some morpho-syntactic and pragmatic conditions.

The Morphosyntax of Numeral System of Mundari and Santali

Kumari Mamta

Jawaharlal Nehru University

This paper talks about the formation of simple and complex numbers, the base, the arithmetic operations involved, the morphology induced in the complex number formation, ordinals, multiplicatives, how numeral 1,2,3 is different from rests and inside it also how numeral ONE behaves differently in most of the Indian languages alongwith the Munda

languages. They both belong to the same family but still show variation in the formation of ordinals. In one hand Santali has decimal base system and on the other hand Mundari has combined base system which includes decimal-vigesimal system but the arithmetic operations they use are same. In Ordinals, we will see the idiosyncrasy of numeral ONE. Alongwith these, it also talks about the formation of multiplicatives and fractions, how they are transcribed. Further this paper deals with the interaction of numerals with syntax which includes demonstrative determiners, definiteness, number and gender with lots of data support, which is really very interesting in these Munda languages.

Endangered Languages of Karnataka – the way ahead

Chaithra Puttaswamy

Indian Institute of Technology, Kanpur

Soliga, Jenu Kuruba, Betta Kurumba, Yerava, Irula, Badaga, Koraga and Hakki Pikki are the endangered languages of Karnataka. The number of speakers of these languages vary between 16,000 Koragas to about 7,00,000 Badagas. They also belong to varied socioeconomic backgrounds. They are dominated by Kannada, the state official languages and locally by Tamil in the Nilgiri hills and Tulu and Konkani on the west coast of Karnataka. Except for Hakki Pikki, all the other languages belong to the Dravidian family and have several structural and lexical similarities with Kannada, Tamil and Tulu. Koraga and Kuruba are listed as critically

endangered languages on the UNESCO list of endangered languages.

This paper suggests some of these measures that can be implemented in the near future to boost the vitality of the endangered languages of Karnataka.

**Endangered Situation of Khortha Language due to
Industrialization: A Sociolinguistic Study of Bokaro
District**

Dayanand Kumar, Prakash Bhadury

University of Lucknow, Indian Institute of Petroleum and
Energy

Khortha is a language spoken in thirteen districts of Jharkhand. Among all the languages of Jharkhand, khortha is the language that is related to sea (near Farakka) and Damodar River. Along with this it is only language of Jharkhand that creates international relation i.e. Bangladesh in relation to its language area or boundary. Languages such as Nagpuri, Khortha, and Panch-Pargana have become lingua franca to facilitate communication between tribal groups as well as between tribals and non-tribals.

This paper, at the outset, endeavors to show how industrialization has affected khortha language and its culture in and around the steel city, Bokaro and soon it might be

turned out to be an endangered language or it might be a dead language like Sanskrit. Ranchi University is conducting Master's Degree in Khortha since 1980s but still it has not received any status.

Compound verbs and serial verbs in Bodo and Dimasa

Daimalu Brahma, Monali Longmailai

Tezpur University

Bodo and Dimasa belong to the Bodo-Garo sub-group of the Tibeto-Burman language family, spoken mainly in northeast India. This paper will look into the verbal morphosyntax of Bodo and Dimas with reference to phrasal verbs. Phrasal verbs here will include compound verbs in terms of morphological distribution and serial verbs in terms of syntactic structure.

It will discuss the kinds of verbal compounds used in Bodo and Dimasa and compare the patterns of phonological and lexical variation in both the languages. The paper will also look into nominalized verbs, denominalized verbs, borrowings and the use of body parts to form verbal compounds as, explicator compound verbs and conjunct verbs. It will also look into the tense-aspect-mood inflections in syntactic constructions. The serial verbs will be discussed in terms of their finiteness and the verbal combinations with free and bound verb stems, adverbial particles and applicative constructions.

A cognitive account of perception verbs in Ladakhi

Konchok Tashi, Neha Garg

Central University of Jharkhand, Banaras Hindu University

The present article is a cognitive semantic account of polysemy in the semantic field of perception verbs in Ladakhi language. It explores why and how our experience and understanding of the five senses constrains and shapes the way in which we create mappings between the physical domain of perception onto more metaphorical and abstract conceptual domains of experience. The different extensions of meaning in these verbs, have not taken place as a result of chance, but are grounded in our own conceptualization of these sense modalities. Therefore, in this article, the focus will be on the analyses of the meanings of perception verbs in Ladakhi and it will show how the study of polysemous categories plays important role in linguistic analysis in terms of prototypes, metaphor and metonymy that are central to cognitive linguistics.

A Preliminary Study on Phong Phonology

Niharika Dutta

This paper is a preliminary description of the phonology of Phong, spoken by a small community of around 4000 speakers known as Phong or Ponthai living in the Changlang and Tirap district of Arunachal Pradesh. The Phong

community is one of the more than thirty-two sub-groups of the larger Tangsa group. This paper presents a first-hand description of the ambiguous segments, syllable structures, consonants, vowels, and tones of Phong. The analysis is based on my fieldwork data collected from the Wagon Ponthai village situated in the Changlang district. The methodology used in collection of the data is the BOLD methodology using Zoom H4n recorder. Consonant and vowel inventories will be presented along with some acoustic measurements to highlight certain phonetic features of consonants and vowels. I will include VOT measurements, F1 and F2 measurements plotted on a vowel space diagram, and vowel length measurements. I will also present the acoustic measurements of the pitch heights and contours of various tones.

**Valence Increasing Operation With Special Focus On The
Causatives And Its Types Found In The Chaparmukh-
Nagaon Variety Of Nepali**

Prangshu Manjul, Gitanjali Chetry
Tezpur University

This paper is an attempt to look into the valence increasing operation and we aim to explore the components of various types of causative constructions, viz. lexical causatives, morphological causatives and analytic or periphrastic causatives with reference to the Chaparmukh-Nagaon variety of Nepali spoken in Assam, India. We collected our data from the Chaparmukh village situated in the Nagaon district of

Assam. Here, one can find three Nepali villages namely, Takalatup, Hatbor and Khariatuli where most of the families speak Nepali as their primary language. In Assam, the Nepali speakers are mostly influenced by the dominant language (Assamese) of this state and the Chaparmukh variety of Nepali too is largely influenced by the Assamese language.

Literacy for Minority Language Speakers in a Multilingual Setting: A case Study on Rabha from Assam

Subhash Rabha, Mouchumi Handique

Gauhati University

This paper is going to focus on how effective Mother tongue based multilingual Education (MTB-MLE) programme can be when applied to a multilingual setting, based on a case study piloted over the Rabha Community children from Assam.

There was a case study of MTB-MLE piloted over some Rabha speaking children and its result was quite positive. Children were noticed to have understood, learned and composed in a much better and easier way than before. Their gradual shifting from L1 (Rabha) to L2 (Assamese) was observed to be easier and smoother when this teaching-learning method was applied. Gradually, they will be able to shift to an L3 or an L4 with the help of this programme. This paper will discuss the MTB-MLE programme, its application to the Rabha Community children and the consequences.

Verbal Morphology in Birhor

Bikram Jora

University of Delhi

Nomadic tribe Birhor belongs to Munda language family of Kherwarian group. It is spoken in Hazaribagh, Ranchi and Singhbhum districts and other small pockets in Jharkhand states and some speakers are also found in Odisha, Chhattisgarh and West Bengal. It has similarities with Santali, Mundari and Ho languages in terms of lexical items and grammatical exponent. However, several authors have claimed a Mundari-esque orientation of Birhor in the Kherwarian language-dialect continuum, grammatically it is clear that it belongs together with Santali.

The present study examine the semantic complexity and syntactic structure of verbal morphology in Birhor. Verbal morphology is very complex system in Birhor as most of the Kherwarian languages shows.

कुरमाली लोकसाहित्य एवं उनकी सामग्रियाँ

मंजय प्रमाणिक

जनजातीय एवं क्षेत्रीय भाषा विभाग, राँची विश्वविद्यालय

कुरमाली भाषा झारखण्ड के राँची, सिंहभूम, धनबाद, बोकारो, हजारीबाग, गिरिडीह, पलामू, संतालपरगना, पश्चिम बंगाल के पुरुलिया, झाड़ग्राम, बांकुड़ा, वीरभूम, उड़ीसा के क्योँझर, बोनई, बामड़ा, मयूरभंज और सुन्दरगढ़ में बोली जाती है। इसके अतिरिक्त पूर्णिया, भागलपुर, असम, मध्यप्रदेश एवं बंगलादेश में भी बोली

जाती हैं, किन्तु इन क्षेत्रों में बोलने वाली की संख्या सीमित है। कुरमाली भाषा बोलने वालों की संख्या लगभग एक करोड़ हो गयी है।

कुरमाली लोकसाहित्य का अपना निजी भरा पूरा इतिहास है। इसकी प्रमुख विधाएँ जनता जर्नादन की अपनी सम्पत्ति हैं एवं सूर्य की किरणों की तरह सर्वत्र व्याप्त है। एवं अपने गर्भ में साहित्यिक विविध विधाओं को संजोये हुए है। परन्तु औद्योगिक एवं तकनीकी के विकास के कारण ग्रामीण जीवन विघटन की ओर है और लोगों में शहरी जीवन जीने की ललक बढ़ रही है। अतः लोकसाहित्य की प्रमुख विधाओं को भाषा-संस्कृति को आधुनिकता के दबाव से बचाने की आवश्यकता है और अपनी भाषा की ओर ध्यान आकृष्ट कराना हमारा पुनीत कर्त्तव्य हो जाता है।

लोक-साहित्य में मुंडारी लोकगीत

सोहन मुण्डा

जनजातीय एवं क्षेत्रीय भाषा विभाग, राँची विश्वविद्यालय

लोक साहित्य में मुख्यतः जीवन की भावात्मक पक्षों की अभिव्यक्ति होती है, इनकी सीमाएँ भी भावों से बनती है। लोक में व्याप्त प्राणियों के क्षण-क्षण की अनुभूतियाँ, हृदयउद्गार, क्रिया व्यापार इसी लोक साहित्य में सजीव एवं साकार होते हैं। इसलिए लोक साहित्य में भाव का विशेष महत्व है। विश्व की समग्रता में सहज एवं सामान्य सत्य रूप है, लोक साहित्य उसकी विवृति करता है। यह देशकाल की सीमाओं से बँधा नहीं रहता बल्कि गतिशील युग की सामान्य चेतना की प्रत्येक गति, जागृति का, धर्म एवं नीति का स्वाभाविक चित्रण इसमें निहित रहता है।

गद्य हो या पद्य, तुकान्त हो अथवा अतुकान्त, गेय हो अथवा कथ्य, श्रव्य हो अथवा दृश्य, पुरुषों से सम्बद्ध हो या नारियों से, ज्ञानवर्द्धन हो या अज्ञानवर्द्धन सभी प्रकार की रचनाएँ मौलिक तथा परम्परागत होने के कारण लोक साहित्य की

परिधि के भीतर समाविष्ट हो जाती है। विभिन्न पर्वों, संस्कारों तथा अन्य अवसरों पर गीत द्वारा मनोरंजन किया जाता है। गाथाओं के द्वारा रसानुभूति होती है। कथाओं द्वारा मन बहलाया जाता है, नाट्यों द्वारा रससंचार किया जाता है। चमत्कारिक लोकोक्तियों द्वारा इच्छित विचारों की अभिव्यंजना की जाती है और प्रहेलिकाओं द्वारा बुद्धि परीक्षण होता है। इस प्रकार लोक साहित्य की ये सभी क्रियाएँ लोकजन, लोक मंगल तथा लोकोन्नयन के अपूर्व साधन हैं।

Plenary Talks

**Language documentation and language revitalisation –
partners or just good friends?**

Peter K. Austin

Linguistics Department, SOAS

University of London

Around 20 years ago a new sub-field of linguistics emerged called Language Documentation (or Documentary Linguistics) with the goal of “compiling a representative and lasting multipurpose record of a natural language or one of its varieties” (Himmelman 1998). Language documentation involves creating archivable audio, video and textual recordings of language use in its social and cultural context, and translating and annotating them, paying proper attention to relevant contextual metadata. This approach emphasises transparency and multifunctionality, arguing that the recordings and analysis should be available and accessible to a wide range of users for a wide range of functions, including community members. There is a growing theoretical and applied literature on language documentation.

Language revitalisation is concerned with increasing the number of speakers of a language and the range of domains within which it is used. This often involves collaboration between researchers (linguists, applied linguists, educators) and community members to create relevant materials and curricula as well as contexts within which the language can be used. The origins of language revitalisation are older than

language documentation; however it has not attracted the same level of funding or recognition. It has also been undertheorised and is often seen as a waste of time by mainstream linguists (Dimmendaal, Blench, Newman), while also failing to engage with mainstream applied linguistics.

This paper explores the relationships between language documentation and revitalisation, arguing that there is a practical and theoretical gulf between them, and calling for better theorization and interaction. Work on both documentation and revitalisation has also failed to pay proper attention to local ethnographies and management of language use, and the crucially important but poorly researched beliefs and ideologies about language and language use held by both speech communities and researchers.

Munda Languages: A Syntactic Typology

Kārumūri V. Subbārāo

University of Hyderabad

The aim of this paper is to present some syntactic structures of the North Munda languages such as Santali, Mundari and Ho and attempt to compare and contrast them with similar structures in the other major three families of the South Asian subcontinent, namely, Dravidian, Indo-Aryan and Tibeto-Burman, which are verb-final and also with select Mon-Khmer languages, which are non-verb-final. Though Munda languages belong to the Austro-Asiatic family, which is

genetically non-verbal, the sub-branch of Munda languages took a U-turn due to language contact and convergence, and ended up as verb-final. Though they acquired many features of the SOV languages due to convergence, there are some residual features left from the ancestral non-verb-final family. We shall attempt to hint at some of those features too in this paper.

The organization of this paper is as follows: (i) some word order universals found in Munda languages; (ii) some significant syntactic features that provide evidence in support of the notion of 'South Asia/India as a linguistic area'; (iii) the parametric variation found in Munda languages; (v) the effects of language contact in Munda languages; and (vi) the unique syntactic features that this subbranch possesses that distinguish this family from the others.

**MITHILANCHAL URDU: A Linguistic Study of newly-
discovered Dialect**

M J Warsi

Washington University in St. Louis

Language does not have any boundaries, nor is it dependent on any boundary. Dialects are the contact languages of particular regions, and they have a deep impact on their cultural heritage. Slowly with time these dialects begin to take shape of languages. The convergence of a dialect into a language is a symbol and pride of the people who speak it.

With its ancient roots and myriad population, India is a linguistic haven. As years change into centuries, the sounds of umpteen languages and dialects have filled the country's air. Their stories, of course, do not follow a similar pattern, some are in extinction mode, and some barely survive while some are flourishing in equal measure.

I am going to talk about the first linguistic study of newly discovered Mithilanchal Urdu, a dialect of Urdu language of Indo-Aryan family, spoken by around four million speakers in Darbhanga, Samastipur, Begusarai, Madhubani, and Muzaffarpur districts of the state of Bihar in India. It has SOV word order and it lacks script and literature. Needless to say, this work is an attempt to document this dialect so that it should contribute in the field of descriptive linguistics. Although, confined to the five districts of northern Bihar, yet highly-popular among the natives, it is the primary mode of communication of the local Muslims.

I will be speaking about the structure of expressions about Mithilanchal Urdu that include the structure of words, phrases, clauses, and sentences. There are clear differences in linguistic features of Mithilanchal Urdu vis-à-vis Urdu, Maithili and Hindi. Though being a dialect of Urdu, interestingly, there is only one second person pronoun *tu* and lack of agentive marker *-ne*. Although, being spoken in the vicinity of Hindi, Urdu and Maithili, it undoubtedly has its own linguistic features, of them, verb conjugation is remarkably unique.

Because of the oral tradition of this link language, intonation has become significantly prominent. I will briefly touch upon the phonology, morphology, syntax, and will go through a sample text of Mithilanchal Urdu to make the audience aware about my findings.

Scripts for Under-Resourced Languages of India

Pramod Pandey

Jawaharlal Nehru University

The present talk addresses issues relating to scripts for under-resourced languages of India. I begin with a brief discussion of the nature of relation between speech and writing, between scripts and writing systems, and a classification of their types. In order to decide on a script for unwritten languages of India, it is useful to take into account some salient features of an efficient writing system: The Alphabetic Paradox, Orthographic Consistency, Morphological Consistency, Readability and Granularity. It is also found to be relevant to consider the educational and cultural factors as contextually relevant in the choice of a script. The rest of the talk is devoted to examining the adaptability of a Brahmi-derived script such as the Devanagari script to the sound systems of the under-resourced languages. I discuss the main aspects of the Devanagari script before looking at the various features of consonantal, vocalic and tonal systems found in Indic languages. I try to point out the efficacies as well as some difficulties in adapting a Brahmi-derived script, based on a

general survey (Pandey, P. 2014, *Sounds and their patterns in Indic languages*, 2 volumes, CUP India). I finally look at some of the received scripts for under-resourced languages in the light of the preceding discussion.

Towards a linguistic prehistory of eastern-central South Asia (and beyond).

John Peterson

Kiel University

In my talk I will summarize initial research into the distribution of morphosyntactic patterns in the languages of South Asia from three different families, namely Austro-Asiatic (Munda), Indo-Aryan and Dravidian, above all in eastern-central South Asia, in an initial attempt to unravel some of the linguistic prehistory of the subcontinent. To this end a small, preliminary morphosyntactic database has been compiled on 29 languages from throughout South Asia based on data from published resources, original field work, as well as questionnaires sent out to researchers working on a number of languages of the region. As I will argue, a NetborNet analysis of the data (using SplitsTree4, Huson & Bryant, 2006, see Figure 1, next page) strongly suggests that there is a strong division within the Indo-Aryan group into eastern and western Indo-Aryan in terms of morpho-syntactic features: Whereas the western Indo-Aryan languages show no particularly close structural relations to any other language group, the eastern Indo-Aryan languages have converged

considerably towards the Munda languages of central and eastern India. On the basis of this finding, I tentatively argue that the present-day Munda languages represent the “residue” of a previous spread of this linguistic group throughout the eastern half of the Indo-Gangetic Plain and that the majority of these pre-Munda Austro-Asiatic speakers switched at an early date to the Indo-Aryan *lingue franche* of the time as speakers of Indo-Aryan languages were spreading further towards the east. As I argue, this wholesale language shift resulted in a clear Austro-Asiatic substrate in eastern Indo-Aryan, which remains visible to this day.

Although still very tentative, the data clearly show areal clusters among the languages of South Asia which allow us to delve deep into the linguistic prehistory of the subcontinent. However, as many of the smaller languages, especially those in the eastern half of the subcontinent, still await detailed linguistic documentation, the study above all illustrates the need for further serious documentation of the languages of the region and also shows how these data can help further our knowledge of the distant past of the subcontinent.

Language documentation in Linguistics and Technology

Dafydd Gibbon

Bielefeld University

The aim of my talk is to 'metadocument' relations between

language documentation in documentary linguistics and in the human language technologies (referred to here as 'language resources').

Documentation of spoken and written authentic texts and speech has a long history. The oldest known documentation is the Sanskrit grammar of Panini, about 2500 years ago, and texts in many ancient languages are much older. But of course this raises the question of what documentation is, and what distinguishes it from collections of texts and speech recordings on the one hand, and fully fledged linguistic descriptions and theoretical argumentation on the other.

Contemporary language documentation has two main sources: the first is traditional corpus linguistics (with roots at least as old as the 18th century) and in phonetic transcription and recording (also with a long history, but with modern methods dating back to the second half of the mid-19th century). The first major technology in language documentation is the printing of texts, of course, some 600 years old. The second major technology did not develop until about 130 years ago: sound recording. The third major technology is the use of computers for both text and speech.

Current procedures in language documentation are heavily dependent on these fairly young technologies, documentation techniques in linguistics and phonetics also derive directly from the speech and language technologies: word processors,

database management systems and the internet were developed independently of linguistics, and documentation infrastructure has been developed rapidly since the 1980s in NLP (Natural Language Processing) and in Speech Technology (for automatic speech recognition and speech synthesis). In fact, the most popular documentation software for speech, Praat, was developed by Paul Boersma in the physics and phonetics laboratory of Louis Pols in Amsterdam, which also produced speech technology experiments for the Russian space lab Mir. Some of the most active creators of language resources for many languages are the big IT companies working on internet search and machine translation.

In my talk I will review these developments, discuss the empirical grounding of language and speech documentation, the need for standards and an integrated approach to documentation, and the utility of automated techniques and machine learning. I will illustrate some of the core issues with reference to two areas in which I have been involved over many years: lexicography, with particular reference to a scale of abstraction between data and theory via documentation, and phonetics, with particular concern for the automation of documentation by automatic annotation (segmentation and transcription), and the automatic analysis of annotations.

The moral of my story: linguists and engineers will benefit from working together in creating 'language documentation' in linguistics and 'language resources' in human language technologies.

**Interdisciplinarity in language documentation:
Documenting language dynamics and multilingualism in
Lower Fungom, Cameroon**

Jeff Good

University of Buffalo

The Lower Fungom region of Northwest Cameroon is noteworthy for its exceptional linguistic diversity. Seven languages are spoken in its thirteen recognized villages, all within an area about the size of the city of Amsterdam. Moreover, its residents are highly multilingual, with adults speaking around five different languages on average. This situation raises the question: What factors have allowed Lower Fungom to develop its extreme linguistic diversity and its speakers to develop competence in so many languages?

This paper considers the ways in which the standard documentary linguistic toolkit has been augmented by an interdisciplinary approach to studying the region, allowing for the creation of a documentary record which covers both the synchronic features of the target languages and offers sufficient ethnographic and historical context to allow us to model its language dynamics and understand its speakers' multilingual lives. In addition to outlining key results of this interdisciplinary research, concrete recommendations are provided for linguists interested in engaging in similar kinds of work.

Beyond Documentation: Language Technology for Endangered Languages

Monojit Choudhury

Microsoft Research India

Natural Language and Speech technologies have rapidly progressed over the last several decades, thanks to development of massive computational infrastructure, sophisticated algorithms and large-scale and united efforts towards language/speech data collection. This allows an average user today to seamlessly interact with the computer or mobile phone through speech (think Microsoft's Cortana or Google's Voice search) or speak in English on one end of the phone while the user at the other end hears Spanish (think Skype translator). Unfortunately though, these magical technologies are available only for a handful of world's languages. There are hardly any available speech or text processing systems for most other languages, including Indian languages such as Telugu or Bengali with a very large native speaker population. The problem though is not due to limitation of technology, but because these technologies require a huge amount of resources – i.e., annotated speech and language corpora, which are typically costly, and effort and time intensive.

Therefore, one cannot even imagine to build large scale speech recognizers or translators for endangered languages using the current technologies. On the other hand, such

technologies would be tremendously useful not only for documentation of endangered languages, but also would directly help the native speakers to use their language on modern smart phones and the Internet. The aspirational value of the native language will increase and the communication barriers would vanish.

In this talk, I will elaborate on the aforementioned issues and show how data and technology across world's languages are extremely skewed. Then I will discuss a few recent attempts on building translation and speech synthesis systems from very little data. I will conclude with some ideas on how we can do medium size data collection for endangered languages and build various useful technologies.

Workshops

Direct palatography for endangered languages

Rajesh Khatiwada

Laboratoire de Phonétique et Phonologie
Sorbonne Nouvelle

Direct or static palatography is one of the oldest phonetic articulatory methods invented in 19th century. But the version we use today was elaborated and popularized by Ladefoged in late '50s. This method consists on collecting articulatory records; such as places of articulation or the tongue patterns during the production - about human speech sounds. Today, though researchers are using sophisticated and dynamic articulatory data capturing tools such as EPG (Electropalatography), EMA (Electromagnéométrie), Echography, or even MRI (Magnetic resonance imaging), we continue to use this method thank to its simple, portable and very economical aspect. It is very easy to learn and affordable for anyone who wants to collect phonetic articulatory data in his/her fieldwork or in laboratory. This is still the perfect tool for linguistic fieldwork for the students and researchers working on the language documentation of endangered languages all over the world.

In this workshop, I am going to demonstrate how to do static palatography: I am using the term "Static Palatography" to designate both palatography and linguography data collecting methods. In these both methods, we study the place of articulation and tongue contact strategies during the

consonant production. This method consists on painting the tongue or palate with a black substance, and taking pictures of the mark left by that substance (made of charcoal, chocolate powder and edible oil), on palate (palatography) or on tongue (linguography). We will equally learn to interpret palatogrammes and linguogrammes using the “dentition plan” proposed by Firth (1948) and “lingual plan” of Catford (1977).

Data organisation, data management, workflows and software tools for language documentation

Peter K. Austin

Linguistics Department, SOAS
University of London

This workshop explores good practices in language documentation with a focus on:

What is documentation data and how do we collect it?

How do we manage the data and metadata we are working with?

How do we analyse the materials and ensure they are transparent, multifunctional, robust and able to be repurposed?

What free software tools exist to support this work?

We will present concrete suggestions about how to answer these questions with examples from our own research on Australian Aboriginal and eastern Indonesian languages. The software tools to be discussed and demonstrated will include ELAN, Toolbox, FLEx, WeSay and SayMore.