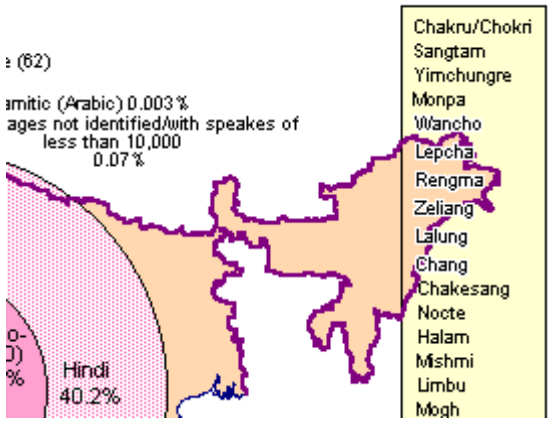
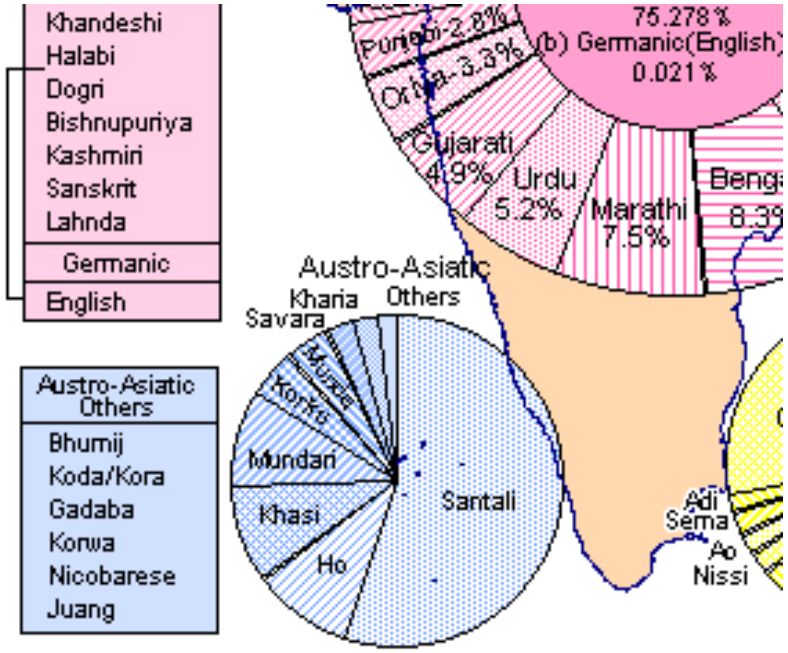


UGC's Centre for Endangered Languages *Project*

First Meeting of the Research Clusters



Presentation by

UDAYA NARAYANA SINGH

Chair, Centre for Endangered Languages &

Professor, Rabindra Bhavana

VISVA-BHARATI, SANTINIKETAN

E-mail: unsciil@yahoo.com

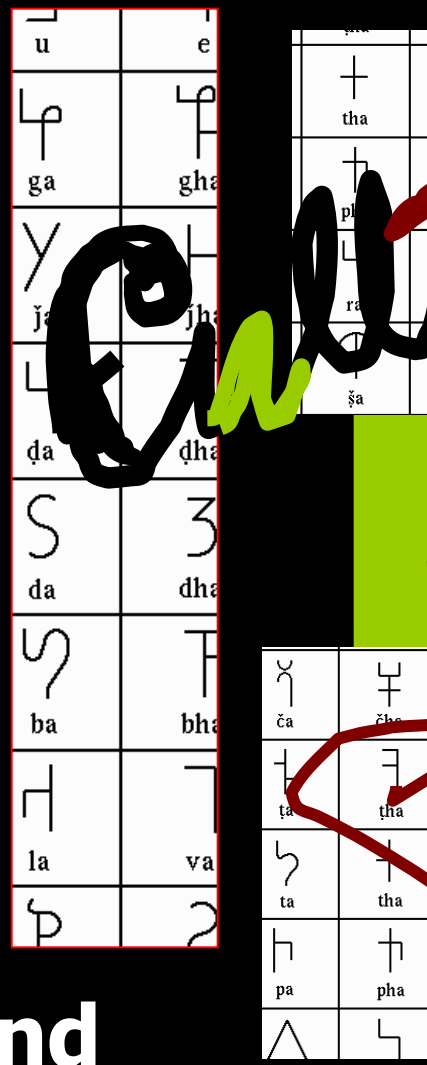
http://www.unesco.org/culture/languages-atlas/

Ground Realities & Task at Hand

- **We are all aware that several cultural traditions vanish every year.**
- **They leave behind only some traits of their once vibrant social life.**
- **We have also heard a number of scholars, politicians, newsmakers and ground-level administrators raising a lot of hue and cry over their death.**
- **This is considered a grave 'danger' literally [Cf. *The MacMillan Dictionary*] because a particular region or a society or community faces "a situation in which harm, death, damage or destruction is possible."**
- **Zepeda and Hill, 1991 had lamented saying that... "The loss of the hundreds of languages that have already passed into history is an intellectual catastrophe in every way comparable in magnitude to the ecological catastrophe we face today".**
- **Chistopher Moosely, 2007 says: "Language has always been a powerful weapon in the subjugation of peoples and nations. Empires have come and gone by the sword, but their true staying power, their lasting influence over many generations, long after the trappings of government and formal administration have disappeared, lies in the power of language." *[Encyclopedia of the World's Endangered Languages]***
- **Many think 'Language Attrition' is inevitable, given man's weakness for a more fashionable life – full of economic opportunities which their mother-tongues somehow fail to achieve for them. But we all here agree that it would be a great loss for humanity if they were to allow to disappear.**
- **So what is it that we can do?**
- **The CFEL Project of the UGC is an attempt to answer this question.**

There are a large number of smaller linguistic groups in India, and all of them need at least the following:

- Grammars for documentation, including Social grammars for registers & contexts
- Primers & Language games
- Graded teaching/learning materials to participate in elementary education
- Writing Systems reflecting their phonetics
- Literacy books for adult learners
- Dictionaries (general purpose)
- Thesauri or WordNet linking up synonymy
- Specialized Glossary for domains & knowledge translation
- Cultural & visual documentation
- Style Manuals
- Encouragements for literary activities



Includes both endangered and potentially endangered

What is the Big Picture like?

- ❑ 1576 rationalized mother-tongues (MTs) & 1796 other MTs
- ❑ 114 languages with 10,000 plus speakers;
- ❑ Variation : Hindi with 337 million to Maram (Manipur): 10,144;
- ❑ 22 Constitutional languages :
- ❑ Large non-scheduled lgs Bhili with 5.57 million speakers;
- ❑ Many minor & minority languages seem to be facing a threat.

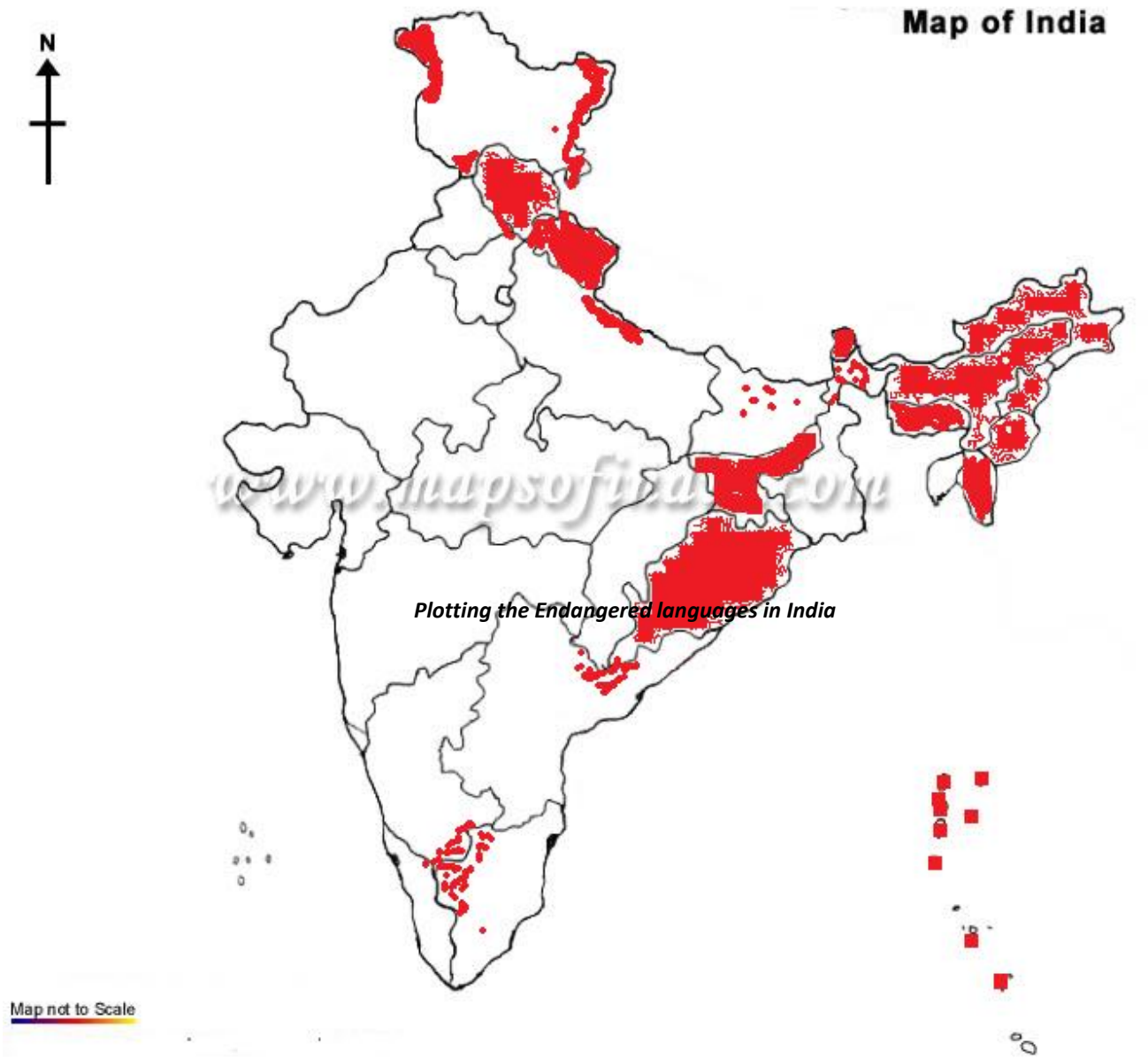
Gut.	DEV	GUJ	PUN	BEN	ORI	TEL	KAN	TAM	MAL	SINH	URD	SIND
k	क	ક	ਕ	ক	କ	క	ಕ	க	ക	ක	ک	ڪ
kh	ख	ખ	ਖ	খ	ଖ	ఖ	ಖ	–	ഖ	ඛ	کھ	ڪھ
g	ग	ગ	ਗ	গ	ଗ	గ	ಗ	–	ഗ	ග	گ	ڳ

- – 146 speech varieties used in radio network now, although before 1939 it used only English, Hindustani and Bangla after which Telugu, Tamil, Marathi and Gujarati were added
- – 47+ languages used in schools (7th Edu Survey, NCERT, 2006)
- – 3954 newspapers in 35 languages as in 1971 (It doubled in '03, with Hindi (2507), Urdu (534), English (407), Marathi & Tamil (395 each).
- – 58 languages with dwindling number of speakers;
- – Highest literary prizes are awarded in 24 languages
- – 96% speak only 20-odd IA & Dravidian languages
- – 14 major writing systems in use but 66 scripts in all.

Language Endangerment & Relevance of the UGC Programme

- **Many sociolinguists (**Pandit 1976; Srivastava 1976**) had claimed that compared to others, for South Asian immigrants, language retention was more natural than language loss.**
- **But in reality, it is seen that the 2nd/3rd generation migrants adopt other tongues/regional languages & are assimilated.**
- **Yet nobody likes the loss of their language & cultural identity**
- **If we look at endangered languages map (cf. Singh 2011), we understand that perhaps certain cluster of pockets arise in the country, and a set of consortia might be best suited to tackle the task, although this is only based on UNESCO Atlas of Endangered Languages.**
- **Notice that we are yet to make independent assessment of language endangerment, and take a call. So, what is needed is a programmed action – and not just resources.**
- **But for that, a common methodology, questionnaire, tools for data gathering, format and archiving, correlating them with GIS database, and a training programme for field work and data handling is required, for which we have accepted the challenge National Resource Centre for this project.**
- **Let us look at the snapshots of endangerment in India and elsewhere :**

LANGUAGE ENDANGERMENT IN INDIA – AS PER UNESCO ATLAS

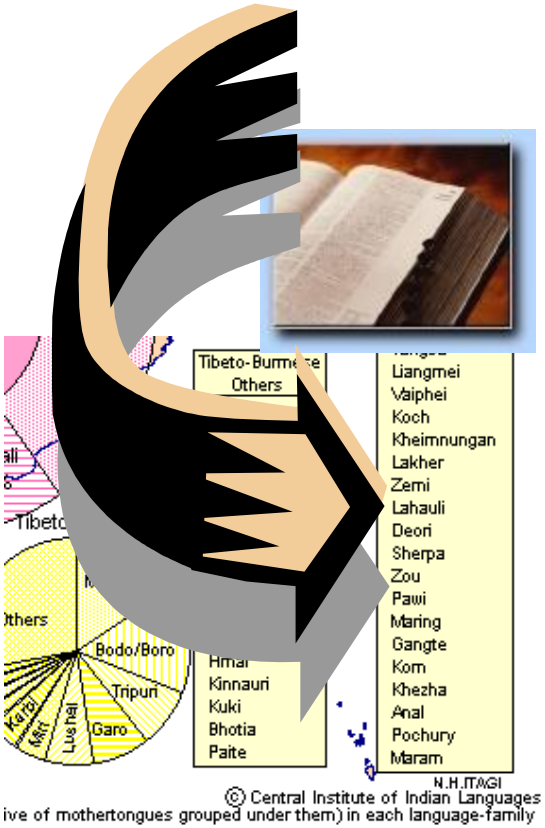


Two-fold Objectives:

- There are **two issues** of importance here:
- The first one is of theoretical importance and also a challenge for all practitioners of language management, namely, **how could we develop smaller languages and their culture in a diverse space - a space where number and economic development seem to be intertwined & important?**
- Secondly, given the profile of such smaller and lesser-known (and often, least cared for) languages and culture of India, **can language technology help in identifying them, changing their status and plight?**

NEWS ITEM: Professor uses Language Technology to **preserve** tribal languages

http://uanews.org/cgi-bin/WebObjects/UANews.woa/wa/MainStoryDetails?ArticleID=9400



- On July 6, 2004, an internet news item popped up in Google:
- Prof **Susan Penfield** with **CRIT** (Colorado River Indian Tribe) and UofA support, and with funding from **Bill & Melinda Gates Foundation**, worked to preserve two Amerindian languages.
- She created and trained tribal members of **Mohave** and **Chemehyevi** in the use of software & internet tools that would support preservation & instruction in these languages.
- For many smaller languages, such timely steps are important as Mohave now has 33 fully fluent speakers & they are of plus 70 age-group, and Chemehyevi has 10 speakers of 60 plus age.
- If we could train members of smaller/tribal communities in joining arms uprising, it should be easier to train them to preserve their culture.

The moral: Lang. technology could be used for threatened languages.

Why Methodology Workshop? Or, why Correlation is important? What Theory?

- **Whenever questions on smaller linguistic groups are raised, I've heard South Asianists raise this question, almost with a vengeance – "Are they dialects or languages?" – meaning – if they are dialects, why worry about them? If you persist, the next predictable question would be: Do they have a script? Even if we get past this question, the next one is: Are they taught in schools?**
- **They look like ever shifting stand of the 'Higgle Piggle Di' (Sukumar Ray: 'Ha-ja-ba-ra-la' where – when confronted, a cat changes its name all the time).**
- **All these issues must be buried now, and one should move ahead. We need a solid theoretical foundation for doing that & we will provide, just as a book on this effort or the viewpoint "from below" – as it were – will also be published.**
- **To start with, a series of methodology workshops would be needed, which should also produce a set of handbooks and field manuals. While the existing texts like Bernard Comrie & Norval Smith's *Lingua Descriptive Studies : Questionnaire* (North-Holland, 1977) or Anvita Abbi's Lincon Europa Book - *A Manual of Linguistic Field Work and Indian Language Structures* (2001) are good texts to begin with, they were not created to correlate linguistic data with language atlases.**
- **What we also need to do is to identify potential leaders from among these communities, train them & involve them in development.**
- **But most importantly, we need a secured Geo-tagged GIS-based Data Collection method with documented data upload on a common format along with data cleaning, transcribing, tagging and corpus-building ready for grammatical analysis and lexicon-building.**

Required – Documentation and Planned Interventions

- **Through UGC's Centers of Endangered Languages, we need to determine the real and potentially endangerment of languages.**
- **For a planned intervention to alter the scenario – what is badly needed is a GIS-based as well as statistics based linguistic mapping of the total scenario – an activity that had begun at CIIL after the 1961 Census but which sadly did not progress much beyond its Literacy Atlas projects.**
- **Although the *Mother-Tongue Survey* of the Census Language Division has covered some states now, they have not correlated Language and Space themselves but based their format and methodology only on Grierson's century-old methodology of word-list and sentence list.**

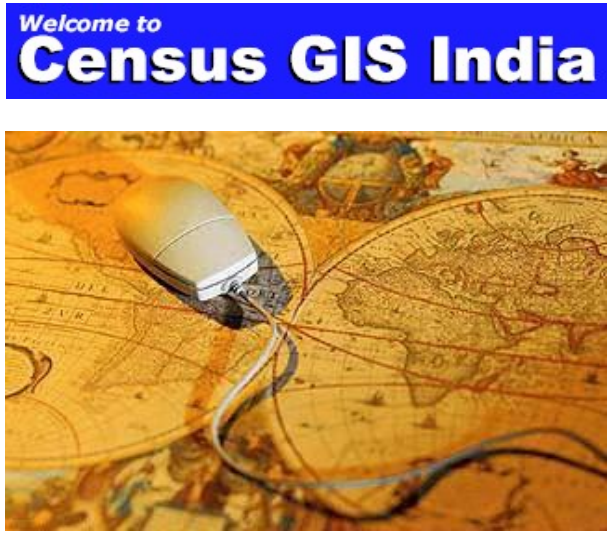
The modern-day technology of Linguistic Geography based on GIS activity undertaken by the MHA for its Census on the web operation should be suitably modified and used here.
The exercise assumes greater importance because language development is related also to the socio-economic development of the 'marginalized' speech groups.

Let's consider figures based on ASI's People of India & other surveys

	STATES	Major lang.	Minor 1 (%)	Minor 2 (%)	Others	Labels
A.	Kerala	96.6	2.1	0.3	1.0	Malayalam
	Punjab	92.2	7.3	0.1	0.4	Punjabi (Hindi, Urdu)
	Gujarat	91.5	2.9	1.7	3.9	Gujarati (Hindi, Sindhi)
	Haryana	91.0	7.1	1.6	0.3	Hindi (Punjabi, Urdu)
	U.P.	90.1	9.0	0.5	0.4	Hindi (Urdu, Punjabi)
	Rajasthan	89.6	5.0	2.2	3.2	Hindi (Bhili, Urdu)
	H.P.	88.9	6.3	1.2	3.6	Hindi (Punjabi, Kinnauri)
	Tamil Nadu	86.7	7.1	2.2	4.0	Tamil (Telugu, Kann.)
	West Bengal	86.0	6.6	2.1	5.7	Bengali (Hindi, Urdu)
	A.P.	84.8	8.4	2.8	4.0	Telugu (Urdu, Hindi)
	M.P.	85.6	3.3	2.2	8.9	Hindi (Bhili, Gondi)
	Bihar	80.9	9.9	2.9	6.3	Hindi (Urdu, Santali)
	Orissa	82.8	2.4	1.6	13.2	Oriya (Hindi, Telugu)
	Mizoram	75.1	8.6	3.3	13.0	Lushai (Beng, Lakher)
	Maharashtra	73.3	7.8	7.4	11.5	Marathi (Hindi,Urdu)
C.	Goa	51.5	33.4	4.6	10.5	Konkani (Marathi, Kan)
	Meghalaya	49.5	30.9	8.1	11.5	Khasi (Garo, Bengali)
	Tripura	68.9	23.5	1.7	5.9	Bengali (Tripuri, Hindi)
	Karnataka	66.2	10.0	7.4	16.4	Kannada (Urdu, Telugu)
D.	Sikkim	63.1	8.0	7.3	21.6	Nepali (Bhotia, Lepcha)
	Manipur	60.4	5.6	5.4	29.6	Manipuri (Thadou, Tangkhul)
	Assam	57.8	11.3	5.3	25.6	Assamese (Beng, Boro)
E.	Arunachal	19.9	9.4	8.2	62.5	Nissi (Nepali,Bengali)
	Nagaland	14.0	12.6	11.4	52.0	Ao (Sema, Konyak)

Mapping multilingualism

- **Even when we look beyond 1951, based on other reports and studies – such as the *People of India* volumes of the Anthropological Survey of India (ASI), we find only very broad parameters; namely, that the number of speakers of minority languages vary from state to state; e.g. in Tripura, over 31% speak minority languages, but in Kerala the figure is only 3.4%, etc. Or, Nagaland and Arunachal Pradesh do not have a majority languages as such (the biggest groups being 14.4% and 19.9%, respectively) .**
- **7 states - Kerala, Punjab, Gujarat, Haryana, A.P., U.P., H.P., Rajasthan, TN & WB have negligible minor speech groups, with 85% speaking a major language. However, considering India's population size, even 15% is a huge number. Are these communities safe or are they threatened?**
- **Further, we know that many minor language groups usually figures elsewhere as a major language but we do not have their details or there are no ways to monitor them.**
- **What is badly needed is a step to map our bilingualism and distribution of languages and speech varieties based on Census 2001 & 2011 reports extensively through Linguistic landscaping work– in clusters like Coastal languages, Central Tribal languages, Western Himalayan tongues, North-eastern languages etc.**



Besides Linguistic Cartography, Where and how to use LT? Some tips

- **Creation of school texts, using 'shell-book' method: Papua New Guinea**
- **Generation of a computational orthography that does justice to the phonetic/phonological nature of the given language, with UNICODE link.**
- **Building up of large and annotated corpora with BIS tagging tools**
- **Appropriate visual and audio documentation.**
- **Setting up of techniques of glossary formation based on such data, and automatic up-dation of the Lexical Resource when more data are added.**
- **Linking it up parallel lexicon of Hindi/English.**
- **Creation of Pictorial glossaries and addition of Cultural material.**
- **Building a bridge material with Web/CD-based or Radio/TV courses.**
- **Pilot Studies, including digital and photo documentation of a few states could be a way to begin.**
- **A model National Archive could be created in the following manner.**

Draft Web-page

c-fel *homepage* Coordinated by Visva-Bharati , Santiniketan for *UGC*

CENTRE FOR ENDANGERED LANGUAGES –
An Initiative of The University Grants Commission

- Aims & Objectives
- About the Activities
- About the Consortia
- UNESCO Report
- Theoretical Foundation
- Presentations
- Recent Reports
- News & Views Clips
- Community Comments
- Grierson's LSI
- Concept of NLSI
- MTSI Project of RGI
- SPPEL Project of CIIL
- PLSI Project
- Other Initiatives
- Central Universities
- State Universities
- Teaching of Linguistics
- Courses on the Web
- Other Initiatives
  - Ministry of Culture (Anthropological Survey, ASI, National Archives, National Museum, Sahitya Akadami, IGNCA)
  - Ministry of Tribal Affairs
  - Ministry of Communications & IT
  - Ministry of Rural Development
  - Min of Social Justice & Empowerment
  - Ministry of Home Affairs (Comms of India & Comm for Linguistic Minorities)
- About Us

India – Languages, Dialects and Mother-tongues

Visual & Oral Texts

News & Announcements

UGC Phase II Univ

Archives

- Phonetic Archives
- Scripts & Fonts
- Word-Book
- Lexical Coinages
- Set Expressions
- Human Resources
- Experts & Analysts
- Bibliographic Archive
- Image archive
  - *Still* *Moving Images*

Endangered Languages GIS

Crossroads - Debates Fora

FOLKLORE

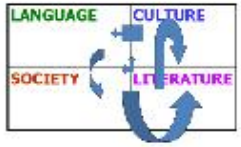
GRAMMARS | LEXICON | ANALYTICS | ETHNOGRAPHY | TYPOLOGY

- * Ontology & Schema
- * Electronic Tools
- * Field Linguistics
- * Data storage &Tables
- * Student Projects
- * Doctoral & Post-Doctoral
- * Graphs & Diagrams
- * Transformation of legacy data into compatible frame

Extracts : People of India

LINGUISTIC GEOGRAPHY

Multilingualism in India
Plurality Square



Surveys on Indian Multilingualism

Foundation for Endangered Languages | NEWS ON THE WEB | LINKS TO OTHER PROJECTS | SUBMISSIONS

Site maintained by Computer Centre, Visva-Bharati
Last Updated on May 14, 2014

SEARCH

BY LANGUAGE
BY AREA
BY FEATURES
BY TOPIC
BY REFERENCES

What Would the National Coordinating Centre do?

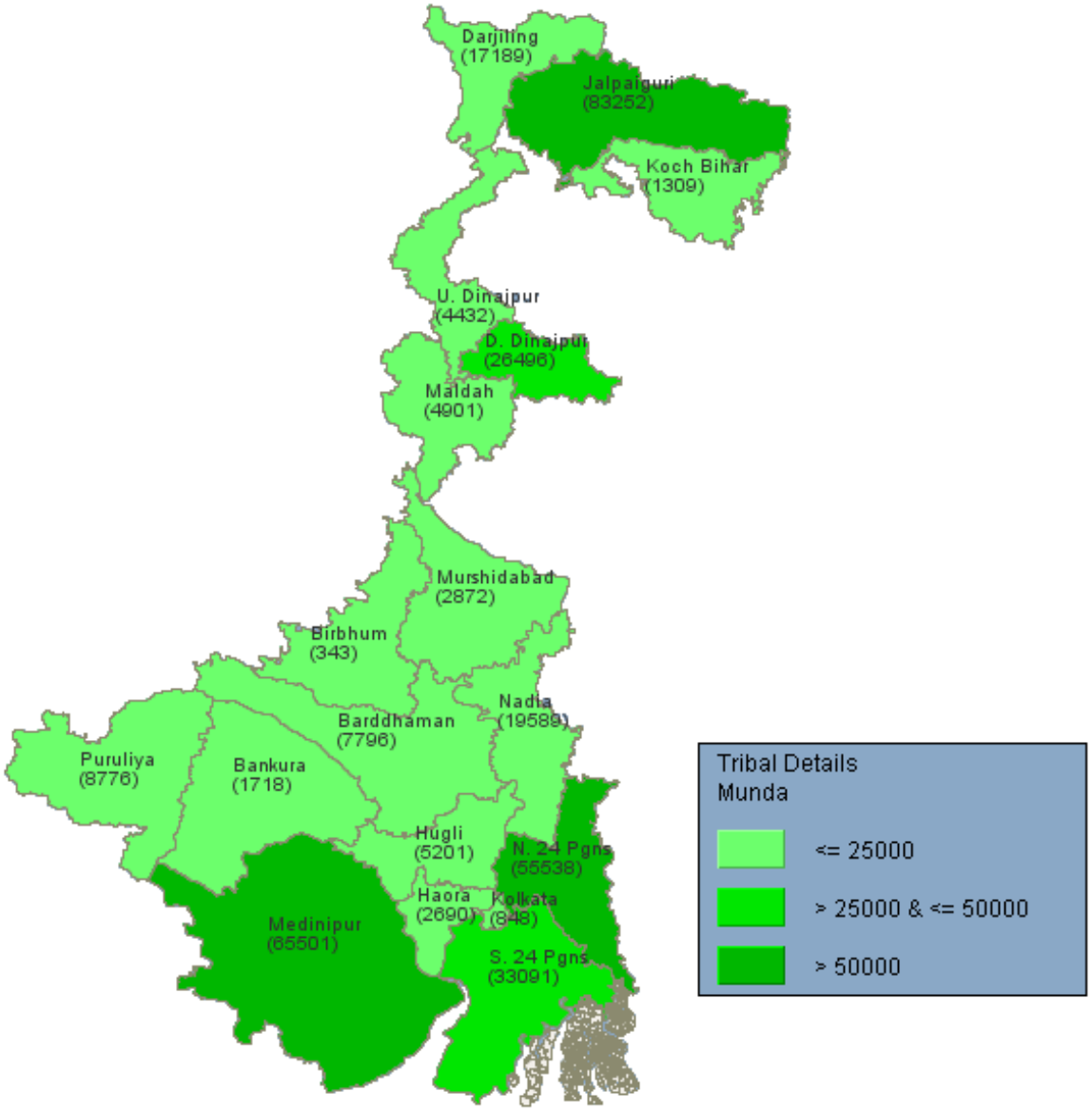
- **Visva-Bharati would set up of a designated server space (also on a Cloud server) for CFEL Project Consortia with a database architecture based on appropriate Database environment and GIS Software to receive Geo-tagged Data & information from remote locations wherever each university group would work.**
- **A Detailed mapping tribal languages & Mother-tongues (State-wise & District-wise, 2001) could be made available to all other Partners for a small cost. Mapping tribal languages & MTs district-wise would be done and upgraded once 2011 Census data is released.**
- **Designing correlatable (with socio-economic data) database with GIS Interface will be taken up.**
- **Preparing village level language atlas for the State of West Bengal could be used as a model for initial assessment of endangerment (Cost for each additional state to be met from other universities' budgets)**
- **Development of customized application for getting information from the villages in these states will be taken up.**
- **All Nine user universities (and not limited number of users) associated with the project would have secured access through Visva-Bharati portal to the huge databases that would get created under this UGC project. See the following for the basic structure:**
- **http://censusindia.gov.in/maps/**

Further Tasks

- **Help developing a geo-referenced map of each district of the focused states to capture the location where data enumerators would be sent**
- **Developing web interfaces for transcription of initial data sent through mobile telephony-server system**
- **Preparing spatial database of location vis-à-vis language based on the data sent from field, and linking the above data with spatial database**
- **Listing of villages (having significant ST population as per 2011) to be visited by enumerators in search of language endangerment in focused states**
- **Identification and help in procuring GPRS-enabled telephonic instruments and recording cards for data collection**
- **Training of enumerators in mobile telephony data capturing and data uploading from remote location. The geo-tagged data through mobile telephony (pre-loaded with specific application) will be transferred to the server(s) where a geo-referenced base map will receive the data.**
- **Help locating transcribers through crowdsourcing (cf: Jeff Howe in *Wired* 2006), and help in training of data transcribers in transcribing descriptive & language data, or in enhancing their skills**
- **Help/advise in training of informants.**
- **Training of master trainers and Technical staff in utilizing the system.**

How Will It Work?

At State level, distribution of different tribes in the districts of the state (which tribe in which district, how many?) would be of immense help (e.g. Say, 'Munda' tribe in West Bengal, how many in which district? Similarly, for all other tribes, there will be a customised GIS for planners. This tool will show the denominator for a tribe in a district.)

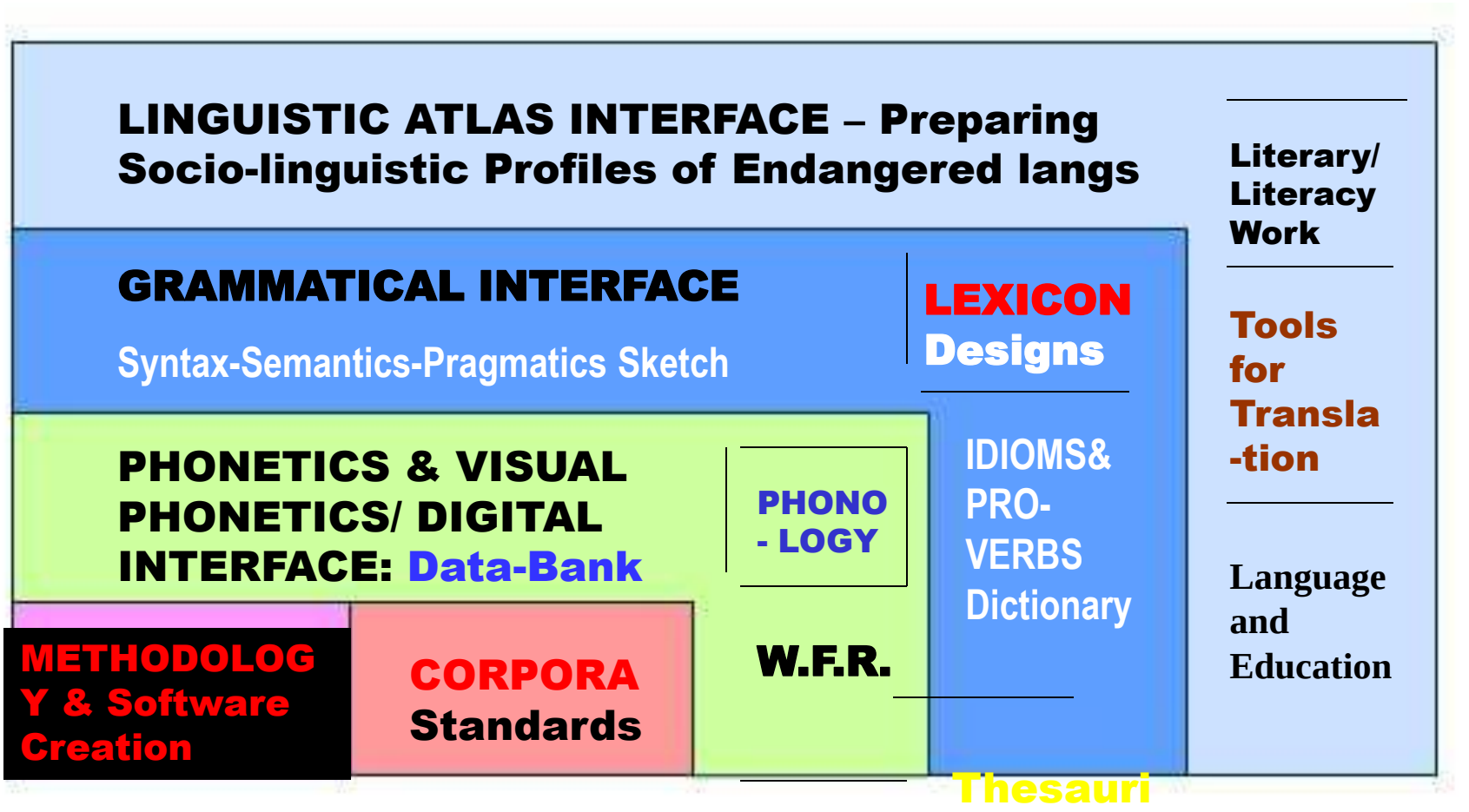


http://72.249.191.129/tribal/pages/map.jsp?strwidth=1366&strheight=670&mapid=19&jar=map.jar

CEFL-GIS: At the Planning Stage

- During planning stage, for each district, the system will provide a district map showing all villages & number of tribal population in each
- This system is for deployment and guiding the field enumerators who will visit every revenue village with tribal population and send information on 'population size of each tribe per village' (first stage)
- The enumerator will upload geo-tagged data thru' mobile telephony to the server where a geo-referenced base map will receive the data, and generate maps.
- May take a enumerator abt 12 to 18 months to complete a district
- During this time, identification of potential Informants will be done by the Initial Team to facilitate later for field visit by researchers.
- There may be districts where more than one enumerator will be required. On the other hand one enumerator may cover two districts where villages to be visited are less in number.
- The GIS system with each district under focus will be portrayed showing all revenue villages and STs in each.
- The system will act as receiver of the field geo-tagged voice file. These voice files will be transcribed for initial info and ported in RDBMS of the system.
- Training costs and schedules for each state enumerators will have to be worked out.

CFEL Grammar Architecture



At Visva-Bharati, we would concentrate on the Bottom-Left tasks in red colour and create/house a Data and Bibliographical archive format as well as training of survey man-power and the UGC's Archive will be enriched by other consortia that will concentrate on Survey and Analysis.

REFERENCES

- ***The MacMillan Dictionary*** **(http://www.macmillandictionary.com/thesaurus/british/)**
- **Singh, Udaya Narayana. 2009a. Rough notes. In Imtiaz Hasnain & Sreesh Chandra Choudhury, eds. *Problematizing Language Studies: Ramakant Festschrift.* Delhi:**
- **----- 2009b. Status of lesser-known languages of India. In Anju Saxena & lars Borin, eds. *Lesser-Known Languages in South Asia* : *Status and Policies, Case Studies and Applications of Information Technology*. Mouton de Gruyter.**
- **----- 2009c. The Sense of Danger: An Overview of Endangered Languages of India. In Kamalini Sengupta, eds. *Endangered Languages of India*. 39-56. New Delhi: INTACH.**
- **----- 2012. Epilogue. In Leslie Farrell & Ram Giri (eds). *English Language Education in South Asia: From Policy to Pedagogy*. Cambridge University Press.**
- **Zepeda, O. and J.H. Hill, 1991. The Condition of Native American Languages in the United States. In R.H. Robins and E.M. Uhlenbeck, Robins (editors). *Endangered Languages.* Oxford: Berg Publishers**
- **http://uanews.org/cgi-bin/WebObjects/UANews.woa/wa/MainStoryDetails?ArticleID=9400**