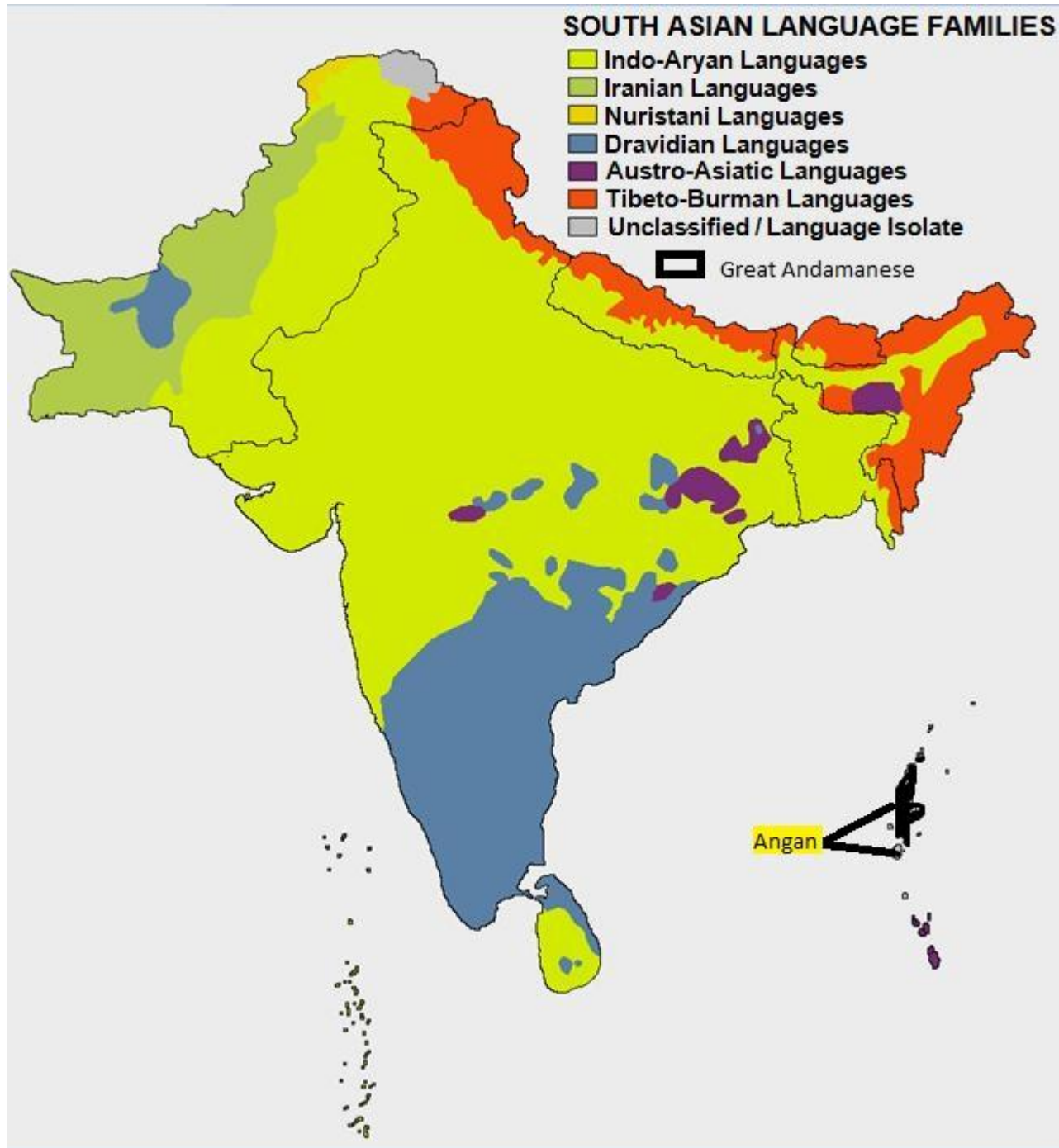Anvita Abbi 2012
anvitaabbi@gmail.com

LANGUAGES OF INDIA AND INDIA AS A LINGUISTIC AREA

Anvita Abbi
Jawaharlal Nehru University, New Delhi

1. India represents six distinct language families spread over a large region and spoken by more than one billion speakers. Though the exact figure of number of languages is not very clear partly because of the fuzzy demarcation between 'language' and 'dialect' question and partly because of shifting language loyalty of the people. However, a rough estimate is that there are more than 1600 languages spoken in the present India. The government of India reports only 122 and recognizes as 'scheduled/official'[1] languages only 22 drawn largely from the Indo-Aryan and Dravidian language family stocks. Consult the Census 2001. A large number of languages are endangered (196 according to the UNESCO).

2. The inclusion or exclusion of languages from the VIII schedule does neither depend on the quantum of its speakers nor it is based on the ideology of fundamental rights. Unfortunately, it is not based on the principle of equality of opportunity or on the ideology of national integration or invasive assimilation. Had it been so a large number of languages from Austro-Asiatic and Tibeto-Burman branch would have found their place in the schedule and Sanskrit with its mere 2500 speakers [reported to have doubled by 1991] would have not.

---

[1] The government of India, in 1956, constituted The Eighth Schedule (ES) and the reorganization of Indian states on the basis of dominant regional languages. The articles 343 to 351 spell out the official language at the level of the Union (i.e. the Central Government and the national polity as a whole), the level of the states and at the level of the judiciary. Originally the ES had 14 languages in its fold.

SOUTH ASIAN LANGUAGE FAMILIES
- Indo-Aryan Languages
- Iranian Languages
- Nuristani Languages
- Dravidian Languages
- Austro-Asiatic Languages
- Tibeto-Burman Languages
- Unclassified / Language Isolate
- ☐ Great Andamanese

Angan

- *Map of Indian language families*

3. **Tribal languages**. It is extremely difficult to define what is a tribe or who is a tribal. The word tribe has stuck since the British rulers introduced it in 1872 to describe a few select communities in India. These tribes are "scheduled" as per Article 342 of the Constitution by the President and

the Parliament. The concept of tribe in India is an administrative, judicial and political concept, which is applied to sections of the population that are relatively isolated. Yet "tribe" in India is a significant reality, characterized by a distinct way of life, rather than by virtue of forming a constituent part of the hierarchical structure of society as in the rest of India. They are outside the 'caste', or '*jati*' system.

4. If India is known for its multiplicity and diversity of languages it is because of our tribal languages. The rate of bilingualism as well as the numerical strength of the distinct varieties of languages is highest among tribal population. Tibeto-Burman and Austro-Asiatic language family are represented by 100% tribal population.
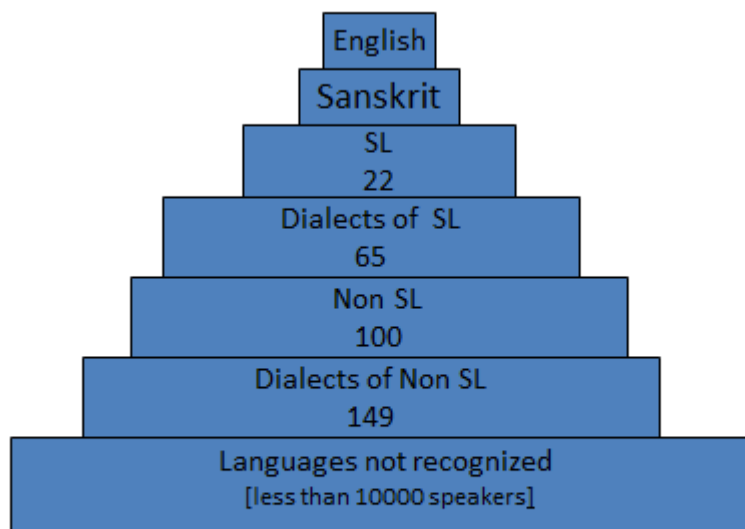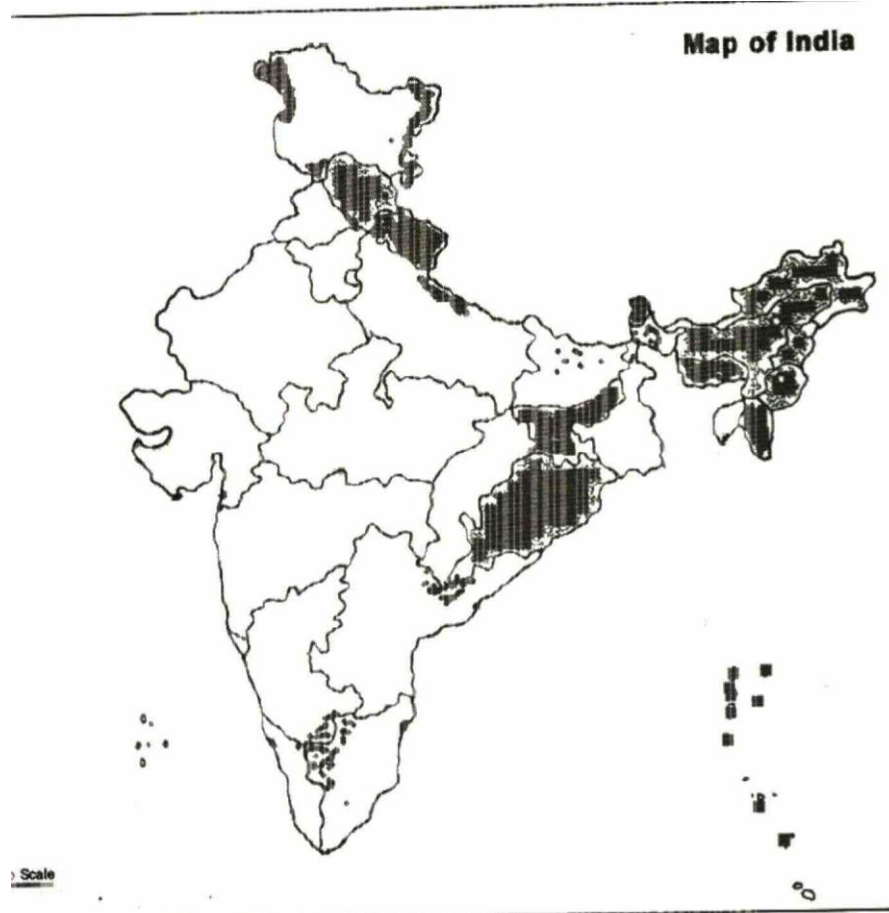

Fig.2. Hierarchical Status of Indian languages and dialects (336) 2001

5. Languages that are vanishing from the langscape of India are represented in the following map.

Plotting the Endangered languages in India

6. **Typologically distinct** languages add to the diversity scene. While Indo-Aryan is highly inflecting, Dravidian is both agglutinative and inflecting, Austro-Asiatic language is highly polysynthetic and incorporating, and Tibeto-Burman is analytic. Great Andamanese is agglutinative and the only head-marking language. The Austronesian languages consisting of Onge and Jarawa are also agglutinative in nature. In the present scenario of linguistic diffusion such compartments can no longer be ascribed to.

7. **Linguistics Area** may be defined as a geographically contiguous area, which is characterized by the existence of common linguistic features shared by genetically non-related language. Hence a Linguistic Area is marked by the "convergence" of linguistic features of various languages spoken in a particular region regardless of the fact that these languages may belong to different families

or subfamilies. For example, India represents a classic example of linguistic area as the languages of the mainland India belonging to four different language families i.e. Indo Aryan, Dravidian, Austro-Asiatic and Tibeto-Burman share several linguistic traits among themselves. The phenomenon of Linguistic Area is also referred to by the terms "Sprachbunde" "diffusion area", "affinité linguistique", "adstratum" "multifamilial convergence (or diffusion) area" (as opposed to multilingual" or "multidialectal convergence area") etc.

6.1 The term Linguistic Area was made popular by M.B. Emeneau (1956) who, ironically, defined with little hesitance[2], the concept of Linguistic Area in a footnote as: "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to other members of (at least) one of the families". He proposed this term till ingenious linguists found a better term. Interestingly, not only that the term became very popular, no one at present gives a second thought to changing it. Moreover, the reverse term that Emeneau thought was not viable to propose, has acquired a dignified, proper, and independent discipline of study namely the "Areal Linguistics".

6.2 The role of **Areal linguistics** is to explain what is not explainable by "genetic" historical linguistics. It is not a substitute but complementary to historical linguistics. It is the "diffusion" that constitutes the body of the study to explain linguistic similarities across distinct language families.

6.3 Indian linguistic area is characterized by common linguistic traits such as retroflex sounds, SOV word order, absence of prepositions, morphological reduplication (expressives), echo formations, reduplicated verbal adverbs,

---

[2] "It is perhaps not quite satisfactory as a technical term, though it has the virtue of having been used previously in the sense by H.V. Valten as a translation of Trubetzkoy's 'Sprachbund' *Pacific Northwest Quarterly* 34, 271-92 (1943).….Among the disadvantages of the term is the lack of an adjective and the impossibility of using the reverse phrase 'areal linguistics', since this is preempted by the Italian neo-linguistic school in another sense. Perhaps, however, it will do for the moment, until some more ingenious scholar invents better terminology" 1956, 1980: 124 fn 28.

explicator compound verbs, use of converbs, oblique marked subjects, morphological causatives among many others.

Phonological

  It is rare to find one common phonological feature across the language families of India. However, Retroflex and voiceless aspirated sounds are widely spread even in those languages that were isolated for thousands of years, e.g. Andamanese.

Great Andamanese

ʈʰu 1SG' : *tʰu* 'born'

Hindi

məʈʰa 'butter milk': *mətʰa* 'churned'

A large number of languages, especially in the Central and Northern part of the country have distinctive voiceless aspirates. Some of them have voiced counterparts as well. The other feature is the presence of syllable-timed than stressed timed words. Central vowel [ə] is also widely present in Indian languages.

Morphological

1. Morphologically derived causatives (Masica 1976). Almost all languages offer morphologically derivable pairs of transitive and causative verbs. Some IA languages have double causatives.

Hindi

pina	'drink'[TR]	*pI-l-ana* 'drink[CAUS1]'	*pI-l-wana* 'drink[CAUS2]'
sona	'sleep'[INTR]	*su-l-ana* 'sleep[TR]'	*su-l-wana* 'sleep[CAUS]'

2. **Expressives.** A kind of Reduplication where iterated syllables are generally onomatopoeic in nature (Diffloth 1976, 1988) and derives its status of a word/lexeme only after it is duplicated as the non -reduplicated syllable does not exist as a word. Languages of the Northeast India abound in expressive morphology for indicating 'manner' of an action. All Indian languages without any exception use expressive morphology to derive modifiers of 'manner', as finite verbs and with CP or a converb construction (Abbi 1991, 1997). Consider:

Anvita Abbi 2012
anvitaabbi@gmail.com

Bangla

1.   *śe*   *huɽmuɽ*   *korch-e*

   3sg   EXPR hurried   do PRS-3sg

   's/he is hurrying up'.

Angika

2.   *matha*   *hən hən-awe-che/ hən hən*   *kər-eche*

   head   expressive-3sg PRS/ expressive   do 3sg PRS

   'Head is aching with heavy and pulsating sense'

Tamil

   taʈataʈa-ttal   'to walk unsteadily with age, to be loose'

   taɭutaɭu-ttal   'to falter or stammer from ecstatic joy, love or other emotion'

3. **Echo Formations.** An EF is a concatenation of a base word followed [in rare cases preceded also] by an echo word. It is partially repeated form of the base word, such that the initial sound [Vowel/Consonant] or the initial syllable of the base word is replaced by another sound or another syllable. The echo word does not change the canonical shape of the base word (Abbi 1991). A repetition of a part of a lexical item carrying a semantic modification, e.g. Hindi $g^h ər > g^h ər\ vər$ 'house etc'. (partial reduplication); *puli-gili* 'tiger etc.' are examples of echo formations where the second morpheme has rigid replacer phoneme/lexeme unique to each language.

4. **Word Reduplication.** This refers to a complete or partial repetition of a word/lexeme. Complete word reduplication is constituted of two identical (bimodal) words. That is, both form and meaning are repeated once. The combined meaning give various modified meanings (Abbi 1991, 1994). The repetition of the entire lexical item, e.g. Hindi $g^h ər$ 'house' $> g^h ər\ g^h ər$ 'house house' 'each and every house' (complete reduplication), or Kharia *bor* 'to ask' $> bor\ bor$ 'begging'.

Syntax

1. **The SOV order** and implicational ordering elements exist in all Indian languages. Exceptions are Khasi, Nicobarese (Austro-Asiatic) and Kashmiri (Indo-Aryan), which maintain SVO. Khasi is a proto typical example of SVO ordering.

Hindi

> Ram-ne   sohan ko   kitab   di
>
> Ram-ERG   sohan-ACC   book   give.FEM.PST
>
> 'Ram gave a book to Sohan'

The sentence above shows postpositions such as *ne* and *ko* attached to nouns, a typical characteristic feature of SOV language.

2. **Explicator compound verb:** A sequence of at least two verbs V1 and V2 where the first is the main or predicating verb and the second member, although, homophonous with an independent verb in the language, does not appear in its primary lexical meaning; V2 only occurs in the sequence to mark the main verb for certain 'grammatical' features. Thus an ECV designates two [or more as in Dravidian] verbs 'acting' as one verb. Schema 2 illustrates the type of verb sequences in compound verb formation that we are going to discuss here. A cross–linguistic comparison of various explicators in South Asian languages reveal that any action or event can be modified by twelve different types of 'manner' (Abbi.1991/1992, Abbi 1992)

Hindi

> 1. sundər   nazara   dekh- kər   vo   ga   **uʈh-i**
>
>   beautiful scene   see-CP   she   sing   ROSE-3FSG
>
>   'Seeing the beautiful scene she sang [suddenly]'.

Marathi

> 2   ammi bas-un   **ghetle**
>
>   we   sit-PP   take
>
>   'we sat down' [assertively], [hurriedly], [deliberately]

Meithei

> 3.   cət   **siŋ**   khre

go substitute/copy pst
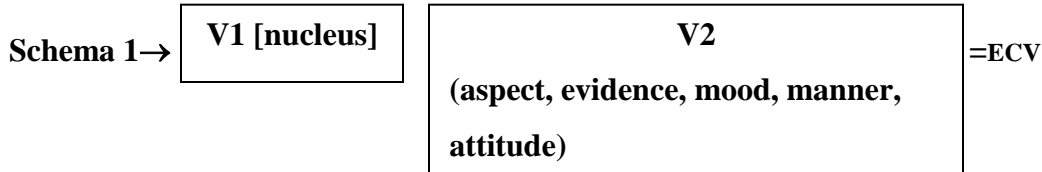
'He went [abruptly]'.

Explicator Compound Verbs

Schema 1→ | **V1 [nucleus]** | **V2** **(aspect, evidence, mood, manner, attitude)** | =**ECV**

Table 1   **Defining Characteristics of Explicators**

Morphological	*Syntactic*	*Semantic*	*Functional*	*Distributional*
Finite, marks TAM and other agreement features	Occupies the final position in SOV language	Delexialized, homophonous with an independent verb in the language	Marks the main verb for aspect, mirative, evidence, attitude, manner, and other adverbial functions	Can be replaced by a simple verb

3. **Converbs/ Conjunctive participles**. The CP constructions are defined as those containing nonfinite verbal forms that "subjoin sentences (usually with the same subject) to the left of the main finite verb" (Masica 1976:108). Various terms such as "conjunctive participles", "adverbial participles", "indeclinable", "gerund", "absolutive", and "incomplete verbs " have been used in the literature according to the specific language characteristic features it could describe. The nonfinite verb can be constituted of either of the following morphological material.

1. Simple Verb root, e.g. *hə̃s kər* = verb + PERF CP = 'laughingly' (Hindi). *kare-du* = verb + PCP = 'having called' (Kannada); *ma:ɖ-utta:* = verb + IMPF CP = 'while doing, making' (Kannada); *gate hue* = verb + IMPF CP= 'while singing' (Hindi); *kɔr-e*= verb + PCP = 'having done' (Bangla); *in-ji* = verb + PCP = 'having said' (Gadaba).

2. Expressives, e.g. *ha ha kɔre* = expressive + PERF CP = 'laughing loudly' (Bangla).

3. Reduplicated verbal adverbs [simultaneous, continuative, iterative], e.g. *hə̃s həs kər* = reduplicated verbal adverb + PERF CP = 'laughingly, sweetly' (Hindi).

Some simple examples:

Hindi

1   *pitaji kʰana kʰa-**kər**   so gəye*

    father food   eat-CP   sleep GO 3MSG HON. PAST

    'Having taken his meals the father went off to sleep'

Kurux

2   *en   naləx **nənn-on-ki** cail kal-**on***

    1SG   work   do-FUT CP   walk go-FUT

    'I will leave after doing the work'

Kannada

3   *maɭe **band-u**   kere   tumbitu*

    rain come-PST-CP   tank fill-PST

    'The tank filled as a result of rain'   (Causal)   (Bhat 2000:197)

The simple verb form with CP is an established areal feature (Masica 1976) and so is the occurrence of the RVADV (Abbi 1991). These have developed over a long period of time due to areal diffusion. Considering the data drawn from all four-language families for two different strategies employed, i.e. explicators in the finite verb construction and the non-finite CP construction in adjunct position, it can be said that the areal diffusion across Dravidian and Indo-Aryan is very strong. The two language families also share a wide variety of the lexical stock in explicator functions[3].

Characteristic features of conjunctive participle constructions

Morphological form	*Morphological markings*	*Syntactic functions*

---

[3] One could reconstruct ECV for Proto Dravidian (Steever 1993, 1997: 26, Krishnamurti *forthcoming*) but not so for the OIA. But that is another matter and open to debate (see Hock 1984).

Nonfinite, Absolutive	+/- tensed, no GNP markings[4]	Subordiantion: conjunction or chaining; Adverbial modifiers: 'manner', 'cause', 'purpose', 'conditional'.

Non Agentive Subjects. Agents of non performative verbs are marked by dative/genitive or accusative in Indo-Aryan, Dravidian and Munda. Semantically the constructions are no-agentive in nature and thus imply an 'involuntary', 'experiencer', 'benefactor', or 'recipient' subjects.

Newari

ji-ta   casu   li-lə
I-DAT   itch   become
 'I am itching'

Meithei

əI   tumni   ge
I-NOM   sleep   AUX
 'I am sleepy'

Bangla

or	lɔjja	hocce
S/he-GEN	Shy	become

 'She is shy'

Availability of finite and non-finite verb forms for encoding 'manner'

	Explicators	**RVADV**	**Expressives**	**V + CP (CONVERBS0**
Indo-Aryan	High	High	Medium	High
Dravidian	High	High	Med-to-High	High
Tibeto-Burman	Medium	High	Very high	High
Mon-Khmer	Low	Low	Very high	High
Munda	Low	High	Medium	High

---

[4] Malto has personal endings on CP constructions (Steever 1997:379)

5. It is not necessary to have bundle of isoglosses to define an area as a linguistic area. Areal linguists like Emeneau (1965, 1980), Ramanujan and Masica (1969), Winter (1973), Masica (1976), Abbi (1985) and Abbi and Mishra (1988) have taken one particular linguistic trait as a diagnostic trait to identify a Linguistic Area. Thus a single areal isogloss may be considered the minimum defining feature. On the basis of single, isogloss or a bundle of isoglosses a Linguistic Area may be defined 'weak' or 'strong' (Campbell, Kaufman and Smith-Stark 1986:532).

6. Since Linguistic Areas are result of convergence of Linguistic features, it implies a simultaneous process of divergence of the languages. When a language 'A' becomes a like a language 'B' because of the influence of the mutual contact, it also stats deviating from the other genetically related languages of its stock. This is inevitable in the process of language change as well as in the emergence of linguistic area. The carriers of this bipolar pull are obviously the bilingual speakers of the language 'A'. It is the multilingualism which is responsible for the genesis of 'South Asia' or 'India as a linguistic area'. Such areas are actually <u>cultural areas</u> and should interest cultural anthropologists and ethnographers.

7. Diffusion across languages give rise to redundancies in grammars and many a time create complex structures. India as a linguistic area gives us robust reasons for writing a common or core grammar of many of the languages in contact. Some of the most interesting results of convergence have fed into the research of Genetics, where scientist get initial clue to look for mixing of population. A striking result of genomic studies had been that "geographical affinity plays a stronger role than does cultural affinity in determining genetic affinity" (Majumdar 1998: 108). The term genetic here is translated as gene flow in the family that indicates intermarriages and intense contact. Conversely, linguistic differences do account for genomic diversity, e.g. Austro-Asiatic tribals show higher genomic diversity than Dravidian tribals.

Pragmatic and Sociolinguistic Systems

a) Negative particle is used as a tag for confirmation

Hindi   tu   gʰər   ja   rahi   hai   nə

2SG   home go   PROG   AUX   TAG

'You are going home, isn't it?'

b) Three way distinction in 2nd person pronouns

Punjabi   tu   tusi   tusĩ

c) Marked +/- honorific feature (in pronoun/verb morphology)

Hindi   ap 2SG.HON; *dekʰiye* 'please see.2SG.HON'

Magahi Verb Forms

ham dekha-l-i	'I saw'	neutral object
ham (okraa) dekha-l-ai	'I saw'	3 Object - Honorific
ham (unkaa) dekha l-i-ain	'I saw'	3P object +Honorific
ham (toraa) dekha-l-i-au	'I saw'	2P object - Honorific
ham (tohraa) dekha-l-i-o	'I saw'	2P object +Honorific

d) Extensive kinship terminology (affinal and non-affinal in all languages.

- **Sub-linguistic Areas**

  Apart from areal features, there are also those features that identify a **micro-area.**

1 Nasalisation (Northern India) Hindi: *hɛ̃* '3plural'

2. Aspiration (Northern India) Hindi: *bʰaraət* 'India'

3. Gender agreement (Western India) *cəŋgi* [ADJ] *kuʈi* [N] 'good girl'

4. Right hand is 'eating hand' (Central India)

5. Quotative Verb 'say' as complementiser (Southern and Northeastern India)

6. Relative-correlative pronoun (Northern India). Hindi:

   jo   kitab aj   khəridi   *vo*   kəl   paʈʰ ̃ũga

   REL   book   today buy.PST.FEMSG   CORREL tomorrow   read.FUT.MSG

   'The book (I) bought today (that book) (I) will read tomorrow.'

'The book that I bought today will be read by me tomorrow.'

7. Most of the languages **are pro-drop** as subject pronoun, direct object pronoun and indirect object pronoun can be dropped from the sentence. Hindi example given above gaps subject pronoun and direct object noun. Consider the following:

Speaker A: *tum-ne   kitab   pəʈʰ   li?*

       2sg.erg   book   read   take

      'Have you read the book?'

Speaker B: *hã   ø   ø   pəʈʰ   li*

       Yes SUBJ OBJ   read   take.pst

      'Yes, I have read it.'

8. Classifiers (Eastern and Central India): Each language has a number of classifier depending upon the shape, size and other qualities inherent to the nominal item (Mahapatra 1997 in Abbi 1997). TB languages are very rich in classifiers.

Tai Khamti Classifiers

    hui   'oval or round shaped'

(1) *tinŋ   nəm   soŋ   hui*

    bucket water two   CL
    'Two buckets of water.'

(1) *meɲi   soŋ   hui*
    orange two   CL
    'Two oranges.'

(2) *mətuŋ   hui   lüŋ*
    ball   CL   one
    'one ball'.

koŋ   'cylindrical'

    (1) koŋ-kʰo   'neck', 'throat'

Malto (Dravidian)

tini jan maler	three+classifier+men	'three men'
tini maq gʰo:ʈo	three+classifier+horse	'three horses'
tini kaʈI ta:li	three+classifier+hair	'three strands of hair'
tini paɽa kaldi	three+classifier+bananas	'three bananas'
tini pula lawa	three+classifier+popcorn	'three popcorns'

Anvita Abbi 2012
anvitaabbi@gmail.com

Methodology

- Bundle of isoglosses mark an area distinct from the other area. This implies that the line which delimits the area has a bundle of F_1.......F_2 shared features. We heavily depend upon the methodology of 'dialectology'.

- Single feature can be mapped to identify a linguistic area provided its share the underlying system and not just the lexicon

- Single feature can be part of two or more linguistic areas. Thus, modifier reduplication is the areal feature of South Asia as well as Southeast Asia linguistic areas.

Results:

- Emergence of an area which is characterized by shared linguistic features among typologically different and genealogically distinct languages.

- Process of convergence implies simultaneous process of divergence of the participant languages in-contact.

- Semantic Unity in linguistic diversity.

References

Abbi, A. 2001. *A Manual of linguistic Fieldwork and Structure of Indian languages.* Lincom Europa.

Abbi, A. 1991/1992.   Edited. India As a Linguistic Area Revisited. *Language Sciences* 13:2. Pergamon Press

Abbi, A. 1991. Identity Crisis of Dative Subjects and Experiencer Nominals in Indian languages. *International Journal of Dravidian Linguistics.* 20.1: 1-50

Abbi, A.1991. *Reduplication in South Asian Languages. An Areal, Typological, and Historical Study.* Delhi, Allied publishers.

Abbi, A. 1994. *Semantic Universals in Indian Languages.* Indian Institute of Advanced Study, Simla.

Abbi, A.1999. What is Indian in Indian Languages? A Case for Grammaticalization of Explicator Verb GO. *PILC Journal of Dravidic studies.* 9:2, 105-120.

Abbi, A. and D. Gopalakrishnan.1992. The Explicator compound verbs : some Definitional issues and Criteria for Identification. *Indian Linguistics* .51.1: 27-46

Abbi, A. and Ahum Victor. Expressive Morphology and Manner Adverbs in Khasi, Tangkhul Naga and Kuki-Chin Languages. In Abbi edited *Languages of Tribal and Indigenous Peoples of India The Ethnic Space.* 1997. Motilal Benarsidass. Delhi.

Anvita Abbi 2012
anvitaabbi@gmail.com

Abbi, A. 1984. Conjunctive Participles of Hindi-Urdu. *International Journal of Dravidian Linguistics* . 13.2: 252-63.

Bh. Krishnamurti, Colin Masica and Anjani K. Sinha. 1986.   *South Asian Languages: Structure, Convergence, and Diglossia.* Motilal Benarsidass, Delhi.

Davison, Alice 1981.   Syntactic and Semantic Indeterminacy Resolved: A mostly Pragmatic Analysis for the Hindi Conjunctive Participle. In *Radical Pragmatics*. Edited by Peter Cole. Academic Press.

Emeneau, M.B. 1980. *Language and Linguistic Area.* Stanford CA: Stanford University Press.

Gumperz, J.J. & Wilson, R. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In Pidginization and creolization of languages, D. Hymes (ed.), 151–168. Cambridge: CUP.

Kachru, Yamuna 1981. On the Syntax, Semantics and Pragmatics of the Conjunctive Participle in Hindi-Urdu. *Studies in the Linguistic Sciences.* 2.2.

Kuiper, Franciscuḇ B. J. 1967. The Genesis of Linguistic Areas. *Indo-Iranian Journal*. 10: 81-102.

Masica, C.P. 1976. *Defining a Linguistic Area: South Asia.* Chicago CA: University of

Chicago Press

Masica, Colin P. 1993.   South Asian Languages. In William Bright edited *International*

Encyclopedia of Linguistics. 38-42. Oxford University Press.

Ramanujan, A. K. and Colin P. Masica 1969.   *Towards a Phonologicval Typology of the Indian Linguistic Area. Current Trends in Linguistics Vol.5* (543-577) ed. By T. Sebeok.
Shapiro, Michael C. and Harold F. Schiffman. 1981. *Language and Society in South Asia.* Motilal Benarsidass. Delhi.

Southworth, Franklin C. and Madhav L. Apte. 1974. Edited. Contact and Convergence in South Asian languages. *International Journal of Dravidian Linguistics.* 3: No.1-2.

Subbarao, K.V. 2012. *South Asian Languages: A Syntactic Typology.* Cambridge University Press.

Verma, M.K. and K.P. Mohanan. 1990 *Expereincer Subjects in South Asian languages.* The Stanford Linguistic Association, CSLI, Stanford.