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’ 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.
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.

5. Languages that are vanishing from the landscape of India are represented in the following map.
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 hesitation\(^2\), 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**

\[t^h_u \quad \text{1SG': } t^h_u 'born'\]

**Hindi**

\[m\dot{a}t^h_u 'butter milk': m\dot{a}t^h_u 'churned'\]

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 \([\alpha]\) 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**

\[\begin{align*}
\text{pina} & \quad 'drink'_{[TR]} & \text{pI-l-ana} & \quad 'drink_{[CAUS1]}' & \text{pI-l-wana} & \quad 'drink_{[CAUS2]}' \\
\text{sona} & \quad 'sleep'_{[INTR]} & \text{su-l-ana} & \quad 'sleep_{[TR]}' & \text{su-l-wana} & \quad 'sleep_{[CAUS]}'
\end{align*}\]

2. **Expressives.** A kind of Reduplication where iterated syllables are generally onomatopoetic 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:
1. Eo Formation. 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^bər^g^bər^v^bər ‘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^bər^g^bər^v^bər ‘house’ > g^bər^g^bər^v^bər ‘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**

\[ \text{Ram-ne sohan ko kitab di} \]

‘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 \( V_1 \) and \( V_2 \) 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; \( V_2 \) 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. \[ \text{sundar nazara dekh-kər vo ga } \text{uṭh-i} \]

‘Seeing the beautiful scene she sang [suddenly].’

**Marathi**

2. \[ \text{ammi bas-un ghete} \]

‘we sat down’ [assertively], [hurriedly], [deliberately]

**Meithei**

3. \[ \text{cət siŋ khre} \]
He went [abruptly].

**Explicator Compound Verbs**

<table>
<thead>
<tr>
<th>Schema 1</th>
<th>V1 [nucleus]</th>
<th>V2</th>
<th>=ECV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(aspect, evidence, mood, manner, attitude)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 **Defining Characteristics of Explicators**

<table>
<thead>
<tr>
<th>Morphological</th>
<th>Syntactic</th>
<th>Semantic</th>
<th>Functional</th>
<th>Distributional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite, marks TAM and other agreement features</td>
<td>Occupies the final position in SOV language</td>
<td>Delexicalized, homophonous with an independent verb in the language</td>
<td>Marks the main verb for aspect, mirative, evidence, attitude, manner, and other adverbial functions</td>
<td>Can be replaced by a simple verb</td>
</tr>
</tbody>
</table>

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. हौस कार = verb + PERF CP = ‘laughingly’ (Hindi).
   कार-दु = verb + PCP = ‘having called’ (Kannada); मास-उत्ता = verb + IMPF CP = ‘while doing, making’ (Kannada); गेट-हू = verb + IMPF CP = ‘while singing’ (Hindi); कृ-ए = verb + PCP = ‘having done’ (Bangla); इन-जी = verb + PCP = ‘having said’ (Gadaba).
2. Expressives, e.g. *ha ha kāre* = expressive + PERF CP = ‘laughing loudly’ (Bangla).

3. Reduplicated verbal adverbs [simultaneous, continuative, iterative], e.g. *hās hās kōr* = reduplicated verbal adverb + PERF CP = ‘laughingly, sweetly’ (Hindi).

Some simple examples:

**Hindi**

1. *pitaji kāna kā-ka so gāye*
   - father  food  eat-CP  sleep GO 3MSG HON. PAST
   - ‘Having taken his meals the father went off to sleep’

**Kurux**

2. *en nālax nān-on-ki cail kal-on*
   - 1SG  work  do-FUT CP  walk go-FUT
   - ‘I will leave after doing the work’

**Kannada**

3. *māle band-u kere tumitu*
   - 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**

<table>
<thead>
<tr>
<th>Morphological form</th>
<th>Morphological markings</th>
<th>Syntactic functions</th>
</tr>
</thead>
</table>

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

+/ tense, no GNP markings

Subordination: conjunction or chaining;

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 non-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 1əjjə hocce
S/he-GEN Shy become

‘She is shy’

Availability of finite and non-finite verb forms for encoding ‘manner’

<table>
<thead>
<tr>
<th>Language</th>
<th>Explicators</th>
<th>RVADV</th>
<th>Expressives</th>
<th>V + CP (CONVERBS0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indo-Aryan</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Dravidian</td>
<td>High</td>
<td>High</td>
<td>Med-to-High</td>
<td>High</td>
</tr>
<tr>
<td>Tibeto-Burman</td>
<td>Medium</td>
<td>High</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>Mon-Khmer</td>
<td>Low</td>
<td>Low</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>Munda</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

---

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 cultural areas 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

\[ tu \quad g^{b}r \quad ja \quad rahi \quad hai \quad n\r \]

\[ 2SG \quad home \quad go \quad PROG \quad AUX \quad TAG \]

‘You are going home, isn’t it?’

b) Three way distinction in 2\textsuperscript{nd} person pronouns

Punjabi

\[ tu \quad tusi \quad tusi \]

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

Hindi

\[ ap \quad 2SG.HON; \quad dek^{b}iye \quad ‘please see.2SG.HON’ \]

Magahi Verb Forms

ham dekha-l-i \quad ‘I saw’ \quad neutral object

ham (okraa) dekha-l-ai \quad ‘I saw’ \quad 3 Object - Honorific

ham (unkaa) dekha l-i-ain \quad ‘I saw’ \quad 3P object +Honorific

ham (toraa) dekha-l-i-aiu \quad ‘I saw’ \quad 2P object - Honorific

ham (tohraa) dekha-l-i-o \quad ‘I saw’ \quad 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: \( h\text{\textacuten} \) ‘3plural’

  2. Aspiration (Northern India) Hindi: \( b^{b}ara\text{\textacuten} \) ‘India’

  3. Gender agreement (Western India) \( c\eta gi \) [ADJ] \( ku\text{\textacuten}i \) [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:

\[ jo \quad kitab \quad aj \quad kh\text{\textacuten}r\text{\textacuten}i \quad vo \quad k\text{\textacuten}l \quad pa\text{\textacuten}r^{b}u\text{\textacuten}ga \]

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 pəṛ-li?
   2sg.erg book read take
   ‘Have you read the book?’

Speaker B: ḥā ो ो pəṛ-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) tīm nəm soŋ hui
   bucket water two CL
   ‘Two buckets of water.’

(1) meŋi soŋ hui
   orange two CL
   ‘Two oranges.’

(2) mətung hui lʊŋ
   ball CL one
   ‘one ball’.

*kʊŋ* ‘cylindrical’

(1) kʊŋ-kʊ ‘neck’, ‘throat’

**Malto (Dravidian)**

* tīn jən mər
  three+classifier+men ‘three men’

* tīn maq gʊ:ro
  three+classifier+horse ‘three horses’

* tīn kətə lə:li
  three+classifier+hair ‘three strands of hair’

* tīn pəɾə kələdi
  three+classifier+bananas ‘three bananas’

* tīn pula lawə
  three+classifier+popcorn ‘three popcorns’
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


