Chapter II

Language Families, Language Contact, And Areal Universals

Indian languages have always been a challenge to linguists for their diversity and multiplicity. In this chapter, we want to concentrate on the characteristic features of each language family of India and then briefly enumerate the areal features prevalent in these languages. We shall attempt to draw the attention of the readers to the omnipresent language contact situation due to heavy bilingualism that we discussed in the last chapter. The chapter ends with an exposition of how the contact situation has given rise to conflicting grammatical patterns and converged languages. You might ask why this in a field manual? Well, we believe that in order to do justice to the data that one collects in the field, and in order to understand why one sometimes finds aberrant patterns during fieldwork, one should be aware of the aspects of Indian languages that we discuss in the following pages. We are trying to prepare you for what you might expect in the field; although of course fieldwork always offers surprises, and perhaps that is the excitement one looks for in the study of linguistics.

2.1 Individual Language Features

India represents five language families, all very distinct from one another. These are:

1. Indo-Aryan
2. Dravidian
3. Austro-Asiatic
4. Tibeto-Burman
5. Andamanese

These five distinct language families are distinguished on a linguistic basis and each family has been studied extensively except the Andamanese. While Indo-Aryan languages are highly inflecting, Dravidian languages are agglutinative, Munda (a major sub-branch of Austro-Asiatic) is of the incorporating and agglutinative type and Tibeto-Burman is of the isolating and agglutinative type. One cannot, however, compartmentalise them in so rigidly.

Let us consider each of these language families in some detail. We shall first describe the geographical location and spread of languages belonging to each of these language families (see also map 1.2 given earlier in chapter 1). Having done that we shall consider the characteristic features of each language family under three major topics, i.e. (1) the sound system, (2) the lexical and morphological system, and (3) the syntactico-semantic system. We have tried to be brief and have restricted the discussion to the main defining features of each language family, rather than discussing in detail individual languages.
2.1.1 The Indo Aryan Languages

2.1.1.1 The Geographical Spread:

Indo-Aryan languages (IA) occupy the largest geographical area in the subcontinent. It stretches from Pakistan in the west, to Bangladesh in the east, from Nepal in the north and to Sri Lanka in the south. A chain of intelligibility can be established from the westernmost tip of India to the easternmost tips of Bangladesh, not, however, from the northern tip of Kashmir to the southernmost tip of Maharashtra. To the west of the IA belt lies the Iranian language family, its sister family from which it separated around 2000 BC. Aryan movement is generally believed to have taken place in successive waves between the period of 1700-2000 BC. Sinhalese, an IA language, dominates the region of Sri Lanka in the south of the Indian continent.

2.1.1.2 The Sound System:

The Indo-Aryan vowel system is as varied as the number of languages in the family. Masica (1991: 109-113) discusses the range from the basic five vowel system of Roma (found in Europe) and a six vowel system of Oriya /i, e, a, u, o, u/ to an “authentic thirteen vowel system” found in Sinhalese, i.e., /i, e, a, u, o, a, o, a, o, o, u, u/ Konkani seems to have the largest number of vowel contrast based on quantity and quality, i.e., fourteen in all. The defining feature of these languages is that the vowel contrast is maintained more on the level of quality than on quantity. Thus, Hindi and various other languages in the Hindi-speaking area (see appendix 14) distinguish vowels on the scale of height and fronting, not according to the length of the vowel. Thus it is common to find /i, i, U, e, e, e, o, o, a, a, o, u/ — a ten-vowel set with length accompanying high vowels interconsonantally in a syllable. However, languages adjacent to Dravidian languages (see below) have incorporated length as the distinguishing phenomenon. Another characteristic feature of modern IA languages is the prevalence of nasalised vowels. Languages such as Hindi and its dialects offer a complete parallel set of nasalised vowels, though the phonemic status of each of them is in dispute (Kaye 1997). Nasalisation in IA languages is predictable in some positions and thus is allophonic, whereas in some positions it is not and thus may be regarded as phonemic. For instance, the plural distinction in Hindi is maintained by nasalisation. Thus: he ‘is’ and he ‘are’ or ayegi ‘she will come’ [f3g] and nyegi ‘they will come’ [fpl].

Indo-Aryan consonant system is identifiable by the aspirated stops prevalent in all the languages. Most of the languages offer four-way contrast in stops, distinguishing voiceless and voiced unaspirated series from voiced and voiceless aspirated series, i.e. /p, pʰ, b, bʰ, t, tʰ, d, dʰ, th, q, qʰ, c, k, kʰ, g, gʰ/. One should, if possible, consult Ladefoged (2001: 123-25) and listen to the four-way contrast recorded in the accompanying CD. There are languages such as Marwari, Sindhi, Gujarati, Marathi, and Nepali that offer aspiration contrast in nasals, i.e., /m, mʰ, n, nʰ/. Some dialects of Hindi and Rajasthani offer aspiration contrast in laterals, as well: /l, lʰ/. For fieldworkers the feature ‘aspiration’ can serve as a strong feature
to identify a language as Indo-Aryan or as having been a language being influenced by it.

Combinations of vowel sounds within a syllable, i.e. dipthongs, are mostly being replaced by a combination of a vowel and a glide y or w, e.g. cay 'tea' in most of the IA languages.

2.1.1.3 The Lexical and Morphological system

The Indo-Aryan languages have derived their lexicon from three main sources. The foremost source is Sanskrit, and these lexical items are generally termed as tatsama words. These words are used in their original Sanskrit forms with morphological endings according to each language. The second source is traditionally known as tadbhava, i.e. the words which were originally derived from Sanskrit but have gone through phonological changes over a period of historical development via the period of Prakrit and Apabhramsa. Thus Hindi hat 'hand' is derived by various phonological rules (and is tadbhava) from Sanskrit hasta 'hand'. Then there are words called desaj 'local', the etymology of which cannot be traced to Sanskrit origins. The fourth type are visesaj, 'foreign' words that are borrowed from Persian, Arabic, French, Portuguese, and English. Although this is a traditional way of classifying IA languages, it proves useful to have this knowledge while conducting field work on a new language, as the rules for phonology and morphology operate differently in each case. For instance, see Ohala 1983 for the schwa deletion rule and its applicational constraints in Hindi.

As far as word formation is concerned, a large portion of the IA vocabulary is derived by expressive morphology and reduplication of words; especially those derived as echo formations. A detailed discussion on these can be found in chapter 5.

Gender in Indo-Aryan languages is both a syntactic and a morphological category. Thus, in languages such as Hindi and Punjabi gender assignment of a noun controls the agreement feature in the rest of the sentence, resulting in each modifier inflecting for the gender. According to Masica (1991: 219): "In the New Indo-Aryan languages that have it, gender is an inherent and classificatory property of one class of words (nouns) and a variable or inflectional property of others (adjectives, certain verbal forms, sometimes pronouns and adverbs, and one extremely important postposition)." On the other hand, in languages such as Bangla some nouns designating professions are classified as masculine and feminine without any syntactic implications. Gender in eastern Indo-Aryan languages has almost disappeared. Languages such as Konkani and Marathi have preserved the old Sanskrit distinction between masculine, feminine and neuter. The common pattern in the rest of the Indo-Aryan languages, however, is the two-way gender system where neuter merges with masculine and feminine.

The other major defining characteristic feature of IA morphology is that the languages in this family are highly inflecting in nature, leading to a large number of portmanteau morphs. Thus in the Hindi form jaalgi 'they [f] will go' ja- is the root, -

1 i.e. genitive such as in Hindi: ram ki jali 'Ram's cap [fsg]' but ram ka ya 'or
Ram's house [msg]'.
e- indicates 3rd person, masalisation indicates plural, -g- is future tense marker and -i- indicates feminine gender; in the word lāṛka ‘boys’ /lāṛkā/ is the root; -e indicates both number and case, [in this case plurality and nominative or oblique]. We will discuss this feature in detail when we discuss word formation processes (Chapter 5).

Agreement is an important feature of this language family, although not all IA languages show it in a uniform pattern. The languages that show agreement prominently are Western Indo-Aryan languages. The agreement is exhibited in a linear fashion, so that the gender and number of the modified noun is encoded in the modifiers and the finite verb, as for example in Punjabi: cangiyā kūṭiyā ‘good girls’, cōnge mūnde ‘good boys’. In Hindi however, while verbs are encoded with the gender and number information of the argument, modifiers encode only gender but not number if the controller is feminine. If the controller is masculine, the modifiers encode number in addition to gender, e.g. acc′e lāṛke ‘good boys’ but acc′i lāṛki/lāṛkiyā ‘good girl/girls’

Another feature which distinguishes languages of this family from others, is that nouns and pronouns change into the oblique form before taking any case marking. Thus, the Hindi word for boy lāṛka changes its form to lāṛke in the oblique form, as in lāṛke-ne ‘boy, agentive’. Pronouns of the first, second and third persons also change their form to the oblique case. Konkani offers the largest number of forms as the language has retained the old Sanskrit pattern of three genders. Consider a small sample of the agentive case (Table 2.1).

| Table 2.1 Konkani III Person Pronouns in Agentive Case |
|---------------------------------|------------------|------------------|
|       | **Singular** | **Plural** |
| Masculine | ta-nē | te-nē |
| Feminine | ti-nē | ta-nī |
| Neuter | ti-nē | ti-nē |

The second person non-honorific pronoun in Bangani tu changes to tāī in oblique case before taking any postposition. Indo-Aryan languages, as stated above, are highly inflecting. Interestingly, various discourse-related features, such as the status of the addressee and the addressee, are incorporated in the verbal morphology of some languages (see the example of Magahi given in Chapter 4 (table 4.3) and the example for Maithili given in chapter 7 (example 7.4)). A small set of 3rd person genitive pronouns is given below to draw attention to the fact that deictic elements are a rather a complex phenomenon to study in Indo-Aryan languages.

| Table 2.2 Maithili Possessive III Person Pronouns |
|-----------------------------------------------|-----------------|-----------------|
| (Adapted from Ramawat Yadv 1996:121) |
| **III person** | **HONORIFIC** | **NON-HONORIFIC** |
| **Singular** | REMOTE | PROXIMATE | REMOTE | PROXIMATE |
| hun-k-ar | hin-k-ar | ek-ar | hin-k-ar | ek-ar |
| **Plural** | hun-k-ar sah−ak | hin-k-ar sah−ak | ek-ar sah−ok | ek-ar-sah−ok |
A very significant Indo-Aryan feature is that languages derive transitive and causative forms by suffixing a long vowel -ar and -va: (or its cognate where vowel and consonant interchange, thus Bhojpuri -a:n) to an intransitive or transitive root, respectively. Thus, Hindi-Urdu lik\textsuperscript{b}+ar: write > lik\textsuperscript{b}+a: - to make someone write > lik\textsuperscript{b}+va:- 'to make someone make someone write'. In the scientific literature, the first causative with long vowel has been termed direct causative, contactive causative or causative 1 by various scholars (see Masica 1976 for a detailed discussion), and the latter form has been called indirect causative, distant causative or causative 2. Since the causative morpheme is an argument-increasing morpheme, an additional argument in the second causative distances the actor from the instigator, hence the terms 'distant causative' or 'indirect causative'. There are languages, however, that do not make a clear-cut distinction between direct and indirect causatives, and therefore an additional argument would not trigger a distinct morphological marking. In other words, the semantic distinction of direct and indirect causation is not overtly maintained by the morphology of the causative. Consider these sentences from Bangani:

2.1(a) a\text{\textacute{u}}\text{\textasciitilde{a}}\text{\textcircled{e}}\text{\textasciitilde{e}}\text{\textacute{u}} 5\text{\textacute{a}}\text{\textacuteslash{u}}
   I he'acc laugh-caus
   'I made him laugh'

(b) a\text{\textacute{u}} guruj\text{\textacute{u}}-ke\text{\textcircled{e}}\text{\textasciitilde{e}}\text{\textacute{u}} 5\text{\textacute{a}}\text{\textacuteslash{u}}
   I teacher-instr he'acc laugh-caus
   'I made the teacher make him laugh'

Typically, Indo-Aryan languages use prefixes and suffixes, the former being used more in word forms of the noun and adjective class than in verbs. The affixes are freely borrowed from Sanskrit and Perso-Arabic (PA) stock. Thus Hindi-Urdu kara\textsuperscript{\text{\textcircled{e}}} 'cause' but a-(Skt) kara\textsuperscript{\text{\textcircled{e}}} 'without cause', p\textsuperscript{\text{\textcircled{e}}}\text{\textasciitilde{a}}\text{\textasciitilde{e}}\text{\textacute{u}} 'educated' but b - (PA) p\textsuperscript{\text{\textcircled{e}}}\text{\textasciitilde{a}}\text{\textasciitilde{u}} 'uneducated'.

2.1.1.4 The Syntactic-Semantic System

Languages of the IA language family exhibit SOV word order. Therefore, in a prototypical case modifiers precede the modified, and adjuncts precede verbs. Kashmiri (see Hook 1976) is an exception and Bangani also prefers verbs in the second position (see Abbi 2000), although verb-final sentences are also used. We shall consider the phenomenon of word order in chapter 6, where we mainly discuss syntactic issues. Although an SOV pattern is maintained, word order is comparatively free in IA languages, and movement of elements is used for strategies such as topic and focus, foregrounding and backgrounding information. Though the verb is an essential component, yet it is not surprising to find verb-less sentences in Eastern Indo-Aryan languages such as Oriya, Sambalpuri, Bangla, and Assamese. Thus in Bangla:

2.2 kaal ke or kaaj φ
   tomorrow-acc he-dat work φ
2. Language Families, Language Contact, and Areal Universals

'he has work tomorrow'

A large number of languages are split ergative (a feature they share with Tibeto-Burman languages), i.e. show ergative features in some tensed-marked sentences, such as in the past tense in Hindi. Eastern Indo-Aryan languages are not ergative in nature. For a detailed discussion on ergativity please see section 5.1.7.2 and 5.3.1.

Another interesting feature of IA languages is that experiential constructions taking typical psychological predicates mark their experiencer nouns with the dative or any other case. These constructions are popular in linguistic literature under the name of 'dative subject', as the dative-marked experiencer noun serves as the subject of the sentence. For details see chapter 6. The following sentence from Hindi will illustrate the issue for the time being.

2.3 radha-to bhu-h lagi he
Radha-dat hunger strike aux
'Radha is hungry'

A typical feature of Indo-Aryan is that negation is always preverbal. Languages in this family generally offer two types of negation, one prohibitive and another for simply negating the assertion. Thus Hindi 
*not jao 'don't go' while vo nahi gaya 'he did not go' type of constructions can be found in almost all IA languages. We shall discuss negation at length in all families in chapter 6.

There are other constructions in the languages of this family that employ what is known as 'conjunct verbs'. The conjunct verb is constituted of two elements, the first of which is generally an adjective or a noun, while the second element is the verb 'to do', 'to be', or 'to happen'. This is a very productive way of forming verbs in these languages - so much so, that a noun or adjective can be borrowed from a foreign language and the indigenous verb 'to do' suffixed to it. For instance Hindi fon karma 'to telephone', or karekt karma 'to correct'. Typically, many languages take a dative subject with conjunct verbs.

Finally, the IA languages typically have relative-correlative constructions, where the modifying clause marked by the relative marker is followed by the main clause preceded by a correlative. Consider the Hindi sentence given below.

2.4 jo bacca baahar krel rehna he vo mere bhai ka he
Rel child outside play prog. Aux 3msg Corel my brother-gen Aux 3msg
'The child who is playing outside is my brother's.

2.1.2 The Dravidian Languages

2.1.2.1 The Geographical Spread

Dravidian languages occupy the southern part of India and a little pocket in West Pakistan inhabited by North Dravidian speakers called Bruhui. In fact, this pocket offers a clue to the directionality of migration of the Dravidian speakers in the
As far as the lexical morphology is concerned, from the scanty data (all examples are from Manoharan 1997) that is available it seems to be an isolating language which is affixal in nature. Thus, a root can take various kinds of formative affixes to derive nouns and verbs. Thus:

/kɔw/

kɔw  ‘bow’

kɔw-bu  ‘umbrella’

kɔw-φο  ‘banana fruit’

ɔt-kɔw-bo  ‘bark of a tree’

ge-ruulu-to-kɔw-bo  ‘your eye lid’ [you-eye-lid]

The genitive always precedes the names for the body part. Thus:

[tɛ-lɛ:]  ‘my naval’

[tɛ-ba:]  ‘my hand’

[tɔː-tə:]  ‘my tongue’

[geli-ruulu]  ‘your (hon) eye’

[akka-φο]  ‘his/her/its mouth’

The pronominal prefixes have two functions. When they occur before a body-part word, they signify a genitive construction, but when they occur before a verb, they assume the subject role. An interesting aspect of the language is that pronominal affixes can be classified into three basic classes on the basis of the phonetic shapes. Thus, though there are three variants of the first person singular pronoun “I”, as in /tɛ/, tɛ-, and ɡɛ-, each of them has a restricted set of verbs that it attaches to. Thus tɛ-otterə:- ‘to tell a story’ but ɡɛ-jo: ‘to come’ and ɡɛ-ja:- ‘to eat’. That is, verbs are classified according to the different types of pronominal prefixes that they take.

Pronominal prefixes and pronouns maintain the distinction between inclusive vs. exclusive first person singular and plural pronouns. Pronouns of both the second and third person are marked by the honorific. Consult chapter 7, table 7.1.

According to Manoharan, the language is highly creolised and multiple affixes for the same function are used simultaneously. “In the process of creolisation the affixes of the different languages spoken earlier are retained resulting in multiplication of the number of affixes” (ibid: 471).

2.2. India as a Linguistic Area

It is not surprising to find distinct languages of India sharing common linguistic features due to areal pressure. Long and stable language contact accompanied by bilingualism has generated many linguistic structures that are common at the structural as well as the semantic levels across languages of various families. It is a known fact that contact-induced structural convergence of languages from different families within a coherent and contiguous geographic area often results in a linguistic area or sprachbund. The current Indian languages share with each other several
features at all levels of grammar, i.e. phonological, morphological, syntax, semantics, pragmatics and sociolinguistics. Since the first attempt of identifying India as a linguistic area by Bloch (1934) and subsequently by Emeneau (1956), many linguists have identified features at all levels, demonstrating that despite distinct language structures of various language families we have a fairly large number of shared areal features. Some of the well-researched areal features of Indian languages, diagnostic of geographic and linguistic contact among the major language families in the country, are the following:

2. Morphologically derived causatives (Masica 1976)
3. Reduplication of words (Abbi 1952)
4. Expressive morphology in psychological verbs (Abbi 1992)
5. Explicator compound verbs (Hock 1977; Masica 1976)
6. SOV word order (Masica 1976)
8. Conjunctive participles (Masica 1976)


As the process of convergence involves a simultaneous process of divergence from the archaic structures of the language family, a pattern of typological homogeneity is witnessed in the languages of a linguistic area. As we will discuss areal features at all levels in the following chapters, we will not go into detail here, but would like to briefly enumerate the salient areal features of Indian languages. We shall undertake this task by exhibiting features at phonological, morphological, syntactic and semantic levels. You will find the details in the chapters 4, 5, and 6.

A very large number of languages share the following features:

### 2.2.1 Sound System

- a) Retroflexion
- b) Aspiration of voiceless stops
- c) Syllable-timed rather than stress-timed

### 2.2.2 Morphological System

- a) Suffixal causatives
- b) Postpositions
- c) Morphological reduplication
- d) Echo formations
- e) Exclusively suffixing agglutinative morphology
2. Language Families, Language Contact, and Areal Universals

2.2.3 Syntactic System

a) OV order of words
b) Conjunctive participles/converbs
c) Explicator compound verbs
d) Syntactic reduplication (aspectual)
e) Dative subject/non-agent experiential constructions

2.2.4 Pragmatic and Sociolinguistic Systems

a) Negation used as a tag for confirmation
b) Three way distinction in 2nd person pronouns
c) Marked +/- honorific feature (in pronoun/verb morphology)
d) Extensive kinship terminology

In addition to the features listed above, there are some ‘areal lexicons’, i.e. words that cannot easily be translated into English, but can be translated very well across Indian languages. We are going to consider some of them in our list of 400 basic words in section 4.5.

2.3 Sub-linguistic Areas

Apart from areal features, there are also those features that identify a micro-area, and it will perhaps be of use to know about these sub-areal features if working in the particular area. We shall briefly enumerate these here for convenience.

1. Nasalisation (Northern India)
2. Aspiration (Northern India)
3. Gender agreement (Western India)
4. Right hand is ‘eating hand’ (Central India)
5. Quotative Verb ‘say’ as complementiser (Southern and Northeastern India)
6. Relative-correlative pronoun (Northern India)
7. Classifiers (Eastern and Central India)

We agree with Masica (1992: 38).

Like all large linguistic areas with a complex history, South Asia shows many sub-areas of more intense or special convergence, or of partial convergence with other areas. It also has features that place it within larger areal configurations.

2.4 Contact and Convergence

When languages are in contact with each other for a very long period of time, they tend to assimilate several features of languages in contact at all levels of grammar and thus deviate considerably from the characteristic features of their own genetic stock. The Dravidian languages, the Munda languages and the Indo-Aryan languages have all been in such close proximity to each other as to induce changes in each of them to such an extent that one can perhaps identify a ‘core grammar of Indian languages’. We shall in this section probe into the various linguistic features that are the result of this language contact. We shall also look into the fact that