
Introduction

What claim do the Indo-Aryan languages have on our attention? Their foremost claim surely has to be the sheer numerical weight of the populations speaking them – possibly one-fifth of mankind. To the plea that these are largely accessible through English, it must be answered that no foreign language can afford full access to the hearts and minds of a people.

Many would say that English, which has been an intimate part of South Asian life for at least a century and a half (in some areas longer), helping to shape the minds of whole generations through schooling, and being shaped in its turn to express many features of South Asian life it was ill-suited to express initially, cannot justly be called a “foreign” language in the area. The whole question of the indigenization of English in South Asia is a fascinating study in itself.¹ Despite such adaptations, however, neither English nor Persian (which held sway in the subcontinent for a much longer period) can be as fully expressive of South Asian cultures as the languages which have been totally formed by those cultures. Both English and Persian have their primary base and formation elsewhere. In any case, knowers of English, however important in terms of absolute numbers and of international impact² and prestige constitute only a tiny minority (average 2.3 per cent³) of these populations.

It is perhaps to the deceptive convenience of English, however, that we owe the popular notion (more prevalent in America than in Britain) that the population of India, for instance, babbles chaotically in “hundreds of dialects,” coupled with ignorance of the very existence of great languages (Bengali, Marathi, Hindi, and half a dozen others) of comparable age and demographic weight to the modern languages of Europe. Another factor in this ignorance is no doubt the degree of political unity the region has succeeded in retaining in the post-colonial period. It has not been Balkanized to any great extent, and most of these languages are accordingly subnational rather than the expressions of independent political entities. (Exceptions are Bengali in Bangladesh, Sinhalese in Sri Lanka – where it must contend with a substantial Tamil-speaking minority, and Nepali in Nepal –

where it is the official language and undoubted lingua franca but again not the mother tongue of all the population. Urdu is the sole official language of Pakistan but it is not the mother tongue of other than a small minority in that country, mostly migrants from North India concentrated in Karachi. It is, however, the preferred *literary* language of the numerically dominant Punjabi speakers.)

The speakers of these languages, in partnership with others brought into a common orbit via the unique cultural achievement that constitutes Sanskrit, have moreover been the creators of one of the great civilizations of the world, which merits the attention of all who would seek to follow and appreciate the human story. This Indic civilization once extended as far as Vietnam⁴ and Indonesia, and contributed important components to the civilizations of China, Korea, and Japan as well, mainly through the vehicle of Buddhism.

The dazzling achievements of Sanskrit literature and thought, providing more than enough to digest as the West continues to make their acquaintance, together with the ignorance referred to above of the very existence of the modern languages, plus the availability of a small but meritorious literature from the area written originally in English (and a voluminous literature *about* the area in English) may account for the fact that the *literatures* of the modern Indo-Aryan languages remain practically unknown to the outside world. Yet many of them do possess flourishing modern literatures in most of the familiar genres, no doubt destined to expand further as literacy increases, as well as important pre-modern literatures in mediaeval and folk-genres meriting attention in their own right.

An aspect of Indo-Aryan that has, for some reason, excited more interest in continental Europe than in the English-speaking world is the fact that the speakers of these languages are our linguistic cousins, fellow members of the great Indo-European linguistic community. For British readers, a substitute has existed in the close historical ties between Britain and the former Indian empire, lately a subject of much renewed interest. (Of course the latter took in considerably more than Indo-Aryan, but Indo-Aryan lay at the heart of the matter, with three out of the four major centers of the empire, namely Calcutta, Delhi, and Bombay, in its territory.)

Apart from these general humanistic, practical, and sentimental concerns, there is a different set of reasons why Indo-Aryan merits attention. These have to do with the scientific study of language.

It is almost a commonplace that modern Western linguistic science took its birth from the discovery that the classical language of India, Sanskrit, is related to the classical and modern languages of Europe. This discovery is usually credited to Sir William Jones. Although Jones was actually neither the first to postulate a common origin for the Indo-European languages nor the first to add Sanskrit to

their company (the former honor seems to go to the seventeenth-century Dutch scholar Marcus Zeurius Boxhorn, and the latter to the sixteenth-century English Jesuit Thomas Stevens),⁵ it may be claimed that it was Jones's publication of his discoveries, which seem to have been largely independent, in the form of his presidential address to the Asiatic Society in 1786, that gave the impetus to Sanskrit studies in Europe, without which Indo-European philology would not have gotten very far.

Indo-European studies still form the backbone of historical linguistics, and the subsequent history of Indo-Aryan as a major branch of Indo-European, and one moreover whose development under the most diverse conditions can be followed almost continuously for 3,500 years, deserves a larger place in such studies than it is usually accorded. In the same address in which we find the much-quoted passage concerning the affinity of Śanskrit, Latin, and Greek, Jones made some remarks concerning modern Indo-Aryan, less often quoted, which may have a bearing on this situation:

and this analogy might induce us to believe, that the pure Hindi, whether of *Tartarian* or *Chaldean* origin, was primeval in Upper *India*, into which the *Sanskrit* was introduced by conquerors from other kingdoms in some very remote age . . . (quoted by Grierson 1927, *LSI* 1.1:11)

In other words, he failed to perceive the relationship of Hindi to Sanskrit, and thought it was basically a pre-Sanskritic language of the "Tartarian" group, indigenous to India. It took a while to get matters straightened out: the first steps of correction overshot the mark, and took the Dravidian as well as the New Indo-Aryan languages to be descended from Sanskrit. All this may have contributed to the slowness with which the later development of Indo-Aryan found a place in Indo-European studies.

Materials for such studies are now relatively abundant, however (although there is naturally always more to be done), thanks to the labors of a remarkable company of scholars over the last hundred years, among which those of Sir George Grierson, especially the unparalleled achievement represented by the *Linguistic survey of India* (1903–27), and of Sir Ralph Turner, culminating in his *Comparative dictionary of the Indo-Aryan languages* (1966), deserve special mention as providing tools with which most other fields are not so conveniently blessed. A place of honor also goes to Suniti Kumar Chatterji (1890–1977), whose monumental *Origin and development of the Bengali language* (1926),⁶ based on his 1921 University of London D.Litt. thesis, is basic reading for anyone interested in historical Indo-Aryan in general, and has served as a model for several

similar studies.⁷ Originally self-taught in European-style historical and comparative linguistics, Chatterji was enabled by a government of India scholarship to study in Europe under such masters as Meillet and Jules Bloch. In England, Grierson took a personal interest in his work.

This interest on the part of foreigners and foreign-trained native scholars joined itself to and was no doubt partly inspired by an indigenous tradition of grammar, phonetics, and lexicography of great sophistication and depth, unmatched in other parts of the world. The full implications of the rule-based Sanskrit grammar of Panini (fourth century B C) could not, perhaps, be properly appreciated in the West⁸ until modern linguistic theory itself had evolved to such a stage, which was only recently, but in the field of phonetics the impact of Panini and of the ancient Indian phoneticians in general on Western linguistics was early (nineteenth century, another service of Jones) and profound (Allen 1953: 3-4). Somewhat later their influence may also be detected in American structuralist morphophonemics, e.g., in Bloomfield's description of Menomini (Allen 1962: 24). In India itself, the first real synthesis of the learning of the pandits and the scholarship of the West is represented by Sir R. G. Bhandarkar (1837-1925), the foremost Indian Sanskrit scholar of the nineteenth century, self-taught in the new philology. He seems not to have gone abroad until 1886, for the Vienna meeting of the International Congress of Orientalists, when he was a well-established scholar. As a measure of the influence of these two men, until recently at least Calcutta (Chatterji's seat) and Poona (Bhandarkar's seat) have remained the centers of historical research in Indo-Aryan.

Indo-Aryan presents special opportunities for the investigation of other linguistic problems also. One of these, which we may call areal or convergence studies, has to do with the phenomenon of linguistic change from a vantage point just opposite, as it were, to that of historical-comparative linguistics. That is, with focus on the results of the process rather than with antecedents, with external rather than internal motivation for change, and with spatial rather than (or in addition to) temporal relationships. As languages diverge from a common ancestor they may also - particularly where extensive migrations are involved - change in type, and come in some degree to conform typologically to new linguistic environments in which they find themselves.

Exactly how and why this happens is not fully understood (although reasonable hypotheses abound), but the history of Indo-Aryan offers ample scope for exploring the question. The fact that typological changes affecting New Indo-Aryan managed to disguise their Indo-European affinity for Jones is indicative of their suitability for this purpose. Moreover, not only is the history of Indo-Aryan itself relatively well documented, but the non-Aryan languages and stocks of the

region are also fairly well documented and studied in comparison with such situations elsewhere. If it should be held that one important factor in the development taken by Indo-Aryan, or a particular branch of it, was the adoption of Aryan speech by non-Aryan speakers (Chatterji for one maintains that this was the case in East Bengal, Assam, Orissa, and South Bihar, among other places), this need not remain in the realm of pure speculation: the same thing can be observed going on today, e.g., in Halbi (the neo-Aryan speech of former Gond speakers in Bastar District in Madhya Pradesh) and in Nagamese (the Assamese-based pidgin of Nagaland in the extreme northeast, now creolizing), to name but two instances. The continuing spread of Nepali among Tibeto-Burman speakers both within Nepal and beyond it to the east is another case in point, complicated by the existence of a written standard.

The effects of super- as well as substrata on linguistic development can also be studied in Indo-Aryan in relation especially to Sanskrit, Persian, and English as prestige languages. To these should be added the influence of Modern Standard Hindi and Urdu on a number of languages and dialects, and indeed of standard literary languages on the spoken languages generally.

This brings us to the whole set of problems coming under the general heading of sociolinguistics. Here Indo-Aryan constitutes a vast laboratory almost without equal, albeit as yet little exploited. On the one hand, there is the complexly compartmentalized traditional caste society of India, socially segregated, occupationally specialized, hierarchically organized, yet interdependent. What effect does this have on language? For comparison, there are the other Indo-Aryan-speaking societies, minimally to maximally different: traditionalist but half-Tibetanoid Nepal, Muslim Pakistan and Bangladesh, Buddhist Sri Lanka, and speaking a language closely related to that of the latter, the Muslim Maldives. Within each of these, but especially in India, there is the contrast between rural and urban subcultures, the latter sometimes ancient, sometimes rapidly modernizing – and typically multilingual. There are striking differences in regional marriage patterns: for example, between the local exogamy of the North of India, bringing in women of different dialectal background, and the local endogamy of the South.⁹

Finally, there are the three quite distinct cases of Indo-Aryan languages taken completely outside the South Asian region: (1) the ancient migration of the Gypsies to the Near East and thence to Europe; (2) the largely nineteenth-century emigration of laborers and/or merchants to Fiji, Mauritius, East and South Africa, Guyana, Trinidad, and other spots in the Caribbean; (3) the mid-twentieth-century movement that has brought upwards of a million seekers of a better life to Britain (largely from the Punjab, in contrast to the nineteenth-

century movement of laborers to the British colonies, mainly from eastern UP and Bihar). To these we should add the still more recent settlement of (until lately) mainly well-educated professionals from all areas in the United States and Canada. (A movement of Punjabi-speaking farmers at the beginning of this century to the west coast of North America, i.e. to British Columbia and California, is also worth noting. Another, very different case is represented by the Parya language of Soviet Tadzhikistan, whose existence also implies a migration, but the date of the latter is quite unknown.)

Modern facilities for travel and communication being what they are, the most recent migrations involving literate speakers are of sociolinguistic interest mainly from the standpoint of ordinary problems of language maintenance and adaptation to new expressive needs. The earlier transplantations of illiterate laborers, subsequently cut off from their roots, involve more specialized phenomena, including the evolution of new Indo-Aryan-based lingua francas to facilitate communication among people of different dialectal and language backgrounds. The arrival on the scene of zealous propagators of Modern Standard Hindi has further complicated the situation.¹⁰ The case of the Gypsies, where a form of Indo-Aryan speech has been jealously guarded for centuries as a secret language of intra-group communication as well as a badge of identity in a semi-nomadic subculture, is unique.¹¹

The multilingual nature of much of South Asian society presents special challenges to the sociolinguist. Participation of linguistically disparate regions in a common civilization, held together by such specific institutions as pilgrimages, as well as requirements of trade, led to the development of lingua francas, of which Hindustani is the most notable recent example. Sanskrit itself could be said to have played this role, as to a limited extent it still does among the traditionally educated elite. As a deliberately standardized and maintained yet flexible medium of elite communication Sanskrit is a fascinating product of the human spirit. In its heyday in the first millennium AD, it linked together – and synthesized elements from – an area much vaster than Indo-Aryan itself or even the subcontinent, and widely separated epochs of time. At the other end of the sociolinguistic spectrum, the expansionist character of Aryan society in a region full of other linguistic stocks has given rise to pidgins and creoles and mixed languages of every description.

As an effect both of superstrata and of a long literary tradition, some Indo-Aryan languages are characterized by a notable degree of *diglossia*. That is, the literary language is different from even the educated colloquial, not only in vocabulary but also in phonology and grammar. This has reached its extreme point in Sinhalese.¹² In some other languages (Bengali, Marathi), there has been

a slowness on the part of well-established traditions to adjust to linguistic change, but such adjustments have eventually to some degree been made. The Sinhalese situation¹³ is approached more closely in some of the non-Aryan languages of the region with long literary histories such as Tamil.

True diglossia, where the literary norm is nobody's spoken language, should be distinguished from the kind of bilingualism or bidialectalism where people are accustomed to use as their literary language a dialect or even a language which is not their own spoken language. This situation is naturally very common in Indo-Aryan South Asia where there are many more major dialects and spoken languages than there are literary languages, with each of the latter, even when it has its own corps of speakers, serving a number of the former.

This situation is hardly unique to South Asia, and the whole question of the formation of standard literary languages is one of general interest. Although it has been studied in other contexts, Indo-Aryan offers not only ample and diverse material for comparison but also processes amenable to concurrent observation. There is also the related question commonly referred to as language "modernization" – that is, "expansion to meet the needs of a developing society." What are these needs, and how are they being met in the meantime? Do some functions lend themselves to "language engineering" more than others? What options are in fact available? Although such questions may not be answered or even properly addressed here, they are sharply posed for further researchers by the sociolinguistic context of Indo-Aryan.

This book is an introduction only to Indo-Aryan, not to the other linguistic stocks in the South Asian region, which are mentioned only when they impinge directly on the development of Indo-Aryan. Moreover, our focus here will be on the modern languages, not on Sanskrit – again, except as necessary background for the former. For Sanskrit as such there are available several excellent books by competent authorities.¹⁴

The modern Indo-Aryan languages and dialects

Before proceeding further, it is appropriate that we identify more precisely the languages to be discussed. A brief survey of the modern Indo-Aryan domain in terms of contemporary political geography is therefore given here. Although this may not be necessary for all readers, it will no doubt be helpful to those less familiar with the area. It is followed by discussion of the problems of language vs. dialect in the Indo-Aryan context, of Hindi-Urdu, and of nomenclature, and supplemented by Map 1, as well as by a comprehensive alphabetical inventory of Indo-Aryan language and dialect names, living or dead, given in Appendix I. Because of the sheer number of names that will be met with in the literature (by those whose interest or work takes them beyond this book), the last is needed for reference purposes in any case: even the specialist is unlikely to be familiar with all of them.

2.1 Indo-Aryan: a bird's-eye view

The Indo-Aryan languages are a sub-branch of the Indo-European family, spoken today mainly in India, Pakistan, Bangladesh, Nepal, Sri Lanka, and the Maldivian Islands by at least 640,000,000 persons (est. 1981). Although they are not the only languages spoken in any of these countries, their speakers in all cases constitute majorities. In the past, Indo-Aryan languages (distinguished here from the Nuristani languages [see Section 2.1.18]) extended also into eastern Afghanistan, where isolated remnants may still exist, and at a more remote epoch (the early centuries of the Christian era), also into Chinese Turkestan (Sinkiang).

The modern Indo-Aryan languages, properly and henceforth called **NEW INDO-ARYAN** (= "NIA", as against "MIA" for the preceding stage of **MIDDLE INDO-ARYAN** [see Chapter 3]), date from approximately AD 1000. The NIA languages are presently distributed as follows (for more details on each language see Appendix I):

2.1.1 A vast central portion of the subcontinent, consisting of the Indian states of Uttar Pradesh, Bihar, Madhya Pradesh, Rajasthan, Haryana, and Himachal Pradesh, plus the Union Territory of Delhi, is known as the "HINDI area", because the official and general written language, that is to say, that of administration, press, school instruction, and modern literature, is Hindi, sometimes called MODERN STANDARD HINDI, and the whole area is heir to the "Hindi literary tradition" – Hindi being used here in a different and wider sense, to refer to pre-modern literatures in **Braj** and **Awadhi**, and often to those in languages proper to Rajasthan and Bihar as well.

While Kellogg could in 1892 describe "**High Hindi**", as he called it, as "understood more or less through all the Hindi-speaking country, but in no place the language of the home," this is no longer accurate: Standard Hindi does have native speakers, especially in urban areas, and is fast encroaching on dialectal forms of speech, to the point where a student of the latter is now sometimes hard put to find "pure" informants.

From this the reader will not incorrectly draw the conclusion that there are other forms of speech "on the ground" in the Hindi area, particularly at the village level (but by no means excluding a good portion of the urban population), over which Standard Hindi is superimposed. These are the so-called **regional languages of the Hindi area**, sometimes less accurately called Hindi "**dialects**". Some of these are fairly closely related to Standard Hindi (and often, confusingly, also loosely called "**Hindi**" by their speakers); some are more distantly related to it. (The situation somewhat resembles that of an earlier historical period in the Italian-, Spanish-, or German-speaking areas of Europe, although the area and population involved in India is much greater, and the role of some of the regional languages or dialects is much larger in the pre-modern literary tradition. Another but looser analogy might be to China.)

The heartland of the Hindi area is the densely-populated Upper Ganges valley, corresponding to the state of Uttar Pradesh (which alone had 110,850,019 people in 1981), minus its hill areas, together with the Haryana region west of Delhi and adjoining areas of northern Madhya Pradesh and perhaps also northeastern Rajasthan. From west to east the regional languages here are: **Haryānvi** (formerly called **Bāngarū**) in most of Haryana State (formerly southeastern Punjab) and rural parts of the Delhi Territory; adjoining it in UP northeastward from Delhi up to the premontane Tarai and as far east as Rampur, and reaching across the Jamuna to include the northeastern portion of Haryana as far as Ambala, there is a form of Indo-Aryan speech with no settled name, despite its importance [see below]: Grierson called it **Vernacular Hindōstānī**; it has often been called **Khari Boli**; since the latter term is applied also to Colloquial Standard Hindi, Bahri

THE INDO-ARYAN LANGUAGES

Nuristani
Area

25

Iranian
languages

Khetri

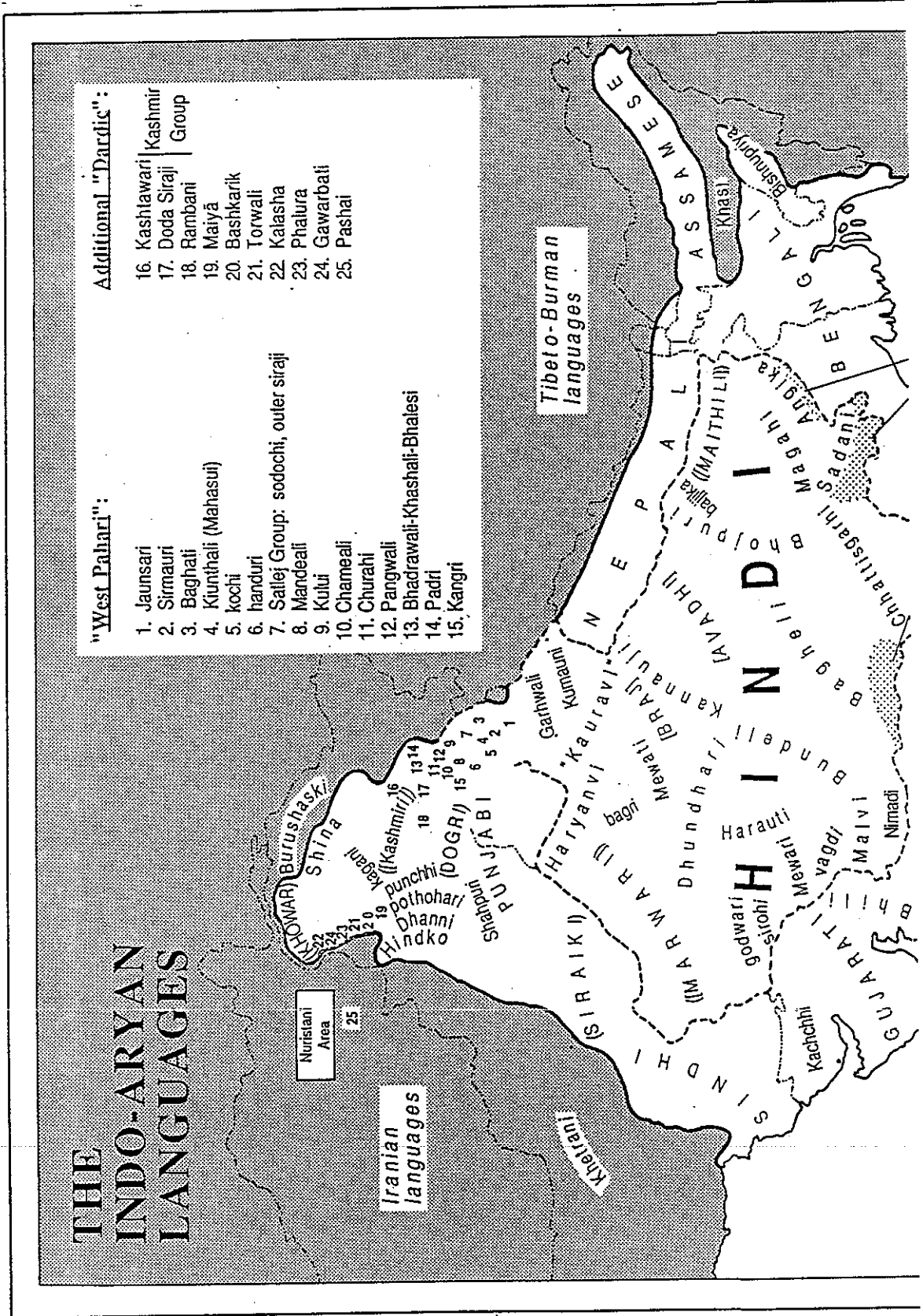
"West Pahari":

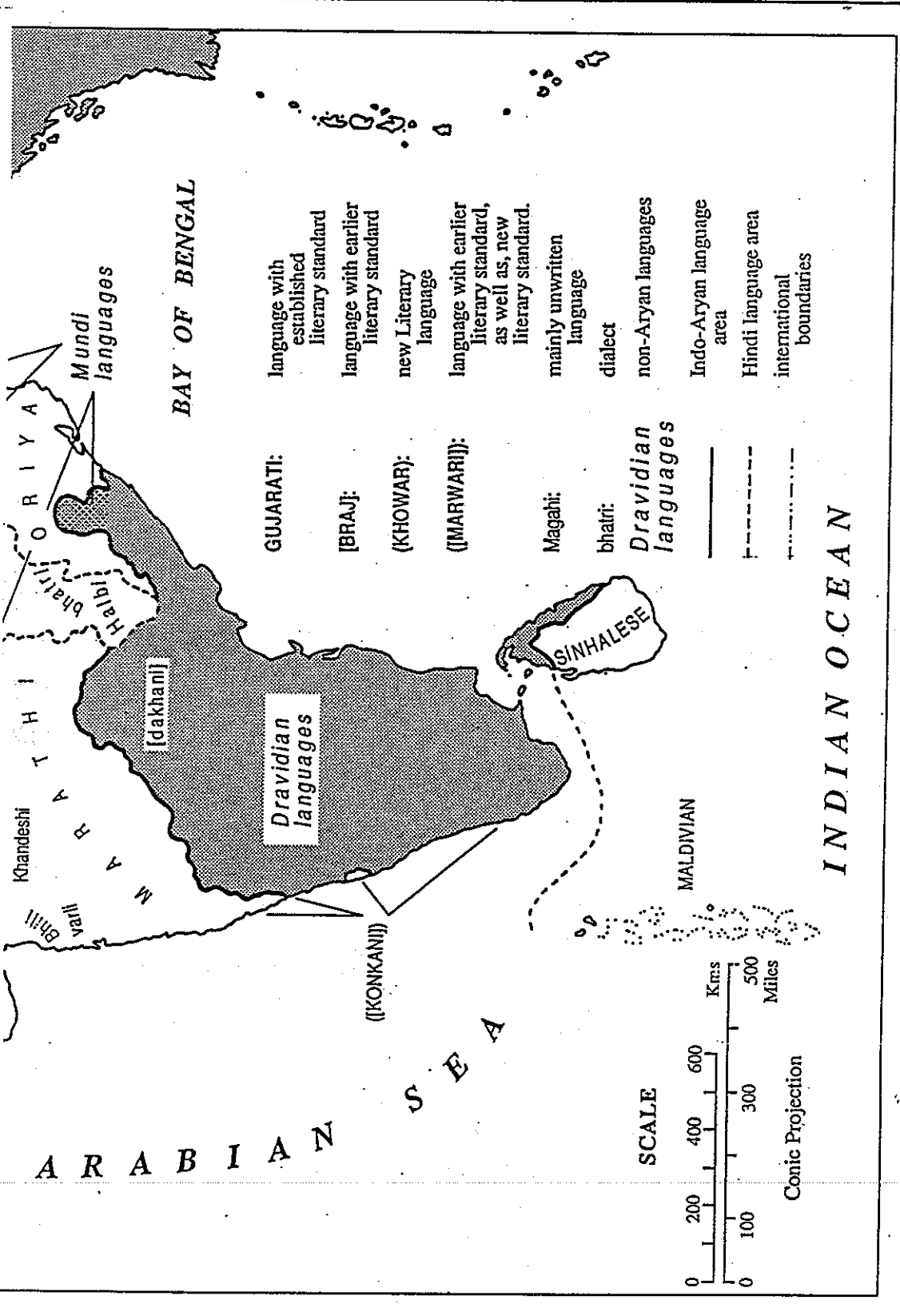
1. Jaunsari
2. Sirmauri
3. Baghati
4. Kiunthali (Mahasui)
5. Kochi
6. handuri
7. Satej Group: sodochi, outer siraji
8. Mandeali
9. Kultu
10. Chameali
11. Churahi
12. Pangwali
13. Bhadravali-Khashali-Bhalesi
14. Padri
15. Kangri

Additional "Dardic":

- | | |
|-----------------|---------|
| 16. Kашtawari | Kashmir |
| 17. Doda Siraji | Group |
| 18. Rambani | |
| 19. Maiyā | |
| 20. Bashkarik | |
| 21. Torwali | |
| 22. Kalasha | |
| 23. Phalura | |
| 24. Gawarbali | |
| 25. Pashai | |

Tibeto-Burman
languages





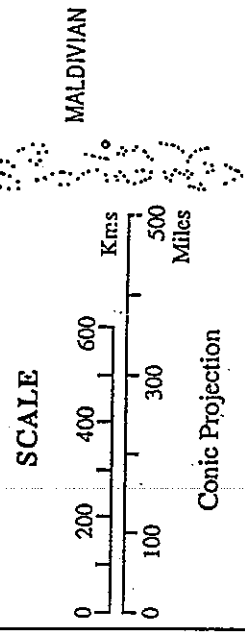
ARABIAN SEA

BAY OF BENGAL

Khandeshi
Marathi
Gujarati
[BRAJ]
[KHOWAR]
[MARWARI]
Magahi
bhatiri
[dakhani]
[KONKANI]
Mundi languages
Sinhalese

language with established literary standard
 language with earlier literary standard
 new Literary language
 language with earlier literary standard, as well as, new literary standard.
 mainly unwritten language
 dialect
 non-Aryan languages
 Indo-Aryan language area
 Hindi language area
 international boundaries

GUJARATI:
 [BRAJ]:
 (KHOWAR):
 ([MARWARI]):
 Magahi:
 bhatiri:
 Dravidian languages



INDIAN OCEAN

(1980) following Rahul Sankrityayan proposes to call it **Kauravī**, after the ancient land of the Kurus; southeast of Delhi, a broad area centering on Mathura but extending northeastward as far as Bareilly is the homeland of **Braj**; in a narrower band to the east, from Etawah and Kanpur up to Pilibhit is the closely allied **Kannaujī**; to the south of these in Madhya Pradesh from Gwalior as far as the tribal hinterlands of Chhindwara and Hoshangabad is **Bundēlī**, also similar to Braj; a more distinct language, **Awadhī**, prevails in east-central UP north and south of Lucknow; a variety of this known as **Baghelī** extends in Madhya Pradesh from Rewa to Jabalpur and Mandla; more isolated and therefore more strongly characterized is the **Chhattīsgarhī** further to the southeast on the borders of Orissa; eastern UP, including Varanasi (Benares), Azamgarh, and Gorakhpur, is occupied by various dialects of **Bhojpurī**, which extend into Bihar (Shahabad and Saran Districts, west of the rivers Son and Gandak respectively, and most of Champaran District).

Grierson classed "Vernacular Hindostani", Braj, Kannauji, and Bangaru (Haryanvi) together as "**Western Hindī**" and Awadhi, Baghelī, and Chhattīsgarhī together as "**Eastern Hindī**", but put Bhojpuri into the more distantly related "**Bihārī**" group. The other principal "Bihari" languages/dialects are **Magahī**, spoken in central Bihar (south of the Ganga and east of the Son) and **Maithilī**, spoken north of the Ganga. The latter has a long literary tradition, the former none. Also in the "Bihari" group are **Sadānī** (or **Nagpurīā**) in South Bihar (Chota Nagpur) centering on Ranchi, **Angikā** in eastern Bihar (Monghyr, Bhagalpur, Santal Parganas, Purnea, according to Pandey 1979: Grierson, who calls it **Chhikāchhikī Bolī**, excludes Purnea), and **Bajjikā** in Muzaffarpur and part of Champaran Districts in northwest Bihar (S. Tivari 1964). Claims of independent status for the latter two, previously taken to be dialects of Maithili, are recent, as are their names, although the dialects themselves are ancient.

Leaving now the North Indian plain with its cultural extensions in the rougher country to its immediate south for Rajasthan, we find the main desert area west of the Aravalli range occupied by various forms of **Mārwārī**, among which the **Bāgrī** of the Haryana border and the **Bhitrautī**, **Sirohī**, and **Godwārī** of the southern Aravalli foothills might be mentioned as distinctive. East of the Aravallis, **Mewārī** in the southeast has been classed as a dialect of Marwari but is also distinctive. (Southeastern Rajasthan south of Udaipur city, as well as the interior of the southern Aravalli range, are occupied by **Bhili** dialects which no one tries to affiliate to either Hindi or Rajasthani. The dialect of the former is known as **Vāgdī**, or **Wāgdī**.) Further northeast lies what Grierson called **Central Eastern Rājasthānī**, with two main representatives, **Dhundhārī** (or **Jaipurī**), centered on Jaipur, and **Hārautī**, centered on the Districts (former princely states) of Bundi

and Kota. In the Alwar District of the extreme northeast, spilling over into the Gurgaon District of Haryana, is *Mewāṭī*. (In the area of Bharatpur, Dholpur, and Karauli just to the south, Braj extends into Rajasthan.) Outside of Rajasthan, the language of western Madhya Pradesh (Ujjain, Indore, Bhopal), *Mālvī*, is also classed with "Rajasthani". A far-southern dialect, *Nimāḍī*, isolated in the Satpura range between the Narbada and Tapti valleys in a tribal area, has developed special peculiarities.

The Himalayan areas of UP, except for the highest elevations, are occupied mainly by two languages (in various dialects), *Garhwālī* and *Kumaunī*, grouped together by Grierson as "**Central Pahārī**" (*Pahārī* = "hill speech"). They are more closely allied to Rajasthani than to the Hindi of the plains. Further west in the mountains, in Himachal Pradesh and beginning already in the western part of Dehra Dun District in UP, lies the highly splintered group of Indo-Aryan dialects collectively known as "**Western Pahārī**". From southeast to northwest the main ones are *Jaunsārī* (in Dehra Dun), *Sirmaurī*, *Baghātī*, *Kiūnthālī* (around Simla, now apparently known as *Mahāsui*), *Handūrī*, *Kulūī*, *Mandeālī*, *Chameali*, *Bharmaurī* (or *Gādī*), *Churāhī*, *Pangwālī*, and (continuing into Kashmir) *Bhadrawāhī*, *Bhalesī*, *Khashālī*, and *Pādrī*. These too bear some Rajasthani affinity, along with characteristic archaisms and innovations that are increasingly marked toward the northwest. Whether because of the complexity of the situation or because of greater linguistic differences, they are less commonly claimed as "dialects of Hindi" (e.g., neither by Kellogg 1938/1892 nor by H. Bahri 1980—although Diack 1896 does indeed title his work *The Kulu dialect of Hindī*), even while Garhwali/Kumauni (and by Kellogg even Nepali) are so claimed. One reason may be the former closer affiliation of these areas politically with the Punjab. There have been reports of an attempt to concoct a "**Himāchalī**" language on the basis of these diverse dialects to serve, in the name of regional identity, as co-official language with Hindi, but it is too early to predict the outcome.

We now leave the complexities of the "Hindi area" to survey, first the remainder of the contiguous Indo-Aryan territory by means of a rough *pradakṣiṇā* (clockwise circumambulation) of the Hindi area, then the non-contiguous languages.

2.1.2 East of "Central Pahari" along the Himalayas lies "Eastern Pahari", that is, *Nepālī*, an independent language by any standard *pace* Kellogg (who had very limited access to it), dominant not only in the kingdom of Nepal but recently also in Sikkim, the Darjeeling District of West Bengal, and parts of Bhutan.

2.1.3 **Assamese** is the language of the Brahmaputra valley in far northeastern India. It was little known in most of the Tibeto-Burman and Khasi-speaking hill areas surrounding the valley, part of the old state of Assam but now largely separated politically as new states and territories. In one of them, however, Nagaland, a pidginized form of Assamese known as **Nagamese** is reported to have become a lingua franca.

2.1.4 Cut off from the Hindi area by the barrier of the Rajmahal hills, and from Assamese partly by the Khasi-Garo hills, both the homes of non-Indo-Aryan-speaking tribes, is the **Bengali** area, basically the great delta of the Ganges, now politically divided between the Indian state of West Bengal and the new country of Bangladesh. Bengali is also dominant in Tripura, an Indian territory to the east of Bangladesh, and Bengali speakers are numerous in Assam. The colloquial standard of Dhaka, the Bangladeshi capital, is different from that of Calcutta. The dialect of Chittagong, in southeast Bangladesh, is different enough to be considered a separate language.

2.1.5 Another Indo-Aryan language of the eastern frontier is **Bishnupriya Manipuri**, formerly spoken in Manipur (on the border with Burma), but driven from that area in the early nineteenth century and presently at home in the adjacent Cachar District of Assam, Tripura, and the Sylhet District of Bangladesh.

2.1.6 Southwest of Bengal, the delta of the Mahanadi is the center of the **Oriyā** language. Much of the state of Orissa is home to non-Aryan-speaking tribal peoples, a large bloc of which separate Oriya from Bengali. The interior Sambalpur lowland has a distinctive dialect. **Bhatrī** is an aberrant dialect of Oriya spoken by former Gond (Dravidian) tribesmen in the northeast of the former Bastar State, now a District of Madhya Pradesh.

2.1.7 **Bhatrī** is transitional to the main Indo-Aryan language of Bastar (where **Gondi** dialects continue to be spoken), **Halbī**. The latter is in turn transitional to **Marathī**, of which it is sometimes considered an aberrant dialect.

2.1.8 **Marāthī** occupies the extensive rolling plateau of the northwestern Deccan from Nagpur to Nasik, Pune (Poona), and Kolhapur, as well as the lowland known as the Konkan below the raised rugged edge of the plateau (= Western Ghats) from north of Bombay (Thana District) to just north of Goa (Ratnagiri District). The dialects of the Konkan are distinct.

2.1.9 These are to be distinguished further from **Konkanī** proper, centered on Goa, but extending slightly to the north (Savantvadi) as well as to the south (coastal North Kanara District of Karnataka State), with an important outlier in South Kanara, centering on Mangalore, and another in Kerala, around Cochin. (For a documented discussion of the "Konkani-Marathi controversy", see Pereira 1971.)

2.1.10 Beginning already in the northern part of Thana District (north of Bombay) and stretching in an arc around the eastern and northeastern periphery of Gujarat is a zone of tribal peoples now speaking, whatever their original language(s), Indo-Aryan dialects mainly grouped together as "**Bhīlī**". As noted earlier, these extend into southern Rajasthan. Their closest affiliation is generally with Gujarati, but the southernmost, such as the **Vārī** of Thana and the **Dangī** of the Dangs District (in southeast Gujarat), are closer to Marathi, and may be regarded as a bridge between the two major languages. (Except in the Nagpur area, the Marathi-Hindi boundary is by contrast a sharp one, marked also physically by the Satpura range, the home of non-Aryan, i.e. Munda-speaking, tribals.) East of Dangs are the Maharashtrian Districts of Dhulia and Jalgaon, formerly known as Khandesh, with a language, **Khāndeshī**, better known locally as **Ahirani**, which is transitional between Gujarati and Marathi.

2.1.11 To the north, **Gujarātī** is the language of greater Gujarat (including the Kathiawar peninsula) and also of an important component of the population of the city of Bombay. Beyond the Gulf of Kutch, however, the language, **Kachchhī**, is more closely related to Sindhi.

2.1.12 Across the Pakistan border, **Sindhī** is the language of the Lower Indus valley, below the narrowing of the valley above the Sukkur dam, and of the desert region to the east. It is more sharply bounded immediately to the west by the Kirthar range that marks the beginning of Baluchistan and Iranian speech. Karachi city, on the margins of the area in any case, is dominated by Urdu-speaking migrants from North India. The center of Standard Sindhi is the city of Hyderabad rather than Karachi.

2.1.13 The valley of the Indus and its tributaries in Pakistan north of Sind up to the Pir Panjal range on the frontier of Kashmir is occupied by a series of dialects known by various local names, and to outsiders first as "**Western Punjabi**". Noting that these – or some of these – had as much in common with Sindhi as with Punjabi, and differed strikingly from the latter in some features,

Grierson bestowed¹ the name “Lahndā” (from a Punjabi word for “western”) on them collectively as a distinct “language”. This has caught on only among linguists (who later began to prefer the feminine form **Lahndī**, matching the usual names of Indo-Aryan languages); it has no currency among the speakers themselves. It will accordingly be used here – for convenience, as there is no ready substitute – always in quotes.

Shackle, who has done more work in the area than any other recent linguist, has challenged (1979, 1980) the “Lahnda” construct even in terms of its convenience, as well as Grierson’s subclassification of the dialects comprising it (which has long been found unsatisfactory), although without presuming to come up with a final scheme himself. The situation is complicated for indigenous scholarship by the rival claims of old (i.e. pan-Punjabi) and new language movements.

In any case, the area concerned is divided, physically and linguistically, into two unequal halves by the great escarpment of the Salt range above Mianwali and Sargodha, which bounds the western Punjab plain on the north. The linguistic self-consciousness of the southern (= Central Pakistan plains) dialects (**Riyāsati–Bahāwalpurī, Mūltānī, Jhangī–Jatki, Thalī**, etc.), centering on the ancient city of Multan and the former princely capital of Bahawalpur, has coalesced around the name **Sirāiki**,² a term unfortunately also applied to a variety of Sindhi (the name is from S. *siro* ‘north, up-river’), doubly confusing because Siraiki is also spoken by many Siraiki settlers in Sind. Affiliated dialects are spoken also by segments of the population west of the Indus where the main language is Pashto. At the north end of the plains area, where linguistic and cultural distance from Multan is maximal, the dialect of Sargodha District, **Shahpurī**, which was taken by Grierson to be “standard Lahnda”, is in fact transitional to Punjabi, if not indeed a dialect of that language (Shackle 1976: 8, 1979: 201). It has been suggested that the non-contiguous dialect **Khētrānī**, spoken by a tribe in northeastern Baluchistan, may be the remnant of a separate language, of “Dardic” affinity (see below).

In the broken hill country to the north of the Salt range are the more diverse dialects of “Northern Lahndā”, Grierson’s pioneering subclassification of which most experts agree is particularly unsatisfactory. The least problematic may be **Poṭhohārī** (*LSI Pōṭhwārī*), the dialect of Rawalpindi and Jhelum Districts (and thus of the southeastern hinterland of the new Pakistani capital of Islamabad).

To the west and north of this, that is primarily in Attock and Hazara Districts, and across the Indus in Kohat and Peshawar, both the dialectal and the terminological picture is much more confusing, with discontinuous dialects (due to migration and invasion), dialects with no settled name, and identical names applied to several different dialects. The worst of the latter is “Hindko”, a term (basically meaning ‘the language of the Indians’ – as contrasted with Pathans) applied not

only to several forms of "Northern Lahnda" but also to the Siraiki dialects of Dera Ghazi Khan and Mianwali Districts (also called **Dērāwālī** and **Thalī** respectively), and of Dera Ismail Khan (Northwestern Frontier Province). Shackle (1980), however, proposes to set up a group called **Hindko proper**, comprising four dialects of Attock District, corresponding more or less to three of its tehsils (**Awāṅkāri** to the southern Talagang tehsil, **Ghēbī** to the central Pindī Gheb tehsil, and **Chachhī** to the northern Attock tehsil, extending to the southernmost Haripur tehsil of neighboring Hazara) plus **Kohātī** of Kohat city beyond the Indus.

The "Hindko" of Peshawar city deserves separate classification according to Shackle, partly due to the influence of Punjabi via the Grand Trunk Road. Despite the fact that a majority of the inhabitants are Pashto-speaking, **Peshāwarī Hindko** has considerable prestige and has been cultivated for literature.

To the east of "Hindko proper" (and west of Pothohari), in western Jhelum District (Chakwal) the dialect is **Dhannī**; to the north of the latter (Fatehjang tehsil, Attock District), in the valley of the Sohan river, is the closely related **Sawain** or **Sohain**. From Abbottabad northward in Hazara District, east of the Upper Indus (in the Northwest Frontier Province), are the northernmost dialects of "Lahnda", also confusingly called "Hindko": Grierson distinguished **Hindkī of Hazara** (the main dialect); **Tināulī** in the southwest; **Dhūṇḍī-Kairālī** in the east. Bailey 1915 described **Kāgānī**, "spoken in the whole of the Kagan Valley" including Mansehra and Abbottabad, and "known as Hindko" – apparently the same as Grierson's "Hindki of Hazara". In the hills and mountains west and southwest of Kashmir (Pir Panjal) are **Chibhālī** and **Punchhī**. This northern area especially stands in need of more work, starting with an up-to-date survey.³ Parts of Hazara are now Pashto-speaking.

2.1.14 This brings us to **Kāshmirī** itself, essentially the language of the Vale of Kashmir, certainly not of the whole state of that name, the greater part of which (Ladakh, Baltistan) is Tibeto-Burman-speaking. Kashmiri influence, however, or the same tendencies that are shown by Kashmiri, are perceptible in bordering Indo-Aryan languages of both the "Lahnda" and "West Pahari" varieties. To the southeast on the Upper Chenab lies the smaller valley of Kishtwar, the language of which, **Kashtawārī**, has been called "the only true dialect" of Kashmiri. Beyond is the Bhadrawahi group of West Pahari mentioned earlier. Other dialects/languages of the Kashmir group lie between Kashmir and Jammu: **Pogulī**, **Doḍā Sirājī**, and **Rāmbanī**.

2.1.15 The language of Jammu itself is **Ḍogrī**, once considered a

“dialect” of Punjabi, now thought to be more closely related to West Pahari, and in any case now claiming independent language status. To the southeast in western Himachal Pradesh is the closely allied **Kāngrī**.

2.1.16 Finally we come to **Punjabi**, on the northwestern flank of the Hindi area, the language not only of Punjab State in India, but also of a major element in the population of Pakistan – some would say the “dominant” element, but this assessment is confused by continued use of the term “Punjabi” by some to cover both Punjabi and “Lahndi” speakers. Grierson fixed the boundary between “Lahnda” and Punjabi, admittedly poorly defined, at a line running north–south through Montgomery and Gujranwala Districts, west of Lahore, that is, well within Pakistan. (Following Shackle, we may call the Punjabi-speaking Lahore–Gujranwala–Sialkot area Central Punjab.)

Whatever validity Grierson’s line may once have had has no doubt been disturbed by the great movements of population associated with Partition. However, H. Bahri seems to have been wrong in his prediction (1962: x) that Partition would have the eventual effect of shifting the uncertain boundary of “Lahndi” eastward to the new international frontier, presumably because Punjabi speakers in Pakistan would be cut off from influence from the main centers of the language in Eastern (Indian) Punjab. The reverse seems to have happened. Not only has Lahore proved to be a sufficiently strong center of Punjabi in its own right (see Shackle 1970), but the position of Punjabi in Pakistan in general has been strengthened by the large number of refugees from Eastern Punjab following Partition, as it had been earlier by the resettlements in the new Canal Colonies. These involved an influx of Punjabi speakers into the Siraiki-speaking area (to which the “Siraiki movement” is in part a reaction).

On the Indian side also, the situation is confused by the increasing identification of “Punjabi” with Sikhism, and the partly successful campaigns of the Arya Samaj to persuade Punjabi-speaking Hindus to return their mother tongue in the census as “Hindi”. This is not to say that many Punjabi-speaking Hindus do not identify with the language also, but the number of speakers and their area of settlement is larger than official statistics indicate. Again, an up-to-date objective survey of the situation on both sides of the border is very desirable, but is unlikely for political reasons to be undertaken in the near future.

2.1.17 Although with Punjabi the circle is completed, there remain to be mentioned a number of Indo-Aryan languages northwest of “Lahndi” and Kashmiri, more or less contiguous with the main group (i.e. except where interrupted by recent intrusions of Pashto) but in important ways outside their

“orbit”, culturally and historically. These are generally grouped together as “**Dardic**”. The most important is **Shina** (Şiṇā), spoken in several dialects in the basin of the Upper Indus (Chilas) and its tributaries (Kishenganga, Astor), beyond the mountains to the north of the Kashmir valley, from Gilgit to Palas in Indus Kohistan, that is to say, mainly in Pakistani territory. West of the Indus in Swat Kohistan are found **Bashkarīk** (= **Gāwrī**, in the Panjkora valley and at the headwaters of the Swat), **Maiyā** (on the right bank of the Indus, with a dialect **Kanyawālī** isolated in Shina territory in the Tangir valley to the northwest), and **Torwālī** (in the Upper Swat valley). Further west again, across another range of mountains, is the large Chitral valley, where the main language is **Khowār**. **Kalasha** survives in side valleys of southern Chitral. **Phalūra**, an archaic dialect of Shina, is or was spoken in some villages in southern Chitral. (Most of these, that is, excluding only Gilgit and Chilas, are presently in the northern reaches of the Northwest Frontier Province, Pakistan.) **Gawar-Bātī** is spoken on the Chitral-Afghan border, centering at Arnawai, where the Chitral and Bashgal rivers unite to form the Kunar. Other, already-fragile linguistic fragments, **Ningalāmī**, **Grangalī**, **Shumāshṭī**, **Kaṭārqaḷā-Woṭapūrī**, **Sāvī**, **Tirāhī**, discovered by researchers in single villages in eastern Afghanistan, in some cases spoken by only a few families (or even a few old men), often a generation or more ago, may no longer exist, but are important for the linguistic history of the region. The encroaching language is everywhere Pashto. A larger collection of now mutually incomprehensible dialects, spoken further into Afghanistan in scattered valleys north of the Kabul river from the Kunar (Chigha Sarai) as far west as the Panjshir, centering in the region known as Laghman, constitutes what is left of the **Pashaī** language, apparently once much more widespread.

2.1.18 [In remote valleys higher up in the Afghan Hindu Kush are several additional languages, before the conversion of their speakers to Islam at the end of the nineteenth century collectively called “**Kāfirī**”, a term now replaced by “**Nūristānī**”, which were once grouped with “Dardic” on the basis of inadequate information. I follow more recent scholarly opinion (Morgenstierne 1961, 1973, Strand 1973, Fussman 1972, Buddruss 1977, Nelson 1986) in treating them as a group separate from Indo-Aryan, but it seems appropriate to mention them here. From east to west, they are: **Katī** (= **Bashgalī**) in the Upper Bashgal valley, with small enclaves in Pakistani Chitral, and the dialect **Kamviri** lower on the Bashgal (Kamdes); **Tregāmī** in three villages (as the name indicates) further to the southwest, between the Pech and the Kunar; **Waigalī** (= **Kalasha-alā**) in the Waigal valley (a northern tributary of the Pech); **Prasun** (= **Veron** = **Wasi-weri**) in six villages in the high valley of the Upper Pech; **Ashkun** between the Pech and

the Alingar. In the drainage of upper tributaries (Kulum, Ramgel) of the Alingar and also of the Ktivi tributary of the Pech, Kati is again spoken, its continuity broken by Prasun. The whole Nuristan area was conquered by the Afghans only in 1896.]

2.1.19 The *non-contiguous* Indo-Aryan languages, that is, those based *outside* the contiguous Indo-Aryan area, may be listed as follows: **Sinhalese**, the principal language of Sri Lanka; **Maldivian** (= **Divēhī**), the related language of the republic of the Maldives (an archipelago in the Indian Ocean southwest of India); **Saurāshtrī**, the language of a community of silk-weavers centered at Madurai in the Tamil country; **Dakhinī**, a southern form of Urdu, insofar as it is centered at Hyderabad in the Telugu area; **Parya**, an Indo-Aryan language recently discovered in Soviet Central Asia (Tadzhikistan). Like the outlying dialects of **Konkanī** (and apparently also **Khetrānī** in Baluchistan) mentioned earlier, all of these are the result of pre-modern migrations of Indo-Aryan speakers – in the case of Sinhalese, as early as the fifth century B.C. More recent migrations (i.e. both overseas and within India and Pakistan) have not yet resulted in distinct languages (and under modern conditions are not likely to), but unique *koinés* have arisen in the course of the colonial experience in Trinidad and Fiji.

2.1.20 There remains a third category of Indo-Aryan languages to be noted, partly overlapping with the above (i.e. in some cases also non-contiguous) – those with *no specific territorial base*. The most important of these is **Urdū**, the language first of the Muslim population, mainly urban, of northern India; now the official language of Pakistan and a second language for all educated persons there; the southern form **Dakhinī**, mentioned above as having a base at Hyderabad, is also found spoken (along with Standard Urdu for formal purposes and by the more educated) by Muslims in cities and towns throughout the Deccan, and in Bombay.

Other such languages are **Gojri** (or **Gujuri**), spoken by semi-nomadic herdsmen found scattered at higher elevations in the hill areas mainly of Jammu-Kashmir (especially Punch District) and adjoining regions of Pakistan and on into Afghanistan; **Lamānī** (= **Banjāri** = **Lambādi**), spoken by another nomadic people (nowadays engaged mostly in construction) found primarily in Andhra Pradesh, Karnataka, and Maharashtra; both are of Rajasthani affinity. Finally there is **Romany**, the language of the Gypsies, not only non-contiguous but extra-Indian as well as non-territorial, although (as with the others) there are marked

concentrations in certain areas, in this case in Eastern Europe (Yugoslavia, Eastern Slovakia).

2.2 "Language" vs. "dialect" in the Indo-Aryan context

We have managed to complete the brief survey above without really confronting a problem which nevertheless did unavoidably obtrude itself from time to time, namely the distinction between a *language* and a *dialect*. A few words may be said about it now.

The problem is that although the distinction is a common and indeed often a useful one, there is no generally accepted criterion for making it. Both terms are used, not only popularly but also by scholars, in several conflicting ways. There are two common senses in which the meaning of one term is linked with the other.

In Sense A, a *dialect* is a subvariety of a larger unit, which is typically a *language*. (It may in turn be subdivided into smaller units, or *subdialects*. These terms have equivalents in Indo-Aryan languages, e.g., H. *bhāṣā* 'language', *bolī* 'dialect', *upabolī* 'subdialect'.)

In Sense B, a *dialect* is unwritten, while a *language* possesses a **written "standard"** and a literature. (This distinction is then undermined by the usage "*the literary dialect*" in situations of diglossia, such as obtain in Sinhalese or Bengali. Inasmuch as this refers to a *subvariety* of a language, even if of a special kind, it may be said to hark back to Sense A.) To be sure, a (non-literary) dialect may also be written down (= transcribed) but this does not turn it automatically into a "language" in this sense: it should also have a (written) literature and a measure of official and cultural recognition, both elastic concepts. It is clear that the entailed status comes and goes, however, and therefore is primarily sociocultural rather than linguistic in nature. In contemporary India and Pakistan several erstwhile *dialects* (Dogri, Siraiiki) are said to be "agitating for *language* status". Meanwhile, one-time literary *languages* such as Braj and Awadhi are said to have "reverted to *dialect* status" (Khubchandani 1983: 27, 168; his term is *vernacularization*).

Even on one side of the unclear boundary between *dialect* and *language* in Sense B, there are differences: one speaks of "developed" and "undeveloped" languages. Such differences are in part linguistic, involving the development of certain specialized registers. While in principle not unquantifiable, such differences are more clinal than absolute. In modern Indo-Aryan every part of the cline is represented, depending on the length of time the *language* has been cultivated and under what circumstances. Thus Modern Standard Hindi, with its official status at two levels (provincial and national), has more developed registers than,

say, Siraiki, until recently cultivated only for religious poetry, or Khowar, which has only recently been cultivated at all.

In view of the slipperiness of Sense B, it might appear that Sense A is the preferable one for the scientific study of language (using the latter word now in a third sense, Sense C). The terms of Sense A, however, are often taken (mainly in academic usage itself) to be purely relative, with different applications at different levels of abstraction: *x* is a dialect of language *L*, which is in turn a "dialect" of construct *G*, etc.; e.g., Sambalpuri is a dialect of Oriya, which is a "dialect" of Magadhan, which is a "dialect" of Indo-Aryan, which is a "dialect" of Indo-Iranian, which (like Germanic, Italic, etc.) is a "dialect" of Indo-European. Although it is the term *dialect* which suffers most, the higher constructs in each case might logically be called "languages", leaving us, it would appear, with no definable level of application for either term.

Even if it is granted that such usages are metaphorical extensions of terms normally and properly applied to a language and its subvarieties, there is unfortunately no universal criterion of linguistic distance for *languages* as against *dialects*, that is, of how different a speech-variety has to be from another to qualify as a separate *language*. Not that attempts to come up with such a measure have not been made. Nigam (1971: xxv-xxvi), for example, perhaps taking a cue from lexicostatistics, suggests that speech varieties sharing 81 per cent or more of **basic vocabulary** should be classed as *dialects*, less than 81 per cent as *languages*. H. Bahri (1980: 1-2), recognizing that "mutual intelligibility" is a relative rather than an absolute concept, suggests a more subtle breakdown: **mutual intelligibility** around 10 per cent = two languages historically related but geographically removed (Punjabi and Gujarati); up to 25 per cent = two languages in long cultural contact (English and French); 25-50 per cent = can be called "languages" or "dialects" (Rajasthani and Hindi); 50-75 per cent = two dialects (Braj Bhasha and Bundeli); around 90 per cent = subdialects (Sargujiya and Bilaspuri).⁴

No one has to my knowledge seriously attempted to apply either Nigam's or Bahri's criteria to problems of language and dialect identification in Indo-Aryan. Mutual intelligibility is an especially tricky concept to apply in a multilingual society such as that of South Asia, where familiarity (i.e. various degrees of "passive bilingualism") as well as purely linguistic distance must be reckoned with. Any attempt to apply it must reckon also with judgments like Morgenstierne's (1962: 21-4) that the Pashai dialects are "decidedly one language" despite their mutual *unintelligibility*, because they are "well-defined through phonetical, and especially through morphological and lexical peculiarities." (Speakers of the geographically fragmented Pashai dialects have few opportuni-

ties for contact and for thus acquiring that degree of passive bilingualism that is often a component even of *interdialectal* mutual intelligibility.)

What happens in practice, of course, is that in parts of the world where a clean slate is not available for these exercises, rather than attempt them even linguists fall back for the most part on the conventional "languages" of Sense B, whatever their mutual linguistic distance, and whether mutually intelligible to a significant extent or not, for identifying the *dialect groupings* that are treated as languages in Sense A. In a continent like Europe, blessed with well-defined peninsulas and islands, and where the nation-state has become the norm, this becomes problematic linguistically only at a few points, such as the Dutch-German and Franco-Italian borders, where there is a true dialectal continuum. Elsewhere *language* in Sense A and Sense B correlates fairly well with geographical and political units.

South Asia, which bears many analogies to Europe, differs from it radically here: it is *shaped* differently. Lacking clearcut geographical units of the European type where dialectal variants can crystallize in semi-isolation, or longstanding political boundaries, the entire Indo-Aryan realm (except for Sinhalese) constitutes one enormous dialectal continuum, where continued contact inhibits such crystallization, and differentiated dialects continue to influence one another. The speech of each village differs slightly from the next, without loss of mutual intelligibility, all the way from Assam to Afghanistan. Cumulatively the differences are very great, but where do we draw the dialect, let alone the language, boundaries?

A careful dialect geography would no doubt show that the subdialectal continuum in fact does not present a uniform gradient, but is punctuated by both smaller (dialectal) and greater (language) bundlings of isoglosses. The *LSI* does not really constitute such a dialect geography, but it is a step in that direction. The region is not totally devoid of natural barriers – for the most part consisting of rough hill country.

Superimposed on this ground pattern are the literary languages of Sense B and their culturally-defined orbits. The relation of these to languages in Sense A is often problematic. Thus the Rajbangsi dialect of the Rangpur District (Bangladesh), and the adjacent Indian Districts of Jalpaiguri and Cooch Behar, has been classed with Bengali because its speakers identify with the Bengali culture and literary language, although it is linguistically closer to Assamese. So has the Chittagong dialect of southeastern Bangladesh, which differs from Standard Bengali more than Assamese itself does. There are limits to this, however: although Urdu is the preferred literary language of Kashmir and of Pakistani Punjab, no one would take Kashmiri or Punjabi to be dialects of Urdu (or of Hindi-Urdu).

As indicated in the preceding section, the real problem is with the vast "Hindi area", defined as the area within which Modern Standard Hindi is today the accepted written language. Are all forms of Indo-Aryan speech within it "dialects of Hindi"? Rejecting this as intuitively too much at variance with the proper scope of a *language* in Sense A (and not having to reckon with the subsequent further consolidation of the status of Hindi), Grierson proceeded to set up, as noted above in section 2.1.1, several artificial constructs at the level of "languages" in Sense A that he felt were needed to make linguistic sense of the situation: "Eastern Hindi", "Western Hindi", "Rajasthani", and "Bihari". (He also used the term "Pahari", but always with reference to a group; never, it seems, in the sense of "a" language.) The first two did not catch on at all; "Rajasthani" and "Bihari" did trickle down to some extent into popular usage – to the annoyance of Nigam who remarks regarding the census of 1961 that the terms should be discouraged: it "is not useful to have a blanket name," which only confuses the statistics. (The first is most often used, however, as a synonym for Modern Literary Marwari and the second for Magahi – usages which may owe nothing to Grierson.) The majority of "Bihari" and "Rajasthani" speakers still report their mother tongues under more specific and traditional names – Maithili, Bhojpuri, Marwari, Dhundhari, etc. – or simply as Hindi.

Another such Griersonian language construct was "Lahnda", discussed in section 2.1.13 above. Elsewhere, "normal" taxonomic problems exist, sometimes complicated by politics, on a scale appropriate to the subcontinent: is Konkani a separate language or a dialect of Marathi? Is Halbi a mixture of Oriya and Marathi, a dialect of Marathi, or a separate language? Is Khandeshi a dialect of Marathi or of Gujarati, or a separate language?

Often such problems correlate with transition areas. Even at the subdialectal level, Grierson tried to distinguish what he regarded as "mixed" and unstable forms of speech characteristic of such areas from "true dialects", presumably part of the underlying gradient. Certain cases might seem particularly to call for such a distinction, but dialect or language mixture has in fact been involved in the formation of most of the major NIA languages to some extent also. It is difficult to know where to draw the line. "Stability" is perhaps the key to the difference, but the sociolinguistic and historical variables involved in such stabilization need further study. Dialectal differentiation in an area geographically like that of Indo-Aryan cannot proceed in a "pure" form, i.e. without the peripheral dialects running up against neighboring languages, in any case, and mixed dialects in the zones of transition between major languages (and mixed subdialects in the zones between major dialects) are an inevitable result.

"Mixed" forms of speech involving *non-Aryan* languages or substrata are

perhaps another matter, calling for special treatment as creoles. Such would include such "dialects" of Bengali as *Chākmā* (spoken in the Chittagong hills presumably by former Chin [Tibeto-Burman] speakers) and *Malpahāriā* (spoken in the Rajmahal hills by former Malto [Dravidian] speakers), as well as the aforementioned Halbi, whose speakers may or may not be former Gond [Dravidian] speakers (Trelang 1966: 359–60), and many others. The matter is complicated by the fact that, except for the first, these *also* typically involve transition zones between Indo-Aryan languages (Bengali/"Bihari" in the case of Malpaharia; Oriya/Marathi in the case of Halbi).

2.3 Hindi and Urdu

The ultimate anomaly in the what-is-a-language dilemma in Indo-Aryan is presented by the Hindi–Urdu situation. Counted as different *languages* in sociocultural Sense B (and officially), Urdu and Modern Standard Hindi are not even different dialects or subdialects in linguistic Sense A. They are different *literary styles* based on the *same* linguistically defined subdialect.

At the colloquial level, and in terms of grammar and core vocabulary, they are virtually identical; there are minor differences in usage and terminology⁵ (and customary pronunciation of certain foreign sounds), but these do not necessarily obtrude to the point where anyone can immediately tell whether it is "Hindi" or "Urdu" that is being spoken. At formal and literary levels, however, vocabulary differences begin to loom much larger (Hindi drawing its higher lexicon from Sanskrit, Urdu from Arabic and Persian), to the point where the two styles/languages become mutually unintelligible. To the ordinary non-linguist who thinks, not unreasonably, that languages consist of words, their status as different *languages* is *then* commonsensically obvious, as it is from the fact that they are written in quite different scripts (Hindi in Devanagari and Urdu in a modified Perso-Arabic).

The latter is a factor of peculiar importance in *language-B* status in South Asia that has not yet been discussed: there is a widespread feeling that a self-respecting language should have a distinctive script (see Chapter 6). Some readers may be drawn to make a comparison with Serbo-Croatian (written in Roman and Cyrillic scripts), but the analogy is not quite apt: there are grammatical differences between these two – for example involving the use or non-use of an infinitive – which are not found in Hindi–Urdu, while on the other hand the lexical differences are not so massive and systematic. The Hindi–Urdu situation is apparently unique in the world.

What, then, is the subdialectal base of these two standard languages? Not surprisingly, it was that of the capital, Delhi, sometimes referred to as *Dehlavi*. It

is often called **Khari Boli** (among various etymologies: < H-U. *kharā* 'standing', hence > 'stand[ard dialect]'). This would be appropriate, if this term were not also frequently applied to the country dialects north of Delhi, which present a number of phonological (/n, l/) and other features not found in the standard of the capital. (As noted in 2.1.1 above, Sankrityayan and Bahri therefore propose to call the former "Kauravi", reserving "Khari Boli" for the Delhi-based standard language, wherever it may be spoken.)

Like urban speech everywhere, and especially that of capitals, the language of Delhi was not based on one dialect in any case, but on a dialectal composite. Thus, along with "Kauravi", Hariyanvi, Punjabi, Rajasthani (Mewati), Braj and other influences have gone into the making of Khari Boli – the last especially during the century (1566–1658) when the imperial capital moved to Agra, in the heart of the Braj country. (Some scholars believe this dialectal fusion took place earlier in the Punjab, i.e. Lahore, which had been under Muslim rule for nearly two centuries, and was then brought to Delhi with the Muslim conquest in 1193, but the evidence for this is very thin from this remote and unsettled period. The proximity of the city to the Punjab is probably sufficient to explain the "Punjabi" elements in Khari Boli.)

This "standard" dialect was moreover not precisely equivalent to the speech of Delhi as such, but more specifically to that of certain *classes* and *neighborhoods* most closely associated with the Mughal court and its predecessors. Although reference to the latter fact is often made, based on statements in literary sources, what precisely this might mean in linguistic terms has not been spelled out. More importantly, a careful *linguistic* analysis of the aforementioned dialectal mixture has yet to be made.⁶ Rai (1984), while fascinating, is deliberately "non-technical". He does make the observation, however, that at least at what he calls the Old Hindi stage (thirteenth–fourteenth centuries?) the contributing dialects

were all in their initial, formative stage, when their identities were not sharply defined – and therefore mixing was easy . . . Any attempt to divide them or to contrapose them one to the other is likely to confuse the linguistic picture of the times altogether and get the researcher tied up in a whole lot of quite intractable problems.

(p. 123)

Tiwari (1961) is more concerned with Hindi in relation to the general history of Indo-Aryan.

Once Khari Boli had taken on a stable shape in the capital, and had spread far and wide as a lingua franca, there were other influences, essentially superficial, on

its *literary* development in later centers of literary activity such as Lucknow (Urdu) and Allahabad (Hindi). These too need to be investigated from a linguistic point of view. In summary, it could be said that although the *sociopolitical* history of Hindi and Urdu has been much studied and commented upon (see also Narula 1955, Brass 1974, Barannikov 1972, Chernyshev 1978), a proper *linguistic* history of them (as distinct from their MIA and OIA antecedents) still needs very much to be written.

Often enough even accounts of sociopolitical history are distorted by the attempts of partisans of one language/style or the other to establish its priority. In this they are aided by terminological confusion. Is Modern Standard Hindi really Urdu in Devanagari script relexified with Sanskrit *tatsamas* (see Chapter 4)? Or is Urdu really Hindi in Perso-Arabic script from which the *tatsamas* have been purged and replaced with Perso-Arabic terms? Both assertions will be found in the literature on the subject.

On the one hand, it is no doubt true that British administrators (and missionaries) played a role in promoting and even creating Modern Standard Hindi at the beginning of the nineteenth century by encouraging the development at Fort William College, in place of the old and limited Braj literary language, of a new prose standard in the Nagari script "*on the basis of Urdu*" – that is to say, on the basis of Khari Boli. This was in recognition of the fact that Urdu had conveniently spread as a lingua franca (as well as preferred language of the Muslim population) wherever in India Mughal influence had been felt,⁷ as well as of the fact that the higher literary style of Urdu had evolved into something remote from Indian life, unintelligible to the masses, and that its script was not originally designed for an Indo-Aryan language, difficult to master, and not suitable for printing.

On the other hand, Urdu was not called *Urdu*⁸ until around 1800.⁹ In fact in earlier Urdu writing itself it was often called *Hindi*! But this is a term, as we have seen, of very different implications for different people. To the aforementioned Urdu writers, *Hindi* or *Hindavi* undoubtedly meant 'the language of India' – which for them happened to be Khari Boli, as contrasted with Persian, the language of the Muslim establishment. For protagonists of Modern Standard Hindi, however, the term includes all the earlier indigenous literary traditions of the present "Hindi area", predominantly in dialects (or languages) other than Khari Boli. Ordinary people in the area, particularly in Uttar Pradesh, also commonly call their non-Khari Boli spoken languages "Hindi". Since *Hindi* as a term is of Muslim (Persian) origin, derived from the river *Sindhu* 'Indus' (and meaning originally simply 'Indian'), it would be interesting to know just when and how they came to do this. In any case, if the linguistic history called for above is to be coherently written, it must have a clear focus – i.e. Khari Boli itself, with

reference to the other languages of the "Hindi area" only insofar as they impinge on it.

Meanwhile, although for the British administration and many others the terms *Urdu* and *Hindustani* were essentially equivalent, *Urdu* in the eyes of some of its protagonists took on a special connotation of stylistic refinement and could not refer to "plain" Khari Boli/Colloquial Hindustani. (Rai [1984] refers to this development as "New Urdu".) Whether Urdu can maintain such a luxury in its new function as the national language of Pakistan remains to be seen, although proponents of this view are not wanting.

Many complex social and political forces, which we cannot go into here, have conspired to pull the two "styles" ever further apart. Their identity as separate languages may now be regarded as a cultural fact, however anomalous linguistically.

2.4 Nomenclature

Although European languages present a few instances of multiple or fluctuating names (e.g. *Ruthenian/Little Russian/Ukrainian*), these have now been largely sorted out. Linguistic nomenclature in the Indo-Aryan field, on the other hand, still constitutes a boulder-strewn path over which one must pick one's way carefully. Nomenclature complicates the Hindi-Urdu situation, as we have seen. (It is in fact even more complicated than just described: besides the once-ubiquitous *Hindustani* (now seldom used), the more specific *Dakani* or *Dakhini*, and the earlier *Hindui* and *Hindavi*, there was also *Rekhta* (< Pers. 'mixed' = 'the Hindustani or Urdu language' [Platts 1965 (1884)]), and its specialized feminine counterpart *Rekhti* '[imitated] women's speech'. "Hindi" in the broader sense, referring to all the speech varieties of the Hindi area, is of course equivalent to a plethora of more specific names.)

Elsewhere in Indo-Aryan, the name for a language or dialect one encounters may be its current official name (*Hariyanvi*), a popular name (*Laria* for Chhattisgarhi), its former name (*Bangaru* for Hariyanvi), a newly emerging name (*Siraiki*, *Angika*), a nickname bestowed by others (*Chhikacchiki Boli*, *Jangli*, *Hakkipikki*), or a name with no popular currency bestowed by a researcher (*Lahnda*, *Central Eastern Rajasthani*). It may be the name, real or fancied, of a community, such as a caste, applied to the language it speaks: *Jatu* for Hariyanvi, *Jatki* for several subdialects of "Southern Lahnda", where Jats or Jatts are numerous; *Ahirani* (from the Ahirs, a caste of dairymen) for Khandeshi. Especially in the case of migrants long-established outside the territory of their mother tongue, there is a tendency to draw their identity from the fragment of their former society they can still see before them – a caste group, an occupation, or a remembered

locality. Thus small groups of Marathi speakers outside Maharashtra return their language under a number of strange names – *Bare*, *Burdi*, *Kamari*, *Koshti*, etc. Some Gujarati speakers in Mysore return their language as *Kshatriya*, while *Kshatri* is a name given to a form of Hindi spoken in Andhra Pradesh. *Khatri* is also an alternative name of Saurashtri, in Tamilnadu. (All three are from *ksatriya*, the warrior caste.)

Many languages/dialects have several names: thus *Hariyanvi/Bangaru/Jatul/Deswali*, and *Khandeshi/Ahirani/Dhed Gujarati*. Political changes often have a surprisingly immediate effect on language names: with the dissolution of the old native state of Keonthal, near Simla, the major Pahari dialect name *Kiunthali* seems to have disappeared, and been replaced by the new coinage *Mahasui*, from the new District of Mahasu (in Himachal Pradesh).

More problematic for census takers is the situation where a single name is used for more than one language/dialect. There are at least four different sub-Himalayan dialects called *Siraji* (see entries in Appendix I). *Dangi* is a dialect of Braj in northeastern Rajasthan and a dialect of Khandeshi (or of Bhili) in south Gujarat. *Thali* is a dialect of Marwari in western Rajasthan, and a northwestern dialect of Siraiki in Pakistan. Significantly, these names are taken from common topographic features: *sirāj* 'mountainous country', lit. 'Shiva's kingdom' < *śiva-rājya*; *thal* 'desert' (also *thar*; cf. *Thareli*, a desert dialect of Sindhi); *dāng* 'heavily forested hill country'. *Pahari* (< *pahār* 'mountain') is another such non-specific topographic term. One must be careful not to jump to conclusions, however. The name *Doabi*, for instance, refers not to the (Braj) dialect of the best known *doāb* ('interfluvium') between the Ganges and Jumna in western UP, but to the (Punjabi) dialect of the Jalandhar *doāb*, between the Beas and the Sutlej.

The census often tells us something regarding the name speakers prefer for their language. For instance, they overwhelmingly prefer the old name *Dhundhari* to the more transparent *Jaipuri*, and *Marwari* to the more grandiloquent *Rajasthani*, although the latter has made considerable headway.

