

Dialects in the Indo-Aryan landscape

Ashwini Deo
Yale University

For The Handbook of Dialectology, ed. by Charles Boberg, John Nerbonne, and Dominic Watt

1. Introduction

The Indo-Aryan language family currently occupies a significant region of the Indian subcontinent, its member languages being spoken in the bulk of North India, as well as in Pakistan, Bangladesh, Nepal, Sri Lanka, and the Maldives. The historical depth of the textual record and the geographical breadth of the Indo-Aryan linguistic area, the diversity of its languages (226 in all), and its many speakers (~1.5 billion in number) all serve to make Indo-Aryan a complex object of linguistic investigation. This chapter offers an overview of the broad structure of the Indo-Aryan family and a classification of its major member languages, tracing briefly the historical record which leads to its synchronic distribution. It is crucial to note that Indo-Aryan is not one language, and a comparative study of the “dialects” of Indo-Aryan necessarily involves a (historical) comparison between several mutually unintelligible languages – many of them with millions of speakers, deep literary records, and complex dialectal differences *within*. The level of resolution at which variation within Indo-Aryan can be considered in this brief survey is thus different from the *intra-linguistic* level at which dialectal variation is usually studied.

The presence of Indo-Aryan (whose closest relatives within the Indo-European language family are the neighboring Iranian languages) in the Indian subcontinent can be dated back to approximately the early second millennium B.C.E. The influx of Indo-Aryan speakers first occurred in the mountainous areas of Afghanistan and Pakistan and the river plains of Punjab, with gradual migrations eastward and southward over the millennia. Within the early phase of Indo-Aryan expansion, we can identify the initial geographical center as the Upper Indus valley (now in Pakistan) and the later center (towards the end of the Vedic period) to be the Gangetic plains of North India.¹ By the time of the Buddha (6th century BCE) most of North India (i.e. north of the Vindhya mountain ranges and the Narmada river) was Indo-Aryan speaking, these groups having displaced the original languages of the region, which included Dravidian and Austro-Asiatic languages as well as languages of unknown stock (see Witzel 1999 for a detailed review of substrate evidence). Gradually, over the next millennium and a half, Indo-Aryan further spread towards the South, occupying areas south of the Narmada river (the region corresponding to the Marathi and Oriya speaking territory in the map in Figure 1.). It is this contiguous geographical territory over which the modern dialectological landscape of Indo-Aryan is to be found. The non-contiguous Indo-Aryan languages (which include, for instance, Sinhala (Sri Lanka), Divehi (Maldives), Parya (Tadjekistan), and Romani (mainly Eastern Europe)) are the result of pre-modern migrations of Indo-Aryan speakers into non Indo-Aryan territory (Masica 1993: 22).

Fifteen of the twenty-two official languages recognized by the Eighth Schedule of the Indian Constitution are Indo-Aryan. This status allows the use of these languages for both educational and regional administrative purposes. Pakistan and Bangladesh recognize only Urdu and

This research was supported by NSF grant BCS-1255547.

¹Most of our evidence for the geographical location and onward trajectory of the Indo-Aryan speaking population from the North West comes from the texts of the Vedic period which mention river names and tribal boundaries.

Bangla respectively as their official languages.² Nepali is the official language in Nepal, Sinhala is the official language of Sri Lanka, and Dhivehi the official language of the Maldives. The scheduled Indo-Aryan languages of India include Assamese, Bangla, Dogri, Gujarati, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, and Urdu. Of these, Hindi (written in Devanagari script) has special status as the official language of the Union of India. The speaker populations for these languages ranges from 422 million (Hindi) to 2 million (Dogri).³ All other Indo-Aryan varieties are classified as “non-scheduled” or unofficial languages, and accorded the status of “mother-tongue” varieties or dialects of the regional standard variety.

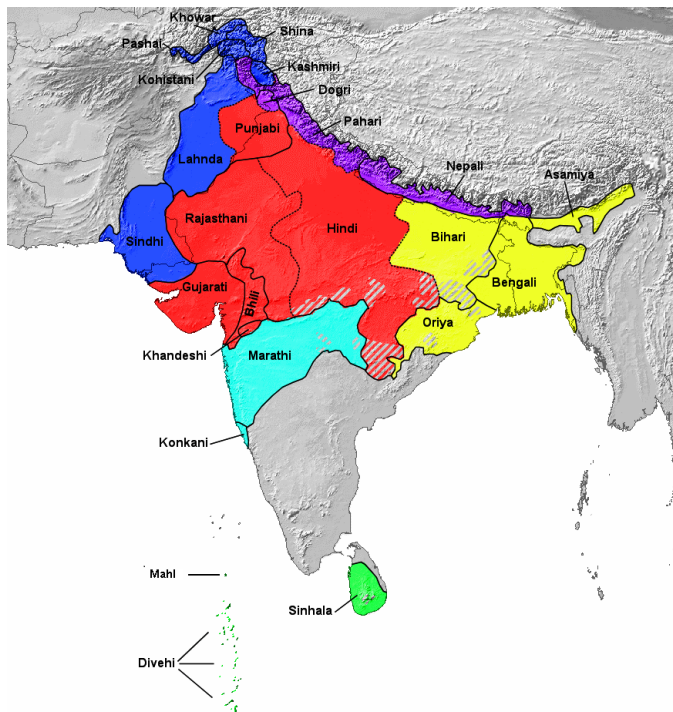


Figure 1: The distribution of the Indo-Aryan languages

2. The historical record of dialect variation

Chronologically, the evolution of the Indo-Aryan language family can be divided into three main stages: the Old Indo-Aryan (OIA) stage, which includes early and late Vedic texts as well as the later Epic and Classical Sanskrit literature; the Middle Indo-Aryan (MIA) Prakrit stage, with the textual record for Prakrit starting around the 3rd century BCE and continuing until the 9th century CE; and the New Indo-Aryan (NIA) stage, whose literatures are attested from the 12th century CE and whose modern forms constitute the current Indo-Aryan landscape.

The earliest clear evidence for regional dialect variation within Indo-Aryan languages

²Despite the presence of sizeable Punjabi, Siraiki, and Sindhi speaking populations in Pakistan, these Indo-Aryan languages do not have official status for educational and administrative purposes.

³http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement4.aspx

comes from the Ashokan inscriptions, which are dated to the 3rd century BCE.⁴ The language of these inscriptions exemplifies the Early MIA stage and is commonly known as inscriptional Prakrit.⁵ On the basis of the distribution of certain linguistic innovations observed in the inscriptions, Bloch (1950) identified three main geographical dialect areas – the Eastern dialect, the Northwestern dialect, and the Southwestern dialect. Southworth (2005) resolves the observed three-way distinction in the Ashokan inscriptions to an earlier binary division between the Northwestern dialect on the one hand and the Eastern and the Southwestern dialects on the other. The shared innovations that have been presented to motivate this classification include:

- (1) a. The three-way distinction between OIA sibilants *s* ([s]), *ś* ([ʃ]), and *ṣ* ([ʂ]) is neutralized to [s] in the Eastern and Western dialects, but retained in the Northwestern dialect.
- b. OIA syllabic *r* (phonetic [r̥]) is vocalized to *a* in the Eastern and Western dialects but to *i* in the Northwestern dialect. (e.g. OIA *kr̥ta* → Eastern MIA *kaṭa*, Western MIA *kata*, but Northwestern MIA *kiṭ(r)a*).
- c. OIA consonant clusters of the form *C_ir* undergo assimilation in the Eastern and Western dialects and change to *C_iC_i*. (e.g. OIA *agra* ‘foremost’ → Eastern MIA *agga*, Western MIA *agga*, but Northwestern MIA *agra*).
- d. OIA consonant clusters of the form *rC_i* undergo assimilation in the Eastern and Western dialects and change to *C_iC_i*; these undergo metathesis in the Northwestern dialect (e.g. OIA *garbha* ‘womb’ → Eastern MIA *gabbha*, Western MIA *gabbha*, but Northwestern MIA *grabha*).

Innovations that are exclusively attested for the Eastern dialect include the following:

- (2) a. The merger of OIA *n* ([n]), *ñ* ([ɲ]), and *ṇ* ([ɳ]) to *n* ([n]) (e.g. OIA *jjātika* ‘kinsman’, *dāna* ‘gift’, *prāṇa* ‘breath’ → Eastern *nātika*, *dāna*, *pāna*). These are retained in the Western and Northwestern dialects.
- b. The full merger of OIA [r] and [l] to [l] is characteristic of the Eastern dialect. (e.g. OIA *rājā* ‘king’, *rūpāṇi* ‘spectacles’ → *lājā*, *lūpāni*.⁶
- c. The masculine/neuter nominative singular ending *-aḥ* changes to *-e* while it is retained

⁴The Ashokan inscriptions contain 33 edicts that are found in scattered locations in modern day India, Pakistan, Nepal, and Bangladesh. These edicts, commissioned by Emperor Ashoka of the Mauryan dynasty (269–232BCE), are standardly taken to define the territorial limits of the Ashokan empire. The edicts have a proselytising purpose and describe in detail Ashoka’s belief in Buddhism, his attempt to spread Buddhism, and his social welfare projects. These edicts have been translated from an original Eastern dialect (Magadhan Prakrit) into other dialects that are taken to represent local Middle Indo-Aryan languages of the period.

⁵Witzel (1989) has attempted to trace dialectal differences even in early OIA – within texts of the post-Ṛgvedic period. This textual study identifies innovations that suggest differences between the Western and Eastern/Southern regions of the Indo-Aryan area of the period. These regions of differentiation can be traced to the path along which the Early Indo-Aryan settlement developed in Northern India.

⁶This particular change has been problematic for linguistic reconstruction in several ways. First, Vedic texts exhibit a partial merger of PIE *r and *l to r. Second, none of the later attested languages (including the modern counterparts in the same region) exhibit the full merger seen in the l-only Eastern dialect. The l–r variation, moreover, was consciously perceived as a sociolinguistic marker, where the use of r distinguished the speech of the Aryan Brahminical society and the use of l was stigmatized as non-Brahminical since the late Vedic period (Bloch 1965, Deshpande 1979, Southworth 2005:161–167). The total merger of r and l that appears in the Eastern inscriptions can be interpreted as the use of this variable to differentiate the Buddhist establishment from its brahminical counterpart. Southworth (2005: 166) suggests that this might reflect a positive stereotype, indicating a rejection of Brahminical speech norms.

(with some exceptions) in its allomorphic form *-o* in the Western and the Northwestern dialects.

Later evidence for dialect variation in MIA comes from traditional grammars and commentaries on the dramatic literature of Classical Sanskrit and Prakrit. These contain valuable evidence about the regional MIA dialects in the first millennium CE (better known as the literary Prakrits), their linguistic features, and their internal sociolinguistic stratification. The more important literary Prakrits include *Śauraseni*, associated with the Northwestern region, *Mahārāṣṭri* (the Prakrit *par excellence*), associated with the Southern region, and *Māgadhi*, associated with the East.⁷

The distinct dialectal divisions identified for Early and Middle MIA have implications for the subgrouping of the modern NIA languages to which I now turn. The question of subgrouping for Indo-Aryan is far from being resolved in a uniform manner. The *Stammbaum* model fails to work satisfactorily for this region. As Masica (1993: 446) describes it, this is a region in which there have been few internal natural barriers, unstable political units have not coincided with linguistic units, and significant internal migrations have taken place. The result has been dialect continua without sharp boundaries separating mutually unintelligible languages. Nevertheless, there are broad geographical divisions that can be identified within the current linguistic landscape and that correspond to the general territories of the modern NIA languages. These major languages include (Western) Hindi, with its multiple dialectal variants, and Nepali, which occupy the North-Central region of the subcontinent; Punjabi, Sindhi, and Lahnda to the Northwest; Rajasthani and Gujarati to the West; Marathi and Konkani to the South; and Bengali, Oriya, Assamese, and Eastern Hindi to the East.⁸

Hoernle (1880) proposed a classification of the Indo-Aryan languages that assumed two broad branches in remote times that gave rise to the contemporary languages – a Southern-Eastern branch (which grouped Marathi together with Bengali, Oriya, and Eastern Hindi) and a North-Western branch (grouping Western Hindi and Nepali with Punjabi, Sindhi, and Gujarati). This hypothesis was further refined by Grierson, whose Inner-Outer hypothesis closely builds on the affinities between the Southern and the Eastern languages observed by Hoernle. Grierson proposed a model involving two distinct waves of immigration into the subcontinent, one of which led to the settlement of Northern India and gave rise to Western Hindi and its dialects – the “inner group” – and another encircling wave that corresponds to languages surrounding this inner group in a semi-circle – the “outer” group. The hypothesis, like Hoernle’s, groups together the contemporary Eastern and the South-western languages (a grouping we already have evidence for from Middle Indo-Aryan dialect classification) and further posits that the North-Western languages such as Sindhi and Lahnda are closer to these than to Western Hindi. Grierson first posited this model *before* setting out on the ambitious Linguistic Survey of India, described in §?? (LSI 1.1: 116–118), and then further revised it after the survey (Grierson 1931–33). The revised model, in addition to the Inner and Outer branches, includes an Intermediate branch, whose members are considered to be “inner” languages superimposed on an “outer” substratum. These include Punjabi, Gujarati,

⁷Within the dramatic literature, we find explicit directions for the usage of different Prakrits. For instance, while male protagonists usually speak in Classical Sanskrit or *Mahārāṣṭri*, *Māgadhi* Prakrit is assigned to people of “lower status” such as (eunuchs, dwarfs, and chamberlains, among others), and *Śauraseni* is assigned to “(cultured) women and their lady friends” (Pischel 1900: 25–28). These examples are merely illustrations of the complex social connotations of dialectal variants observed in the first Millennium CE. For more detailed descriptions of the sociolinguistic connotations of the observed dialects see Deshpande (1979), Jain (1981) Pischel (1900) among others.

⁸The Dardic languages to the North (which include Kashmiri and Shina) form a separate subgroup that is sometimes considered to be an independent branch of Indo-Iranian and sometimes located within Indo-Aryan.

Nepali, Eastern Hindi, and Rajasthani. Grierson supports this picture of the linguistic geography of Indo-Aryan by providing a number of grammatical and phonological criteria along which variation patterns:

- (3) a. The Outer languages form their perfective aspect forms with an *-l* affix (first attested in late MIA) while the Inner languages retain the perfective paradigm derived from OIA.
- b. The Inner languages tend to be more analytic and retain fewer inflectional features than the Outer languages.
- c. The Inner languages preserve the distinction between OIA sibilants *s* ([s]), *ś* ([ʃ]), and *ṣ* ([ʂ]), which is reduced in different ways in the Outer languages.

Grierson's criteria, and indeed the entire Inner-Outer hypothesis, were severely criticized by Chatterjee (1926), according to whom Grierson had in some cases inaccurately represented the geographical distribution of features and in others was describing changes of relatively recent origin or cases of independent development.⁹ Later proposals, including Chatterjee's own, all described in Masica (1993), assume different divisions, which emerge out of varying criteria being privileged over others. Masica describes in some detail the incompatible classifications implicitly or explicitly provided by Turner 1975, Katre 1968, Cardona 1974, Nigam 1972 and comes to the conclusion that:

Perhaps a wiser course would be to recognize a number of *overlapping genetic zones*, each defined by specific criteria. (Masica 1993: 460)

...We might therefore be well-advised to give up as vain the quest for a final and "correct" NIA historical taxonomy, which no amount of tinkering can achieve, and concentrate instead on working out the history of various features. (Masica 1993: 460)

While this is a sensible caution against striving for sharp, Stammbaum-like classifications, Southworth's revival of Grierson's original construal of Indo-Aryan regional divisions deserves mention here. Southworth (2005: 135 ff) introduces linguistic evidence overlooked by Grierson and later scholars to adduce support for a Griersonian picture of dialect divisions.

- (4) a. The usage of the OIA gerundive in *-tavya* to mark future meaning is observed the Eastern and (South)-western languages in contrast to the North-Central languages.
- b. The North-Central languages have preserved some instances of phonemic contrast between long and short high vowels ([i]/[i:] and [u]/[u:]) while length in the Eastern and the South-western languages is entirely positionally conditioned (allophonic).
- c. Although there are exceptions, the Eastern and the South-western languages typically exhibit initial word accent. This contrasts with the pattern in the North-Central languages in which stress accent typically falls on the rightmost heavy syllable (modula final extrametricality). The North-Central pattern is taken to be the conservative one inherited from Late OIA and MIA. (Turner 1916: 47, Chatterjee 1926: 280).
- d. The lateral [l] changes to [n] in several lexical items in the Eastern and the South-western languages while the change is far less common in the North-central languages.

⁹See Masica 1993: 449–450 for an overview of Chatterjee's critique.

Southworth synthesizes his linguistic arguments with textual and archaeological evidence to build a convincing picture of the routes of expansion of Indo-Aryan speaking populations into the sub-continent and the resulting dialect space that has generated the patterns of distribution observed in the NIA languages.

3. Studies and Sources

The linguistic study of the NIA languages was first undertaken by the scholar-administrators of the East India Company/British Government – initially in the context of training British officers who came to the colony and later as part of the larger Orientalist pursuit. In addition to grammars of individual NIA languages that were written in the 19th century, Beames (1872–1879 [1966]) undertook a comparative investigation of seven major languages – Hindi, Punjabi, Sindhi, Gujarati, Marathi, Oriya, and Bangla. Hoernle (1880) represents the other main 19th century contribution to the regional taxonomy (and its historical implications) of Indo-Aryan.

The first systematic study of contemporary spoken Indo-Aryan languages and dialects was undertaken by George Abraham Grierson, through the monumental project, The Linguistic Survey of India (LSI), a comprehensive survey of the languages and dialects of British India.¹⁰ The survey was planned as a collection of specimens in which a standard passage was “to be selected for comparison and this was to be translated into every known dialect and subdialect spoken in the area covered by the operations” (LSI 1.1: 17). In order to access idiomatic language, this translation was to be complemented by a piece of folklore or some other passage in narrative prose or verse. There was further a list of words and sentences whose equivalents in each language were sought.

The results of the project, which was conducted by the (colonial) Indian Government between 1894 and 1928, is an 11 volume set of publications, of which 7 are exclusively dedicated to the Indo-Aryan languages. Grierson’s work depended crucially on the cooperation of local government officers in identifying dialect communities and collecting/correcting samples obtained from a large part of the subcontinent. For each regional variety, the LSI makes available a translation of *The Prodigal Son*, and for most varieties, also a brief (often single page) grammatical sketch that identifies their phonological and morphological properties. For many of the varieties, we also find the narrative passage and word/sentence lists. The survey distinguishes between standard and non-standard varieties of individual languages and further also identifies dialect continua that cannot be considered to be subsumed under a distinct regional language. The classification in the LSI builds on Grierson’s ideas of prehistoric dialect regions discussed in §?? but can also be taken to be a neutral synchronic description of the linguistic space (Grierson’s dialect regions are expressed through color coding in Figure 1). Bengali, Assamese, Eastern Hindi (Magahi), and Oriyā are assigned to the Eastern group, Marāthi and its dialects to the Southern group, Sindhi, Lahnda and the Dardic languages (including Kashmiri) to the North-Western group, while Western Hindi, Punjabi, Rajasthani, Gujarati are assigned to the Central group (which corresponds to the Outer group of the original proposal). This is also the grouping assumed in the main by the SIL Ethnologue.

An important outcome of the LSI was the identification and basic description of languages and dialect continua that could not be easily assimilated within any of the major regional languages.

¹⁰As Grierson describes it, the assembled scholars at the Oriental Congress in Vienna in 1886 passed a resolution urging upon the Government of India to undertake ‘a deliberate systematic survey of the languages of India’ (LSI 1.1: 17).

The Bhīl and Khāndeśī languages in the South-West and the Pahārī and Gujurī languages in the North-West are notable examples, to which are dedicated entire sub-volumes of the LSI (LSI 9.3 and LSI 9.4 respectively). Both of these language groups are assigned to the Central Group of Indo-Aryan and are represented by dozens of distinct, sometimes mutually unintelligible varieties.¹¹ The data generated by the LSI has been invaluable in terms of providing a high-resolution picture of the Indo-Aryan language and dialect space in the first decade of the 20th century and has not been fully replicated in its scope or linguistic rigor to date. The entire contents of the 11 volumes are available in searchable form at <http://dsal.uchicago.edu/books/lsi/>.

Another invaluable resource is Sir Ralph Lilley Turner's *A Comparative Dictionary of the Indo-Aryan Languages* published in four volumes (1962–1966). Turner, an army officer turned comparative philologist, assembled this dictionary based on his own and other scholars' work over a period of fifty years. The main dictionary contains 14,189 lexical entries, either attested in OIA or MIA or reconstructed, and provides their reflexes (as applicable) in a number of standard and non-standard NIA languages. The dictionary is available in searchable form at <http://dsal.uchicago.edu/dictionaries/soas/>.

For individual languages, there were some smaller-scale dialect surveys undertaken in the post-independence period. To mention a few: Ghatage (1962–68) documents several dialects of Marathi (including those spoken by diasporic communities) while Gill (1973) is a Linguistic Atlas of Punjabi dialects. For Bengali, there are descriptions of isolated dialects (Chaudhuri 1939, 1940; Sen 1972; Goswami (1939)) but no unified survey apparatus that identifies sub-regional dialect patterns and relevant linguistic variables. Gusain (2000–08) is a series of grammars of Rajasthani dialects that are compiled using a set template enabling comparability and identification of points of variation. The dialectological resources for Hindi are quite scattered, especially because of the fact that Hindi is an umbrella term covering the Western and Eastern varieties (which quite certainly have very distinct grammars) both of which contain several sub-varieties that are arguably distinct languages as well. The best resource compiling work on standard and non-standard Indo-Aryan languages and their dialects is still Masica (1993) and I refer the reader to its bibliography, which has been organized by language and puts together available material for each language and its dialects in one location. Shapiro (2007) also provides some references for work on distinct varieties of Hindi.

Two ongoing efforts at documenting the languages of India must be mentioned in this context. The first is the People's Linguistic Survey of India, <http://peopleslinguisticsurvey.org/>, an ambitious survey of the entire country's languages initiated in 2010 by the Bhasha Research and Publication Center. Spearheaded by Ganesh Devy, the project "envisions the creation of a Linguistic Survey rooted in people's perception of language."¹² The survey, carried out by native speakers of the distinct language communities, is articulated as a political movement for the identification and self-assertion of distinct linguistic varieties that exist in the country. The survey consists of a basic grammatical template and also asks for information about the geographical and linguistic contexts of the speech community, and samples of oral literature. Each description is in the form of a book-chapter. The multi-faceted goals of this survey project and its reliance on native-speaker researchers make it a politically empowering project. The results of this survey are in the process of being published; for Indo-Aryan, the data on the languages of four states –

¹¹In the Bhili and the Khandeshi case, I know from personal fieldwork in the linguistic region that mutual intelligibility often has to do more with long-term contact and multilingualism than the grammatical properties of different languages, which differ greatly.

¹²<http://peopleslinguisticsurvey.org/aboutus.aspx?page=PLSI>

Maharashtra, Rajasthan, Jammu & Kashmir, and Uttarakhand – are available, with the volume on Gujarat about to be published at the time of this writing.

The second effort is the web-based Language Information Service (LIS)-India, <http://www.lisindia.net/>, initiated by the Central Institute of Indian Languages (CIIL). This project seeks to make available linguistic data including information about census, intra-linguistic dialect variation, and historical evolution for both scheduled and non-scheduled Indian languages.

4. Language and dialect in the Indo-Aryan context

In 1956, following regional demands and widespread agitations for linguistic states, the Indian Government implemented a major reform of the boundaries of India's administrative states along linguistic lines by passing The States Re-organisation Act of 1956.¹³ This drawing of linguistic-political boundaries, which were further articulated over the next several decades through the creation of more and more linguistic states, has been central to defining how linguistic identity is “officially” established in the Indian context. Movements for regional autonomy (and “statehood” status) have gone hand-in-hand with movements for official linguistic recognition as was evidenced in the case of the formation of the states of Maharashtra (Marathi) and Gujarat (Gujarati) in 1960, Punjab (Punjabi) in 1966, and Goa (Konkani) in 1987.¹⁴

The distinction between “language” and “dialect” is an especially fraught one in a situation where the political and administrative recognition of regional languages has served to obliterate sub-regional linguistic identities, creating new hierarchies and threatening the loss of the historically stable differentiation between varieties (Khubchandani 1991, Annamalai 2001, Dasgupta & Sardesai 2010). Historically, India has been tolerant of multilingualism with distinct varieties co-existing side by side as well as long-term retention of native languages by diasporic communities in non-native geographical contexts. The multilingual nature of most social interactions, widespread diglossia, and non-discrete boundaries between speaker communities have created fluid zones of linguistic interaction, where linguistic identity can often be a shifting notion determined by sociopolitical goals. In modern India, the fluid boundaries between language and dialect have been used systematically in both delineating new identities and in suppressing priorly recognized identities.

Shapiro and Schiffman (1983: 5) provide a clear example of the first case with Punjabi, which is a dialect that turned into a language within a span of a few decades. Punjabi was widely considered a dialectal form of Hindi before India's independence. Punjabi was used at home but Hindi or Urdu was resorted to for purposes of education. Post-independence, strong demand for a Punjabi-speaking state led to “changing popular and official attitudes towards the code”. The formation of Punjab in 1966 in turn led to the language becoming a medium for education and broadcasting and to a corresponding increase in publication, thus crystallizing its linguistic identity and distinctness from standard Hindi.

¹³The Sri Lankan Government passed the Official Language Act in 1956, making Sinhala the only official language of a country with a large Tamil-speaking minority.

¹⁴See Brass (1974) and Windmiller (1956) for overviews and perspectives. This theme of linguistic regionalism extends beyond the Indian border within the sub-continent. Pakistan, formed in 1947, had two major linguistic contingents – Urdu and Bangla. Differences between these first led to the creation of a separate provincial ministry for the Bangla speaking region, and ultimately, to the separation of Pakistan into two distinct countries – Pakistan (Urdu-speaking majority) and Bangladesh (Bangla-speaking majority) in 1972.

The second situation is more easily observed. The pressure for linguistic homogeneity through public media and education within administrative-linguistic states, which contain significant proportions of minority linguistic populations, has led to relationships of inequality between dominant and minority speaker communities (Emeneau 1962, Khubchandani 1991). Thus, minority languages like Bhili (9.5 million in 2001), Khāndeśī (2 million in 2001), and Halbi (0.59 million in 2001), while assigned “mother tongue” status in census counts, are being subsumed under the regional standard determined by political boundaries.

A historical example of the suppression of previously dominant linguistic varieties comes from the Hindi dialect continuum and the emergence of Khari Boli as the prestige dialect in the eighteenth and nineteenth centuries (King 1994: ch.2–5). The Hindi belt, a contiguous region which encompasses several dialect varieties and often distinct languages as well, occupies a large central region of the subcontinent.¹⁵ The boundaries of this broad dialect continuum, following Masica (1993), often include Bihari, Magahi, and Maithili, which are the Eastern-most varieties, the last of which was, in fact, declared an official language in 2003. Rajasthani (with its dialects) is at the western boundary of this dialect continuum. It is recognized as a language for literary purposes, but does not have official status yet.

Gumperz (1957, 1958) distinguishes between three hierarchical forms of speech within this continuum – local level village dialects, regional dialects found in small market centers which are relatively uniform over a large area, and Standard Hindi, whose primary base of native speakers is found only in large cities like Delhi, Agra, and Lucknow. Within the Hindi continuum, there are some regions where the three forms are mutually unintelligible, while in others, the three are relatively close to each other.¹⁶

Some of these varieties, such as Braj and Awadhi, as the dominant languages of the North Indian pre-modern literary traditions, were recognized as independent languages for much of the New Indo-Aryan linguistic period. However, in post-independence India, these languages, together with Kanauji, Bundeli, Bagheli, Bangaru, and several others, are subsumed under Hindi as regional varieties, with their literatures co-opted within the great Hindi literary tradition. The reason for this language to dialect transition over a century has to do with the rise of Khari Boli as a literary dialect and the lingua franca in the 19th century. It acquired this status in part due to its role in training of British officers in colleges for administrators (e.g. Fort William College in Calcutta).¹⁷ The rise of this dialect was accompanied by a growing social movement that sought to differentiate between Hindi and Urdu (King 1994). This polarization, which was underpinned by socioeconomic as well as political motivations, further consolidated the status of Khari Boli as the authoritative dialect across North India and ensured that this variety of Hindi emerged post-independence as the official language of the Union of India. In the next section, I elaborate further on the emergence and blurring of regional and cultural differences in the Hindi/Urdu dialect region.

¹⁵Hindi speakers are concentrated in the Indian states of Uttar Pradesh, Uttaranchal, Madhya Pradesh, Chattisgarh, Bihar, Jharkhand, Haryana, Rajasthan, Himachal Pradesh, and Delhi.

¹⁶This hierarchical pattern is also found in the Bhili and Khandeshi dialect continuum where the highest standard is either Gujarati or Marathi depending on the geographical location of the community and usually mutually unintelligible with the members of the dialect continuum.

¹⁷Most of the publication of textbooks and grammars for Hindustani and Urdu by Fort William College was undertaken in the Khari Boli dialect (King 1994).

5. The Hindi belt: dialect differences and emergence of the standard

Determining the dialectal varieties of Hindi becomes a complicated issue for several reasons – variation among scholars regarding the boundaries of the Hindi belt and varieties that can/should be subsumed under the Hindi umbrella; inconsistency in language planning policies and census practices over time; and diglossia with varying degrees of control over and fluid socially-determined movement between several varieties of the language.¹⁸ The one point of consensus concerning the dialectology of Hindi is that there are two sets of dialects – the Western and the Eastern – that are subsumed under the Hindi umbrella (Masica 1993, Shapiro 1989: 3–5). The Western dialects of Hindi include Braj, Bundeli, Harianvi/Bangaru, Kanauji, and vernacular Hindustani (also called Khari Boli). Eastern Hindi has three major dialects – Awadhi, Bagheli, and Chattisgarhi (Shapiro 2007: 251). Some of the salient linguistic differences between the Western and the Eastern Hindi dialects include the following:

- (5) a. The Eastern varieties form the future with the suffix *-ba*, inherited from the Old Sanskrit gerundive in *-tavya*, while the Western dialects have an innovated future paradigm based on the OIA present stem.
- b. The perfective paradigm in the Eastern dialects has an added *-l* affix which is absent in the western dialects.
- c. The ergative clitic *-ne*, characteristic of western Hindi, is absent in the Eastern dialects.
- d. Lexical accent tends to be word-initial in Eastern dialects while the Western dialects retain older stress patterns.

Note that many of these differences correspond to the broader and historically deep-rooted Eastern/North-Central divide for Indo-Aryan languages in the Grierson/Southworth model discussed in §??.

The Hindi language region, for much of the history of the language, has been a loosely-knit network of only partially mutually-intelligible dialect varieties. The regional literary traditions developed in some of these varieties, notably Braj, Avadhi, and Maithili, did not remain confined to the region of their origin but spread broadly across Northern India, reinforcing speaker-competence in and intelligibility between multiple dialectal varieties in the broad Hindi region. Until the mid-nineteenth century, there were no clear assimilatory trends among dialectal varieties in favor of any standard in the Hindi speaking region. The current situation, however, is markedly different, with the clear recognition of a prestigious Hindi standard among speakers of all varieties and varying degrees of competence in this standard dialect relative to the more local varieties. This trend of assimilation, effected over the past one and half centuries, is best understood as the consequence of another process of conscious *differentiation*, the polarization of Hindi and Urdu.

5.1. The evolution of Modern Standard Hindi

Modern Standard Hindi, the official national language of India and the language used across the country in both education and administration (in addition to English and regional vernaculars), takes as its dialectal base Khari Boli, the dialect spoken in and around Delhi. This language is virtually identical in grammatical structure to Urdu, the official language of Pakistan, and one of the scheduled languages of India. This differentiation of a single dialect into two socioculturally distinct entities historically emerged through a conscious fashioning of the two linguistic identities

¹⁸There are more than 49 varieties of Hindi listed in the 2001 Census of India.

and their harnessing for religious and nationalistic expression in the 19th and the 20th centuries (King 1994, Trivedi 2003). The polarization between Hindi and Urdu and the crystallization of each linguistic identity have had far-reaching effects on the social dialectology of the Hindi belt in India as well as on language policy and education in the multilingual South Asian context. In the early 19th century, Urdu or Hindustani (as it was called by the colonial British administrators) was the Khari Boli dialect with Perso-Arabic influence, spoken predominantly by Muslim speakers, but also used by literate, elite Hindus. There was also another version of Khari Boli, the so-called Hindu version, which relied on Sanskrit for its learned vocabulary and was written in the Devanagari script – a script shared by several Indo-Aryan languages. In 1837, the British colonial government passed an act (Act 29) which replaced the use of Persian with Urdu as the official language for judicial and administrative purposes in the North West Provinces and parts of the Central Provinces of India. This sparked off a movement in several regions of North India demanding that Devanagari, the script in which Hindi is written, be used as an alternative script alongside the modified Persian script (in which Urdu is written), for administrative and judicial businesses in Northern India (King 1994). This initial demand transformed into an organized movement in the literary and sociopolitical domains in the second half of the 19th century – the Hindi-Nagari movement. While Urdu (with its accompanying Persian-based script) already had a significantly well-established literary tradition in poetry, Hindi prose and poetry only began to emerge and develop in the second half of the 19th century. As both Urdu and Hindi cultivated distinct literary styles of the same Khari Boli dialect base, the reliance on distinct classical languages (Perso-Arabic and Sanskrit respectively) increased and the languages increasingly became markers of religious and political identity. King (1994), describing this evolution in the second half of the 19th century, calls this a process of “multi-symbol congruence” by which the Hindi language, the Devanagari script, and Hindu religion became more closely identified and opposed to Urdu, its Persian-based script, and Islam. The Hindi-Nagari movement, driven by the need to define Hindi as distinct from (and superior to) the culturally, literarily, and politically dominant Urdu, was underpinned by this multi-symbol congruence (Brass 1974, King 1994, Dalmia 1997). The Indian Independence movement, in particular political bodies like the Indian National Congress, adopted Hindi/Hindustani as the language of choice, focusing on the religion-neutral grammatical core of Hindi and Urdu, and its intelligibility to large swathes of the North Indian population. Despite political efforts to amalgamate the two sociocultural entities and dissociate Hindustani from issues of script, Perso-Arabic vs. Sanskritic influence, and religion, the Hindi language movement and the literary and linguistic norms it established in opposition to Urdu prevailed (King 1994, Dasgupta and Sardesai 2010). After independence and the concomitant separation of colonial India into India and Pakistan, the Indian constitution of 1950 provided that Hindi, written in Devanagari script, was to be the national language of the Indian republic. Over the coming decades, Hindi was increasingly “cast in a heavily Sanskritized mould”, in education as well as print media and television (Shapiro 1997: 256).

This top-down state-imposed linguistic norm of Modern Standard Hindi (MSH) has had far-reaching effects on the dialectal situation in the Hindi belt. Until constitutional sanction for Hindi, there were relatively few native speakers of this language in either its deliberately crafted literary version, or its dialectal base, Khari Boli. The majority of the so-called Hindi/Hindustani speaking population spoke one of the regional Eastern or Western dialects, and depending on access to education, had some familiarity with the standard. The current situation is markedly different. Khubchandani (1997), describing the code fluidity of the Hindi belt, divides the population into five categories:

- (6)
 - a. Bilinguals of the north-central region who view their own primary speech in terms of substandard variations of the prestigious MSH standard.
 - b. Bilinguals who use primary varieties in their intimate rural milieu and MSH in modern settings.
 - c. Bilinguals who use primary varieties for all oral communication and relegate MSH to written use.
 - d. Illiterate monolinguals who only speak primary varieties but say they belong to the Hindi fold.
 - e. “Real” users who speak and write MSH and nothing but MSH.

There are two points to note here. First, distinct varieties that are controlled by bilingual speakers are often perceptually assimilated into the larger Hindi fold. That is, bilingual speakers see themselves as monolingual speakers, where MSH and the variety they control (e.g. Braj or Awadhi) are taken to be stylistic variants of a single language. Second, the last category of speakers – monolingual MSH speakers – has increased dramatically in the Hindi belt as a consequence of educational policies and exposure to media. Thus, both in terms of perception as well as in terms of actual linguistic behavior, we see assimilatory trends as the Hindi standard is imposed over a complex dialect continuum.

6. Concluding remarks

The Indo-Aryan situation is sociolinguistically rich and complex, characterized by plurilinguality and dialect continua spread over large regions spanning multiple languages. Super-imposed on this intrinsic heterogeneity are homogenizing linguistic standards – regional (official state languages) as well as national (Hindi) – whose influence increases with increased literacy efforts and media exposure. The effect of this organization is a layering of several languages – the home and intra-community languages, local lingua francas (which may differ from the official state language), official state languages, and finally, the national standard. Cross-cutting these differences is the presence of old and new written traditions, which do not necessarily correspond systematically with official language status.

One question that emerges in such a plurilingual ethos is the degree to which it creates a hierarchical situation where language varieties are categorically placed with respect to each other as ‘low’ or ‘high’ varieties. Despite the presence of standards and recognition of state-imposed linguistic norms, much of the linguistic situation in the Indo-Aryan (and the larger South Asian) context can be described as ‘horizontal’, where distinct speech varieties coexist within individuals and communities and are seen as available resources for communication without particular social meanings being attached to such choices (Satyanath 2015). Researchers have emphasized the fluidity of codes, observing how the inherent flexibility in communication patterns in South Asian society leads to dialect-switching, dialect-mixing, and bilingualism as an integral part of the linguistic experience of speakers (Khubchandani 1985, 1997). Much more research on the grammatical features that characterize and differentiate dialects of a single language or sub-regions within dialect continua is necessary in order to get a real understanding of the “facts on the ground”, but it is critical that such descriptions factor in the effects of code fluidity and flexible usage that underpin language use in the South Asian context.

References

- Beames, John: 1872-79, A Comparative Grammar of Modern Indo-Aryan Languages of India, Munshiram Manoharlal, Delhi, Republished 1966.
- Brass, Paul R.: 1974, Language, Religion and Politics in North, Cambridge University Press, Cambridge, Republished in 2005 by Backinprint.
- Cardona, George and Dhanesh Jain: 2007, The Indo-Aryan languages, Routledge, New York.
- Dasgupta, Probal and Madhavi Sardesai: 2010, 'Sociolinguistics in South Asia', In Ball, Martin J. (ed.), The Routledge Handbook of Sociolinguistics Around the World, 81–88, Routledge, Oxford.
- Emeneau, Murray B.: 1956, 'India as a linguistic area', Language **32**, 3–16.
- Ghatage, Amrit: 1962-68, A Survey of Marathi Dialects (vol 1-9), The Maharashtra State Board for Literature and Culture, Bombay.
- Gill, Harjeet Singh: 1973, Linguistic atlas of the Punjab, Dept. of Anthropological Linguistics, Punjabi University, Patiala.
- Grierson, Sir George Abraham Grierson: 1903–1928, Linguistic Survey of India, volume I–XI, Office of the Superintendent of Government Printing, India, Calcutta, Reprinted 1967 by Motilal Banarsidass.
- Gumperz, John T.: 1957, 'Language Problems in the Rural Development of North India', The Journal of Asian Studies **16.2**, 251–259.
- Gumperz, John T.: 1958, 'Dialect Differences and Social Stratification in a North Indian Village', American Anthropologist **60**, 668–682.
- Gusain, Lakhan: 2000-2008, Bagri, Shekhawati, Mewati, Marwari, Mewari, Dhundhari, Harauti, and Wagri grammars, Lincom Europa, Munich.
- Hoernle, A. F. Rudolf: 1880, A Comparative Grammar of the Gaudian Languages, Trübner and Co., London, Reprinted 1975.
- Khubchandani, Lachman: 1991, 'India as a sociolinguistic area', Language Sciences **13**, 265–288.
- Khubchandani, Lachman: 1997, 'Indian diglossia', In Revisualizing boundaries: a plurilingual ethos.
- King, Christopher: 1994, One Language, Two Scripts: The Hindi Movement in Nineteenth Century North India, Oxford India, Delhi.
- Masica, Colin P.: 1993, The Indo-Aryan Languages, Cambridge Language Surveys, Cambridge University Press, Cambridge.
- Mitra, A. and R. C. Nigam: 1971, 'Grammatical Sketches of Indian Languages with Comparative Vocabulary and Texts', Language Monographs (1961 Series), No. 2, Census of India Publication.
- Satyanath, Shobha: 2015, 'Language variation and change: The Indian experience', In Smakman, Dick and Patrick Heinrich (eds.), Globalizing sociolinguistics: Challenging and expanding theory, 107–122, Routledge, Oxford.
- Shapiro, Michael: 2007, 'Hindi', In Cardona, George and Dhanesh Jain (eds.), The Indo-Aryan languages, 250–, Routledge, New York.
- Shapiro, Michael and Harold Schiffman: 1983, Language and Society in South Asia, Foris, Dordrecht.
- Southworth, Franklin: 2005, Linguistic archaeology of South Asia, Routledge.
- Trivedi, Harish: 2003, 'The progress of Hindi, Part 2: Hindi and the nation', In Pollock, Sheldon

- (ed.), Literary Cultures in History, 958–1022, University of California Press, Berkeley.
- Turner, Ralph L.: 1962–66, A comparative dictionary of Indo-Aryan languages, Oxford University Press, London, Republished by Motilal Banarsidass Publishers, New Delhi.
- Witzel, Michael: 1999, ‘ Substrate Languages in Old Indo-Aryan (Ṛgvedic, Middle and Late Vedic’, Electronic Journal of Vedic Studies (EJVS) **5**, 1–67.