Sims-Williams, Nicholas (ed.)
  2002   *Indo-Iranian Languages and Peoples.* (Proceedings of the British Academy 116). Oxford: Oxford University Press.
Singh, Ram Adhar
  1991   Lexicography of New Indo-Aryan. In: Hausmann et al. (eds.), 2507−2521.
Strandberg, Elisabeth
  1991   Lexicography of Middle Indo-Aryan. In: Hausmann et al. (eds.), 2497−2507.
SWTF: See Waldschmidt.
Vogel, Claus
  1979   *Indian Lexicography.* (A History of Indian Literature V/4). Wiesbaden: Harrassowitz.
Wackernagel, Jacob
  1896   *Altindische Grammatik.* Band I. *Lautlehre.* Göttingen: Vandenhoeck & Ruprecht.
Waldschmidt, Ernst (founding ed.)
  1973−2015   *Sanskrit-Wörterbuch der buddhistischen Texte aus den Turfan-Funden und der kanonischen Literatur der Sarvastivada-Schule.* Göttingen: Vandenhoeck & Ruprecht.
Witzel, Michael
  1999   Aryan and non-Aryan Names in Vedic India. Data for the linguistic situation, c. 1900−500 B. C. In: Johannes Bronkhorst and Madhav M. Deshpande (eds.), *Aryan and Non-Aryan in South Asia. Evidence, Interpretation and Ideology.* Cambridge, MA: Harvard Univ., Dept. of Sanskrit and Indian Studies, 337−404.
Witzel, Michael
  2006   Early loan words in western Central Asia. Indicators of substrate populations, migrations, and trade relations. In: Victor Mair (ed.),*Contact and Exchange in the Ancient World.* Honolulu: University of Hawaii Press, 158−190.

Melanie Malzahn, Vienna (Austria)

30. The dialectology of Indic

1. Preliminary remarks

A thorough linguistic history of each of the dialects of Indic is challenging in a single book and impossible in a single chapter. Rather than providing such a treatment, this chapter aims to outfit the internal history of Indic provided by Oberlies (this handbook) with an external history, furnishing each language with details relevant to its regional history. This chapter is best used in consultation with an atlas. We shall proceed chronologically through Old Indic, Middle Indic, and New Indo-Aryan, concluding with a brief discussion of the status of Nuristani.

2. Old Indic

There are four regions of Old Indic dialect, each with its own peculiarities and each with its own *śākhās* or schools of Vedic priesthood. These schools orally composed and committed to memory texts in a hieratic language called Vedic Sanskrit. These four regions span northwest India and Pakistan. Although dialects of Vedic were very similar, the speech of Gandhāra, the Panjab/Haryana, western Uttar Pradesh, and eastern Uttar Pradesh each had a number of distinct local features.

2.1. Pre-Vedic

Because of the highly archaic nature of the *Ṛgveda*, it is taken as a kind of pre-Kuru Vedic dialect, and dialect idiosyncrasies will be innovations against the backdrop of the *Ṛgveda*. Much of its internal imagery seems to situate the composition of at least a portion of its individual hymns near the confluence of the Sutlej and the Beas rivers in the Panjab. It bears mentioning however, that while the *Ṛgveda* is archaic both in terms of grammar and content, it has been filtered through the phonetics of the redaction of the text, which is believed to have taken place further to the southeast near the modern state of Haryana between the Ghaggar-Hakra river and the Yamuna. Thus while the reconstructed text is "pre-Vedic" in many respects, it has surface features of the Western dialect proper to the eastern edge of the Panjab and Haryana. It is in this area that a federation of tribes emerged near the end of the 2nd millennium BCE. In this pastoral polity, kingship was not hereditary, but rather sovereignty was bestowed upon a *viśpati* 'clan-lord' through the Soma sacrifice. It is from this region and time period that the early recensions of the *Ṛg*, *Sāma*, and *Yajurveda* emerge as anthologies of verse, melody, and ritual phrases used in the political rituals of this tribal confederacy.

   The majority of the work done on Vedic dialect has been by Michael Witzel. Witzel localized the myriad priestly *śākhās* by careful consideration of environmental details contained in the canonical texts of each school, particularly the direction in which rivers flowed. In younger text strata, the Vedic schools expanded eastward into the Gangetic basin. By the Middle Vedic period, the *śākhās* were situated around four centers of political power: Kurukṣetra 'the field of the Kurus' in Haryana, Pañcāla in western Uttar Pradesh, and Kosala and Videha in eastern Uttar Pradesh.

2.2. Western Vedic

A number of dialect features are peculiar to the region of Kurukṣetra located in modern day Haryana and the eastern Panjab. One of these is the development of intervocalic /ḍ(h)/ to [ḷ(h)] including final /ṭ(h)/ which could be voiced through sandhi. Another feature of Kurukṣetran is the archaic retention of Pre-Vedic *-śc- where this has else-where been simplified to <cch> as in Classical Sanskrit *gacchati*. The Kathas, situated in the Eastern Panjab, routinely employ <śch> for historical *-śc- as do the vulgate chapters of the Paippalāda *Atharvaveda*. The Śākalya recension of the *Ṛgveda* and the Maitrāyaṇi school of the *Black Yajurveda*, both located in Haryana, use <ch> but this

always scans long, which Witzel (1989: 6.1) argues is indicative of the fact that a fricative preceded the affricate.

2.3. Central Vedic

To the east, the Pañcālan dialect of Vedic is host to a distinct set of dialect features. Pañcāla was located in the Doab, the region in western Uttar Pradesh bounded by the confluence of the Yamuna and the Ganges. The Black Yajurvedic school of the Taittirīyas and the Sāmavedic school of the Jaiminīyas were located there in the Middle Vedic period, before they later migrated South. Because the *Ṛgveda* is a metrical text, its metrical deviations can be corrected. This process has led to the discovery that surface forms like *svàr* 'sun' can be restored to the phonetics of their era of composition, that is *súvar*. While the Kurukṣetran dialect underwent vowel syncope, distorting the meter of the *Ṛgveda*, the Pañcālan dialect of Vedic did not; Jaiminīya and Taittirīya texts attest a *súvar*. Where Pre-Vedic originally had a nominal suffix *-iya-, developing in other dialects of Vedic to [-ya-], Pancālan Vedic innovated an [-īya-], which resulted in doublets such as *śunāsīryà* and *śunāsīrīya-*. Another feature of this region is a fem. gen. sg. in [-āi]. Both fem. gen. sg. [-āḥ] and fem. dat. sg. [-āi] produce the same sandhi outcome when preceding a word which begins with a vowel: [-ā]. From this [-ā], a new fem. gen. sg. in [-āi] could be hypercorrectly formed on the basis of the dative, and in Pañcāla country that is most likely what happened.

2.4. Eastern Vedic

Another region with a distinct dialect is Kosala in eastern Uttar Pradesh. Kosala was located east of the Gomatī along the Sarayu. Videha lay still further to the East in what is now western Bihar. Kosala and Videha were home to Kāṇva and Mādhyandina schools of the *White Yajurveda*. The two constitute one dialect region: Eastern Vedic. Both share unique innovations, including the development of the perfect into a narrative past tense and the *bhāsika* accent, which reduced the three tone system of Pre-Vedic to two. In cases where Kosala and Videha disagree, it is the more easterly Videha which patterns with the Western schools. Witzel (1989: 5.1) notes that the Aśvalāyana *śākhā* of the *Ṛgveda* was supplanted by the originally westerly Śākalya school. A late migration from far west to far east may explain this pattern and suggests that the political fortunes of the East were on the rise if they were attracting peripatetic priests.

2.5. Northwestern Vedic

This leaves one dialect region still unaccounted for in the Vedic period, and this was to become the source of Classical Sanskrit. Neither the Vedic of Kurukṣetra, Pañcāla, Kosala, or Videha can be the direct ancestor of Classical Sanskrit. The language studied and codified by Pāṇini in the 4th century BCE near Taxila, must have developed from the

Old Indic of the surrounding region: Gandhāra, situated in northern Pakistan. According to Patañjali, proper Sanskrit is spoken in the northwest. He mocked and derided eastern speech, despite the fact that the Gangetic basin was the new cultural center of Ancient India; attitudes about language prestige continuously looked to the West. We have no direct attestation of Gandhāran Vedic, but its immediate descendant is likely Classical Sanskrit.

2.6. Other Sanskrits

This leaves the origins of other forms of Sanskrit contemporaneous with early Classical Sanskrit unexplained. Epic Sanskrit must have developed out of a courtly or bardic *lingua franca* spanning Uttar Pradesh, eventually to buckle under the social pressure of Classical Sanskrit. It has many features, including its use of case, which suggest a kind of semantic syncretism seen in older Middle Indic (Oberlies, this handbook) but already underway in the Pañcāla dialect of Vedic. It makes use of the imperfect narrative past, as Western Vedic would, but it also employs the perfect as a narrative past, just as Eastern Vedic would. Unlike Classical Sanskrit, which from its inception was codified by Pāṇini and preserved by a dedicated community of grammarians (the antecedents of the *vyākaraṇa* tradition), there is no evidence that Epic Sanskrit had such institutional policemen. As such it exhibits a great degree of internal diversity. This is far more true of the *Mahābhārata* than the *Rāmāyaṇa*. The former, being much larger and containing many older parts which clearly predate a standardized *śloka*, has more independence from the prosodic norms of Classical Sanskrit. The *Rāmāyaṇa* on the other hand, like the youngest parts of the *Mahābhārata*, has already come under the influence of Classical Sanskrit. These growing similarities in rhetoric and language are perhaps why the *Rāmāyaṇa* is referred to as *ādikāvya* 'the first *kāvya* [poetry]', and indeed its aetiology of the *śloka* is a charter myth of Classical Sanskrit poetry and drama. The *Mahābhārata*, on the other hand, is classified as *itihāsa* 'history' rather than poetry. Deviation from Classical Sanskrit is even greater in Buddhist Sanskrit and Aiśa, a Śaiva Sanskrit. It may be that early Buddhist Sanskrit originated as a number of independent translations of early Buddhist texts into a vernacular Sanskrit which retained many Middle Indic grammatical sensibilities but became a liturgical language of Buddhism. While never being standardized, Buddhist Sanskrit achieved a kind of hybrid grammar.

3. Middle Indic

Middle Indic dialects can be treated chronologically and regionally, but it is important to remember that like Vedic, which was the professional language of Vedic priests, our earliest Middle Indic sources often become, whatever their vernacular origins, the doctrinal language of a religious sect. In part, this must be because for a Middle Indic language to have remained grammatically fixed rather than continue its development to New Indo-Aryan, it required a dedicated tradition of oral transmission or scribal copyists. No doubt many varieties of Middle Indic existed which vanished without a trace since they lacked this form of institutionalization.

3.1. Inscriptional Middle Indic

One institution that was highly instrumental in the preservation of Middle Indic texts was government. In the mid-3rd century BCE, Aśoka Maurya, after consolidating his empire, commissioned numerous edicts inscribed on rock pillars and in caves. These edicts, when read aloud by a literate agent of the emperor, communicated the *dhamma* to those living in his dominion. This *dhamma* constitutes the set of ethical principles governing his empire which seem to be based on a lay understanding of the Buddha's doctrine. The Aśokan inscriptions form a massive ring around Northern India. They are most concentrated in the northeast, in Bihar, Jharkhand, and Bengal, but extend as far south as Erraguḍi in Andhra Pradesh, as far southwest as Girnar in Gujarat, as far west as Khandahar in Afghanistan, and as far north as Mānsehrā near the Khyber Pass in Pakistan.

These inscriptions were composed in the administrative language of Magadha, the kingdom Aśoka inherited, and translated into the administrative languages of the distant lands he conquered. In areas which already used writing, the inscriptions were commissioned in native orthographies. In Khandahar, for example, the inscription is in both Greek and Aramaic and uses their respective orthographies. In Pakistan, the inscriptions are in Kharoṣṭhī. Elsewhere, however, Aśoka uses a script called Brāhmī. The Brāhmī script, apparently specifically designed for Middle Indic, is the source of all native orthographies of India as well as Tibet and much of Southeast Asia. Its own origin is much more contentious. Georg Bühler first proposed a Semitic origin for it, due to parallels with Phoenician and Aramaic orthography. Indeed, it is unclear where the genre of the imperial edict inscribed in stone could have come from other than the easternmost Aramaic inscriptions of the Achaemenid Empire. Aramaic seems to be the source for Kharoṣṭhī script, and the Middle Indic word *lipi* 'a writing' seems to be a borrowing of Old Persian *dipi* itself borrowed from Elamite. Other theories, however, argue that Brāhmī was invented *ex nihilo* by Aśoka or modeled after the Indus Valley seals.

The Middle Indic languages preserved in the Aśokan inscriptions reflect a number of areal features in a dialect continuum with a great deal of local variation; however, linguistic details are often obscured by the orthography. Rock Edict VII at Shāhbāzgaṛhī in Pakistan attests a form *devanaṃpriyo*, the Kharoṣṭhī script preserving the cluster [pr-]. The Rummindei pillar does not, attesting instead an inst.sg.m. *devānapiyena*. In the pillar inscription, the usual Middle Indic shortening of vowels in heavy closed syllables (here co-occurring with the loss of vowel nasalization as well [*devānām > devāna]) perhaps reflects that this syllable is still metrically heavy despite the absence of either vowel nasalization or a consonant cluster, suggesting an underlying stem *ppiya- which makes position. Another feature of the inscription at Rummindei is the merger of /l/ and /r/ resulting in /l/. Let us compare these two inscriptions again. The Shāhbāzgaṛhī inscription in the far West attests a form *raja*, equivalent no doubt to nom. sg. *rājā* without marking vowel length. At Rummindei in the East the form *lājina* is an inst.sg. agreeing with *devānapiyena*. Note that, in comparison with Sanskrit *rājñā* an [l] appears, the consonant cluster is separated by epenthetic vowel [i], and the final [ā] is shortened. Another difference is that at Shāhbāzgaṛhī the nom.sg. *devanaṃpriyo* ends in [-o], while the Rummindei pillar tells us that *hida bhagavaṃ jāte* 'here the lord (Buddha) was born', indicating that the nom.sg. of a-stems ended in [-e].

3.1.1. Eastern Inscriptional Middle Indic

The merger of [r, l] in favor of [l] and the nom.sg. [-e] of a-stems would become the most iconic elements of the Māgadhī Prākrit used in Classical Sanskrit drama, so named after Aśoka's homeland, Magadha. In fact, this Eastern Inscriptional Middle Indic has other features which distinguish it regionally. The eastern Aśokan inscriptions attest to a merger of [ś, s, ṣ], which is represented by a single sibilant <s> and a merger of [ñ, ṇ, n] represented by one character <n>. Consonant clusters are more often reduced or resolved by insertion of an epenthetic vowel, like in *lājina*. Where one finds Sanskrit [kṣ], Eastern Inscriptional Middle Indic has [kh]. Other differences include a present middle participle in [-mīna-], and loc. sg. masc./nt. in [-(s)si]. Although Patañjali evidently despised the Eastern *bhāṣā* 'patois', it must have been very prestigious in its day, as it was the language of Aśoka's capital Pāṭaliputra. Because this dialect is also the language of the Erraguḍi edict in Andhra Pradesh, and it seems unlikely that the dialect of Pāṭaliputra was spoken as far south as Andhra, the Aśokan inscriptions must represent not the vernacular but an elevated political register of Middle Indic deemed suitable for imperial proclamation; Aśoka's own Pāṭaliputra dialect was the default used for all pillar edicts and minor rock edicts with other versions of Middle Indic appearing only on the Western and Northwestern frontier. It was the official administrative language of the Mauryan dynasty and a dialect bound to the political fortunes of that empire; for all inscriptions in this dialect are Mauryan, and none post-date its fall.

3.1.2. Western Inscriptional Middle Indic

The western dialect of Middle Indic is best represented by the Aśokan inscriptions at Girnār in Gujarat. This dialect also attests to the merger of the Old Indic sibilant series, but it retains the distinction between [l] and [r] and its nasals remain distinct. Where one finds Sanskrit [kṣ], Western Inscriptional Middle Indic generally has [cch]. Furthermore, Western Inscriptional Middle Indic more often retains clusters rather than adding an epenthetic vowel, especially if these clusters involve a semivowel. Western Inscriptional Middle Indic features a loc. sing. m./n. in [-e] or [-mhi], and a gerund in *-tvā* > [-tpā]. Salomon (1989: 74) notes that this Western dialect often differs from eastern and northwestern inscriptions in vocabulary; for example, Girnār attests a *paṃthesū* 'along the roads' instead of *mag(g)esu* 'id.'.

3.1.3. Northwestern Inscriptional Middle Indic

The Shāhbāzgaṛhī and Mānsehrā inscriptions, found in Afghanistan and Pakistan respectively, are believed to represent an early form of Gāndhārī and constitute a third dialect of early Middle Indic. This dialect retains the respective distinctions between sibilants, nasals, and liquids. While the precise pronunciation of [kṣ] is unknown, it is represented by a distinct character, which suggests it did not merge with another phoneme. This dialect preserves internal consonant clusters, although when [r] is first in such a cluster, it often metathesizes with a preceding vowel, **dharma* > *dhrama*. Special developments

involving sibilants include [ṣy] merging with [ś] and [sv, sm] > [sp] (e.g. future stem *manuśa-* < *manuṣya, and pronominal loc. sg. m./n. in [-spi] < *-smin).

3.1.4. Post-Mauryan inscriptions

The institution of kingship which preserved Middle Indic in inscriptional forms did not end with the fall of the Mauryas but was continued by the polities which followed. The Yuga Purāṇa tells us that the Śuṅga empire was founded when the last Maurya emperor, Bṛhadratha, was assassinated in 185 BCE by his *senānī* 'army commander' Puṣyamitra Śuṅga. Whether this is historically true or not, a Puṣyamitra did leave behind Middle Indic inscriptions which proclaimed that he had completed two horse sacrifices, suggesting he had a public investment in Vedic ritual. Note that Puṣyamitra uses neither the Eastern Inscriptional Middle Indic of the Mauryas nor Sanskrit, which was still very much a hieratic language in the 2nd century BCE. A more westerly dialect of Middle Indic, which Émile Senart (1881, 2: 488) dubbed "Monumental Prakrit", remains the default language of these imperial proclamations. The earliest Sanskrit inscription is found at Ayodhyā. It is dated to the first century BCE on the basis of its claim that this inscription was commissioned *senāpateḥ puṣyamitrasya ṣaṣṭhena* 'by the sixth descended from General Puṣyamitra.' Note that Classical Sanskrit would have probably used an abl.sg. *puṣyamitrāt** rather than the gen.sg. The Junāgaḍh inscription of Rudradāman in 150 CE is a turning point from the Pre-Classical Sanskrit style to the poetic Sanskrit of the Classical period and constitutes the first *praśasti*. Even so, Epigraphical Hybrid Sanskrit, a mixture of Sanskrit and Middle Indic, remained the dominant language of inscriptions until the 3rd century CE. With the rise of the Guptas, however, Classical Sanskrit would become the standard for political discourse, scholastic texts, and the literary arts.

3.2. Scriptural Middle Indic

Besides the state, other institutions of power include the monastic orders of Buddhists and Jains. The oldest strata of texts preserved by these orders are believed to have originated as oral compositions which were at first transmitted orally and then translated into a variety of literary Middle Indic languages. It merits pointing out that these are "scriptural" dialects of Middle Indic because they are best known from Buddhist and Jain scripture, but by no means were they used exclusively for scripture.

3.2.1. Buddhist Middle Indic

The oldest Buddhist texts are written in a script called Kharoṣṭhī, which, because it is derived from Aramaic, does not distinguish vowel length. Kharoṣṭhī script is primarily used to record a Middle Indic language called Gāndhārī, named after the region in which it was found. Ancient Gandhāra constituted the territory around the Indus, Swat, and

Kabul river valleys in Pakistan and Afghanistan. Its capital, Taxila, is believe to be the home of Pāṇini, the creator of Classical Sanskrit, and Kharoṣṭhī may have been the *lipi* to which *Aṣṭādhyāyi* 3.2.21 refers. Under the patronage of Kuṣāṇas, Kharoṣṭhī spread along the Silk Road: northwest into Bactria and northeast into the Tarim Basin.

Gāndhārī was a major language of Buddhist literature, the best represented genre being *sutra* texts such as the *Dharmapada*, but commentaries, devotional songs, and scholastic treatises are also well represented. It was the administrative language of Gāndhāra, but it was also a literary one into which old texts were translated and in which new ones were composed. Gāndhārī undergoes many changes during its period of use (2nd c. BCE–4th c. CE), and appears to be a more advanced stage of the language of the Northwest Middle Indic found in the Aśokan inscription at Shāhbāzgaṛhī. In Gāndhārī, intervocalic consonants are sometimes voiced, becoming fricatives. Consider the form <bosisatva> which is derived from *bodhisattva* and likely pronounced [bozizatva]. This lenition is often hard to detect, as the spelling is under progressively greater influence from Sanskrit. Salomon remarks that Gāndhārī *sat*[*t*]*a* 'seven' is "corrected" later to *sapta* due to the influence of Sanskrit. The quality of final vowels was evidently neutralized in light of the diverse finals of the m. and n. a-stem [-e, -o, -u, -a], with all variants potentially appearing within a given text. In coda position, [r] within clusters is sometimes metathesized into a preceding onset; compare Sanskrit *durgati*, Pāli *duggati*, and Gāndhārī *drugadi* 'bad fate.' While Gāndhārī maintains its set of sibilants, [s, ś, ṣ], it gradually loses the distinction between [n] and [ṇ] as well as the distinction between aspirated and unaspirated consonants.

Thomas Burrow (1937) believed that the Aśokan edicts at Shāhbāzgaṛhī and Mānsehrā represented two distinct dialects of Middle Indic. The former, originally on the eastern side of the Indus, was marked by Old Indic [-as] > [-o], as attested by gen.sg.m. *raño* from *rājñas at Shāhbāzgaṛhī. The latter dialect, on the west side of the Indus, was marked by Old Indic [-as] > [-e]; compare gen.sg.m. *rajine* from the edict at Mānsehrā. For Burrow, the former was the direct ancestor of Gāndhārī while the latter the direct ancestor of Niya. Salomon (1998: 78), on the other hand, notes that final-vowel marking in Gāndhārī is highly inconsistent and not a probative distinction. A better model, perhaps is to consider Gāndhārī as the Northwestern Middle Indic that stayed in Gandhāra, while Niya is Northwestern Middle Indic exported along the Silk Road into the Tarim Basin where, in the 3rd century CE, it became the administrative language of the oasis kingdom of Kroraina. It shares more features with the Northwest Middle Indic of the Aśokan inscriptions than it does with Gāndhārī. In addition, however, it has independent innovations as well; Niya uses a single ending [-a] for nom. and acc. of both sing. and pl. There is a tendency in Niya to confuse voiced and voiceless stops, and to deaspirate aspirates. A suffix [-tu], which Burrow (1937: 49) believed to be taken from the pronoun, marks the second person of all tenses of the verb.

Pāli, the liturgical language of Theravada Buddhism, *appears* to be the most archaic Middle Indic language. The Pāli canon, or *tipiṭaka*, comprises three 'baskets' of texts: *sutta*, *vinaya*, and *abhidhamma*. The *Abhidhammapiṭaka* 'basket of higher dharma' consists of scholastic texts on topics of psychology, philosophy, and metaphysics. They are attributed to the *arhat*s 'worthies' or chief disciples of the Buddha. The *Vinayapiṭaka* 'basket of discipline' constitutes the system of monastic codes. The *Suttapiṭaka* contains teachings mostly attributed to the Buddha in his own words as well as other collections like the *Theragāthā* and *Therīgāthā*, which are anthologies of songs composed by elder

monks and nuns. The Pāli canon was exported to Śrī Laṅkā, continuing there its life as a literary language. The *Visuddhimagga*, believed to have been composed by the great Theravada commentator Buddhaghosa in 430 CE, is a comprehensive manual which explains and systematizes the Buddha's teachings and would critically contour the Theravada doctrine as it spread throughout Southeast Asia. The *Mahavaṁśa*, an epic which chronicles the legendary history of of Śrī Laṅkā, is also composed in Pāli.

That many of the *sutta*s are attributed to the Buddha is linguistically problematic. The narrative provided by the Pāli canon is that the Buddha was born and preached in Magadha some two hundred years before the birth of Aśoka. While the Pāli canon is archaic, it does not have features which resemble the language of the Aśokan inscriptions from Magadha. Rather, it resembles more the language of the Girnār inscription in the West or the "Monumental Prakrit" which proliferated after the fall of the Mauryas. For one thing, it maintains distinct nasals, does not merge [l] and [r], and resolves final *[-as] as [-o] and not as [-e]. Unlike the Girnār inscription, Pāli loses consonant clusters, including even those with an [r]; compare Sanskrit *pūrva* with Pāli *pubba*. Yet, in some passages Pāli does attest Eastern features. For example, whenever the Buddha directly addresses the monks, he uses the voc. pl. *bhikkhave* rather than *bhikkavo*, which shows the Eastern reflex of *bhikṣavaḥ.

This suggests some core material may be of an easterly origin, subsequently translated into a more western dialect. Warder (2000: 284) argues that Pāli was spoken in Avanti, an ancient kingdom believed to have been in the Malwa region in western Madhya Pradesh and southeastern Rajasthan. Hirakawa and Groner (2007: 119), on the other hand, place Pāli in the ancient kingdom of Śūrasena, which lay north of Avanti but south of Kurukṣetra and Pañcāla. Pāli cannot be the direct descendant of any attested form of Vedic Sanskrit. Compare Pāli *jhāyati* 'burns' with Sanskrit *kṣāyati*. The Sanskrit outcome [kṣ] is the result of a thorn cluster *dʰgʷʰ-. The Old Indic from which Pāli is descended evidently deleted the initial dental, resolving *dʰgʷʰ- into *gʷʰ-. This would indicate that if Pāli had a homeland, it was not one of the regions of Vedic dialect: Gandhara, the Panjab, Haryana, and Uttar Pradesh.

Another theory, however rejects the assumption that Pāli ever had a regional origin. Keeping in mind the rapid spread of Buddhism from far East to the Northwest, Pāli may have begun life as a *lingua franca* of trade routes. In this model, a pre-canonical Buddhist vernacular would have no homeland but rather be what Helmer Smith (1952: 178) dubbed a *koine gangétique* which absorbed forms from all over the trade routes. From *kṣana*, for example, Pāli receives both western *chaṇa* 'leisure, festival' and eastern *khaṇa* 'instant', not because one is more original, but because a community of traders and peripatetic ascetics would have had a translocal vocabulary. According to this theory, what began as a language accessible all over North India was then artificially re-engineered as a liturgical language. This accounts for Pāli forms like *brāhmaṇa* when *bamhaṇa* is the expected outcome. The process of transforming a translocal Middle Indic into an archaic liturgical language produced hypercorrections, for example the name Yamataggi in place of the Vedic seer Jamadagni.

3.2.2. Jain Middle Indic

Three Middle Indic languages are associated with particular Jain sects. Ardhamāgadhī, also called Ārṣa, is the language of the canonical texts of the Śvetāmbaras. Jain Māhā-

rāṣṭrī is used by Śvetāmbara Jains for non-canonical compositions. A third language, Jain Śauraseṇī, is the language of the canonical texts of the Digambaras. This threefold division mirrors the three primary Dramatic Prakrits, which are conceived of as the dialects of three regions of North India: Mahārāṣṭra in the West, Magadha in the East, and Śūraseṇa in the center. The extension of this nomenclature maps Jain texts to East, Center, and West. Śvetāmbara Jains assert that Ardhamāgadhī is the language spoken by Mahāvīra, the founder of Jainism, long ago in Magadha. An origin for Ardhamāgadhī in Magadha itself seems unlikely, as the language differs from the eastern Aśokan inscriptions. Its voicing, frication, and loss of intervocalic stops is more progressed than Pāli but less so than in the Dramatic Prakrits. It shares one iconic feature with Māgadhī Prakrit: the nom.sg.m. a-stem is in [-e]; but unlike Māgadhī it has both [l] and [r], and for that reason has been dubbed "half Māgadhī". Helmer Smith (1952: 178) argued, however, that Ardhamāgadhī, like Pāli, was the normalization of a translocal Middle Indic *koine gangétique*, and has no regional affiliation. Jain Māhārāṣṭrī is closely related to Māhārāṣṭrī Prakrit; but as Pischel (1957: 20) notes, "it is in no way fully identical to it", pointing out that Jain Māhārāṣṭrī has clearly been under the influence of Ardhamāgadhī and gained some of its peculiarities such as a t-stem nom. in [-m], an infinitive in [-ittu], and an absolutive in [-ttā]. One of the earliest examples of a Jain Māhārāṣṭrī text, the *Paumacariya*, is a telling of the *Rāmāyaṇa* dated to the 3rd or 4th century CE.

Jain Śauraseṇī shares only superficial features with the Dramatic Prakrit known as Śauraseṇī. Both Dramatic Śauraseṇī and the canonical language of the Digambaras have a nom.sg.m. a-stem in [-o]; because this language is neither Ardhamāgadhī nor Māhārāṣṭrī, it is assigned to the only remaining option. However, Pischel (1957: 21) notes that "… even a preliminary investigation of the dialect will show it has such forms and words as are altogether foreign to Śauraseṇī." He points to its loc. sg. in [-mmi] which it shares with Māhārāṣṭrī as well as its absolutive in [-ttā], a feature of all Jain Middle Indic dialects. While Dramatic Śauraseṇī has *karedi* < *karati, Jain Śauraseṇī, Jain Māhārāṣṭrī, and Ardhamāgadhī all attest a *karadi*. Findings suggest this Jain Śauraseṇī may be more closely related to Ardhamāgadhī than previously imagined. Dundas (1992: 80) argues that "everything points to the existence of an original and ancient shared Jain textual tradition which gradually bifurcated." Both Śvetāmbaras and Digambaras believe in an ancient lost body of Jain literature called the *pūrva*s 'ancient (texts)'. If this lost textual transmission existed, was it in a common ancestor to Ardhamāgadhī and Jain Śauraseṇī? Or another Middle Indic dialect altogether?

For the Śvetāmbaras, this lost material was located in the third chapter of the lost final limb of a twelve-limb canon. This twelfth limb was called *Dṛṣṭivāda* and the third chapter *Pūrvagata*. As *Dṛṣṭivāda* means 'Disputation about Views' the *Pūrvagata* may have been the opening arguments by adherents of heretical doctrines, much like the *pūrvapakṣa* in Indian philosophical texts. While some variation exists between Śvetāmbara sects, the most important texts are the eleven surviving limbs (*aṅga*s) and the twelve subsidiary limbs (*upāṅga*s). The Digambara textual tradition is much less well-known than the Śvetāmbara, and consequently, Jain Śauraseṇī is less well understood. The Digambaras reject the Śvetāmbara canon, believing the original twelve-limb canon to be long lost. According to tradition, by the time of Dharasena, the 33rd teacher in succession after Mahāvīra, there was only one *aṅga* remaining. This limb would be lost too, but Dharasena would transmit two texts: the *Ṣaṭkhaṇḍāgama* 'Scripture of Six Parts' and

the *Kaṣāyaprābhṛta* 'Treatise on Passions'. The Digambaras maintain that this is all that remains of the lost *pūrva*s.

3.3. Dramatic Prakrits

When one speaks of Prakrit, Māhārāṣṭrī constituted both the aesthetic ideal and the descriptive standard; the Prakrit grammarians explain the other Dramatic Prakrits as deviations from the Māhārāṣṭrī norm. It may have arisen as the living language of the northwestern Deccan or as the courtly language of the Sātavāhanas, an empire which covered much of central India from 230 BCE to 220 CE**.** The compilation of the *Gaha Sattasai*, an anthology of 700 Māhārāṣṭrī poems, is attributed to Hāla, a Sātavāhana king. Weber produced the first critical edition of the *Sattasai* in 1881. Based on seventeen manuscripts, this edition contains 964 poems in total, but only 450 of these were common to all manuscripts. The text is generally dated to the early first millennium CE and was well-known in literary circles in India by the late first millennium.

Early reference to Prakrit is found in the *Nāṭyaśāstra*, a dramaturgical text dated to the beginning of the first millennium. The *Nāṭyaśāstra* provides rules for the appropriate use of seven Dramatic Prakrits on the theatre stage: Māgadhī, Āvantī, Prācyā, Śaurasenī, Ardhamāgadhī, Bāhlīkā, and Dākṣiṇātyā; of these, only Māgadhī and Śaurasenī seem to have been institutionalized in Classic Sanskrit theatre. Sanskrit is used for speech and song by the gods and culturally elite human males; thus Sanskrit dominates the play. Śaurasenī is used only for the speech of the *vidūṣaka*, the king's jester, and female cultural elites. When these women sing, they sing in Māhārāṣṭrī. Māgadhī, on the other hand, is the language of ascetics or working-class characters. A number of dialects (Śākārī, Cāṇḍālī, Śābharī, Ḍhakkī) appear to be occupation-specific variants of Māgadhī.

All Dramatic Prakrits are subject to the typical Middle Indic reduction of the vocalic inventory and of consonants in clusters. Śaurasenī and Māhārāṣṭrī both use the [-o] nom.sg. ending for the a-stem and merge all sibilants into dental sibilant [s]. They both undergo successive stages of voicing, spirantization, and elimination of intervocalic stops leaving vowels in hiatus for most forms. Śaurasenī patterns with Māgadhī, however, by restoring dental stops; Compare Sanskrit nom.sg. *prākṛtaḥ* 'Prakrit', Māhārāṣṭrī *pāuo* 'id.', and Śaurasenī *pāudo* 'id.'. Voiced aspirates typically lose occlusion and are reduced to [h]; compare Sanskrit nom. sg. *prabhṛtaḥ* 'offering', Māhārāṣṭrī *pāhuo* 'id.'*,* and Śaurasenī *pāhudo* 'id.'. Māgadhī operates along the same principles, but its nom.sg. a-stem is in [-e], and it has a single sibilant [ś]; compare Sanskrit nom. sg. *puruṣaḥ* 'man' with Śaurasenī *puriso* 'id.' and Māgadhī *puliśe* 'id.'. Māgadhī tolerates [ś] before consonant clusters, compare Sanskrit nom.sg. *śuṣkaḥ* 'dry' with Māgadhī *śuśke* 'id.', and it replaces [cch] with [śc]; compare Sanskrit *gaccha* 'go!' with Māgadhī *gaśca* 'id.'. While this last form looks archaic on the surface, it is important to note that [śc] is very likely a secondary development. Consider that Sanskrit *pakṣa* 'wing' is cognate with Māgadhī *paśka* 'id.', presumably via a Proto- Māgadhī **pakkha*. In Māgadhī, both [y] and [j(h)] are captured by a character <y(h)> which may have a [ž] or [z] quality, compare Sanskrit *jāyate* 'is born' with Māgadhī *yāyade* 'id.'. The most striking feature of Māgadhī, however, is one shared with the eastern Aśokan inscriptions: the conversion of all [r] sounds to [l].

Of these three, only Māhārāṣṭrī has a literary life beyond its prescribed use in drama. Other important works in Māhārāṣṭrī include the *Setubandha* and the *Gauḍavaho*. It is important to note that whatever their spoken origins, the Dramatic Prakrits as we know them are highly artificial languages produced by applying transformation rules to Classical Sanskrit. These transformational rules are codified by grammars like the *Prākṛtaprak-āśa* of Vararuci and the *Siddhahemaśabdānuśāsana* of Hemacandra Sūri. The Dramatic Prakrits are no more the living vernacular than Classical Sanskrit but rather dramatized depictions of Prakrits. Indeed, it is hard to imagine a speech community that would tolerate the polysemy that loss of intervocalic stops produces without producing new words or compounds to disambiguate meaning.

Paiśācī or Cūlikapaiśācī is a Dramatic Prakrit known only from grammarians. A lost anthology of stories called the *Bṛhatkathā,* attributed to Guṇāḍhya, was supposedly composed in this language. Sadly, no complete work in Paiśācī survives, although there are fragments. Bhāmaha, in his commentary on Vararuci, calls Paiśācī *bhūtabhāṣā*, which is generally taken to mean 'the language of ghosts.' Andrew Ollett (2014: 406) argues Paiśācī's name is something of a misinterpreted literary joke, interpreting Daṇḍin's use of *bhūtabhāṣā* as simply meaning a 'dead language', not the 'language of the dead'. It would be Uddyotanasūri's comical placement of *bhūtabhāṣā* in the mouths of ghosts that would give Paiśācī a new literary life.

The most iconic feature of this Prakrit is the apparent devoicing of intervocalic stops (Compare Sanskrit *bhagavatī* with Paiśācī *phakkavatī*). The grammatical rules at work in Paiśācī could simply be the reverse application of the voicing rules applied to produce the other Dramatic Prakrits. For von Hinüber (1981), however, the supposed devoicing in Paiśācī is actually a fiction of orthography. According to his theory, at some point in the development of Middle Indic, the character <g> no longer represents voiced velar stop [g] but rather voiced velar fricative [ɣ] due to lenition. After this shift, the character <k> is repurposed to mark [g]. For von Hinüber, the odd appearance of Paiśācī is due to the distorting lens of this orthographical shift.

3.4. Apabhraṁśa

Patañjali describes *gāvī*, *gonī*, and *gotā* as *apabhraṁśa* 'fallen' perversions of the proper Sanskrit *go* 'cow'. The *Nāṭyaśāstra* characterizes Apabhraṁśa as marked by the ending [-u], presumably the historical outcome of Sanskrit *[-aḥ]. The text claims it is the language of the Ābhīras. Little is known about these Ābhīras, but Samudra Gupta records them on the Allahabad pillar as one of the nations he conquered, and it is generally believed they were a nomadic people who lived west of Mathura up to the Rann of Kutch. It is clear that up to and during the Gupta reign, Apabhraṁśa was a pejorative term for some Indic vernaculars. While Kālidāsa provides certain songs in Apabhraṁśa, it is best to consider this "Dramatic Apabhraṁśa" a stylized dramatic representation of language like the other Dramatic Prakrits. Consider that in Kālidāsa's *Vikramorvaśīyam*, King Purūravas sings in Apabhraṁśa only after Urvaśī has vanished and he is madly searching for her, asking the forest animals for her whereabouts. Apabhraṁśa then, is portrayed as the language in which madmen communicate with animals.

The literary prestige of Apabhraṁśa, however, would rise in the centuries following the Guptas. Between the 5th and 12th centuries CE, Apabhraṁśa was used by Jain poets.

Epic literature, biographies, and more secular poetry were composed in Apabhraṁśa during this period as well. Abdul Rahmān's *Saṁdeśa Rāsaka* is an example of a literary Apabhraṁśa overlapping chronologically with compositions in early New Indo-Aryan, sometime in the 12th or 13th centuries CE, and the influence of Apabhraṁśa literature on early compositions in New Indo-Aryan blurs the boundaries between the two. Like other literary languages of India, Apabhraṁśa was heavily theorized. The 12th century grammarian Kramadīśvara articulates a threefold categorization of Apabhraṁśa as Nāgara, Upanāgara, and Vrācaṭa. Rather than referring to specific languages, this threefold division may have been a way of conceptualizing the continuum of vernacular speech within a given region as proper to an urban, suburban, or rural milieu.

4. New Indo-Aryan

New Indo-Aryan, or NIA, refers to the Indic languages which emerged in medieval and early modern India and are spoken today. Many are attested already by inscriptions in the first centuries of the 2nd millennium CE. While the linguistic features and literary histories of each of these languages cannot be exhaustively presented here, a few notes will be made about the languages in the Indic dialect continuum. Note that the following divisions do not recapitulate diachrony but rather geography. An in-depth linguistic study of a given language should be pursued with an appropriate language-specific treatment, for example Thomas Oberlies' *A Historical Grammar of Hindi,* in conjunction with Colin Masica's *The Indo-Aryan Languages.*

As Masica (1991: 454) notes, "just about every conceivable way of carving up the NIA pile has been advocated by one scholar or another." In part, this is due to the prevalence of polyglotism in India. Frequently, a speaker knows a home language as well as the *lingua franca*. This results in the proliferation of non-genetic areal features which blurs the linguistic history of a particular dialect. It is also difficult for field linguists to determine if two languages are mutually intelligible when both the informant and the translator share complete or partial knowledge of another language. When the informant's native language is endangered, this is the typical scenario.

The following geographic designations have been used to divide up New Indo-Aryan: Upper and Central Gangetic Indo-Aryan comprises Hindustani, Bihari, and Rajasthani; West Indo-Aryan comprises Gujarati, Maraṭhi, and Konkaṇi; Northwest Indo-Aryan comprises Sindhi, Panjabi, and Dardic; Greater Himalayan Indo-Aryan comprises Western, Central, and Eastern Pahari; East Indo-Aryan comprises Oḍia, Bangla, and Asamiya; and, as non-contiguous NIA, Sinhala and Řomani are each treated independently.

There are a few supra-regional tendencies worth noting here. Initial [v-] > [b-] is generally an areal feature which extends from eastern Rajasthani and Kumauni all the way to Asamiya and Nepali. Note that in Marwari, a dialect of Rajasthani spoken west of the Aravalli mountains, initial *[v-] has also become [b-], but because *[b-] has become a voiced bilabial implosive [ɓ-], the old phonemic contrast is preserved. Another supra-regional tendency is post-nasal voicing, which seems limited to Northwest Indo-Aryan and Greater Himalayan Indo-Aryan. Excluded from both these supra-regional tendencies, West Indo-Aryan preserves both initial [v-] and post-nasal voiceless stops.

4.1. Upper and Central Gangetic Indo-Aryan

These languages are located in the Upper and Central Gangetic Plain. Mountain ranges surround this region, separating it from other NIA subgroups. The Himalayas form a natural boundary to the north, the Satpura and the Vindhya mountain ranges to the south, the Thar desert to the west, and to the east the Chota Nagpur Plateau and the Rajmahal hills.

4.1.1. Hindustani

The plains of Haryana, Uttar Pradesh, Madhya Pradesh, and Chhattisgarh are home to a number of Hindustani dialects. "Western Hindi" consists of Haryanvi, Braj, Bundeli, and Kannauji. Haryanvi, spoken in Haryana, is both the westernmost and northernmost dialect of Hindustani. Braj is spoken in the area around Mathura and Vrindavan. Braj Literature begins in the 14th century, but its most renowned work is the 16th century *Sursagar* by Surdas. Bundeli is spoken to the south, beginning around Gwalior and continuing as far as Chhindwara. Kannauji is the easternmost "Western Hindustani" dialect, as Kannauj is roughly 130 km away from Lucknow. "Eastern Hindi" consists of Avadhi, Bagheli, and Chhattisgarhi. Avadhi is centered around Lucknow. Like Braj, it has a literary history which dates back to the 14th century. Maulānā Dāūd's *Candāyan* may be the earliest work in Avadhi, but Tulsidas' *Ramcharitmana* is perhaps its most famous. Bagheli is very similar to Avadhi, but is spoken in southeastern Madhya Pradesh. Chhattisgarhi is both the easternmost and southernmost Hindustani dialect, spoken in the state of Chhattisgarh. Hindi and Urdu originated as Khariboli, the dialect of Hindustani spoken around Delhi. This developed first into a prestigious urban dialect and from there into the *lingua franca* of the Indo-Gangetic plains. Both Hindi and Urdu emerged from this Khariboli koine. While Hindi became the national language of India, Urdu became the national language of Pakistan. The differences between Hindi and Urdu are stylistic. While Hindi borrows heavily from Sanskrit and is written in Devanāgarī, Urdu borrows from Persian and Arabic and is written in a form of the Perso-Arabic abjad. Dakhini is the dialect of Urdu spoken around Hyderabad in Telangana.

4.1.2. Bihari

Further to the East is the Bihari group, a designation which predates the breakup of Bihar and Jharkhand but includes several NIA languages geographically located in both states. The languages in the Bihari group with the largest populations of speakers are Bhojpuri, Magahi, Maithili, and Bajjika. Bhojpuri is spoken in eastern Uttar Pradesh as well as western Bihar, and was initially categorized as "Eastern Hindi" by Beams (1872) on the basis that it lacks the complex verbal system of Magahi or Maithili. It is named after the dialect spoken in Bhojpur, just as the dialect spoken near Varanasi is often called Banarasi. Northern Bhojpuri is spoken in Deoria and eastern Gorakhpur. Dialects of Bhojpuri spoken east of the Gandak river are called Madhesi. Nagpuria Bhojpuri is the dialect spoken in the South near Ranchi, the capital of Jharkhand. It is not to be

confused with Nagpuri-Sadri which is a separate Bihari language spoken in Jharkhand. Magahi is spoken south of the Ganges and primarily in southern Bihar and northern Jharkhand. "Eastern Magahi" collectively designates the many dialects of Magahi spoken in southeast Bihar and northeast Jharkhand as well as along the western borders of Orissa and Bengal. Maithili, spoken north of the Ganges in Bihar and in Nepal, has a long literary history, with the poems of Vidyapati in the 14th century considered to be a high watermark. Bajjika is spoken in north-central Bihar. Standard Bajjika is the dialect spoken around Vaishali and Muzaffarpur. Dialects of Bajjika show the influence of Bhojpuri in the west, Maithili in the east, and Magahi in the south. Finally, Aṅgika is spoken on the border shared by Bihar and West Bengal. Aṅgika has sufficient affinities with the East Indo-Aryan subgroup to defy easy categorization.

4.1.3. Rajasthani

Rajasthani is spoken in India's largest state by area. Much of Rajasthan is the vast Thar desert, bordered on the West by the Rann of Kutch and on the East by the Aravalli mountain range which cuts a diagonal from the southwest to the northeast. The main dialects of Rajasthan are Marwari, Mewari, Ḍhunḍhari, Mewati, Harauti, and Nimadi. Marwari is spoken west of the Aravalli range, and thus does not really belong to the Gangetic Plain. Marwari has a series of voiced implosive stops. Shekhawati, the dialect of Marwari spoken in the northeastern districts of Churu, Jhunjhunu, and Sikar, is reported to have contrastive tone. Mewari is spoken on the eastern side of the Aravallis, while Ḍhunḍhari is the dialect spoken around Jaipur, the state capital. Mewati is spoken on the Haryana border, a dialect of which, Gujri, is spoken in Jammu and Kashmir. Harauti is the dialect spoken in eastern-central Rajasthan, from Bundi and Kota up to Madhya Pradesh, while Malvi is a dialect of Rajasthani spoken in the western parts of Madhya Pradesh itself. Another dialect of Rajasthani not spoken in Rajasthan proper is Nimaḍi, spoken further south in the Satpura range in the Nimar district, which is also home to Nahali, a language isolate. South of Udaipur are a number of Bhili dialects which are thought to be more closely related to Gujarati or Marāṭhi. Finally, the dialects called Lambani or Banjari seem to have originated as a western dialect of Rajasthani but have spread all over India, especially in the Deccan. The Banjaras are nomadic merchants and craft specialists whose culture shows numerous sociological parallels to that of the European Řomani, to whom they are not directly related.

4.2. West Indo-Aryan

West Indo-Aryan languages are all spoken from the Rann of Kutch to the Konkan. They are primarily spoken along the coast of the Arabian Sea, although Marāṭhi penetrates deeply into the interior as well. While these languages may form a genetic group, this is difficult to determine, because Gujarati has been influenced by Hindustani to the east and Rajasthani, Persian, and Sindhi to the north. Marāṭhi and Konkaṇi, on the other hand, were spoken from their inception in an area where the Satpura mountains and the Deccan plateau served as physical barriers to language contact. Even if they form only

a geographical group and not a genetic one, they certainly are not better categorized in any other subgrouping.

4.2.1. Gujarati

Gujarati is the official language of the state of Gujarat, but speakers of Gujarati are found all over the world. The standard dialect is spoken in the area north of Vadodara and Amdavad. Kathiawadi is the dialect spoken around the Kathiwar peninsula. There is a distinction between Hindu and Parsi dialects of Gujarati, with the latter admitting many Persian borrowings. Sauraṣṭri is a dialect of Gujarati spoken in Madurai by a weaver community believed to have migrated from the Kathiawar peninsula to Tamil Nadu a millennium ago. Gujarati has a long literary history; its earliest text is Śālibhadra-sūri's *Bharateśvarabāhubali* in the 12th century, but the most famous work of Old Gujarati is the *Vasantavilāsa* probably from the 14th or 15th century. Gujarati inscriptions from the Kacch and Kathiawar regions date back to the 15th century, but Gujarati features can be seen influencing earlier Sanskrit inscriptions as well. These inscriptions are usually written in Devanāgarī or a local script called Boriyā. Today Gujarati is written in its own script related to Devanāgarī.

4.2.2. Maraṭhi

The Maraṭhi dialects are located primarily in Mahārāṣṭra state. It is believed that Maraṭhi descends directly from Mahārāṣṭrī Prākrit and Mahārāṣṭrī Apabhraṁśa. It has a rich literary tradition, and, among the NIA languages, the most abundant epigraphical legacy. Yādavas of Devagiri and the Śilāhāras of northern Konkan commissioned hundreds of inscriptions in Maraṭhi as early as the 11th century. Maraṭhi literature dates to about the same period, when the astrological text *Jyotiṣratnamala* is thought to have been composed. The *Līḷācarītra* is a 13th-century biography of the peripatetic Chakradhar Swami; it is a particularly interesting text for linguists as it contains the reported colloquialisms of the many places he traveled. Other significant texts from the 13th century include the works of the bhakti poets Dñāneśwar and Nāmdev.

The chief dialects of Maraṭhi are Deśi, Varhaḍi, and Jhadi Boli. Khandeśi, spoken in the valley of the Tapti river, is sometimes treated as a dialect of Maraṭhi, Gujarati, or a separate language like the Bhili dialects. Standard Maraṭhi is a literary language, but it is most similar to the Deśi dialect spoken from Marathwada up to Pune in the eastern interior regions of the Konkan coast. There is also a dialect of Maraṭhi called Konkaṇi which is spoken further west on the coast itself. This dialect is not to be confused with the separate and distinct Konkaṇi language spoken in Goa. Another major dialect is Varhaḍi, which is spoken in the northeastern Vidarbha district of Mahārāṣṭra as well as in neighboring Madhya Pradesh and Chhattisgarh. It has been heavily influenced by Hindustani, and one important phonetic feature which distinguishes it from the standard is that Standard Maraṭhi [l] surfaces as [y] in Varhadi. Jhadi Boli is spoken in the forest regions of east-central Mahārāṣṭra. Thanjavur Maraṭhi is spoken in Tamil Nadu. Finally,

there is also a dialect of Marāṭhi with heavy Hebrew and Aramaic borrowings, typically dubbed Judeo-Marāṭhi, spoken by the Bene Israel, a Jewish ethnic minority in India.

4.2.3. Konkaṇi

Most speakers of Konkaṇi reside in Goa. There was, however, a significant diaspora following the Portuguese invasion, and Konkaṇi speech communities are found in neighboring states as well. In Goa proper, there is a Goa Hindu Konkaṇi, spoken everywhere in the state, and two regional dialects spoken by Christian communities. Bardes Christian Konkaṇi is spoken in the *talukas* 'counties' of Bardes and Tiswadi north of the Zuari river. Saxtti Christian Konkaṇi is the dialect spoken south of the Zuari river in the *talukas* of Saxtti and Mormugao. Together these regions constitute the "Old Conquests" seized by the Portuguese in the early 16th century. The rest of Goa was seized in the 18th century and occupied until 1961. Outside of Goa, Konkaṇi dialects are also sectarian. Saraswat Brahmans in coastal Karnataka and Kerala speak Southern Saraswat Konkaṇi, while Christians speak Karnataka Christian Konkaṇi.

4.3. Northwest Indo-Aryan

The catalogue of Northwest Indo-Aryan languages is immense, in part because of the overwhelming physical barriers of the region. Northwest Indo-Aryan is spoken along the Indus river valley, all the way up to the intimidating heights of the Hindu-Kush and the Karakoram mountain ranges. Historically, speakers have been relatively isolated in their inaccessible valleys producing one of the most diverse linguistic areas in the world.

4.3.1. Sindhi

Most Sindhi speakers are in the Sindh and Balochistan regions of Pakistan, where their language has been influenced by the administrative languages of Persian, Arabic, and Hindustani, as well as by its linguistic neighbors Balochi, Brahui, and Gujarati. Sindhi is also spoken outside of Pakistan in parts of Gujarat and Rajasthan. The five major dialects of Sindhi are Vicholi, Lari, Lasi, Thari, and Kachhi. Four dialects are spoken within the borders of Sindh itself. Siraiki, in Upper Sindh, is not to be confused with the Punjabi language of the same name. Vicholi, considered the standard dialect, is spoken in central Sindh, while Lari is the dialect in southern Sindh. Lasi is spoken on the western frontier of Sindh and in Balochistan. The Sindhi spoken in the Thar desert of the Jaisalmer district of Rajasthan is called Thari. In Gujarat, Kachhi is spoken along the Rann of Kutch and in the Kathiawar peninsula.

The most striking aspect of Sindhi phonology is its series of voiced implosives, articulated with ingressive air-stream mechanism, believed to be the outcomes of geminated voiced stops. Compare Sanskrit *padma* 'lotus', Pāli *pabba* 'id.', and Sindhi [paɓuṇi] 'lotus plant fruit'. The number of voiced implosives differs from dialect to dialect, but

all have at least one, and curiously none have a dental. In Sindhi an historical dental + [r] > retroflex; compare Sindhi ṭe 'three' with Hindi tīn 'id.' which has lost all trace of an initial cluster *[tr].

4.3.2. Panjabi

The Panjab, from Persian panj āb 'five waters', is a region which encompasses the area of the five tributaries of the Indus. George Grierson (1916) mistook the influence of Hindustani on eastern dialects of Panjabi and categorized it as a far western dialect of Hindustani while grouping western dialects of Panjabi, Hindko, and Saraiki as "Lahnda" languages, a word which simply means 'western' in Panjabi. Today, Panjabi is considered to be merely one language in a "Panjabi language group" which contains the separate languages Saraiki, Hinko, and Panjabi.

Panjabi is the official language of both the Pakistani province of Panjab and the Indian state of the same name. In India, the language is written in a form of nāgarī called Gurmukhī 'from the Guru's mouth', while in Pakistan it is written in a form of the Perso-Arabic abjad called Shahmukhī 'from the King's mouth'. Majhi, spoken around Lahore and Amritsar, is considered to be the standard dialect. Other dialects of Panjabi include Doabi, spoken in the region between the Beas and the Sutlej. Malvai and Puadhi are spoken south of the Sutlej along the boundary of the Haryanvi language area. Panjabi has a very old literary history going back to the 12th century. In the late 15th century, Guru Nanak composed the foundational texts of Sikhism in his native Panjabi, influenced by previous Sufi and Bhakti poets who composed in Persian, Hindustani, and Maraṭhi. The Arya Samaj, a Hindu nationalist group, has used the association of Panjabi with Sikhism to successfully persuade many Panjabi-speaking Hindus to return to their "mother tongue" of Hindi as an act of solidarity.

Many languages in the Panjabi group have tone. This is not the inherited pitch-accent of Indo-Iranian, but an innovation. One of the major differences between Panjabi and Hindko is the number of tones. Standard Panjabi has a two-tone system. The low tone is the result of the loss of aspiration in syllable onset; if this aspirate was word-initial, it became devoiced. Compare Panjabi kə̀r 'house' with Hindi ghər 'id.'. A high tone is the result of loss of aspiration in syllable-coda position. Although there are exceptions, typically the lack of historical aspiration results in a third option: neither high nor low tone. Pothohari, spoken on the Pothohari Plateau, shares this system with Panjabi as does the Western Pahari languages of Dogri and Kangri.

Hindko is spoken in parts of northern Panjab and is the majority language in the province of Khyber Pakhtunkhwa. As opposed to Panjabi, Hindko has a one-tone system. The eastern dialects of Hindko in the Hazara division have a high tone produced like the Panjabi high tone. Hazara Hindko lacks a contrastive low tone, however, and it retains aspirated onsets. Western dialects of Hindko, spoken in Peshawar, also lack a low tone. Peshawari Hindko generates a high tone in two ways, by deaspiration in syllable-coda position and by devoicing in syllable-onset. Compare Standard Panjabi tì 'daughter' with Hazara Hindko dhi and Peshawari Hindko tʼi 'id.'.

Saraiki appears to be a transitional language between Panjabi and Sindhi. Spoken in Upper Sindh as well as the southern Panjab, it is sometimes considered a dialect of

either Sindhi or of Panjabi due to a high degree of mutual intelligibility. Like Sindhi, it possesses a set of implosive voiced stops and lacks contrastive tone. Notice Saraiki *vədh* 'more' retains voiced aspiration while Standard Panjabi *vɔ́d* 'id.' loses it but generates a high tone. There is a political movement in Pakistan to declare Saraiki the administrative language of its own region.

4.3.3. Dardic

Scattered throughout the isolated mountains and valleys of the Hindu-Kush and Karakoram, the Dardic languages elude conclusive proof of their unity. Their similarities are often due to shared archaisms and not innovations. Indeed, the Dardic languages are the most archaic of all NIA languages. One shared innovation which is suggestive of common ancestry is the retroflex affricate series [c̣, c̣h, j̣, ẓh], which are the result of Old Indic consonant clusters. Another similarity is that most Dardic languages show the loss or partial loss of aspiration, often resulting in tone; compare Sanskrit *dhūma* 'smoke' with Pashai *dū̃ᵘm* 'id.'. This similarity is likely to be areal, not genetic, however. Joan Baart (1997: 20) observes that in Kalam Kohistani aspiration is in the process of evolving into a tonal system. Our knowledge of Dardic is often out of date due to the rise of the Taliban and the subsequent war in Afghanistan. Not only is it difficult for linguists to do new fieldwork, but more importantly, war is devastating traditional ways of life. Languages with only a few thousand speakers easily vanish due to the death or relocation of its speakers. In addition, all these languages are under pressure from Pashto and Urdu, the administrative languages of Afghanistan and Pakistan, respectively.

   George Grierson (1919) originally conceived of Dardic as a third branch of Indo-Iranian and included in it the Kafiri languages, a term derived from the Arabic word for 'infidel'. Georg Morgenstierne, whose fieldwork documented many of the languages of the Hindu-Kush for the first time, separated Dardic and these so-called Kafiri languages. Morgenstierne (1961) argued that Dardic was properly Indo-Aryan, while Kafiri was, in fact, a third branch of the Indo-Iranian family. The designation Kafiri has been abandoned in favor of the term Nuristani, coined by Richard Strand (1973). Dardic comprises six groups of languages: Kashmiri, Shina, the Chitral group, the Kohistan group, the Kunar group, and the Pashai group.

   The Kunar and Pashai language groups are spoken primarily in eastern Afghanistan but also in parts of Chitral, Pakistan. The Kunar group of languages is located for the most part in the lower Hindukush in and around the Kunar river valley in east Afghanistan and Pakistan. Gawar-Bati, Shumashti, and Grangali-Ningalami seem somewhat more closely related. It is not clear if Dameli, spoken in the Damel valley on the left bank of the Chitral river, belongs to the Kunar group or to Nuristani. The Pashai group of languages is spoken further west, north of the Kabul River in Afghanistan, in four mutually unintelligible languages with dialects typically named after localities. All Pashai languages, however, have a number of shared features; for example, feminines in [-c] and masculines in [-k].

   The Chitral and Kohistan groups are primarily spoken in the Malakand and Hazara divisions of the Khyber Pakhtunkhwa province of Pakistan. The lingua franca of Chitral prior to the Soviet-Afghan war was Khowar. In a part of the world where many of the

languages documented have no more than a few thousand speakers, Khowar stood out with over 300,000 speakers. The two best known languages of the Chitral group, Khowar and Kalasha-mun, are considered to be the most archaic Indo-Aryan languages spoken today. Whereas most other NIA languages have developed ergativity or split-ergativity, Khowar and Kalasha-mun retain the nom-acc system of Old Indic, as well as the verbal augment. Kalasha-mun retains the Old Indic voiced aspirate series, whereas Khowar has lost it and instead produced a pitch accent; compare Skt *bhūmi* 'earth' with Khowar *buúm* 'id.'. Kalasha-mun has a number of other archaisms, including preservation of the augment, as evident in forms *kārim* 'I do' and *akārim* 'I did', alongside fascinating peculiarities such as a series of retroflex vowels.

The Kohistan group is divided into two language groups: Indus Kohistani and Swat-Dir Kohistani. Indus Kohistani, also known as Maiyã, is spoken primarily in the Upper Kohistan and Lower Kohistan districts in Pakistan. Some important dialects on its fringe are Chiliso Gabar, Bhatise, and Kanyawali. Chiliso and Gabar are dialects spoken on the east bank of the Indus in Kohistan with heavy borrowing from Kohistan Shina. Bhatise is on the east bank of the Indus opposite Besham. Instead of the pitch accent system of other varieties of Indus Kohistani, Bhatise has a complex system of interacting tones and stress accents. Finally, Kanyawali is a dialect of Maiyã spoken in the Tangir valley. Swat-Dir Kohistani is spoken in the districts of Swat, Upper Dir, and Lower Dir in the Malakand division of Pakistan. Kalam Kohistani and Dashwa are spoken in northern Swat, while Rajkot/Patrak and Kalkot are spoken in Dir. Torwali is a language spoken in the Swat valley north of Madyan. Outside of Pakistan, Tirahi is spoken around Jalalabad in Afghanistan. Although influenced by surrounding Pashai, it appears to be more closely related to Kohistani. In light of some "Lahnda"-type features, Morgenstierne (1965: 138−139) suggested it may be a transplant from the Peshawar district. Another Kohistani language documented in Afghanistan is Wotapuri-Katarqalai, now believed to be extinct.

Shina is the majority language in Gilgit-Baltistan, the northernmost administrative territory of Pakistan, but it is also spoken in the Kashmir valley and Ladakh. The prestige dialect is Gilgiti, centered around the capital city. Dialects of Shina are typically named after the valley in which their speakers dwell. Thus, Astori speakers are in the Astore district, and Kohistani Shina is the dialect spoken to the south in Upper and Lower Kohistan. Except for Brokskat, spoken in eastern Baltistan and Ladakh, a tone or pitch accent is common to all dialects of Shina. In Gilgiti, a long vowel is analyzed as having two morae. If the accent falls on the first mora, the result is a high falling pitch on the vowel. If the accent falls on the second mora the result is a low rising pitch. This system seems to be similar to the Burushaski and the Khowar pitch accent. There are a few dialects of Shina outside of the contiguous Shina area. Palula is an archaic dialect originally from the Chilas district transplanted to Lower Chitral. Sawi is a transplant of the same archaic dialect but to Kunar in Afghanistan. Ushojo is spoken in the Bishigram valley near the Swat river; it has similarities to Kohistani Shina. All dialects of Shina retain three contrasting sibilants: [s, ṣ, š].

The Kashmir valley is divided from the Western Pahari languages on the east by the Greater Himalayas and from the rest of Dardic on the west by the Pir Panjal range. Kashmiri, the language of this valley, is distinctive within Dardic for two reasons. First, it has a long attested literary history and second, it has administrative status in the Indian state of Jammu and Kashmir. The earliest specimen of Kashmiri is the *Chummāsaṃketa-*

prakāśa, which is a Sanskrit commentary on brief aphorisms in Old Kashmiri. Its date is uncertain, but it predates Śitikaṇṭha's *Mahānayaprakāśa*. The *Mahānayaprakāśa* is a tantric text in Old Kashmiri whose dating is also debated. Grierson (1929: 73−76) believed it to be a work of the 15th century, archaized by Śitikaṇṭha's knowledge of Kashmiri Apabraṃśa. Sanderson dates the text closer to the 12th century, noting that the poetry of Lāl Ded, a Śaiva mystic firmly dated to the 15th century, is far closer to modern Kashmiri than Śitikaṇṭha's text. Another consideration, however, is that Kashmiri orthography was not standardized at this time, and there is no real critical edition of her work. Lāl Ded's poetry has remained perennially popular in Kashmir to this day, and it is a distinct possibility that later forms may have crept into the texts.

The prestige dialect of Kashmiri is spoken in Srinagar, and it is this dialect which is the written standard. The local script of Kashmir, Śāradā, developed directly from the Gupta script and has been in use since the 10th century. Today, Śāradā is used only by paṇḍits; most use a form of Devanāgarī or Perso-Arabic abjad with additional diacritics. Kashmiri has a set of central vowels [ɨ, ɨ̄, ə, ə̄] and V2 syntax that distinguishes it from other Indo-Aryan languages. Regional dialects of Kashmiri inside the valley include Maraz in the south and southeast and Kamraz in the north and northeast. Outside of the valley, Poguli is spoken in the Pogul and Paristan valleys to the west. Kashtawari, spoken in the Kashtawar valley to the southeast, has retained archaisms that standard Kashmiri has lost.

4.4. Greater Himalayan Indo-Aryan

Pahari means 'hill speech', and thus from the outset the Pahari languages were geographical rather than genetic designations. Western Pahari is primarily spoken in Himachal Pradesh, Central Pahari in Uttarkhand, and Eastern Pahari in Nepal.

4.4.1. Western Pahari

Western Pahari languages have more affinities with the Northwestern group of NIA languages than with Central or Eastern Pahari. The Dogri-Kangri dialect chain, located on the borders of Jammu and Himanchal Pradesh, constitute the best documented Western Pahari languages. Kangri and Dogri were once considered dialects of Panjabi, as they possess the same two-tone system. Pothohari, spoken further northwest on the Pothohar Plateau, is still classified as a Panjabi language because it has the same two-tone system, although it resembles a Western Pahari language in other respects. The designation of each of these languages as "Panjabi" has been predicated on the assumption that the two-tone system is a genetic feature of Panjabi rather than an areal one. Western Pahari languages do borrow heavily from their neighbors, but they are more similar to each other than to Panjabi, Rajasthani, or Dardic. The eastern limit of Western Pahari is Jaunsari, spoken in the Dehradun district of Uttarkhand but containing many Central Pahari affinities. Mandeali is spoken primarily in the Mandi valley. Some have attempted to standardize a "Himachali" from the dialects of this region, but the official administrative language of Himachal Pradesh is Hindi.

4.4.2. Central Pahari

The mountains of Uttarkhand are home to two major languages, both of which are vanishing due to the pressure of Hindi: Garhwali, in the northwest of the state, and Kumauni in the southeast. A third language, Baṅgani, whose status is controversial, is also found in the northwestern tip of Uttarkhand.

Kumauni is splintered into a number of regional variants: Central Kumauni in the districts of Almora and northern Nainital, Northwestern Kumauni in Pithoragarh, and Southeastern Kumauni in southeastern Nainital. Western Kumauni is spoken west of Almora and Nainital in the Garhwali division. Garhwali was once the official language of the Kingdom of Garhwal, with medieval inscriptions surviving from the 14th century. The standard dialect of Garhwali is Srinagariya, spoken around Srinagar in the Pauri district. Other regional dialects of Garhwali include Majh-Kumaiya, along the border of Garhwal-Kumaon and in the Kumayun hills; Badhani, in the Chamoli district; Nagpuriya, in Rudraprayag; Tihriyali, in Tehri Garhwal; and Ranwalti, in the Yamuna valley of Uttarkashi.

Another language spoken in Uttarkashi is Baṅgaṇi, which began to receive scholarly attention when Claus-Peter Zoller (1988) argued that, unlike the rest of Indo-Iranian, Baṅgaṇi was a centum language. He pointed out that its old lexicon contained many forms like *kɔpɔ* 'hoof' (compare Skt. *śapha* 'id.') and *dɔkɔ* 'ten' (compare Skt. *daśa* 'id.'). Later, van Driem and Sharma (1996) stated that they were unable to elicit Zoller's "kentum" forms from their informants. In follow-up fieldwork, Abbi (1997) confirmed the existence of Zoller's forms and found other peculiarities, including forms which had not undergone the RUKI rule, such as *muskɔ* 'bicep' from *mūs 'mouse', a semantic development paralleled by Latin *mūsculus* 'little mouse', the source of French *muscle*. This suggests that an Indo-European but non-Indic speech community switched to an Indic language preserving a core set of lexical items. Linguists have yet to agree on a compelling scenario for this phenomenon, and thus the origins of this aberrant core vocabulary in Baṅgaṇi remain mysterious.

4.4.3. Eastern Pahari

Nepali is the best known language of the Eastern Pahari group. It is the national language of the Federal Democratic Republic of Nepal as well as a *lingua franca* of the Himalayas. It is spoken in India as far west as Kashmir and as far east as Arunchal Pradesh. Nepali speakers can also be found in Myanmar, Bangladesh, and Bhutan. Nepali is characterized by a number of interesting linguistic features, including a complex system of honorifics and infixes which mark verb stems as negative, impersonal, transitive, or causative. The two major dialects of Nepali are Gorkha and Jumli. Gorkha is the standard dialect of the Kathmandu valley. The Darjeeling-Kalimpong dialect, spoken in Darjeeling, is very similar to Gorkha. Jumli is the best known of the western dialects, spoken around Baitadi and Doti. It has many affinities with Kumauni, which is spoken in southeast Uttarkhand. Palpa, now extinct, was the dialect spoken around Lumbini, the birth place of the Buddha. Like Jumli, it had features of Kumauni and Nepali.

Nepal has a rich epigraphical history, but most of it is Sanskrit. These Sanskrit inscriptions often show the influence of Nepali or Newari, the Sino-Tibetan language

indigenous to the Kathmandu valley. The Nepali inscriptional record begins under the western Mallas kingdom in the 13th century. Nepali translations of Sanskrit texts on mathematics and astrology, such as the *Khaṇḍakhādyaka,* begin appearing by the late 15th century. Due to the dominance of the Sanskrit and Newari tradition, however, the Nepali literary arts had a slow start. Nepali poetry is considered to have begun in the 19th century, with Bhanubhakta's adaptation of the Rāmāyaṇa.

4.5. East Indo-Aryan

East Indo-Aryan languages are shielded from the Central Gangetic Plains geographically by the Rajamahal hills and the Chota Nagpur plateau and demographically by Muṇḍa-speaking populations. East Indo-Aryan is bound on the North by the Himalayas and on the East by the Patkai range. Bangla, Asamiya, and Oḍia each use an orthography which developed from the Gauḍī script, which itself is the eastern development of the Siddha-mātṛkā script. The *Charyapada,* an anthology of poems in the Vajrayāna Buddhist tradition, collects materials from between the 8th and 12th centuries CE. The poems seem to capture the transition from a late Apabhraṁśa to early forms of Bangla, Asamiya, and Oḍia. Some of the poems already bear linguistic innovations of Bangla and Asamiya, suggesting that Bangla, Asamiya, and Oḍia were already distinct by this period. The easternmost language in this subgroup is Bisnupur Manipuri. Formerly spoken in Manipur, the language is now dispersed throughout Assam, Tripura, and northeast Bangladesh.

4.5.1. Oḍia

Oḍia is the official language of the state of Oḍisha, although many Dravidian and Muṇḍa languages are spoken in the region as well. The language is sometimes referred to as Oriya and the state Orissa because voiced retroflex stops surface as flaps intervocalically and word finally. In the Oḍia script, however, the phonemic spelling is preserved. After Maraṭhi, Oḍia has the most abundantly attested inscriptional record. The earliest of these is dated to 1051 CE, but Sanskrit inscriptions from previous centuries already contain traces of Oḍia features. The *Mādaḷāpāñji,* 'the Drum Chronicle', represents the collected records of the Jagannath Temple in Puri. Its oldest stratum is the earliest prose literature in Oḍia, dating back to the 12th century. The dialect of Oḍia spoken around Puri is taken to be the standard, while the Northern, Western, and Southern regional dialects show influence from Bangla, Hindi, and Telugu, respectively. Bhatri is a dialect of Oḍia spoken by former Gond tribesmen in Bastar district in southern Chhattisgarh. Halbi, also spoken in Bastar, has features of both Oḍia and Maraṭhi.

4.5.2. Bangla

The official language of the Indian state of West Bengal and the nation of Bangladesh is Bangla, for the term *bāŋāli* properly refers to a member of the speech community and

not the language itself. Bangla has a long literary history but is better known for its recent literature. Rabindranath Tagore, the "Bard of Bengal", was the first non-European to win the Nobel Prize in Literature; he composed the national anthems of both India and Bangladesh.

Bangla has many striking phonological differences from other NIA languages. Its vowel harmony and lack of contrastive vowel length make it similar to the Muṇḍa languages which surround it. Bangla also has a reduced sibilant inventory; while it technically possesses both /s/ and /š/, they contrast infrequently. /š/ is the most frequent sibilant, while /s/ regularly surfaces in clusters /sk, st, sp, sr, sn, sl/. The eastern dialects of Bangla have an alveolar series rather than a retroflex series contrasting with the dentals. Chatgaya, a distinct language spoken around Chittagong in southeast Bangladesh, is related to Bangla but has developed contrastive tone.

4.5.3. Asamiya

Asamiya is spoken primarily in the Brahmaputra and Barak valleys in the state of Assam, separated from Bangla-speaking areas by the Khasi-Garo hills. Asamiya has lost its retroflex stop series, which is unusual for a language of India. Even more unusual, the Old Indic sibilant series /s, ṣ, ś/ have merged into a single velar fricative /x/, while Old Indic palatal /c(h)/ has become alveolar /s/. Kinship terms in Asamiya always specify for seniority or juniority. Another interesting feature of Asamiya is the use of enclitics which categorize the size and shape of the nouns to which they are bound. Central and Eastern Asamiya dialects have medial stress, while in Western Asamiya stress is word-initial. Differences in word stress, speech intonation, vowel quality, and degree of palatalization can strain intelligibility between the eastern and western dialects of Asamiya. Western Asamiya is spoken around Guwahati, Darrang, and Goalpara while the Eastern dialect is spoken primarily in the districts of Sivasagar and Lakhimpur as far west as Sonitpur and Nowgong. The Central dialects span the intermediate regions, although dialects of these regions more often agree with the eastern language. Literary Asamiya developed during the 15th and 16th centuries under the playwright Śaṅkaradeva. Kaviratna's translation of the *Bhāgavata Purāṇa* introduced many Sanskrit borrowings into prose Asamiya. By the 17th century, administrative documents, known as the *Burañjī*s, had introduced Arabic and Persian borrowings into Asamiya. In the neighboring state of Nagaland, the *lingua franca* is Nagamese, a stable creole of Asamiya and the languages of the Naga tribes which are Sino-Tibetan.

4.5.4. Sinhala

Sinhala is spoken in Śrī Laṅkā with no close kinship to any other NIA language, save perhaps Dhivehi, spoken in the Maldives. The Sinhala script is a variety of the Southern Brāhmī which developed under the Pallavas in the 6th century CE. Pāli, also written in this script, is the literary language that accompanied the first Indic-speaking migrants to Śrī Laṅkā. The ancestor of Sinhala is attested in the inscriptional record by the late 3rd or early 2nd century BCE. The oldest inscriptions are in a Sinhala Prakrit. This language

developed in the first millennium into Eḷu, which Suniti Kumar Chatterji (1926: 15) described as "sort of a Sinhalese Apabhraṁśa". Sinhala Prakrit already attests to the major phonological changes which make Sinhala unique. Final [-e] in some forms has been explained by affinity with Eastern Inscriptional Middle Indic, but this seems to be the only feature in common. For example, Eḷu retains the distinction between [r] and [l]. These inscriptions indicate that Sinhala maintained contrastive [ṇ] and [ḷ] centuries longer than mainland Indic. It is a very real possibility that early migrants to Śrī Laṅkā did not constitute a homogenous speech community but rather spoke a variety of Middle Indic dialects.

Sinhala is an Indic enclave surrounded by Dravidian languages. Contact with Dravidian has deeply influenced Sinhalese syntax. Much like Tamil, Sinhala has considerable diglossia; Literary Sinhala differs from Colloquial Sinhala in every respect save phonology. Sinhala phonology, however, cannot be wholly attributed to contact with Dravidian and seems either to have developed independently or to have been influenced by a substrate language which no longer exists. One clue as to what this language may have been like is possibly to be found in the Vedda language. Initially thought to be a dialect of Sinhala, Vedda appears to be a creole of Sinhala and an unknown aboriginal language of Śrī Laṅkā. Perhaps it is from contact with this unknown language family that Sinhala lost aspiration. Compare Sanskrit *dhanus* 'bow' with Sinhala *dunna* 'id.' and Dhivehi *duni* 'arrow'. Sinhala and Dhivehi also both possess a series of prenasalized stops [^{m}b, ^{n}d, ^{n}ḍ, ^{ñ}j, ^{ŋ}g] which contrast with nasal + voiced stop. Compare Sinhala *a^{ŋ}gə* 'horn' and *aŋgə* 'features, components'. Sinhala has a number of phonotactic processes worth noting. All non-high short vowels in medial position undergo reduction to [ə]. The glides [y] and [w] break up vowel hiatus following front and non-front vowels, respectively. The fricatives [s] and [h] alternate medially, with [h] becoming [s] word-finally and in gemination. Sinhala features a few interesting morphophonological processes as well. One is grammaticalized umlaut, in which certain specific morphological processes trigger vowel fronting. Sinhala also has grammaticalized gemination, once again triggered by certain specific morphological processes. In cases of gemination, the prenasalized stop series becomes nasal + voiced consonant. An example which illustrates both grammaticalized umlaut and gemination is Sinhala *bænda* 'tie-PAST' beside *ba^{n}dinəwa* 'tie-PRES'.

4.6. Řomani

A study of NIA would not be complete without a discussion of the "Gypsy languages": Řomani, Domari, Lomavren, and the nearly extinct Ḍumāki, the only one of these languages to remain, broadly speaking, in situ. The root of the names of each of these languages is cognate. Although often designated by the offensive term "Gypsies", they call themselves Řom, Dom, and Lom, respectively, all of which derive from a common root *[ḍom]. These languages appear to be a form of Indic which was born in the Upper Gangetic Plains but matured among the Northwest Indo-Aryan languages. Although some are now settled, speakers of all of these languages historically practiced commercial nomadism, specializing in metalwork, crafting, and music. Řomani, Domari, Lomavren, and Hunza valley Ḍumāki lack the shared innovations that would suggest a unified

speech community. Rather, the first three must be the products of independent migrations out of India.

Řomani dialects are found throughout Europe, but they all share hundreds of roots from the period antedating their entry into Europe, the majority of which are Indic. After Indic, Greek has left the greatest impact, as there are perhaps 250 Greek roots common to all Řomani dialects, which share Iranian and Armenian roots as well. The absence of Arabic roots in the lexicon suggests that Proto-Řomani acquired its Iranian loans prior to the rise of Arabic as an administrative language in Syria and Persia, which probably means prior to the establishment of Damascus as the capital of the Umayyad Caliphate in the late 7th century.

Domari, on the other hand, has a heavy Arabic influence, as it never entered Europe but spread throughout Egypt and the Middle East. Its best known dialect is Palestinian Domari or "Syrian Gypsy", which is spoken by the Dom community in Jerusalem.

The four major dialect groups of Řomani are Balkan Řomani, Vlax, Central Řomani, and Northern Řomani. Balkan Řomani is an extremely conservative group which spans Turkey, Macedonia, Greece, Albania, Kosovo, and southern Bulgaria, although Ursari is spoken further north in Romania. The Drindari-Kalajdži-Bugurdži subgroup is spoken in northeastern Bulgaria and is less conservative than its southern neighbors. All Balkan dialects are marked by greater Greek influence as well as Turkish influence. Abruzzian, Calabrian, and Molisean Řomani are spoken in the south of Italy, but appear to be dialects of Balkan Řomani. The Vlax branch is divided into Northern Vlax and Southern Vlax. Northern Vlax is spoken primarily in Romania, while Southern Vlax has spread outside of Romania into the Balkans. Kalderaš is the best documented variety of Northern Vlax and has had a considerable global diaspora. Lovari is another Northern Vlax language which emerged from Translyvania and spread to Hungary where it is the dominant Řomani dialect. Central Řomani is spoken primarily in the Czech Republic, Slovakia, and Hungary. Northern Central Řomani comprises West Slovak Řomani and East Slovak Řomani which, following the extinction of Bohemian Řomani, have become the dominant Řomani dialects in the Czech Republic. North Central dialects are spoken as far north as southern Poland and as far east as Transcarpathian Ukraine. Southern Central Řomani, called *ahi* dialects due to their imperfect/pluperfect suffix, show considerable Hungarian influence. Vend is spoken in western Hungary, Premurkje in northern Slovenia, and Romungro in eastern Hungary and Slovakia. The western limit of this group is Roman, which is spoken in the Burgenland district of eastern Austria. Northern Řomani is divided into four subgroups: Northeastern, Northwestern, British, and Iberian Řomani. Northeastern Řomani consists of Xakaditka in Russia, Polska in central Poland, and Čuxny in Latvia and Lithuania. Northwestern Řomani consists of Manuš spoken in France and Sinti in Germany, Austria, and the Netherlands. The Finnish dialect of Řomani is closely related to Sinti-Manuš. Laiuse Řomani, a dialect once spoken in Estonia, may have been related to Finnish Řomani, but it was extinguished by Nazi genocide. The British Řomani and Iberian Řomani subgroups are now extinct. British Řomani was known from the Kååle dialect in Wales, and Iberian Řomani survives only as a secret vocabulary in the Spanish dialect of Caló and the Basque dialect of Errumantxela. When Řomani survives total extinction, remaining as a secret dialect, a secret vocabulary furnished with the grammar of another language, it is called Para-Řomani. Lomavren, the secret language of the Lom of Armenia, underwent a comparable development; it has the grammar of an Armenian dialect but an Indic root lexicon.

Řomani, Domari, and Lomavren all show developments which Ralph L. Turner (1926) compared to Śauraseṇī Prakrit, suggesting an origin in the Upper Gangetic Plains. Namely, Old Indic */r̥/ > /i, u/, */sm/ > /mh/, */tv/ > /pp/, /kṣ/ > */kkh/, */t(h), d(h)/ > /l/, and */-m-/ > /-v-/. Compare Sanskrit *bhūmi* 'earth' with Řomani *phuv* 'id.'. Like Northwest Indo-Aryan, however, these languages retain initial dental + /r/ clusters. Compare Sanskrit *trīṇi* 'three' with Řomani *trin* 'id.' and Domari *taran* 'id.'. Domari and Lomavren, however, do not retain initial labial + /r/ clusters as seen when comparing Sanskrit *bhrātar-* with Řomani *phral*, Lomavren *phal*, and Domari *bar*. Note that the extinct dialect of Řomani once spoken in Wales lost this [r]; Kååle *phal* 'brother' is the source of English *pal*.

In Domari the voiced aspirates were deaspirated, while in Řomani and Lomavren they were devoiced independently, as can be seen by comparing Sanskrit *dugdha* 'milk' with Řomani *thud* 'id.' and Lomavren *luth* 'id.'. That is, Lomavren devoices a preform *dudh while Řomani has a preform *dhud, suggesting Řomani transferred aspiration before devoicing its voiced aspirates. Among the three, only Řomani changes initial /v-/ to /b-/, as can be seen by comparing Sanskrit *viś-* 'to enter' with Řomani *beš* 'sit' Domari *wēs-* 'id.' and Lomavren *ves-* 'id.'. The borrowing of Iranian *veš* 'forest' into Řomani suggests /v-/ > /b-/ occurred prior to the departure from the subcontinent. As least, it appears transfer of aspiration occurred after /v-/ > /b-/. Compare Sanskrit *vr̥ddha-* 'old man' with Řomani *phuřo* 'id.' and Domari *wuda* 'id.'. These forms show that in Řomani initial /v-/ became /b-/ giving a form *buḍho. Next, transfer of aspiration produced *bhuḍo and only then did devoicing of voiced aspirates occur. Domari, on the other hand retained *wuḍho* and deaspirated the /ḍh/. These forms also show that Domari merged its retroflex and dental series, while Řomani produced a uvular /ř/ which remained distinct from its dentals.

Finally, Ḍumāki is the nearly extinct language spoken by the Ḍoma of the Hunza valley in northwestern Pakistan. Most Ḍoma speak Shina as well; thus Ḍumāki grammar looks very Dardic, and its linguistic history is difficult to recover. Ḍumāki does not have the archaism of Řomani, Domari, and Lomavren, preserving neither intervocalic dentals nor dental clusters. Compare Sanskrit *madhu* with Ḍumāki *mō* 'wine' and Řomani *mol* 'id.'.

5. Nuristani

Up until the end of the 19th century, the peoples of the Hindu Kush in northeast Afghanistan and north Pakistan had resisted conversion to Islam, for which reason the area was called Kafiristan 'the land of the infidels'. Morgenstierne astutely distinguished within these Kafiristani languages an eastern moiety of Indo-Aryan, the Dardic subgroup, and a western moiety which was something else. When the western half of Kafiristan fell in 1895 to Abdur Rahman, the Emir of Kabul, the region was renamed Nuristan 'the land of light'.

The Nuristani languages are Ashkun, Kati, Tregami, Prasun, and Waigali, which is also known as Kalasa-ala. The relationship of Nuristani to Indic and Iranian is controversial. Like Iranian, these languages do not have aspirated stops, but the loss of aspirated stops could have occurred independently. Like Indic, Nuristani retained Proto-Indo-Euro-

pean /s/ long after Iranian had lost it. The Nuristani reflex of the double dental law patterns with Indic, producing /-tt-/ rather than Iranian /-st-/. In its palatal and labio-velar reflexes, Nuristani looks remarkably similar to what is reconstructed for Proto-Indo-Iranian, allowance made for the loss of aspiration. The palatal *[k̂], which yields sibilants in Indic and Iranian, retains occlusion in Nuristani. Compare Sanskrit *śvan-* 'dog' with Avestan *span-* 'id.' and Waigali *cū̃* 'id.'.

Morgenstierne argued that Nuristani does not apply the RUKI rule after */u/ based on Kati *mū̃sə* 'mouse' and Ashkun *musā* 'id.'. Irén Hegedűs (2012: 149) adds that RUKI does not operate in thorn clusters or historical */k̂s/, instead yielding Proto-Nuristani */c/. Compare Sanskrit *r̥kṣa-* 'bear' with Waigali *ōċ* 'id.'. Her hypothesis is that */k̂/ had become affricate */c/ before the RUKI sound law occurred. The sequence */k^(w)s/ has a distinct outcome as well. Compare Sanskrit *kṣura-* 'razor blade' and Waigali *čür* 'large knife'. She points out that RUKI following */i/ may be a later development as well. /l, r/ does retroflex /s/, but then is lost; compare Sanskrit *varṣa-* 'rain' with Waigali *waṣ* 'id.'. Retroflexion of /s/ behaves quite differently in Nuristani than it does in Indic or Iranian, and on that basis Hegedűs (2012: 145) argues that "Nuristani was the earliest sub-branch to split off from the Aryan branch of PIE and as such had a phonotactic context quite different from that in Indic and Iranian." Such an argument effectively positions RUKI as the shared innovation of Indo-Iranian with Nuristani preserving the Proto-Aryan state of affairs. It bears mentioning, however, that RUKI in Iranian and Indic is triggered by an [r] which was historically *[l] in addition to an original *[r]. The sound change of *[l] > [r] post-dates the breakup of Indic and Iranian by virtue of dialects of Vedic which maintain the distinction between [l] and [r]. Another possibility is that Indo-Iranian RUKI, like Nuristani, was originally triggered by both *[r] and *[l], but in Indic, [l] ceased to trigger RUKI and subsequently the dental [s] allophone was analogically restored. Because *[l], *[r] > [r] in Iranian, no trace of [l]-RUKI would be detectable, and instead Iranian would appear to have undergone RUKI after the merger of *[l] and *[r]. Thus, whatever species of RUKI did occur at the common Indo-Iranian period, its history is obscured by independent developments in Indic and Iranian. Even if Indic, Iranian, and Nuristani are only equally archaic, there is a great deal more to learn about Indic and Iranian from Nuristani, and future fieldwork is a desideratum.

6. References

Abbi, Anvita
   1997   Debate On Archaism of Some Select Bangani Words. *Indian Linguistics* 58: 1−14.
Adhikary, Surya Mani
   1988   *The Khaśa kingdom: a trans-Himalayan empire of the Middle age*. (Nirala History Series 2). Jaipur: Nirala Publications.
Baart, Joan L. G.
   1997   *The Sounds and Tones of Kalam Kohistani*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.
Beams, John
   [1872] 1966   *A Comparative Grammar of the Modern Aryan Languages of India.* Delhi: Munshiram Manoharlal.

Bühler, Georg
    1898   *On the Origin of the Indian Brāhma Alphabet.* (Indian Studies No. III). 2nd edn. Strass-
            burg: Karl J. Trübner.
Burrow, Thomas
    1937   *The Language of the Kharoṣṭhi Documents from Chinese Turkestan.* Cambridge: Cam-
            bridge University Press.
Caillat, Colette (ed.)
    1989   *Dialectes dans les littératures indo-aryennes.* Paris: Collège de France.
Cardona, George and Dhanesh Jain (eds.)
    2003   *The Indo-Aryan Languages.* Routledge Language Family Series. New York: Routledge.
Chatterji, Suniti Kumar
    1926   *The Origin and Development of the Bengali Language.* 2 Vol. Calcutta: Calcutta Univer-
            sity Press.
van Driem, George and Suhnu R Sharma
    1996   In search of kentum Indo-Europeans in the Himalayas. *Indogermanische Forschungen*
            101: 107−146.
Dundas, Paul
    2002   *The Jains.* 2nd ed. New York: Routledge.
Fussman, Gérard
    1989   Gāndhārī écrite, Gāndhārī parlée. In: Colette Caillat (ed.), *Dialectes dans les littératures
            indo-aryennes.* Paris: Collège de France, 433−501.
Gair, James W.
    1982   Sinhala, an Indo-Aryan Isolate. *South Asian Review* 6: 51−64.
Grierson, George Abraham
    1916   *Linguistic Survey of India, Vol. IX. Indo-Aryan familiy, central group; Pt.* I. *Specimens
            of Western Hindī and Pāñjābī.* Calcutta: Office of the Superintendent of Government
            Printing.
Grierson, George Abraham
    1919   *Linguistic Survey of India, Vol. VIII. Indo-Aryan familiy, North-western group; Pt.* II.
            Specimens of the Dardic or Piśācha languages (including Kāshmīrī). Calcutta: Office
            of the Superintendent of Government Printing.
Grierson, George Abraham
    1929   The Language of the Mahā-naya-prakāśa: an Examination of Kāshmīrī as Written in the
            Fifteenth Century. *Memoirs of the Asiatic Society of Bengal* 11: 73−130.
Hegedűs, Irén
    2012   The RUKI-rule in Nuristani. In: Benedicte Nielsen Whitehead, Thomas Olander, Birgit
            Anette Olsen, and Jens Elmegård Rasmussen (eds.), *The Sound of Indo-European: Pho-
            netics, Phonemics, Morphophonemics.* (Copenhagen Studies in Indo-European, Vol. 4).
            Copenhagen: Museum Tusculanum Press, 145−167.
von Hinüber, Oskar
    1981   Die Paiśācī und die Entstehung der sakischen Orthographie. In: Klaus Bruhn and Al-
            brecht Wezler (eds.), *Studien zum Jainismus und Buddhismus: Gedenkschrift für Ludwig
            Alsdorf.* Wiesbaden: Harrassowitz, 121−127.
Hirakawa, Akira and Paul Groner
    2007   *A History of Indian Buddhism: From Śākyamuni to Early Mahāyāna.* Delhi: Motilal
            Banarsidass.
Kielhorn, Franz
    1962   *The Vyākaraṇa Mahābhāṣya of Patañjali.* 3rd edn. Pune: Bhandarkar Oriental Research
            Institute.
Masica, Colin
    1993   *The Indo-Aryan Languages.* Cambridge: Cambridge University Press.

Matras, Yaron
  2002   *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
Morgenstierne, Georg
  1926   *Report on a Linguistic Mission to Afghanistan*. Oslo: Instituttet for Sammenlignende
          Kulturforskning, Serie CI−2.
Morgenstierne, Georg
  1932   *Report on a Linguistic Mission to Northwestern India*. Oslo: Instituttet for Sammenli-
          gnende Kulturforskning.
Morgenstierne, Georg
  1961   Dardic and Kāfir Languages. *The Encyclopaedia of Islam, New Edition, Vol. 2, Fasc. 25*.
          Leiden: E. J. Brill, 138−139.
Morgenstierne, Georg
  1973   *Indo-Iranian Frontier Languages*. 2nd rev. edn. Vols. 1−4. Oslo: H. Aschehoug.
Morgenstierne, Georg
  1975   *Irano-Dardica*. (Beiträge zur Iranistik, Band 5). Wiesbaden: Reichert Verlag.
Oberlies, Thomas
  2005   *A Historical Grammar of Hindi*. Graz: Leykam.
Ollett, Andrew
  2014   Ghosts from the past: India's undead languages. *Indian Economic and Social History*
          Review 51: 405−445.
Pischel, Richard
  1965   *Comparative Grammar of the Prakrit Languages*. 2nd edn. Translated by Subhadra Jhā.
          Delhi: Motilal Banarsidass.
Salomon, Richard
  1998   *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and Other*
          Indo-Aryan Languages. New York: Oxford University Press.
Salomon, Richard
  2001   "Gāndhārī Hybrid Sanskrit": New Sources for the Study of the Sanskritization of Bud-
          dhist Literature. *Indo-Iranian Journal* 44: 241−252.
Sanderson, Alexis
  2007   The Śaiva Exegesis of Kashmir. In: Dominic Goodall and André Padoux (eds.), *Mélan-*
          ges tantriques à la mémoire d'Hélène Brunner). (Collection Indologie 106). Pondicher-
          ry: Institut français de Pondichéry, 231−442.
Saxena, Anju
  2004   *Himalayan Languages: Past and Present*. (Trends in Linguistics. Studies and Mono-
          graphs 149). New York: Mouton de Gruyter.
Senart, Émile
  1881   *Les inscriptions de Piyadassi*. 2 vols. Paris: Imprimerie Nationale.
Sharma, R. N.
  1987   *The Aṣṭādhyāyī of Pāṇini*. New Delhi: Munshiram Manoharlal Publishers.
Sims-Williams, Nicholas (ed.)
  2003   *Indo-Iranian Languages and Peoples*. Proceedings of the British Academy. Oxford/New
          York: Oxford University Press.
Singh, Pashaura
  2012   Revisiting the Ārya-Samāj movement. In: Johannes Bronkhorst and Madhav Deshpande
          (eds.), *Aryan and Non-Aryan in South Asia*. New Delhi: Manohar Publishers & Distribu-
          tors.
Smith, Helmer
  1952   Le futur moyen indien. *Journal Asiatique* 240: 169−183.
Strand, Richard F.
  1973   Notes on the Nūristāni and Dardic Languages. *Journal of the American Oriental Society*
          93: 297−305.

Turner, Ralph L.
  1926   The position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society*, 3rd series 5: 145−189.
Warder, A. K.
  2000   *Indian Buddhism*. 3rd rev. edn. Delhi: Motilal Banarsidass.
Witzel, Michael
  1989   Tracing the Vedic Dialects. In: Colette Caillat (ed.), *Dialectes dans les littératures indo-aryennes*. Paris: Collège de France, 97−265.
Zoller, Claus Peter
  1988   Bericht über besondere Archaismen im Bangani, einer Western Pahari-Sprache. *Münchener Studien zur Sprachwissenschaft* 49: 173−200.

Caley Smith, Cambridge, MA (USA)

31. The evolution of Indic

1. Old, Middle, and New Indic
2. Phonology
3. Morphology
4. Abbreviations
5. References

1. Old, Middle, and New Indic

It is possible to trace a steady development of Sanskrit from the Ṛgveda through the later Vedic texts. The grammar was gradually simplified, mostly by eliminating archaic forms and by reducing the rich varieties of nominal and verbal categories. Side by side with the evolution of Sanskrit the popular vernacular which co-existed with the Vedic "high speech" developed into what is called Middle Indo-Aryan (MIA). Its rise as a literary language coincides with the foundation of the new religions of Buddhism and Jainism in the middle of the first millennium BCE. The first accurately datable documents of this linguistically developed stage of Indo-Aryan are the inscriptions of King Aśoka. MIA can be divided into three linguistic, albeit not strictly chronological, stages − Old, Middle and New Middle Indo-Aryan − covering a period ranging from approximately 500 BCE to 1000 CE. Old MIA is represented by Aśokan Prakrit, Pāli, and Ardha-Māgadhī. The next stage comprises younger Prakrits such as Jaina Māhārāṣṭrī. And the final phase of MIA is instantiated by Apabhraṃśa, which evolved from Prakrit under the influence of the much more developed vernaculars. Nor did this process stop at a particular point, as was the case with Sanskrit, but it transformed MIA in its entirety into what would become New Indo-Aryan.

Vedic texts are composed in a − deliberately − archaic form of Sanskrit. The then *spoken* language was already, it seems certain, far more developed. From it quite a number of features intruded into the hieratic "high speech" of the Veda, where a number