



Linguistic history and language diversity in India: Views and counterviews

SONAL KULKARNI-JOSHI

Deccan College, Pune, India

(Email, sonal.kulkarni@dcpune.ac.in)

This paper addresses the theme of the seminar from the perspective of historical linguistics. It introduces the construct of ‘language family’ and then proceeds to a discussion of contact and the dynamics of linguistic exchange among the main language families of India over several millennia. Some prevalent hypotheses to explain the creation of India as a linguistic area are presented. The ‘substratum view’ is critically assessed. Evidence from historical linguistics in support of two dominant hypotheses – ‘the Aryan migration view’ and ‘the out-of-India hypothesis’ – is presented and briefly assessed. In conclusion, it is observed that the current understanding in historical linguistics favours the Aryan migration view though the ‘substratum view’ is questionable.

Keywords. Aryan migration; historical linguistics; language family; Out-of-India hypothesis; substratum

1. Introduction

The aim of this paper is to lend a linguistic perspective on the issue of human diversity and ancestry in India to the non-linguists at this seminar. The paper is an overview of the major views and evidences gleaned from the available literature.

The paper is organized as follows: Section 1 introduces the linguistic diversity and language families of India. It concludes by stating two contesting hypotheses regarding the peopling of India. Section 2 briefly presents the methods of historical linguistics, some relevant data and inferences based on the data. Section 3 proceeds to examine data made available to address the Indo-Aryan Migration Problem. A summary of the evidence and a conclusion are presented in Section 4.

1.1 Languages and language families of India

India is home to a wide variety of languages. The estimates of the actual number of languages vary. Ethnologue (accessed on 10 September 2018) lists 462 individual languages (these include 448 living and 14 extinct). According to the Census of India (2011), the total number of languages stands at 121: 22 scheduled and 99 non-scheduled (Source: *Language: India, States and Union Territories* (Table C-16), Office of the Registrar General, India. Part 1, p. 21). The estimates differ partly because it is not easy to define language as distinct from dialect – these are entities defined on

the basis of social, political and cultural criteria more than linguistic criteria.

This vast number of languages is classified into four (or six) language families or genealogical types: Austro-Asiatic (Munda), Dravidian, Indo-Aryan (IA) and Tibeto-Burman; more recently, two other language families have been reported – Tai-Kadai and Great Andamanese. South Asia is also home to languages with no known genealogical associations – these are referred to as language isolates; of these, Nihali is an isolate spoken in central India while Burushaski is spoken in Gilgit-Baltistan, Kusunda in Nepal and Vedda in Sri Lanka.

Speakers of the diverse languages and language families have closely interacted with each other over millennia. Exchanges among speakers of all these groups have resulted in convergence across languages and the genesis of India as a linguistic area (Emeneau 1956; Masica 1971).

1.2 Documentation of Indian languages

Systematic documentation of languages,¹ linguistic affinities and classification of Indian languages was undertaken for the first time in the British colonial era. This exercise was largely motivated by the need to understand the culture and social organization of their colony in India better through its

¹ George Grierson’s *Linguistic Survey of India* (1905–1926), which documented 364 languages, remains an indispensable resource for linguists in India.

languages. Their ultimate goal was to consolidate their administrative position in the colony (Kulkarni-Joshi and Hasnain, forthcoming). The knowledge-creation about Indian languages and society was the result of partnerships between British Orientalists and Indian pandits under colonial conditions (Trautmann 1997: 136). British Orientalist discourse on India paved the way for the concept of the superior Aryan race and scientific racism in the twentieth century (Trautmann 1997). This development can be traced back to William Jones's 'discovery' of Sanskrit and his claim for the monogenesis of Sanskrit and of Greek and Latin. Jones, a judge in British Bengal and an Orientalist, made observations regarding the similarity of the Indian language Sanskrit and the European classical languages, Greek and Latin, in his third anniversary lecture at the Asiatic Society of Calcutta on 2nd February 1786.

'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 Gothick and the Celtick, 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.'

(Sir William Jones, Third Anniversary Discourse delivered 2 February 1786)

This discovery was fundamental to the subsequent development of philology, which led to the postulation of historical relations among languages, their classification and, eventually, language families. The discovery by Jones led to the postulation of an Indo-European language family, which became the foundation for comparative philology and the standard of historical linguistics. The Indo-European language family (of which Sanskrit, Greek and Latin were members) is the most extensively studied of these families; the shared ancestry of the speakers of European languages and those of Persian and IA languages was now established. This interest encouraged the comparative philology enterprise, culminating in the proposal for shared parentage and a 'family' of languages which included Indic and European languages. The philological label 'Indo-Aryan' was manoeuvred to construct the superior 'Indo-Aryan' race/ethnic group, members of which had (presumably) originally lived in Central Asia. Thus, the Aryan language speakers in the subcontinent, like their British colonizers, were 'outsiders'. As is well documented, this discovery of philological research was instrumental in proposing that speakers of the

Aryan languages in northern India had migrated into the Indian subcontinent in prehistoric times and, by virtue of their superior knowledge of advanced metallurgical techniques, well-developed rituals, military arts and horsemanship (Southworth 1971: 258), had assumed a socially dominant position, subjugating, in the process, 'original' inhabitants of the subcontinent. Subsequently, The Calcutta-based Orientalists (guided by the doctrine of the linguistic unity of India) had advocated the theory that the south Indian languages were descendants of the Sanskrit language too. However, colonial officers stationed in south India (Madras Presidency) realized the distinctiveness of the south in terms of land systems, languages, ryotwari settlement² and district administration, legal traditions relating to inheritance of property, history and culture (Venkateswarlu 2012: 41). This led to the birth of the Madras school of Orientalism, which was to have an important role in identifying a distinct family of languages in southern India, unrelated to Sanskrit, the genitor of the northern, modern IA languages. The distinctiveness of languages such as Tamil, Kannada, Malayalam and Telugu, spoken in southern India, was noticed. Ellis (1816) and Caldwell (1856) contributed to the consolidation of a separate, Dravidian language family (Trautmann 2006; Venkateswaralu 2012). This was a significant development in the understanding of philology.

In the present day, the majority of Dravidian languages are concentrated in the southern part of the country; there also exist a few northern 'outlier' languages belonging to this family: Brahui (in the Baluchistan province of Pakistan), Kurukh, Malto, Pengo (in the east-central parts of India; see map in figure 1). The IA languages, on the other hand, spread from the north-western parts of India to the north-east; they are also spoken in western India. Two other major language families in India are the Austro-Asiatic (the Munda sub-branch of this family is represented by about a dozen languages in central India and the Khasian sub-branch, primarily in Meghalaya) and the Tibeto-Burman family of languages in the north-east.

The philological enterprise was usurped by indologists interested in the prehistory of the Indian subcontinent.

1.3 India as a linguistic area

The languages in India (the Indian subcontinent), although belonging to different language families, show shared linguistic traits. These include retroflex sounds, SOV word order, absence of prepositions, morphological reduplication (expressives), echo formations, reduplicated verbal adverbs, explicator compound verbs, use of converbs, oblique marked subjects, morphological causatives among many others. Hence, India has been described as a *linguistic area* (Emeneau 1956; Kuiper 1967; Masica 2005).

² This was a system introduced in the nineteenth century in British India in which land ownership rights were given to the peasants and taxes were collected directly from the peasants.

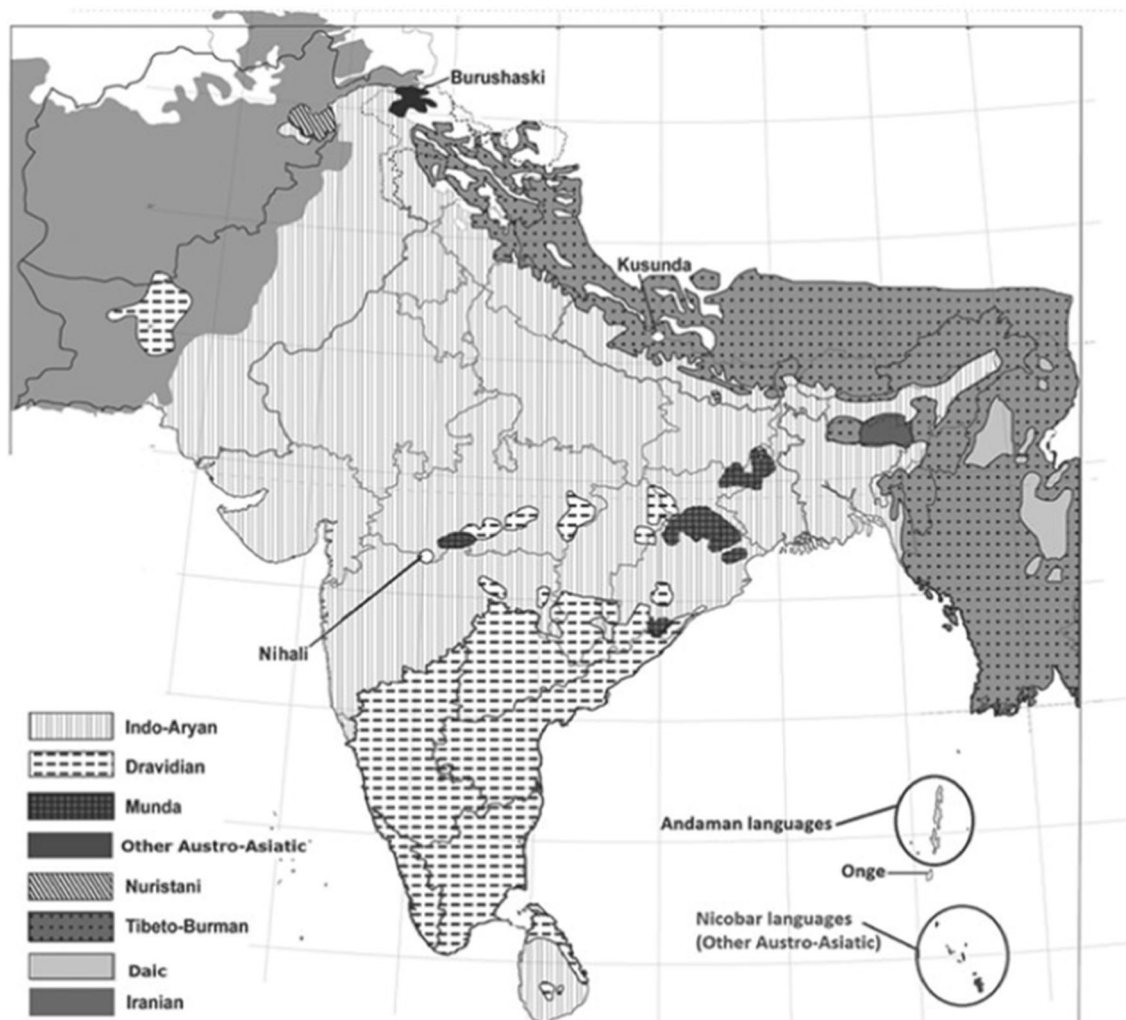


Figure 1. Language families in the Indian subcontinent. (Source of the map used in this figure could not be determined; the map was obtained from the internet)

1.4 Explaining the non-Aryan element in IA languages

Linguists have grappled with non-Aryan elements observed even in Vedic Sanskrit, the earliest attested stage of IA in the Indian subcontinent. These elements cannot be reconstructed for most other branches of the Indo-European language family or the Indo-Iranian sub-family. The general opinion has been that IA must have acquired these features from indigenous languages of the land, the Dravidian and Munda groups of languages. Emeneau (1956, 1970) and Kuiper (1967, 1990), for example, ascribed the occurrence of retroflex sounds, non-finite verb forms and the quotative construction *iti* in Rgvedic Sanskrit to the influence of Dravidian languages. Evidence of borrowed vocabulary and linguistic structure between languages generates an inference of contact between the speech communities. While the *content* of the borrowed vocabulary often provides direct evidence for the cultural areas of contact, the *quantity* and *types* of borrowed material may provide evidence of the *intensity* (social

integration among groups in contact), *range* (proportion of each group which is involved in contact) and *dynamics* (processes of change in the context of the interrelations among the speech communities in contact) of contact (Southworth 2005a: 98–99). The presumed diffusion³ of linguistic traits from early Dravidian (as well as Munda and other languages which no longer exist) into early IA generated speculations about the nature of the sociolinguistic context of interaction among speakers of early IA and the ‘indigenous’ languages. The favoured explanation was that speakers of indigenous languages, in the process of shifting to the socially prestigious IA languages, carried over features

³ *Diffusion*, a one-way process, is distinguished from *convergence* which is the gradual reduction of distance between languages as a result of the diffusion of features through the medium of bilingual speakers.

⁴ The substratum model refers to a situation in which one language intrudes into the territory of another, typically as a result of migration. ‘Substrate’ is a language with lower social power or prestige than a ‘superstratum’ language.

Table 1. Establishing regular sound correspondences in the comparative method

Sanskrit	Latin	Greek	Old English	Old Norse	Old German	Meaning
pitar	pater	pater	faeder	fathir	vater (<i>v=f sound</i>)	father
pad- prthu (broad)	ped- planus (flat)	pod- platus (flat)	fot feald (field)	fotus fold (field)	fuss feld (field)	foot field/flat/broad

of the shifting languages into IA. This is referred to as the *substratum*⁴ effect. The social setting for substratum effect is one of social inequality. The indigenous Dravidians were compelled to learn Sanskrit, the language of the IAs, presumably by the latter's conquest of the land and social prestige. In the course of the shift, the structures of the latter's language were carried over to the Sanskrit language. (A modern parallel to this change could be the influence of Indian languages on the English language brought to this land by the British colonizers.)

The existence of Brahui of the Dravidian family away from the main concentration of the family⁵ was seen to be consistent with regard to the theory that before the Aryan conquest, Dravidians occupied a much greater area including considerable portions of Northern India (Southworth 1971, 1974, 2005a is a strong supporter of this view). Extensive influence of Dravidian on Sanskrit vocabulary and structure since Vedic times is also seen as strengthening this hypothesis (Burrow 1955: 378).

Observation of shared linguistic structures across language families in the subcontinent led to speculation about the relative social status of the linguistic groups (families) in prehistoric times. Early indologists had equated language groupings with racial groupings. Thus, a linguistic observation gained racial overtones. It is important to note that most anthropologists agree it is impossible to claim that speakers of languages belonging to one family all belong to a particular race.

We find a sharp critique of the substratum/subversionist position in Trautmann (1997), who writes: 'By century's end a deep and lasting consensus was reached respecting India, which I call the racial theory of Indian civilization: that India's civilization was produced by the clash and subsequent mixture of light-skinned civilizing invaders (the Aryans) and dark-skinned barbarian aborigines (often identified as Dravidians). The racial theory of Indian civilization has proved remarkably durable and resistant to new information, and it persists to this day' (p. 4).

1.5 Peopling of the Indian subcontinent: Hypotheses

Two major hypotheses are proposed to account for the present-day geographical distribution of language families in India. These are explained below.

⁵ Elfenbein (1987: 215) presents evidence to propose an alternative hypothesis, viz. that the Brahuies are a recent migration to their present habitat from Central India.

Table 2. Historical reconstruction of the proto-form – An illustration

	Latin	Greek	Sanskrit	Gothic	PIE
	<i>decem</i>	<i>deka</i>	<i>daśa</i>	<i>taihun</i>	* <i>dekm</i>
1	d	d	d	t	
2	e	e	a	ai (= e sound)	
3	c (= k sound)	k	ś	h	
4	em	a	a	un	

1.5.1 *Traditional/established view*: Speakers of Aryan languages entered the Indian subcontinent at the beginning of the second millennium B.C. Speakers of the original ancestor language lived in Central Asia.

1.5.2 *Alternative view*:

- Sanskrit was the ancestor language of the Aryan languages in India and in Europe.
- Speakers of the ancestor language lived in India.

Before we proceed to examining some of the linguistic evidence available in support of each of the views, we will briefly introduce some key constructs and methods in the field of historical linguistics. An important objective of this seminar was to understand the basics of the methods of the diverse disciplines represented there.

2. Historical linguistics: Methods, inferences and caveats

One of the important methods used in historical linguistics is the comparative method (the other method being 'internal reconstruction'). The method entails comparing forms (words or morphological endings) in two or more languages with a view to discovering regularities of correspondence. For example, on comparing native words (not borrowed words) meaning 'father', 'foot' and 'field/flat/broad' in Sanskrit, Latin, Greek, Old English, Old Norse and Old German, one can see that where Sanskrit, Latin and Greek have /p/, Old English, Old Norse and Old German have /f/. Such regular sound correspondences establish relatedness of the languages (table 1).

The next question to answer is whether /p/ or /f/ was the original sound. Here general knowledge about linguistic/phonetic processes offers an answer. One could argue that

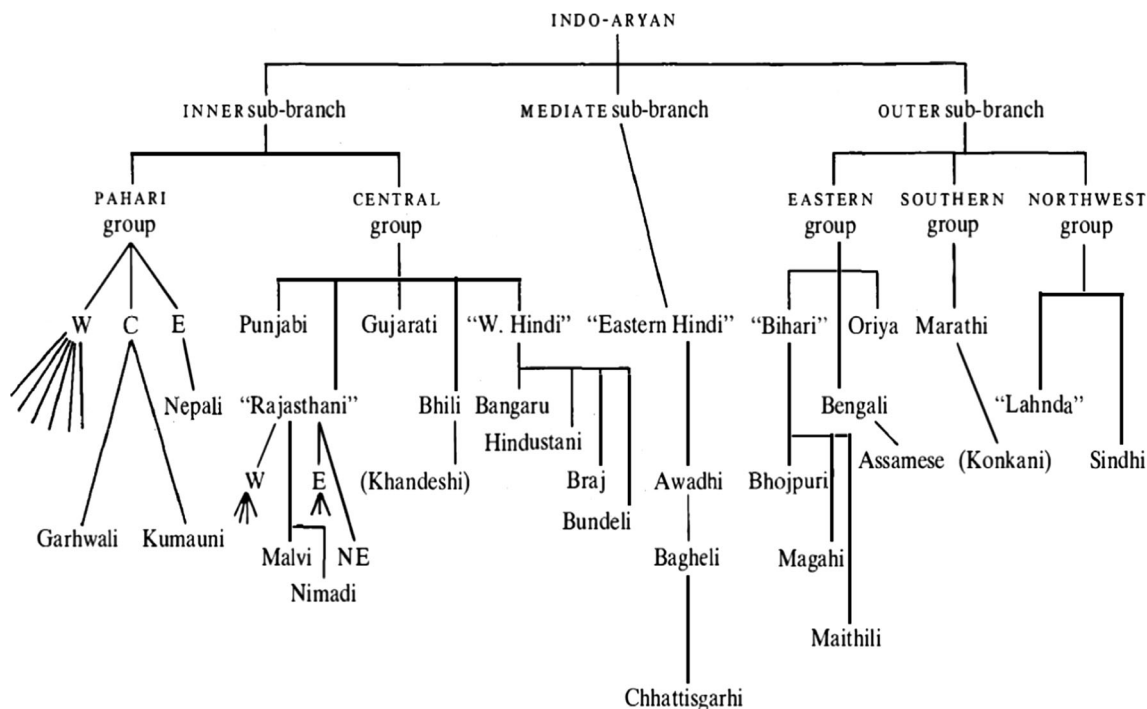


Figure II.2 NIA subclassification according to Grierson (LSI): tree-diagram

Figure 2. The IA language family (Source: Masica C 1991 *The Indo-Aryan Languages* (CUP) p 449).

the fricative is more likely to develop from a stop through a general process of weakening, rather than vice versa.

Thus, the comparative method establishes the relatedness of two or more languages which have presumably descended from a common ancestor language. The evidence can now be used to reconstruct the proto-form in the parent language (shown in table 2). Here, cognates meaning ‘ten’ in Latin, Greek, Sanskrit and Gothic are compared and the form **dekm* is reconstructed for the Proto-Indo-European (PIE) language.

This method enables historical linguists to establish language families (the genealogical model of language change). In figures 2 and 3, we see tree representations of the IA language family and the Dravidian family, respectively.

One caveat that must be borne in mind in this context is that (as was noted above) membership of a language family cannot be equated with membership of a single race. To give just one example, members of a number of non-Aryan tribal groups in India today speak an IA language.

3. Applying methods of historical linguistics to the Indo-Aryan migration problem

In this section we examine linguistic evidence presented to vindicate two competing hypotheses to explain the linguistic affinity between a large majority of Indian languages and languages spoken to the west, in Iran and in Europe. The hypotheses are:

[A] The Aryan Migration Hypothesis [substratum-based model]

[B] The ‘Out-of-India’ Hypothesis

3.1 Early linguistic observations which led to the Aryan migration hypothesis

The non-Aryan element – both lexical and structural – in Rgvedic Sanskrit which was referred to above has been examined by historical linguists since the late nineteenth century.

3.1.1 *Non-Aryan element in Rgvedic Sanskrit: Lexical evidence:* The lexical evidence includes cultural vocabulary as well as vocabulary referring to core domains. Kuiper (1955, 1991) identified 383 words (i.e. 4% of the Vedic vocabulary) as being without Indo-European etymology. (See Burrow 1955, Chapter VIII for detailed examples.)⁶ These loan-words span all of village life and cover local flora and fauna, agriculture and artisanship, terms of toilette, clothing and household. Dancing and music are particularly prominent, and there are some items of religion and beliefs. These borrowed words are said to have increased from the Early Vedic to the Later Vedic and the Classical periods of Sanskrit, but to have declined in the middle IA and modern IA

⁶ However, see Bryant (2012: 70–71) for a critique of Kuiper’s methodology.

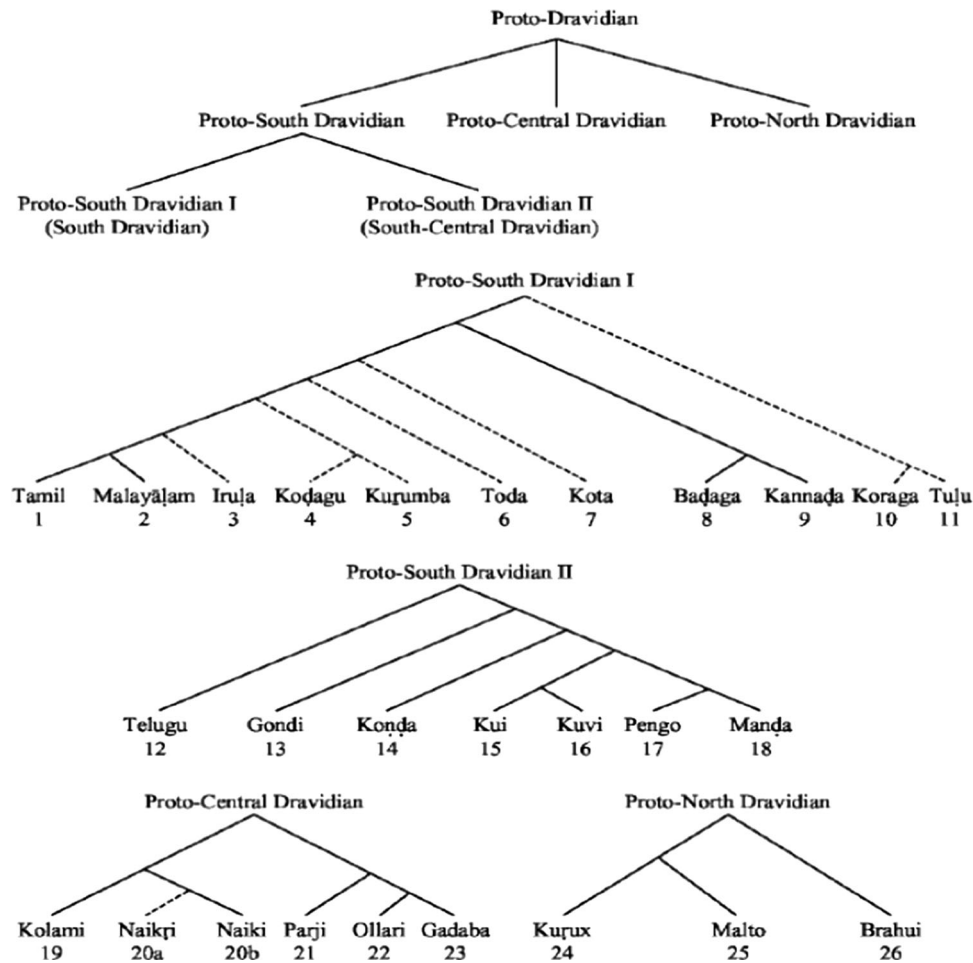


Figure 3. The Dravidian language family (Source: Krishnamurti B 2003 *The Dravidian Languages* (CUP)).

periods. A dominant view among historical linguists is that these non-Aryan loanwords are of either Dravidian or of Munda origin; some lexical items, whose etymology is untraceable, are ascribed to lost substrate languages of the land. Borrowings led to lexical alternatives in Sanskrit: for example, *mahiṣā-* and local words such as *kāsara-*, *luḷāya*, *sailibha*, *heramba-* for ‘buffalo’ (Burrow 1955: 375).

3.1.2 Non-Aryan element in Rgvedic Sanskrit: structural evidence: The two main structural traits which have received attention from (Sanskrit) historical linguists are (i) the presence of retroflex consonants and (ii) the quotative *iti* construction in earliest Vedic texts. We will now briefly outline arguments in favour of and against ascribing a foreign origin to these features.

3.1.2.1 Retroflex consonants At the phonological level, the development of a dental: retroflex contrast in consonants (Skt. *dī-* ‘shine; *ḍī* ‘fly’) in the earliest attested stage of IA has been ascribed a Dravidian origin (Burrow 1955: 96; Emeneau 1980; Kuiper 1991). (The retroflex consonants are articulated with the tip of the tongue turned towards the roof of the mouth.) The contrast is reconstructed for

Proto-Dravidian while the other Indo-European languages are said to have lacked it. Early Dravidian is said to have had a three-way contrast – dental: alveolar: retroflex/post dental.

Kuiper and Emeneau propose that contact with Dravidian led to the redistribution of dental-retroflex allophonic variation in pre-IA. In other words, Dravidian may not have been a direct source of retroflex consonants in Early IA, but served as a catalyst for the change to come about. Historical linguists such as Hock (e.g. Hock 1996 and 2001), who are opposed to the substratum view but see a greater role for bilingualism and convergence in the development of shared linguistic features in the subcontinent, consider the development of retroflexion in IA and in Dravidian as a parallel innovation.

Deshpande (1979) claims that the Sanskrit dental: retroflex contrast developed in post-Rgvedic. He makes the following observation:

‘In the case of Sanskrit, the origin of retroflexion lies not so much in the Aryans’ borrowing this trait from Dravidians in early times as in Dravidians’ adapting Aryan speech to their native phonology. As can be seen from the cultural history of India, by the time of

the *Brāhmaṇa* period, the speakers of the Sanskrit language were not pure Vedic Aryans but were already a mixed people.’ [...] ‘However, to the speech of the Rgvedic poets, retroflexion was most probably still a foreign habit.’ (pp. 297–8)

Hock (1993: 3), who is of the opinion that retroflexion was pre-Rgvedic, speculates that the dental-retroflex contrast may have been a joint, convergent innovation of IA and Dravidian and not a case of subversion (substratum effect). He points to several structural differences between IA and Dravidian in the occurrence of retroflexion (1993: 78). Cumulatively, the evidence does not support the ‘Dravidian as source of retroflexion in IA’ view. He also points out that the development of retroflexion is not an unusual phonological trait in the Indo-European language family – Norwegian and Scandinavian have developed the feature and these were not as a result of language contact.⁷

3.1.2.2 Quotatives The quotative construction is used widely in all Dravidian languages today to mark off direct speech (see Tamil example in 2). The same syntactic construction is also noted in the language of the Rgveda (while other Indo-European languages predominantly use complementizers to mark quoted speech; e.g. English: *He said that ...* and Hindi: *usne kahā ki ...* in which the main clause precedes the quoted clause). (The following examples are drawn from Hock 2001.)

1. Rgvedic

<i>[tvām stoṣāma ...]</i>	<i>iti tvāgne ...</i>
<i>you.sg.acc praise.pl.1.aor.subj.</i>	<i>Q you clit.= Agni.sg.voc</i>
<i>ṛṣayo</i>	<i>‘vocaṇ</i>
<i>seer.pl.nom.m</i>	<i>say.pl.3.aor.</i>

“‘We will praise you ...’ **thus** the seers said to you, Agni’

(RV 10.11 k 5.8-9)

2. Tamil

<i>nāṇ [[avaṇ iṅkē vantāṇ]</i>	<i>eṇru] niṇaikkirēṇ</i>
<i>I.sg.nom he.sg.nom here come.sg.3m.past</i>	<i>Q think.sg.1pres</i>
‘I think [(that) [he has come here]].’	

Hock (2001) re-examines Vedic and early-European and non-literary, northern Dravidian languages as well as of the older southern Dravidian. The evidence suggests a remarkable similarity in the overall syntactic typology of early Dravidian and IA/Indo-European. Both sets of languages agree on the ‘strict SOV’ restriction that sentences have only one finite verb. Hock opines that the quotative (for marking cited discourse) had developed independently in both sets of

languages as a result of typological pressure of ‘Rule A’ which permits them to accommodate subordinate structures to the strict ‘SOV restriction’ by resorting either to non-finite subordination or to relative-correlative structures. Thus, contact, convergence and less so substratum effect have resulted in these constructions.

3.2 Aryan–Dravidian contact: Latter-day parallels

Linguists have examined latter-day parallel contact settings in prehistoric times involving IA and Dravidian languages to reconstruct sociolinguistic processes as a result of which linguistic features come to be shared. Of the modern IA languages, Marathi is today at the frontier between IA and Dravidian. In the writings of Southworth we find a concerted effort to examine the Marathi language to reconstruct a prehistoric contact situation between IA and Dravidian. Southworth (1971) anachronistically uses linguistic data from present-day Marathi to make a case for Dravidian substratum for Marathi. In subsequent writings (1974, 2005a, b) he draws on lexical data from Old Marathi and ethnonyms (place names) in the Marathi-speaking area to make a case for widespread presence of Dravidian language speakers in the region in prehistoric times. For instance, core (i.e. non-cultural) vocabulary items including words for ‘left’, ‘vomit’, ‘side’ and verbs for ‘sprinkle’, ‘pinch’, ‘scorch’ in Marathi are shown to have Dravidian etymology; *kaḍa* is used both as locative postposition and as a marker of possession as in Dravidian. He observes 20% less retention of Sanskrit vocabulary in Marathi as compared with Hindi; he ascribes 5% of the core Marathi vocabulary to Dravidian origin. He (2005) examines ethnonyms and finds *oli* < *vali* < Proto-Dravidian **paḷli* ‘hamlet in the mountains’ to be a commonly occurring suffix in Marathi place names. He finds a relatively heavier concentration of such place names in the south and south-west of the Marathi-speaking region. The evidence suggests that one or more Dravidian languages were earlier spoken in the area now known as Maharashtra. Thus, he favours the Dravidian substratum model as explanation of the Dravidian element in Marathi.

Kulkarni-Joshi (2017) re-examines some of the lexical and structural evidence discussed in Southworth’s writings. She examines lexical data from written sources for Old Marathi – Dnyaneshwari (dated: 1290 AD), Liḷacharitra (dated: late thirteenth century) and inscriptions (dated from circa 1063 to 1298 AD). While she finds the presence of core as well as some cultural vocabulary of Dravidian origin in old Marathi, there is a near absence of core vocabulary of Dravidian origin in Gāthā Saptashati and Lilāvai (texts representing Mahārāṣṭrī Prākṛit, an earlier developmental stage of the language). Further, she reports near absence of structural influence of Kannada on Old Marathi. She notes lexical doublets in Old Marathi such as Kan. *ārogaṇā* vs Mar. *jevaṇa* ‘eat’ and Kan. *poṭ* vs. Skt. *udara* ‘stomach’. In the first example, the verb *ārogaṇā* is used in the text

⁷ Talageri (1993 Chapter 12) is of the opinion that retroflexion was acquired by the other language families of India from Indo-Aryan.

Liḷacharitra when reference is made to Chakradhar swami and the verb *jevaṇa* is used when reference is made to others. She observes a diglossic⁸ relationship in Old Marathi between lexical items of Sanskrit and Kannada origin. She therefore proposes an alternative model/hypothesis based on borrowing with diglossia to account for the Dravidian element in Marathi. She also reports supporting anecdotes received in the contact area in the course of fieldwork in southern Maharashtra: Sanskrit was the language of Gods; Kannada, the language of kings and Marathi is the language of common people. (Here the reference is to 600 years of rule by Kannada-speaking dynasties in the Marathi-speaking region between 543 A.D. to 1334 A.D.)

3.3 Linguistic evidence for the ‘Out-of-India’ hypothesis

In the preceding discussion we noted that there is no consensus among historical linguists regarding the historical processes which resulted in shared linguistic features among languages in India. However, they largely support the view that speakers of the IA language entered the subcontinent in the second millennium B.C. A relatively smaller group of scholars including SS Misra, S Talageri, and D Frawley has contested this proposal. These are the opponents of the Aryan Migration Hypothesis. Of these, some dismiss linguistics as being a pseudo-science, while Misra uses the methods of historical linguistics to assess the available evidence. We will base our review mainly on the discussion in Misra’s (1992) *The Aryan problem: A linguistic approach*.

The important questions that he examines using linguistic evidence are:

- i. Can OIA/Sanskrit be regarded as the mother of all Indo-European languages?
- ii. Can South Asia/India have been the original homeland of the speakers of PIE?

Misra’s main arguments are summed up below.

3.3.1 *The Mitanni language*: The discovery of a treaty dated c. 1400 B.C. between the kings of Hittites and Mitannis of Asia minor which mentions the Vedic gods Mitra, Varuna, Indra and Nasatya as witnesses to the treaty provided an important clue to understanding the prehistory of the Indo-Aryans. A horse training manual was also discovered in Mitanni which mentions numerals (e.g. aika ‘one’ and šatta ‘seven’). Misra, however, argues that the language of the Mitanni is later than Vedic Sanskrit. The Mitanni numeral

for seven is šatta, a form which looks like Prakrit satta rather than like Sanskrit sapta. This suggests to Misra that Mitanni language is at the relatively late chronological stage of Middle IA. He therefore places Vedic Sanskrit at 5000 BC. Hock (1993: 3) argues that, alternatively, the geminate in Mitanni šatta may be derived from šinta/šitta, the word for seven in the non-IE Hurrian language of the subjects of the Mitanni.

3.3.2 *The Proto-Indo European vowels*: The vowels *a-*, *e-*, *o-* in Greek and Latin correspond to *a-* in Sanskrit. Historical linguists find it easier to set up a process of reduction from the vowel system in IE languages in the west to the IA language which was spoken in India.

Skt. <i>a-</i>	:	Gk. and Lat. <i>a-, e-, o-</i>
	ě	
	ǣ	> ä
	ō	

Misra, however, argues that dialects of Gypsy [Romany] language, which emigrated from India much later, have developed the three-way vowel distinction after its speakers left the Indian subcontinent. In Misra’s view a similar change could be posited for the western branches of Indo-European which presumably migrated out of India.

3.3.3 *The kentum/satem divide*: The Indo-European languages are classified into two branches, the centum languages in the west and the satem languages in the east depending on the initial sound (k- or s-) in the words meaning ‘one hundred’. It was assumed that the velars k,g were changed to palatals c,j, via palatalization, a unidirectional process.

3.3.4 *Indo-Iranian palatalization*: The Indo-Iranian merger of the three PIE vowels interacted with another Indo-Iranian sound change: merger of palatalized PIE velars and labiovelars into plain velars. (The following examples are from Hock 2012.)

a.	*deḱm	>	daša ‘10’	(Gk. déka)
	*es-ti	>	as-ti ‘is’	(Gk. esti)
b.	*aǵ-e-ti	>	aj-a-ti ‘drives’	(Gk. ágō)
c.	*oktō	>	aštau ‘eight’	(Gk. oktō)

*k^w, g^w, g^wh } *k, g, gh > *č, j, j h (before y and “front” vowels
i and e)

Reconstructing a three-vowel system for PIE makes it possible to posit the phonetic environment (preceding ‘y; and front vowels *i* and *e*) to account for the process of palatalization undergone by the Indo-Iranian branch. The successive stages in the derivation from Proto IE to Indo-Iranian/Sanskrit are shown below.

⁸ Diglossia refers to a situation in which a community used two languages or two varieties of the same language exclusively for high functions (literature, formal conversation, education, etc.) and low functions (e.g. everyday informal conversation).

- a. Proto-Indo-European *g^we- g^wom-e ‘he, she, it went’
 Velar-Labiovelar merger *ge-gom-e
 Palatalization *j.e-gom-e
 Vowel merger **ja-gām-a** (in Skt.)
- b. Proto-Indo-European *g^we-g^wom-tAe ‘you went’ (2 sg)
 Velar-Labiovelar merger *ge-gom-tha
 Palatalization *j.e-gom-tha
 Vowel Merger ja-gan-tha (in Skt.)

In the view of Indo-Europeanists, therefore, Sanskrit must be a daughter of PIE, like Greek and Latin (Hock 2012: 6).

3.3.5 *The kentum/satem divide and the Bangani language:* The discovery of Bangani, a kentum language, in the midst of satem languages in India also gives fillip to the claim that speakers of the western branches migrated out of India (i.e. India as homeland of Indo-European). (For details of the case of Bangani, see Zoller 1993 and van Driem and Sharma 1996.)

Other arguments by Misra which, as Hock (2012) points out, are difficult to defend on linguistic grounds. Misra argues that if a language shows the results of changes which are found in Middle Indo-Aryan, but not in Sanskrit, then the language must be chronologically at the same level as Prakrit, and therefore younger than Old IA. Thus, Indo-Iranian in Misra’s view is like an early MIA. Misra’s view that Sanskrit is more archaic than all early IE languages is not supported by the available evidence.

Regarding the second important question, was India the original home of the speakers of PIE: historical linguists find it more difficult to refute this hypothesis (since it is not based on linguistic evidence). Hock (2012) remarks that this view would require special pleading and idiosyncratic reasoning. He describes the ‘PIE in India’ hypothesis as a failed attempt.

4. Summary and conclusion

The overview in this paper focused on presenting evidence from historical linguistics used in reconstructing population movements and contact situations of the past in India. The point of departure was the construct of ‘language family’ (the genealogical model of language classification) based on the comparative method in historical linguistics. The first systematic documentation and genealogical classification of Indian languages was traced back to the British colonial administrative practices in the country. Problems in equating language families (a linguistic construct) with racial groupings were noted.

The presence of non-genetic linguistic traits in the language families of India triggered a discussion of the sociolinguistic processes as a result of which linguistic material may have been transferred. We noted that a majority of

historical linguists favour the unidirectional process of language shift/substratum effect to explain the transfer. This model entails a claim for unequal social relations. Our overview proceeded to examining the implications of this linguistic evidence for South Asian (Indian) linguistic pre-history. Of the various migrations/population movements the Indian subcontinent has experienced, the putative in-migration of the Aryan-language speakers has received inordinate attention within historical linguistics. The Aryans (i.e. speakers of Aryan languages) are said to have arrived not in a linguistic vacuum but in a space which was inhabited by speakers of various other languages belonging to either Dravidian or Munda or some lost linguistic groups. We assessed the ‘Dravidian substratum’ view to explain the non-Aryan element in Rgvedic Sanskrit. However, there is no consensus among scholars in the field either on the lexical or the structural evidence. Bryant (2012: 80), for example, concludes his reassessment of the substratum view thus: ‘The apparent “evidence” of a linguistic substratum in Indo-Aryan, in and of itself, cannot be used as a decisive arbitrator in the debate over Indo-Aryan origins.’ Alternative models are being proposed based on fresh linguistic evidence which suggest that the Aryans and other language speakers in India interacted on equal social footing and that the transfer of linguistic material may have resulted from bilingualism and a bidirectional transfer rather than shift to a socially more prestigious Aryan language. Explanation based on bilingualism and bidirectional transfer is a minority view.

We then turned our overview to collating views of historical linguists on the origins of the Aryan language speakers in India. Two competing hypotheses were examined: the Aryan migration view and the Out-of-India hypothesis. Linguistic evidence such as sound changes can be subjected to critical analysis and thus presents stronger evidence. The current understanding of the linguistic evidence does not support the ‘Sanskrit (OIA) as PIE’ view. Further, the ‘Out-of-India’ hypothesis which does not appear to be based on linguistic evidence is difficult to verify or falsify.

To conclude our overview based on secondary sources in the field of historical linguistics: in the current state of our understanding of the linguistic evidence, the question regarding the linguistic ancestry of India cannot be answered with confidence.

References

- Bryant EF 2012 (1999) Linguistic substrata and the indigenous Aryan debate; in Aryan and non-Aryan in south Asia (eds) J Bronkhorst and MM Dehshpande (Delhi: Manohar) pp 59–83
 Burrow T 1955 *The Sanskrit language* (London: Faber and Faber)
 Caldwell R 1856 *A comparative grammar of the Dravidian or south Indian family of languages* (Madras: University of Madras)

- Deshpande MM 1979 Rgvedic retroflexion; in Aryan and non-Aryan in India (eds) MM Deshpande and PE Hook (Ann Arbor: University of Michigan Press) pp 235–315
- Driem GV and Sharma SR 1996 In search of Indo-Europeans in Himalayas. *Indogermanische Forschungen* **101** 107–146
- Elfenbein JH 1987 A periplus of the ‘Brahui Problem’. *Stud. Iranica* **16** 215–233
- Ellis FW 1816 Note to the introduction of A.D. Campbell’s in *A grammar of the Teloogoo language* (Madras). [Reprinted with an editorial note by N. Venkata Rao in *Annals of Oriental Research of the University of Madras*, 12: 1–35. Madras: University of Madras.]
- Emeneau M 1956 India as a linguistic area. *Language* **32** 3–16
- Emeneau MB 1970 *Dravidian comparative phonology: a sketch* (Annamalainagar: Annamalai University)
- Emeneau MB 1980 The Indian linguistic area revisited; in *Language and linguistic area* (ed.) MB Emeneau (Stanford: Stanford University Press) pp 197–249
- Hock HH 1993 Subversion or convergence? The issue of pre-Vedic retroflexion reexamined. *Stud. Linguist. Sci.* **23** 73–115 [published October 1996]
- Hock HH 1996 Subversion or convergence? The issue of pre-Vedic retroflexion reexamined. *Stud. Linguist. Sci.* **23** Fall 1993
- Hock HH 2001 Typology vs. Convergence: The issue of Dravidian/Indo-Aryan similarities Revisited; in *The Yearbook of South Asian Languages and Linguistics* (ed) P Bhaskararo and KV Subbarao (London: Sage Publications) pp 63–99
- Hock HH 2012 Out of India? The linguistic evidence; in *Aryan and non-Aryan in south Asia* (eds) J Bronkhorst and MM Deshpande (Delhi: Manohar) pp 1–18
- Kuiper FBJ 1967 The genesis of a linguistic area. *Indo. Iran. J.* **10** 81–102
- Kuiper 1990 *The Aryans in the Rigveda (Leiden Studies in Indo-European)* (Leiden: Brill)
- Kuiper FBJ 1991 *Aryans in the Rigveda* (Amsterdam: Brill)
- Kulkarni-Joshi S 2017 Substratum effect and the Dravidian element in Marathi: Towards an alternative model. *Int. J. Dravidian Linguist.* **XLVI** 18–54
- Kulkarni-Joshi S and Hasnain SI (forthcoming) Northern perspectives on language and society in India in colonial linguistics, knowledges and epistemes (eds) Deumert A and Storch A (OUP)
- Masica C 1971 *A study of the distribution of certain syntactic and semantic features in relation to the definability of an Indian linguistic area* (Chicago: University of Chicago)
- Masica C 2005 *Defining a linguistic area: south Asia* (Orient Blackswan)
- Misra SS 1992 *The Aryan problem: A linguistic approach* (New Delhi: Munshiram Manoharlal)
- Southworth F 1971 Detecting prior creolisation: an analysis of the historical origin of Marathi; in *Pidginization and creolisation of languages* (ed) Hymes D (Cambridge: CUP) pp 255–274
- Southworth F 1974 Linguistic stratigraphy of north India *IJDL* **3**(2) 201–223
- Southworth FC 2005a *Linguistic archaeology of south Asia* (London: Routledge-Curzon)
- Southworth FC 2005b Prehistoric implications of the Dravidian Element in the NIA lexicon, with special reference to Marathi *Int. J. Dravidian Linguist* **34** 17–28
- Talageri SG 1993 *The Aryan invasion theory: A reappraisal* (New Delhi: Aditya Prakashan)
- Trautmann T 1997 *Aryans and British India* (Berkeley: California University Press)
- Trautmann T 2006 *Languages and nations: conversations in colonial south India* (Berkeley University of California Press)
- Venkateswaralu K 2012 *Colonialism, orientalism and the Dravidian languages* (New Delhi: Routledge)
- Zoller CP 1993 A note on Bangani. *Indian Linguist. Col.* **54** 112–114