

# Indic Language Families and Indo-European

Subhash Kak

*Yavanika*, Number 6, 1996, pp. 51-64

As the science of language, historical linguistics in the early 19th century saw itself as providing a framework for studying the history and relationships of languages in the same manner as biology describes the animal world. But whereas biology has been revolutionized by the discovery of the genetic code, no similar breakthrough has brought new illumination to linguistics. Over the protestations of its many critics, mainstream historical linguistics has remained within the parameters of 19th century thinking. In the meanwhile, archaeological discoveries have altered our understanding of ancient Eurasia (e.g. Renfrew 1987, Feuerstein et al 1995). The Indo-Europeans are seen to be present in Europe a few thousand years earlier than was supposed before. The Indian evidence, based on archaeology as well as the discovery of an astronomy in the Vedas, indicates that Vedic Sanskrit is to be assigned to the 4th and the 3rd millennia BC, if not earlier. The Indian cultural area is seen as an integral whole. The Vedic texts are being interpreted as a record of the complex transformations taking place in the pre-2000 BC Indian society (Shaffer and Lichtenstein 1995).

But the whole edifice of historical linguistics related to the Indo-European family is based on the assumption that Hittite around 2000 BC is the earliest member of the family and Vedic Sanskrit belongs to the period 1200-1000 BC. A major effort is needed to put together a new framework to understand the pre-history of the Indo-European language family. In this note, I consider a few random linguistic questions of interest to the readers of *Yavanikā* that demand fresh examination.

## Language of Paradise

We all understand how the 19th century construction of the Orient by the West satisfied its needs of self-definition in relation to the Other. To justify its ascendancy, the Other was defined to be racially mixed and inferior; irrational and primitive; despotic and feudal. This definition was facilitated by a selective use of the texts and rejecting traditional interpretations, an approach that is now called Orientalism. The terms in the construction were not properly defined. Now we know that to speak of a “pure” race is meaningless since all external characteristics of humans are defined in a continuum. In the 19th century atmosphere of European triumphalism, what obtained in Europe was taken to be normative. With hindsight it is hard to believe that these ideas were not contested more vigorously.

Although this was the age which marked the true beginnings of modern science, old myths continued to exercise great power. When it was found that the languages of India and Europe were related in structure and vocabulary, the West responded with “a tissue of scholarly myths. These myths were steeped in erudition, informed by profound knowledge of Hebrew and Sanskrit, fortified by comparative study of linguistic data, mythology, and religion, and shaped by the effort to relate linguistic structures, forms of thought, and features of civilization. Yet they were also myths, fantasies of the social imagination, at every level. The comparative philology of the most ancient languages was a quest for origins, an attempt to return to a privileged moment in time when God, man, and natural forces still lived in mutual transparency. The plunge into the distant past in search of ‘roots’ went hand in hand with a never forgotten faith in a meaningful history, whose course, guided by the Providence of the one God, could be understood only in the light of Christian revelation. As scholars established the disciplines of Semitic and Indo-European studies, they also invented the mythical figures of the Hebrew and the Aryan, a providential pair which, by revealing to the people of the Christianized West the secret of their identity, also bestowed upon them the patent of nobility that justified their spiritual, religious, and political domination of the world.” (Vernant 1992)

Although the term Aryan never had a racial connotation in the Indian texts, the scholars insisted that this was the sense in which the term ought to be understood. It was further assumed that Aryan meant European by race. By doing so Europe claimed for itself all of the “Aryan” texts as a part of its own forgotten past.

The West considered itself the inheritor of the imagination and the mythic past of the Aryan and the idea of the monotheism of the Hebrew. This dual inheritance was the mark of the imperial destiny of the West. Despite his monotheism, the poor Jew, since he lacked Aryan blood, should have seen “the dark silhouette of the death camps and the rising smoke of the ovens.” (Vernant 1992).

On the other hand, the Asiatic mixed-blood Aryan had no future but that of the serf. He could somewhat redeem himself if he rejected all but the earliest core of his inheritance, that existed when the Aryans in India were a pure race. For scholars such as Max Müller this became ultimately a religious issue. Echoing Augustine, Müller saw in his own religious faith a way for progress of the Asiatic. We would smile at it now but he said, “Christianity was simply the name ‘of the true religion,’ a religion that was already known to the ancients and indeed had been around ‘since the beginning of the human race.’” (see Olender, 1992) But ideas— bad and good— never die. Müller’s idea has recently been resurrected in the guise that Christianity is the fulfillment of Vedic revelation! (e.g. Panikkar, 1977).

A linguistic “Garden of Eden” called the proto-Indo-European (PIE) language was postulated. Europe was taken to be the homeland of this language for which several wonderful qualities were assumed. This was a theory of race linking the Europeans to the inhabitants of the original homeland and declaring them to the original speakers of the PIE. By appropriating the origins, the Europeans also appropriated the oldest literature of the Indians and of other IE speakers. Without a past how could the nations of the empire ever aspire to equality with the West?

Indian literature was seen to belong to two distinct layers. At the deepest level were the Vedas that represented the outpourings of the nature-worshiping pure Aryans. At the next level, weakened by an admixture with the indigenous tribes, the literature became a

narrative on irrational ritual.

## Science and Pseudoscience

In scientific or rational discourse the empirical data can, in principle, falsify a theory. This is why creationism, which explains the fossil record as well as evolution by assuming that it was placed there along with everything else by God when he created the universe in 4004 BC, is not a scientific theory: creationism is unfalsifiable. Building a scientific theory one must also use the Occam's razor, according to which the most economical hypothesis that explains the data is to be accepted.

Bad intent should not turn anyone away from good science. Why isn't PIE good science? It looks reasonable enough: If there are biological origins then there should be linguistic origins as well. And why don't we believe that the nature of language tells us something about culture? If Europeans have been dominant in recent history, then why don't we accept it as a characteristic of the European? If Europe was dominant in ancient times then the origin of the PIE must be in the European sphere from where the energy of its early speakers carried them to the far corners of Asia and allowed them to impose their language on the native speakers.

There are several problems with the idea of PIE. It is based on the hypothesis that languages are defined as fixed entities and they evolve in a biological sense. In reality, a language area is a complex, graded system of several languages and dialects of a family. The degree of homogeneity in a language area is a reflection of the linkages, or interaction within the area. For a language distributed widely in the ancient world, one would expect several dialects. There would be no standard proto-language.

It is clear that language families belong to overlapping groups (Figure 1), because such a view allows us to represent better the complex history of the interactions amongst their ancestor languages. Such an overlap need not imply that the speakers of either group intruded into the overlapping region.

We note further the warning by N.S. Trubetskoy (1939) that the presence of the same word in a number of languages need not suggest

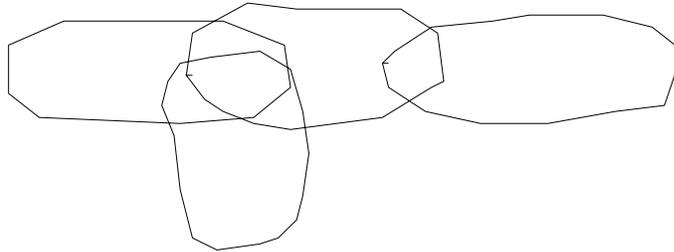


Figure 1: Overlapping language families

that these languages descended from a common parent:

There is, then, no powerful ground for the assumption of a unitary Indogerman protolanguage, from which the individual Indogerman language groups would derive. It is just as plausible that the ancestors of the Indogerman language groups were originally quite dissimilar, and that through continuing contact, mutual influence and word borrowing became significantly closer to each other, without however going so far as to become identical.

The evolution of a language with time is a process governed by context-sensitive rules that express the complex history of interactions with different groups over centuries. The changes in each region will reflect the interaction of the speakers with the speakers of other languages (most of which are now extinct) and various patterns of bilingualism.

There is no evidence that can prove or disprove an original language such as PIE. We cannot infer it with certainty since the historically attested relationship between different languages could have emerged from one of many competing models. If one considers the situation that prevailed in the New World when Europeans arrived as typical, the ancient Old World had a multitude of languages. It is from this great language diversity that a process akin to biological extinction led to the currently much smaller family of languages. The metaphor of something perfect or pure leading to large diversity must be replaced by the metaphor of a web (Robb, 1993). This be-

comes clear when we consider biological inheritance: as we go back in time we have more and more ancestors.

The postulation of PIE together with a specific homeland in Europe or Turkey does violence to facts. There is no evidence that the natives of India for the past 8,000 years or so have looked any different from what they look now. The internal evidence of this literature points to events that are as early as 7000 years ago (Kramrisch, 1981) and its geography is squarely in the Indian region.

If there was no single PIE, there was no single homeland either. The postulation of an “original home”, without anchoring it to a definite time-period is to fall in the same logical trap as in the search for invasions and immigration. Tree or animal name evidence cannot fix a homeland. In a web of languages, different geographical areas will indicate tree or animal names that are specific to these areas. When the European side of the IE languages are examined, the tree or animal names will favour those found in its climate and when the Indian side of the languages are examined, the reference now will be to its flora and fauna.

Colin Renfrew (1987) has pointed out how a circular logic has been used by linguists to justify what has already been implicit in their assumptions. Speaking of the work by Paul Friedrich (1970) on “Proto-Indo-European trees”, Renfrew reminds us that the starting assumption there is that PIE was current in western Caspian and the Carpathians during the fourth millennium and the first centuries of the third millennium and then Friedrich proves that this was the PIE homeland! Reminds Renfrew:

[Friedrich’s] assumption is highly questionable. So complete an adoption of one specific solution to the question of Indo-European origins is bound to have a considerable impact upon his analysis of the origins of tree-names, and the historical conclusions he reaches. It is scarcely surprising if his theory harmonizes with the historical reconstruction upon which it is based. It is perhaps reasonable that the historical linguistics should be based upon the archaeology, but that the archaeological interpretation should simultaneously be based upon the linguistic

analysis gives serious cause for concern. Each discipline assumes that the other can offer conclusions based upon sound independent evidence, but in reality one begins where the other ends. They are both relying on each other to prop up their mutual thesis.

### Aryan and Dravidian

It was Bishop Caldwell (1875) who suggested that the South Indian languages of Tamil, Malayalam, Kannada, and Telegu formed the separate Dravidian family of languages. He further suggested that the speakers of the proto-Dravidian language entered India from the northwest. Other scholars argued against this Dravidian invasion theory. Scholars have argued that this attempt to see both the North and the South Indian languages coming to the subcontinent from outside (West Asia) as another example of the preoccupation with the notion of the “Garden of Eden”. In reality, the problem of what constitutes an Aryan or a Dravidian, in the biological or cultural sense in which it is generally posed, is insoluble.

The problem of Aryan and Dravidian is a conflation of many categories. Indian texts do not use the term Arya or Aryan in a linguistic sense, only in terms of culture. There is reference in the *Manu Smṛti* where even the Chinese are termed Aryan, proving that it is not the language that defines this term. The South Indian kings called themselves Aryan as did the South Indian travelers who took the Indian civilization to Southeast Asia.

One may have posed the problem in terms of the anthropological ‘distinction’ between the speaker of the North and the South Indian languages. But the anthropologist tell us that there is no difference.

When linguists in the last century insisted that the term “Aryan” be reserved for the North Indian languages alone, it was inevitable confusion would emerge (Kak, 1994). The definition of Aryan and Dravidian are extrapolated from the culture of the speakers of the North and the South Indian languages. But the cultures of the North and the South are the same as far back as we can go. (There is some minor difference in kinship rules.) There is even a mirroring of the

sacred geography. The North has Kashi and Mathura; the South has Kanchi and Madurai. Who is to say what was the original? If there is no cultural difference then the use of the term “Aryan” as defining the culture of just the speakers of the North Indian languages is misleading.

This following example puts the absurdity of the terminology in focus. There exist texts that state that Tamilian Hindus came and settled in Kashmir in the early 15th century in the liberal reign of Bada Shah. We don’t know how many people came, but that is the nature of such textual evidence anyway. Now what does that make a Kashmiri? An Aryan or a Dravidian?

Some scholars have claimed a Dravidian substratum for Marathi, but how do we know that prior that Dravidian substratum there was not some other language that was spoken there? And maybe there has been more than one shift back and forth.

Let’s imagine that everyone in India originally spoke Dravidian and then due to some process of “elite dominance” most people in the North started speaking Indo-Aryan and they kept their old traditions and legends. The new speakers will still be culturally Dravidian and certainly they would be so “biologically”, if that could ever mean anything. If this is what happened in India then are the Aryans actually Dravidians and, by implication, are the Dravidians also Aryans? There could be two groups of people speaking two different languages who culturally belong to the same tradition like the modern-day Hungarians and Czechs.

We don’t know who were the authors of the Vedas. They could have been bilinguals who knew ‘Dravidian’ and ‘Vedic’; maybe their first language was really Dravidian even though they had Sanskrit names as has been true in South India for much of historical times; or they were purely Sanskrit speaking. No rhetoric or ideology can resolve this question.

The use of a language in literature does not even mean that the speakers are a dominant elite. Let’s consider the use of Urdu in Pakistan. The Punjabi speaking Punjabis are the dominant group but Urdu is used for official work purely due to some historical factors. In fact, the only Urdu-speaking ethnic group in Pakistan, the Mohajirs, feel they are at the bottom of the totem pole.

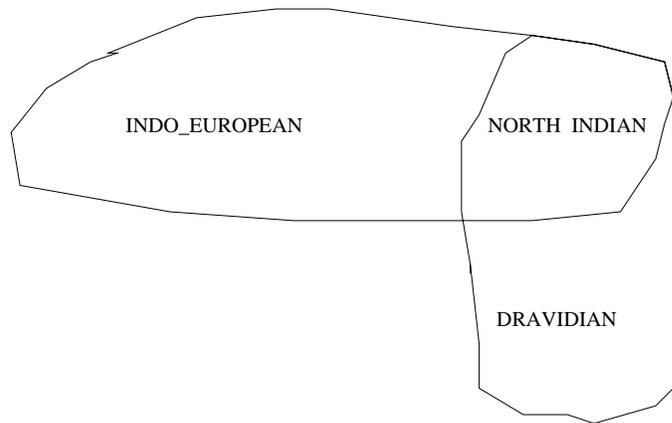


Figure 2: Indo-European and Indic families. The Indic family has the sub-families of North Indian and Dravidian

The texts cannot reveal the ethnic background just as Indians in the US who have adopted American names cannot be identified as ethnically Indian from their writing. The lesson is that the term “Aryan”, misused by so many different parties, should be retired from academic discourse.

### Several kinds of families

The Indian linguistic evidence requires the postulation of two kinds of classification. The first is the traditional Indian classification where the whole of India is a single linguistic area of what used to be traditionally called the Prakrit family. Linguists agree that based on certain structural relationships the North and the South Indian languages are closer than Sanskrit and Greek (Emeneau, 1980).

Second, we have a division between the North Indian languages that should really be called North Prakrit (called Indo-Aryan by the linguists) and the South Indian languages that may be called South Prakrit (or Dravidian) (Figure 2).

There is also the Indo-European family to which the North Prakrit languages belong. Likewise, Dravidian has been assumed to belong to a larger family of agglutinative languages.

This classification will allow us to get rid of the term Aryan in the classification of languages which is a good thing because of the racist connotation behind its 19th century use. Its further virtue is that it recognizes that language families cannot be exclusive systems and they should be perceived as overlapping circles that expand and shrink with time.

### **Back to Ancient India**

Some Indologists driven by the old race paradigm have stood facts upside down to force them to fit their theory. We know that the internal evidence of the Indian texts shows that the Vedas precede the Puranas. But since the Puranic themes are shown in the iconography of the Harappan times (2600-1900 BC), the Puranic material is taken to precede the Vedas so that the Vedas could be placed in the second millennium BC.

I think the only logical resolution of all the archaeological and textual evidence is to assume that the Indic area became a single cultural area at least around 5000 BC. The Indian civilization was created by the speakers of many languages but the language of the earliest surviving literary expression was Vedic Sanskrit, that is itself connected to both the North and the South Prakrit languages.

This idea is supported not only by the internal evidence that shows that the Indic tradition from 7000 BC onwards is an indigenous affair, but also from the new analysis of ancient art (Kak, in press). For example, David Napier (in press) shows how the forehead markings of the Gorgon and the single-eye of the cyclops in Greek art are Indian elements. Although he suggests that this may have been a byproduct of the interaction with the Indian foot soldiers who fought for the Persian armies, he doesn't fail to mention the more likely possibility that the influence was through the 2nd millennium BC South Indian traders in Greece. This is supported by the fact that the name of the Mycenaean Greek city Tiryns — the place where the most ancient monuments of Greece are to be found— is the same as that of the most powerful Tamilian sea-faring people called the Tirayans.

## Greece and India

Since the 2nd millennium interaction between Greece and India is becoming clear only now, it is appropriate to ask if their languages were frozen into fixed categories wrongly by the 19th century historical linguists.

Consider the *centum/satem* divide in which Greek belongs to the *centum* group and the North Indian languages belong to the *satem* group. The old tree model is used to divide the PIE into these two sub-classes with the *centum* group representing the western branch and the *satem* group representing the eastern branch. The discovery of Tocharian as a *centum* language was seen as an example of heroic a movement of *centum*-speaking people from the west. But now the discovery of Bangani, a *centum* language in India itself has make the whole idea of a tree-like division suspect.

Consider also the question of our knowledge of the vocabulary of various languages. For some languages, this knowledge was primarily obtained in quick field-work done decades ago by scholars who were not native speakers. Could it be that they missed out on vital evidence?

Mallory (1989, page 114) informs us that the word *\*mori* “seems originally to have meant swamp, marsh land or lake, rather than a large body of open water. [I]t is found only in European languages and not in Indo-Iranian other than Ossetic – an Iranian language contiguous to Europe although originating further to the east.” This “fact” has lent itself to endless theorizing. But this “fact” is a result of incomplete surveys. The word *mār*, a cognate, is a common Kashmiri term for a swamp or even a lake. We see this word in the formation of Kashyapmār from which the word Kashmir is derived. Even Kannada has a cognate.

Also, many Hindi speakers pronounce the word for “hundred” as *sainkarā* rather than *saikarā*, which the field studies tell us. Does that make Hindi a *centum* language?

Certain assumptions regarding provenance and chronology were used to devise the vocabulary of PIE. The assumptions regarding chronology were shown wrong by the discovery that the Ṛgveda should be dated prior to at least 1900 BC.

### **Concluding Remarks**

The archaeological findings from India and the discovery of the astronomy of the Vedic period are fatal for the constructions of historical linguistics that arose in the 19th century and are still being followed in schoolbooks in India although textbooks in the West have begun to present the new picture. While the general language categories seem reasonable, the concept of overlapping families seems essential to obtain better conceptual clarity. The Indic family is an example of such overlapping families.

The breakdown of the old paradigm calls for considerable effort to create a new one to take its place. In particular, the emerging chronological framework can be used to examine the relationships between Sanskrit and other ancient Indo-European languages. Etymological dictionaries should be revised to take note of the antiquity of Vedic Sanskrit. If PIE did not exist, can we extrapolate from the earliest layer of Vedic Sanskrit for correlations with life in prehistoric Harappan India?

## Notes and References

- Caldwell, R. 1875. *A Comparative Grammar of the Dravidian Languages*. 2nd edition. London.
- Emeneau, M.B. 1980. *Language and Linguistic Area*. Stanford University Press.
- Feuerstein, G. Kak, S. and Frawley, D. 1995. *In Search of the Cradle of Civilization*. Wheaton.
- Friedrich, P. 1970. *Proto-Indo-European Trees*. Chicago.
- Kak, S. 1994. On the classification of Indic languages. *Annals of the Bhandarkar Oriental Institute*, 75, 185-195.
- Kak, S. in press. Indic ideas in the Graeco-Roman world. *Indian Historical Review*, in press.
- Kramrisch, S. 1981. *The Presence of Siva*. Princeton University Press.
- Mallory, J.P. 1989. *In Search of the Indo-Europeans*. London: Thames and Hudson.
- Napier, D. in press. "Masks and metaphysics in the ancient world: an anthropological view." To be presented at the *International Seminar on Mind, Man and Mask*, Indira Gandhi National Centre for the Arts, New Delhi, Feb 24-28, 1998.
- Olender, M. 1992. *The Languages of Paradise: Race, Religion, and Philology in the Nineteenth Century*. Cambridge: Harvard University Press.
- Panikkar, R. 1977. *The Vedic Experience*. Berkeley.
- Renfrew, C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.
- Robb, J. 1993. A social prehistory of European languages. *Antiquity*, 67, 747-760.

Shaffer, J.G. and Lichtenstein, D.A. 1995. "The concept of 'cultural tradition' and 'palaeoethnicity' in South Asian archaeology." In *The Indo-Aryans of Ancient South Asia.*, G. Erdosy (ed.). 126-154. Berlin.

Trubetskoy, N.S. 1939. In Renfrew (1987) page 108.

Vernant, J.-P. 1992. In Olender (1992).