

On The Classification Of Indic Languages

Subhash Kak*

1 Introduction

Language, as part of human expression, may be viewed in analogy with genetic expression. Evolution of language is a result of complex temporal and spatial processes where, if one could aggregate the processes, one may speak in terms of parent traits and the resultant descendent traits. Insights from the theory of non-linear dynamics indicate that the multitude of interactions amongst speakers would lead to the formation of just a few languages. Strongly interacting systems of very many components, like assemblies of neurons or human speakers, have only a few stable interaction states, called attractors, associated with their behaviour,¹ and these, for speakers, are the various languages. In evolving systems, the nature of these stable states will also change. This is how isolated languages can be seen to change. But more significant than this process is the change due to interaction with other languages. With this background it is clear that a correct view of language evolution is within the framework of other interacting languages.

But for about one and a half centuries, language evolution has been studied using models inspired by early, mechanistic physics. Like a physical system that evolves due to radiation and other incident forces, languages were taken to change spontaneously. The spread of languages was explained by another mechanistic metaphor, namely, that of transfer of populations and invasions. This led to models of language families. The German philologist August Schleicher pioneered the tree approach in the 1860's which assumes that when populations are isolated their speech get increasingly differentiated until they become distinct languages; this assumption allows one to set

*Annals of the Bhandarkar Oriental Research Institute, vol. 75, pp. 185-195, 1994

up a family tree of languages. Representation of language families is predicated on an assumed chronology of evolution. Soon after Schleicher, another German linguist, Johannes Schmidt, theorized that linguistic changes spread in “waves” leading thereby to a convergence amongst languages that might have been dissimilar to begin with. In 1939 the Soviet linguist N.S. Trubetskiy suggested that the similarities among the Indo-European languages were due to the wave model of Schmidt. Scholarly opinion has generally dismissed “wave advance” theories and languages are generally characterized in terms of family trees.

2 On Language Families

But language family representation that does not consider the previous history of interactions cannot be reliable; even in the case of an isolated population it is too simplistic. Using the analogy of biological family trees, the daughter language must carry characteristics of the parent languages, where the parents aggregate the influence of all dissimilar languages and dialects. If language grammar and vocabulary is likened to the genes of a biological organism, the daughter language picks up genes from both the parents. But since a language is defined by the interaction and behavior of diverse speakers across space and time, the actual inheritance in the daughter language is a chance phenomenon. Nevertheless, genetic classification of languages routinely speak of a single parent language. For example, Spanish, Catalan, French, Italian are seen to be the daughter languages of Latin without defining the other parents.

Theories of language evolution arose in the heyday of mechanistic physics, before the laws of genetics and quantum mechanics had come to be known. Since the discovery of these laws, no successful attempt has been made to establish a rational basis for inheritance of characteristics in languages.² Recent theories do claim to provide “genetic” classification, but the term “genetic” is used in an unscientific manner. It is used in a meaning equivalent to the old tree classification diagrams or in the operative sense of “random mutations”. However, random mutations in biological evolution are supposed to represent the cumulative effect of complex interactions. Furthermore, significant mutations are seen only after many, many generations. The historical records related to languages exist over a time span that is relatively very

brief and no convincing evidence exists that defines processes, over such a brief period, that are truly analogous to biological random mutations.

The current state of linguistics is due, in part, to the central place the study of Indo-European languages has had on the subject. Implicit in such a study has been the Eurocentric notion of the special place of the hypothesized Proto-Indo-European (PIE) language and thereby its homeland. Circular arguments were used to postulate IE forms and then the words in the various IE languages were derived from it. The languages were related in terms of tree diagrams without considering the history of their interactions with other languages. Another recent tendency is to derive all languages from the same ancestor. Here the motivation is to use models that describe the genetic diversity of human populations. But I believe that we simply do not have the data at this point to determine whether language arose before the postulated early human migrations from the original single homeland of the humans. Neither do we know if there was a single such homeland.

The *comparative method* that has been used to reconstruct features of ancestral languages may be compared to a sieve. Using a sieve of a certain size to find diamonds in dirt, one may theorize that such diamonds have a certain minimum size. But such a theory does nothing more than declare the limitations of the sieve! This is not to say that languages are not related, but that the relatedness is much more complex than the techniques used in historical linguistics indicate. No wonder then that linguists have reached seemingly contradictory conclusions: (i) There is such typological commonality between the Indo-Aryan, Munda, and the Dravidian languages that these languages should be considered a single super-group and India considered a “linguistic area,”³(ii) Sanskrit and Old-Indo-Aryan are strikingly similar to Old Iranian, a language taken not to have been influenced by Dravidian, so that the Avestan texts can almost be read as Vedic Sanskrit.⁴

With the backdrop of the above points, we take up the question of the classification of the Indic languages to illustrate the pitfalls of current theories. We argue that based on *genetic* classification, both the Indo-Aryan and Dravidian languages have had common parents and these languages share many typological categories.

3 Indo-European and Dravidian

We first consider the wider question of the relationship between Indo-European and Dravidian. Three decades ago the Soviet linguists Vladislav M. Illich-Svitych and Aron Dolgopolsky proposed that a number of Eurasian language families including Indo-European, Dravidian, and Afro-Asiatic belong to a superfamily which they called Nostratic,⁵ derived from the Latin for “our (language)”. Although the notion of the superfamily is sometimes taken to imply a common ancestor, it appears that a more reasonable assumption is that in the remote past the speakers of these languages interacted strongly resulting in many shared characteristics amongst the languages.

The idea of the superfamily has been increasingly accepted in recent years. The spread of these languages has been ascribed to various mechanisms. One mechanism is the “wave of advance” model of Ammerman and Cavalli-Sforza,⁶ according to which the surplus produced by agriculture led to rapid increase of population density over earlier hunter-gatherer communities. The second popular model is that of elite-dominance; here the spread is generally ascribed to invasions.

It has been suggested that the ancestors of these three families may have lived in some proximity in Western Asia around 7000 B.C. Colin Renfrew sees the ancestors of the Indo-Europeans in Anatolia, those of the Afro-Asiatics in Jericho, and those of the Dravidians in the Zagros.⁷ If one postulates that early farming arose in these regions of Western Asia then the spread of farming by the “wave of advance” mechanism took their languages and genes into other areas. Although, the presence of Indo-European languages in Iran and India is explained by Renfrew as a later expansion by an elite that forced its language on the Elamite and the Dravidian speaking people, this is not convincing. This is a restatement of the theory articulated earlier by Childe⁸ and others which has no archaeological evidence to support it.⁹ There is no explanation for why suddenly hordes from Anatolia decided to push in the southeast direction and how they were able to impose their language on an area which was already heavily populated.¹⁰

There are other theories for the spread of the Indo-European languages, amongst which the most prominent is the “kurgan” theory of Marija Gimbutas¹¹ which is, however, concerned mainly with Europe. According to this theory kurgan warriors from north of the Black Sea invaded Europe in waves

over the period 4300 to 2800 B.C. and imposed their languages on the indigenous Europeans. The expansion into Iran and India in the Gimbutas scheme is taken to be the old intrusive model as has been described by Mallory.¹²

The spread of the Indo-European languages is thus related to the problem of the location of their original homeland. But as J.P. Mallory summarizes:

Since the 19th century, attempts to resolve the problem of Indo-European origins have included evidence drawn from physical anthropology. This may be broadly divided into four traditions—pigmentation, cranial index, the correlation of physical types (based on multivariate analysis) and archaeological cultures, and genetics. None of these have satisfactorily determined the location of the Indo-European homeland.¹³

The various choices for the homeland of the different language groups is quite arbitrary. It is foolhardy to associate a language to a reconstruction of an ethnic type based on archaeological records.

If one considers the astronomical references in the Vedic literature, then one can postulate the presence of Indo-Europeans in Northwest India in the fourth millennium B.C. and earlier.¹⁴ The priority of the Indic literature makes Northwestern India as another candidate for the homeland of the Indo-Europeans. But the question of the location of the homeland is in many ways an inappropriate question to ask with the current state of knowledge. The choice of the homeland and the original physical type is strongly correlated with the nationality of the proponent! Many North European scholars thus argued that the original Indo-Europeans were blond. It is not surprising then that most Western scholars did not consider Northwestern India as a viable candidate.

Whatever model one might choose, the relationship amongst the Nostratic languages is ascribed to proximity about eight thousand years ago. In turn these languages are taken to be derived from a yet earlier parent or to have picked up their shared characteristics from their early interaction.

The characterization of the Nostratic superfamily is based on the assumption that the relationship was defined at the pre-expansion phase. Such an assumption is inherent in a tree classification.

The search for a single superfamily of all languages is driven by the assumption that language arose only at one place. This hypothesis cannot be proved or disproved, so its discussion falls outside the purview of science. Since there do not exist any isolated populations there is no way to determine if the commonality being seen now is a result of historical interaction or is to be explained as a remembrance of the common origins.

In reality a tree classification is a misnomer. There is a further implicit assumption that the languages diverge from each other because their speakers are in societies undergoing different changes and are interacting with speakers of different languages.

4 On Language Identity and Societal Processes

Societal processes and organization determine how long a language will maintain its identity as the speech of a minority group. Thus Murray Emeneau reminds us that Saurashtran weavers in Tamil Nadu appear to have preserved their language for a period that could be more than a thousand years.

After a period of at least fifteen centuries of migrations, Saurashtran still survives as the domestic language of the immigrant silk weavers of Madura. The historical events of their migrations were certainly very complex. The sequence, partly known from their traditions, brings them from Saurāṣṭra (Lāṭa-viṣaya) to Mandasor in Rajasthan prior to the fifth century A.D. (inscriptions there record the building of a temple in A.D. 437-438 and its repair in 473-474), then to Devagiri of the Maharashtra Yādavas (fl. thirteenth century), to Vijayanagar (Telugu-speaking; fl. fourteenth-sixteenth centuries), and finally to Madura. Whatever degree of exactness may be attributed to this tradition and history, the language certainly has traits that point to all the linguistic areas involved, but yet has been preserved over these many centuries of sojourn away from its place of origin. In every place the weavers were probably lower in the social structure than at least some of the neighboring communities (in spite of their present brahmanical pretensions), but there was no

American-like pressure for total linguistic conformity with these neighbors.¹⁵

There are other examples that can be given from India. In contrast, minority groups have tended to lose their language within a generation or two in the United States. Language stability in India has been ascribed to stratification of society according to caste.

Nevertheless, languages will influence each other. The question to ask is: How might the encounter between two languages take place? The answer to this would depend on whether the two languages come face to face suddenly as would happen if invaders brought a different language or if two languages grow together in vicinity. In other words the nature of the encounter depends on whether the languages meet as equals or if it is one-sided. For example, the interaction between Spanish and the American Indian languages has been one-sided. In a one-sided encounter the language of the conquering invaders is likely to be influenced little by the second language.

The similarities between Indo-Aryan and Dravidian are well known. It is interesting that one of these similarities, namely reduplication of words which is generally assumed to have been borrowed by Indo-Aryan from Dravidian, is also to be found in the European languages. Thus in English we have words such as pooh-pooh, choo-choo that have identical reduplication; examples of a different type are chitchat, chiffchaff, knickknack, riffraff, tick-tack, zigzag, hodge-podge, and thingy-wingy. Reduplication in the Indian languages is much more common than in the European languages.

Considering the borrowings between Indo-Aryan and Dravidian, Eme-neau says:

[T]he languages of the two families, Indo-Aryan and Dravidian, seem in many respects more akin to one another than Indo-Aryan does to the other Indo-European languages.¹⁶

For this reason India is considered a linguistic area with “languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families”.¹⁷ This indicates that the encounter between Indo-Aryan and Dravidian must have been a long and an equal one.

Nevertheless, the limitations of the philological approach are apparent if one considers that this analysis has led to the conclusion that the conser-

vative caste system was adopted by the Indo-Aryans from the Dravidians. Emeneau says:

We are almost forced to a hypothesis that the Dravidians whom the Indo-Aryan invaders met in the riverine plains of North India had a caste system with linguistic traits mirroring it, which they shared with the Dravidians of the plains of the south.¹⁸

This raises a very thorny question. If the caste system and social stratification are to be invoked for the persistence of the Saurashtran language in South India for more than a millennium, and if the Dravidians had a caste system in the north before the arrival of the Indo-Aryans, then why was there no trace of the Dravidian language in the centuries before Christ in North India which was not too long after the supposed Aryan invasion?

At the same time scholars have argued that all ancient Indo-European societies had classes that might have been the forerunner to the caste system.¹⁹

But if the caste system was adopted by the Indo-Europeans from the Dravidians, then the original homeland of the two groups must have in proximity and they must have interacted amongst each other. Emeneau proposes that the North Indians themselves were originally Dravidian speaking and they adopted Indo-Aryan after a long period of bilingualism. But Emeneau's proposal does not have facts to back it. There are social practices and other features that show that Marathi speakers represent a region where bilingualism of Indo-Aryan and Dravidian was once prevalent. But such features are not to be found in the region of the Indus, Sarasvati, and the Ganga valleys.

The only way out appears to question the traditional classification of the Indic languages and the models of their evolution.

5 A Scenario Based on the Current Archaeological Evidence

The difficulty with most language classification models is that they do not do justice to the linguistic and archaeological evidence from the Indian sub-continent. To get over the contradictions where the current models lead us, one may propose the following scenario: Around 7000 B.C. the Indo-Europeans were located in the Indus-Sarasvati valleys, northern Iran, and

southern Russia; the Afro-Asiatics were in West Asia; and the Dravidians were located just south of the Indo-Europeans in a belt stretching from South India to southern Iran. Their existed many trading links between the groups. The Vedic period is to be seen as following a long interactive era between the Indo-Aryans and the Dravidians.²⁰The proof of this comes in many Dravidian features of the Vedic language.

This scenario does not address or answer the question as to the original homeland of the Indo-Europeans or the Indo-Aryans. It has the virtue of explaining the astronomical evidence from the Vedic literature as well as explaining the deep structural commonality shared not only between Indo-Aryan and Dravidian but also between European languages and Dravidian.

This scenario also explains the striking resemblance between Vedic form and a head unearthed at Nevali Çori in Anatolia by Harald Hauptmann.²¹The site of Nevali Çori dates to about 7500 B.C. The striking thing about the head is that it is clean shaven except for a long tuft at the top that looks strikingly similar in style to the *śikhā* that a student wore in the Vedic times. B.G. Sidharth²²has taken this similarity to mean that this Anatolian civilization was Vedic. Our model, that considers the Indo-Europeans to be already spread from Anatolia to Northwest India at the time of Nevali Çori, is consistent with such an identification.

An important implication of our model is that there is no need to force the placement of events of the Vedic texts and the epics Ramāyaṇa and Mahābhārata, that are clearly defined by their contexts in Indian locales, to places outside India where they cannot be reconciled to other evidence.

6 Conclusions

The structural relationships amongst the Indo-European family of languages are well known. Not equally well known are the structural connections between the Indo-Aryan, the Dravidian and the Munda languages. These languages may be said to belong to the Prakrit family of languages. We use the label “Prakrit” since it has been traditionally used to describe all Indian languages.

In other words we argue that in general one might speak of membership of a language to more than one family. We believe such a usage is more accurate than the term “linguistic area” used earlier by Emeneau.

In recent years studies have been made to correlate genetic background of populations with languages.²³ These studies have had some success in describing the spread of languages. It is significant that on many counts the vast majority of the Indian population, in North as well as South India, is classed as a single group.

The evolution of the Prakrit family over millennia through prolonged interaction of the populations explains structural as well as biological commonality. The attested migrations of the Indo-Iranians into Europe explains the presence of several Dravidian features in the European languages.

7 Notes

1. See for example Kak (1993).
2. See reviews such as Baldi (1983) and Ruhlen (1987).
3. Emeneau (1980).
4. See, for example, Baldi (1983), page 63.
5. Kaiser and Shevoroshkin (1988).
6. Ammerman and Cavalli-Sforza (1984).
7. Renfrew (1989).
8. Childe (1926).
9. For example see Shaffer (1992), Kak (1994a,b).
10. Kak (1987).
11. Gimbutas (1985).
12. Mallory (1992).
13. Mallory (1989).
14. Kak (1992, 1994a,b).
15. Emeneau (1980), p. 43.

16. Emeneau (1980), pp. 119-120.
17. Emeneau (1980), p. 1.
18. Emeneau (1980), pp. 237.
19. See for example Dumézil (1988).
20. See Kak (1994b) for details.
21. Hauptmann (1993).
22. Sidharth (1992).
23. See for example Cavalli-Sforza (1991), Sokal *et al* (1992).

8 References

- A.J. Ammerman and L.L. Cavalli-Sforza. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton: Princeton University Press.
- P. Baldi. 1983. *An Introduction to the Indo-European Languages*. Carbondale: Southern Illinois University Press.
- L.L. Cavalli-Sforza. 1991. "Genes, Peoples and Languages." *Scientific American*, November, pp. 104-110.
- V.G. Childe. 1926. *The Aryans*. London.
- G. Dumézil. 1988. *Mitra-Varuna*. New York: Zone Books.
- M.B. Emeneau. 1980. *Language and Linguistic Area*. Stanford: Stanford University Press.
- M. Gimbutas. 1985. "Primary and secondary homeland of the Indo-Europeans." *Journal of Indo-European Studies*, 13, pp. 185-202.
- H. Hauptmann. 1993. "Ein kultgebäude in Nevali Çori." In *Archaeologica Anatolica et Mesopotamica Alba Palmieri dedicata*, M. Frangipane, H. Hauptmann, M. Liverani, P. Matthiae, M. Mellink (eds.).
- S.C. Kak. 1987. "On the chronology of ancient India." *Indian Journal of History of Science*, 22, pp. 222-234.
- S.C. Kak. 1992. "The Indus tradition and the Indo-Aryans." *Mankind Quarterly*, 32, pp. 195-213.

- S.C. Kak. 1993. "Feedback neural networks: new characteristics and a generalization." *Circuits, Systems and Signal Processing*, 12, pp. 263-278.
- S.C. Kak. 1994a. *India at Century's End*. New Delhi: Voice of India Publications.
- S.C. Kak. 1994b. *The Astronomical Code of the R̥gveda*. New Delhi: Aditya.
- M. Kaiser and V. Shevoroshkin. 1988. *Annual Review of Anthropology* 17. 309-329.
- J.P. Mallory. 1989. *In Search of the Indo-Europeans*. London: Thames and Hudson.
- J.P. Mallory. 1992. "Human populations and the Indo-European problem." *Mankind Quarterly*, 33, pp. 131-154.
- C. Renfrew. 1989. "The origin of Indo-European languages." *Scientific American*, October. 106-114.
- M. Ruhlen. 1987. *A Guide to World's Languages*. Stanford: Stanford University Press.
- J.G. Shaffer. 1992. "The Indus valley, Baluchistan, and Helmand traditions: neolithic through bronze age", In J. Ehrlich (ed.), *Chronologies in Old World Archaeology (3rd Edition)*. Chicago.
- B.G. Sidharth. 1992. "A Lost Anatolian Civilization—Is It Vedic?" *Research communication, Birla Science Centre, Hyderabad*.
- R.R. Sokal, N.L. Oden and B.A. Thomson. 1992. "Origins of the Indo-Europeans: Genetic evidence." *Proc. Natl. Acad. Sci. USA*, 89, pp. 7669-7673.