



## Commentary

# On Linguistic history and language diversity in India: Views and counterviews *By* Sonal Kulkarni-Joshi

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

There is an immanent tension in the word ‘ancestry’. Gaining significance at crucial historico-political junctures, it disengages the present from the past, pushing us into a hungry quest for unearned heirlooms, achievements and conquests, with the false hope that it will absolve us from our current battered minds and identities. When ‘ancestry’ takes center stage in academic discourse, it is, therefore, best to tread carefully and to avoid the overshadowing of the present by the past. Scientific discourse – whether it be in the domain of linguistics or genetics – thus needs to focus on enriching the present by, first, illuminating the multitudinous, intricate and invisible web of relations that define us and our present and, second, by acknowledging the possible drawbacks of scientific methods and constructs, and the near-impossibility of exact and precise results in any domain of knowledge. Sonal Kulkarni-Joshi’s paper, which aims to ‘lend a linguistic perspective on the issue of human diversity and ancestry in India’ fairs well on both counts.

Using the lens of languages, Kulkarni-Joshi presents the current Indian subcontinent as a land whose linguistic identity (or identities) has been shaped and is continuously being shaped by its multitude and their multiple languages. The present also reflects the past. Several millennia ago, people’s languages, irrespective of their origins, were not isolates, but mingled and assimilated with each other, engendering new and interesting lexical and structural overlaps in their lexicons and grammars. Evidence for contact between old Dravidian and Old Indo-Aryan/OIA languages abounds in shared vocabulary items found in the ancient Rg Veda verses, covering different cultural, economic and ecological aspects of life (Kuiper 1967, 1991). There is also evidence of phonetic and syntactic assimilation, such as the case of the retroflex sounds and quotative constructions that apparently made their way into the earliest Vedic texts from old Dravidian. But history is not just about facts, but also the historian’s take on those facts. So it is with linguistic facts and language histories. Here too, there are multiple theoretical and explanatory narratives or inferences. Contradicting the contact and diglossic narratives about OIA and old Dravidian, Hock (1996) and Hall (1997) build their explanations away from contact situations, and directly into grammar-internal processes. Hall, for instance, suggests grammar-internal repair strategies to avoid impossible phonemic inventories as possible reasons for the rise to retroflex sounds in any language, even when there is no positive evidence of such sounds in the speaker’s linguistic environment. The same can be said of quotative constructions, which are not always results of borrowing from neighboring languages; a recent illustration being the rise of the new quotative marker ‘like’ in some current, non-standard varieties of English.

The ‘one fact/multiple explanations’ narrative gains further ground when Kulkarni-Joshi presents her case for the origins of the Indo-Aryan people and their languages. Three major pieces of evidence are adduced in favor of the ‘migration into India’ story: (i) the mention of Rg Vedic gods, concepts and words in Mittani (Egyptian) texts, dating back to 1400 BC, (ii) the economical process that reduces three vowels *e*, *a* and *o* in Greek and Latin to a single vowel *a* in Vedic Sanskrit and (iii) the unidirectional process of change from velars *k*, *g* to palatals *c*, *j*. All three facts, taken together with many others, strongly suggest that the lineage of people speaking the IA languages lies somewhere outside India. Once again, and unsurprisingly, there is no unanimity on this point. Hock (1993) points to the possibility that some of these said OIA-like features in the Mittani language may have come from contact with non-Indo European Hurrian languages. Misra (1992) likewise questions

the proposed unidirectional, developmental route from a three to a one vowel system, drawing on evidence of a three vowel system that developed in the Indo-Aryan Gypsy (Rumany) language, long after its people emigrated outside India.

Kulkarni-Joshi does a fair and impartial presentation of varied accounts of language classification and ancestry in the Indian subcontinent. However, the dubiety of inferences and results force us to ask deeper questions about the adequacy of constructs and methods adopted by historical linguists. For many decades, historical linguists have related languages primarily on the basis of lexical correspondences, cognates and sound changes (cf. Greenberg 1987, 2000). Lexical items or words are often crucial to shallow processing or comprehension, and hence are the most likely candidates for language comparison. Nonetheless, there are problems related to the lexical comparative approach (see Longobardi and Guardiano 2009 for an elaborate discussion). Lexical items and cognates fail to provide precise measurements of distances between languages, because there is no known or well-accepted trajectory for word development that can inform the development and time scales of different languages within a language family. Moreover, since meaning is arbitrarily associated with each lexical form, even minimally different phonological items, in reality, are not comparable. To illustrate, the relatedness of German and English cannot be inferred from phonetically similar forms such as German *klein* (small) and English *clean*, because they are different words with different meanings. Neither can we compare German *hund* with English *hound*, since there is only a partial correspondence of meaning – the former has a more generic meaning than the latter. Such and other problems with the conventional lexical-taxonomic methods were not unknown to historical linguists. Greenberg's alternative to conduct multilateral or mass lexical comparison across a large number of languages was aimed specifically at countering such cases of accidental homophony. However, despite the grand scale of comparison, a lexical approach falls way short of the target due to its insistence on treating lexical properties as vertically transmitted from one language to its daughter languages. It is well known that lexical borrowings also happen horizontally among genealogically unrelated but geographically proximate languages. Therefore, the clearest indication that phonetically and semantically minimal units make is regarding the close proximity of two languages, without committing to any genealogical relations between them. An example of horizontal lexical borrowing is current-day standard Hindi that has rampantly borrowed words from another commonly used language in the subcontinent – (Indian) English.

To avoid problems associated with the lexical comparison method, Longobardi and Guardiano (2009) suggest considering abstract structural parameters (e.g. the possible orders of adjectives and determiners vis-à-vis the noun) as potential comparanda. Each parameter has a binary 0/1 (=off/on) setting, which means that the results are more precise when comparing distances between languages and dialects. There are also fewer chances of accidental comparisons while using formal parameters since they are dissociated from semantics and, hence, are less vague. Interestingly, the results of the structural parametric comparisons are not radically different from the ones produced via the lexical methods, although language diversification is modeled in terms of networks rather than as singular points in a tree. The structural parametric approach also isolates some outliers within individual language families that could never be found through a comparison of lexical items. Therefore, while the structural parameter study reconfirms the presence of macro-families such as the Indo-European, Uralic and Altaic, as well as specific language isolates such as Basque, it produces negative results for Farsi in the Indo-Iranian group and Bulgarian in the Slavic group that fail to display behavior expected of their respective family members.

Historical linguists are increasingly acknowledging the advantages of using structural techniques, alongside the conventional lexical methods. But there is more ground to cover. In its current state, the structural parametric approach may still generate false-positives and -negatives. To illustrate, a comparison of abstract schemata for case-markers on nouns in Braj perfective sentences places it closer to Punjabi than to Standard Hindi and KhaRi Boli. Not only do both current-day Punjabi and Braj dialects display pronoun-specific (1st, 2nd versus 3rd) case-markers, but similar patterns are also attested in Old Braj texts (Surdas' 14th century *Sursagar*). The Punjabi–Braj relation, however, can be questioned when we observe similar pronoun-specific patterns in the Australian aboriginal language Dyirbal, language-isolate Basque as well as the Romance language, French. These similarities, theoretical linguists contend, arise from abstract principles that underlie all languages, though they surface morphologically in some languages (across families), subject to certain other constraints. Abstract schemata of the kind that Longobardi *et al.* suggest therefore do not suffice as explanatory comparative tools. It is imperative for historical linguists to complement their results about shared lexical and structural items with knowledge from theoretical linguistics about universal principles about the language organ, and the complex network of interactions between the computational system of human language and the interfaces of sound and meaning (cf. Chomsky 1995).

The inquiry into language relations and ancestry presents a strenuous task. There is no single and simple mechanism, using lexical and structural signatures, which can give us definite answers to language lineages. Kulkarni-Joshi's paper on views and counterviews on the ancestry of the Indian people and languages presents the first seed of doubt. There is a lot more evidence that can be adduced to substantiate the uncertainty that surrounds all ancestry-related answers available in the linguistic literature.

### References

- Chomsky N 1995 *Minimalist Program* (MIT Press)
- Greenberg J 1987 *Language in the Americas* (Stanford University Press)
- Greenberg J 2000 *Indo-European and Its Closest Relatives* (Stanford University Press)
- Hall TA 1997 The historical development of the retroflex consonants in Indo-Aryan. *Lingua* **102** 203–221
- Hock HH 1996 Subversion of convergence? The issue of pre-Vedic retroflexion reexamined. *Stud. Linguist. Sci.* **23.2** 73–115
- Kuiper FBJ 1967 The genesis of a linguistic area. *Indo-Iran. J.* **10** 81–102
- Kuiper FBJ 1991 *Aryans in the Rigveda* (Amsterdam)
- Longobardi G and Guardiano C 2009 Evidence for syntax as a signal of historical relatedness. *Lingua* **119** 1679–1706
- Misra SS 1992 *The Aryan problem: A linguistic approach* (New Delhi: Munshiram Manoharlal)

PRITHA CHANDRA  
*Indian Institute of Technology Delhi,*  
*Hauz Khas, India*  
*(Email, pritha@hss.iitd.ac.in)*