
**Genesis of Proto-Indo Languages in the
Light of Neo-Nationalism in India: Problem,
Challenges and Possibilities**
Santosh Prakash Patil

Lecturer in English, Department of English, KLE Society's G.I.
Bagewadi Arts, Science and Commerce College, Nipani.

DOI: <https://doi.org/10.5281/zenodo.18328487>

ABSTRACT:

India's multilingual landscape, characterized by profound linguistic diversity and historical stratification, has long been structured around persistent binary oppositions that shape both linguistic practices and academic discourse. Prominent among these are the hierarchical distinctions between Sanskrit (as the refined, classical, and "cultured" language of sacred texts and elite knowledge) and Prakrit (the "natural," vernacular forms associated with everyday speech and the masses), as well as the postcolonial tensions between English and indigenous/regional languages. This article interrogates these oppositions not as natural or fixed categories but as constructed hierarchies embedded in cultural, political, and ideological discourses, revealing their instability, mutual interdependence, and role in perpetuating exclusionary structures within Indian languages and knowledge production.

KEYWORDS:

Indian Languages, Proto, Genesis, Sanskrit, Prakrit, Aryan, Dravid, Federal State, Nationalism.

.....

Jacques Derrida's theory of binary oppositions, central to his deconstructive philosophy, critiques the way Western metaphysics organizes meaning through hierarchical pairs where one term is privileged as primary, pure, or superior while the other is subordinated as secondary, derivative, or inferior. Classic examples include speech/writing, presence/absence, or nature/culture. In the context of Indian languages, this framework illuminates the longstanding Sanskrit/Prakrit opposition, where Sanskrit (refined, grammatical, "perfected," associated with sacred texts, elite knowledge, purity, and cultural authority) is privileged over Prakrit (natural, vernacular, "ordinary," spoken by the masses, linked to everyday life, Jain/Buddhist canons, and emotional immediacy). This binary—rooted in ancient Indian linguistic traditions and reinforced through colonial philology and postcolonial nationalist discourses—mirrors Derrida's speech/writing hierarchy: Sanskrit embodies a logocentric "presence" of refined origin and eternal truth, while Prakrit is cast as a fallen, supplementary deviation. Deconstruction reveals their mutual contamination—Prakrit as the "natural" source from which Sanskrit is codified and abstracted, yet Sanskrit depending on Prakrit's vitality for its historical evolution—exposing how such privileging perpetuates exclusionary power structures in language, literature, and knowledge production, while inviting a rethinking of Indian linguistic identity beyond rigid hierarchies.

The genesis (origin or beginning) of any group of languages—commonly called a language family—is the historical process by which a set of related languages descends from a single common ancestor language, known as a proto-language. In historical linguistics, this is the standard explanation for why certain languages show systematic similarities in vocabulary, grammar, sound systems, and other features. These similarities are not due to chance, borrowing (words taken from contact with other languages), or coincidence, but because the languages evolved from the same original source through gradual divergence over centuries or millennia.

How Does a Language Group/Family Come into Existence?

A single ancestral language exists — at some point in prehistory, a community speaks one relatively uniform language (the proto-language). This could be spoken by a relatively small or large group, often in a specific homeland region. Geographic or social separation occurs — Groups of speakers migrate, settle in different areas, become politically divided, or are otherwise isolated from one another (e.g., due to rivers, mountains, conquest, or simply distance). Regular communication breaks down. Gradual, independent changes accumulate — Each separated group changes its language naturally over generations: Sounds shift systematically. New words are invented or old ones gain new meanings.

Grammar evolves. These changes are mostly regular and follow patterns that linguists can trace using the comparative method. The family tree forms — Linguists represent this as a branching tree diagram (similar to a biological family tree or phylogenetic tree), where:

1. The root = the proto-language.
2. Main branches = major subfamilies.
3. Twigs = modern languages or dialects.

The genesis (origins) of groups of languages in India follows the same principle as elsewhere: most belong to distinct language families, each descending from a reconstructed proto-language through gradual divergence due to migrations, geographic separation, cultural contact, and natural linguistic change over millennia. India's linguistic landscape is exceptionally diverse, with over 780 languages (and thousands of dialects) belonging to several major families. The 22 scheduled languages (8th Schedule of Constitution) and 6 classical languages (Tamil, Sanskrit, Kannada, Telugu, Malayalam, Marathi and Odia) reflect this layered history. This diversity makes India one of the world's most linguistically complex regions—much like a living museum of human migration and evolution through language.

The major languages of India fall into at least four distinct, unrelated language families, each with independent historical origins and phylogenetic reconstructions. This position is grounded in comparative–historical linguistics, which relies on systematic correspondences in phonology, morphology, syntax, and core lexicon to establish genetic relationships (Campbell, 2013; Campbell & Poser, 2008).

The principal families represented in modern India are as follows:

Indo–Aryan: This family accounts for approximately 78% of the population and includes languages such as Hindi, Bengali, Punjabi, Marathi, Gujarati, and Odia. All Indo–Aryan languages descend from Proto–Indo–Aryan, which itself derives from Proto–Indo–Iranian and ultimately from reconstructed Proto–Indo–European (PIE), spoken circa 4500–2500 BCE in the Pontic–Caspian steppe region (Mallory & Adams, 2006). The arrival of Indo–Aryan speakers in northwestern South Asia is conventionally dated to circa 2000–1500 BCE.

Dravidian: Representing roughly 19–20% of speakers, primarily in southern India, this family includes Tamil, Telugu, Kannada, and Malayalam. Dravidian constitutes an independent family whose proto–language (Proto–Dravidian) is estimated to have existed approximately 4,500 years ago (Kolipakam et al., 2018). No demonstrable genetic link exists between Dravidian and Indo–European; proposed connections remain speculative and lack support from the comparative method. Extensive areal contact has occurred, however, resulting in shared phonological features.

Austro–Asiatic (primarily the Munda branch): Spoken by smaller tribal populations in central and eastern India (e.g., Santali, Mundari, Ho), this family shows clear affinities with languages of Southeast Asia (e.g., Khmer, Vietnamese). Its presence in India predates or is independent of Indo–Aryan expansion, reflecting distinct migratory histories (Sidwell, 2011).

Sino-Tibetan (mainly Tibeto-Burman branch):

Predominant in northeastern India and the Himalayan region (e.g., Bodo, Manipuri/Meitei, Mizo, and various Naga languages), this family traces its origins to East Asia, with maximal diversity suggesting ancient presence in or near northeastern India.

While intensive language contact across millennia has produced significant areal convergence—such as shared retroflex series, quotative constructions, and echo-word reduplication—these phenomena reflect diffusion within a linguistic area (Sprachbund), not descent from a common ancestor (Masica, 1991). Hypotheses positing a single deep-time “Proto-Indian” or macro-family encompassing all four groups (e.g., variants of Nostratic or other long-range proposals) remain highly controversial and are rejected by the majority of specialists due to insufficient regular sound correspondences and reliance on superficial typological similarity rather than the rigorous application of the comparative method (Campbell & Poser, 2008).

The Problem of Dravidian Languages

While early hypotheses posited external origins, contemporary research increasingly supports an indigenous development within the Indian subcontinent, potentially linked to ancient civilizations such as the Indus Valley Civilization (IVC), though this association is tentative and contested. This expansion draws on comparative linguistics, phylogenetic modeling, archaeology, and archaeogenetics to delineate the proto-history of Dravidian, emphasizing its estimated antiquity of around 4,500 years and its role in shaping regional linguistic diversity.

Linguistic Reconstruction and Chronology

Proto-Dravidian, the reconstructed ancestral language from which all modern Dravidian varieties descend, is estimated to date to approximately 4,500 years before present (ca. 2500 BCE), based on Bayesian phylogenetic analyses of cognate-coded lexical data. These studies, utilizing first-hand elicited vocabulary from native

speakers across subgroups, employ statistical models (e.g., relaxed-clock frameworks) to infer divergence times, yielding consistent results across multiple evolutionary simulations. The family subdivides into four main branches—North (e.g., Brahui, Kurux), Central (e.g., Kolami, Parji), South I (e.g., Tamil, Malayalam), and South II (e.g., Telugu, Kannada)—with some uncertainty in inter-branch relationships, diverging from earlier classifications like those of Krishnamurti (2003).

The earliest attested Dravidian evidence appears in the 2nd century BCE with Tamil-Brahmi inscriptions in southern India, but linguistic substrata in Old Indo-Aryan texts (e.g., Rigvedic Sanskrit, ca. 1200 BCE) indicate a broader prehistoric distribution. Loanwords from Proto-Dravidian into Old Indo-Aryan, such as terms for flora, fauna, and cultural artifacts, suggest contact during the mid-Rigvedic period, implying Dravidian presence across northwestern India prior to Indo-Aryan dominance. Reconstructed Proto-Dravidian vocabulary, including agricultural and settlement terms (e.g., *ūr* ‘village’, *kōṭṭai* ‘fort’), points to a sedentary, agrarian society, potentially aligned with Neolithic developments in the subcontinent.

Archaeological Correlations: The Indus Valley Hypothesis

A prominent hypothesis links Proto-Dravidian to the IVC (ca. 3300–1300 BCE), centered in modern-day Pakistan and northwestern India. Proponents argue that the undeciphered Indus script may encode a Dravidian language, supported by etymological analyses of substrate words in later Indo-Aryan (e.g., elephant-related terms like *pīlu* ‘elephant/ivory’, derived from Proto-Dravidian *pal* ‘tooth’).

However, this connection remains unproven due to the script’s undeciphered status and alternative proposals (e.g., Para-Munda or an isolate ‘Language X’). Krishnamurti (2003) tentatively dates Proto-Dravidian to the early third millennium BCE, coinciding with IVC’s mature phase, while Parpola (2015) adduces ethnographic parallels like Dravidian kinship systems in IVC-descended societies.

Genetic and Migration Hypotheses

Archaeogenetic studies complement linguistic data, suggesting Dravidian speakers' ancestry derives from a mixture of indigenous South Asian hunter-gatherers and ancient Iranian farmers, with a putative 'Proto-Dravidian' component emerging around 4,400 years ago near the Iranian plateau-Indus interface. Consensus holds that Dravidians were likely autochthonous to the subcontinent, present before Indo-European speakers, with subsequent contractions due to Indo-Aryan expansion.

Challenges and Future Directions

Despite advances, uncertainties persist: long-range proposals lack methodological rigor, and the family's isolation underscores the need for interdisciplinary integration. Future research, incorporating expanded genomic datasets and computational decipherment of the Indus script, may clarify dispersal patterns and external affinities.

In conclusion, the linguistic evidence firmly indicates multiple independent origins for the languages of India, shaped by distinct population movements, followed by prolonged contact and convergence. Claims of a single source are not substantiated by current historical-comparative linguistics.

Nationalism, in short academic terms, is a political principle or ideology that holds that the boundaries of the nation (a cultural or ethnic community) and the state (the political unit) should be congruent—meaning each nation should ideally have its own sovereign state, and political legitimacy derives from aligning governance with national identity.

Federalism and nationalism represent two distinct yet often intersecting concepts in political theory and practice. Federalism is a system of government in which power is constitutionally divided between a central (national/federal) authority and constituent sub-national units (states, provinces, regions, cantons, etc.), with each level having sovereign authority in specified spheres. It

institutionalizes shared sovereignty and multilevel governance, often to accommodate diversity, promote efficiency, or prevent excessive centralization.

India is constitutionally a federal union because the Constitution of India establishes a system with key federal characteristics, while also incorporating significant unitary features. This makes it a unique model often described as “quasi-federal” or “federal in structure but unitary in spirit.” Legislative powers are clearly divided between the Union and States via the Seventh Schedule.

In the Indian Constitution, the subject of “language” (in the context of legislative powers and division between the Union and States) is not listed as a distinct entry in the Seventh Schedule (which divides powers into Union List, State List, and Concurrent List). In essence, language as a subject of legislation is not confined to one list in the Seventh Schedule. Core aspects (official Union language, promotion) lean toward the Union, while states have significant autonomy in adopting and using languages for state purposes, and education is concurrent. This reflects India’s federal design balancing national unity with linguistic diversity.

Neo-nationalism (also called new nationalism or neo-populist nationalism) is a modern form of nationalism that emerged mainly in the mid-2010s, globally. It’s a right-wing populist reaction against globalization, open borders, multiculturalism, free trade agreements, and supra-national bodies. Indian neo-nationalism based on the triad Hindu-Hindi-Hindustan refers to a prominent strand of contemporary Hindu nationalism (often called Hindutva or neo-Hindutva in modern contexts), which fuses religious, linguistic, and territorial identity into a vision of India as a predominantly Hindu nation. This formulation draws from early 20th-century Hindutva ideologue Vinayak Damodar Savarkar, who articulated the slogan “Hindi, Hindu, and Hindustan” to advocate for: Hindu as the core cultural-religious identity (Hindutva = “Hindu-ness,” defining

true Indianness through shared Hindu civilizational roots, ancestry, and holy land), Hindi as the unifying national language (promoting Sanskritized Hindi over English, Urdu, or regional languages to forge a common linguistic bond), and Hindustan as the sacred homeland (the land from the Indus to the seas, historically tied to Hindu heritage, with non-Hindus expected to assimilate or accept subordinate status).

Hindu-Hindi-Hindustan encapsulates a vision of reclaiming India as a unified Hindu homeland with one dominant religion, language, and cultural narrative—a core driver of India’s current ruling ideology and a prime example of neo-nationalism in the Global South.

India’s secular framework, as articulated in its Constitution, is predicated on the principle of equal respect for religious, cultural, and linguistic diversity, ensuring state neutrality amid a mosaic of identities. However, the emerging discourse of portraying all Indian languages as belonging to “only one family” or deriving from a “same source”—exemplified by initiatives like the Bharatiya Bhasha Pariwar—presents a veneer of unity that masks deeper ideological agendas. This note offers a critical academic examination of how such homogenization efforts, often rooted in neo-nationalist ideologies, undermine India’s secular ethos by privileging majoritarian cultural narratives, reinforcing hierarchies, and eroding federal pluralism. Drawing on postcolonial theory and linguistic anthropology, it argues that this narrative repurposes colonial legacies to centralize power, potentially transforming India’s secular democracy into a culturally monolithic entity. India’s Constitution recognized this diversity through the Eighth Schedule (initially listing 14 languages, now 22) and Articles 29–30, which safeguard linguistic minorities, reflecting a commitment to secular pluralism over homogenization.

Yet, contemporary reframing like Bharatiya Bhasha Pariwar—promoted via the National Education Policy (NEP) 2020 and

Bharatiya Bhasha Samiti—claim to decolonize linguistics by emphasizing “shared civilization roots” and “mutual influences,” often invoking Sanskrit as a common progenitor. This “one family” trope, however, inherits colonial binaries: it privileges Indo-Aryan languages (e.g., Hindi, Sanskrit-derived) as the “core,” marginalizing Dravidian (e.g., Tamil, seen as “demarcated” despite prestige) and indigenous tongues (e.g., Adivasi languages as “distant relatives”). Such classifications echo the 19th-century European “march from synthesis.”

Critical Analysis: Homogenization as a Threat to Secularism

The “one family” or “same source” narrative poses multifaceted challenges to Indian secularism, which, as Rajeev Bhargava elucidates, is not a strict church-state separation but an “equal respect” model accommodating diversity without favoritism. First, it fosters linguistic nationalism that communalizes identities, as seen historically in the Hindi-Urdu divide, where “secular” compromises like Hindustani devolved into religious lines. By positing a singular source—often Sanskrit-centric—this discourse implicitly aligns with Hindutva, portraying non-Indo-Aryan influences (e.g., Persian-Arabic in Urdu) as “foreign,” thereby marginalizing Muslim and other minority communities. This violates secular neutrality, transforming language policy into a tool for majoritarian assimilation.

Second, it enables centralization and homogenization, undermining federalism—a cornerstone of Indian secularism. Linguistic states, formed post-1956 reorganization, embody regional autonomy (Article 345), allowing states to adopt official languages without Delhi’s imposition. Critics argue this mirrors colonial administrative consolidation, where language homogenization facilitated state control, now repurposed for neo-nationalist ends. For instance, NEP 2020’s emphasis on “Bharatiya Bhasha” as a unified entity risks eroding multilingualism, exacerbating inequalities: urban elites benefit from Hindi-English bilingualism, while regional speakers face cultural erasure.

Third, this narrative intersects with broader assaults on secularism, such as the “Indianized” variant that integrates religion into public spheres, blurring state–religion boundaries. By framing languages as kin in a “civilization” family, it enables cultural consolidation that reinscribes hierarchies—Indo–Aryan at the apex—while masking ideological homogenization. This has real–world implications: anti–Hindi agitations (1960s Tamil Nadu) and recent protests against Hindi medical curricula highlight how such policies foment identity clashes, turning language into a “problem” solvable only through majoritarian dominance. From a developmental lens, linguistic diversity drives social cohesion and innovation; homogenization stifles it, hindering equitable participation and amplifying caste–religion intersections (e.g., Sanskrit’s Brahminical ties excluding Dalits and minorities).

Conclusion: Reaffirming Pluralism against Homogenizing Myths

The problem of “wrapping” secular India in a “one family” or “same source” linguistic narrative lies in its subversion of diversity: it repurposes colonial tools for contemporary centralization, privileging majoritarian identities at the expense of minorities and regions. To counter this, India must recommit to multilingual federalism, expanding protections for non–dominant languages and decoupling policy from ideological homogenization. As A.R. Venkatachalapathy asserts, India’s strength resides in linguistic diversity, not uniformity; embracing this ensures secularism’s vitality amid neo–nationalist pressures. Failure to do so risks unraveling the pluralistic tapestry that defines Indian democracy, replacing it with a monolithic illusion of unity. Hence, making Proto–Indo group of Languages is problematic and challenged linguistically & culturally.

References:

1. Campbell, L. (2013). *Historical Linguistics: An Introduction* (3rd ed.). MIT Press.
2. Campbell, L., & Poser, W. J. (2008). *Language Classification: History and Method*. Cambridge University Press.
3. Emeneau, M. B. (1956). India as a linguistic area. *Language*, 32(1), 3–16.
4. Kolipakam, V., et al. (2018). A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science*, 5(3), 171504.
5. Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge University Press.
6. Mahadevan, I. (2021). Ancestral Dravidian languages in Indus Civilization: Ultraconserved Dravidian tooth-word reveals deep linguistic ancestry and supports genetics. *Humanities and Social Sciences Communications*, 8(1), 193.
7. Mallory, J. P., & Adams, D. Q. (2006). *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford University Press.
8. Masica, C. P. (1991). *The Indo-Aryan Languages*. Cambridge University Press.
9. Narasimhan, V. M., et al. (2019). The formation of human populations in South and Central Asia. *Science*, 365(6457), eaat7487.
10. Pagani, L., et al. (2024). Novel 4400-year-old ancestral component in a tribe speaking a Dravidian language. *Scientific Reports*, 14(1), 1934.
11. Parpola, A. (2015). *The Roots of Hinduism: The Early Aryans and the Indus Civilization*. Oxford University Press.
12. Southworth, F. C. (2005). *Linguistic Archaeology of South Asia*. RoutledgeCurzon.

Funding:

This study was not funded by any grant.

Conflict of interest:

The Authors have no conflict of interest to declare that they are relevant to the content of this article.

About the License:

© The Authors 2024. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.