JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DECODING DIALECTS WITH ARTIFICIAL INTELLIGENCE: REIMAGINING INDIAN LANGUAGE PEDAGOGY THROUGH INDIGENOUS LINGUISTIC TRADITIONS AND EMERGING TECHNOLOGIES

Prof Dr DEVANANDAN K V, Prof Dr R LAKSHMI, Principal, KPPM College of Teacher Education, Malappuram, Kerala, India, Institute of Advanced Study in Education, Thrissur, Kerala, India

ABSTRACT

India is a land of diversities and pedagogical complexity due to its rich linguistic diversity, which includes both classical languages and regional dialects. This study explores the potential of Artificial Intelligence (AI) to transform Indian language teaching by referencing traditional linguistic frameworks like Yāska's Nirukta and Pāṇini's Ashtādhyāyī. Particularly in underserved contexts, it criticises the shortcomings of standardised curricula, insufficient teacher preparation, and the unavailability of dialect-inclusive resources. The study investigates how dialect-sensitive, individualised instruction can be supported by AI tools such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and adaptive learning platforms. The study focusses on learner-centric AI applications that take into account regional linguistic identities, echoing the individualised mentoring of the Gurukula tradition. The study highlights new developments that bridge dialectal divides using illustrative examples from EdTech startups, government-led initiatives like Bhashini, and Indian research labs. The integration of AI supports the objectives of the National Education Policy (NEP) 2020, which include fostering multilingualism, inclusion, and contextual learning, despite barriers such as a lack of dialectal datasets, infrastructure issues, and ethical concerns. In order to promote fair, AI-enabled Indian language instruction that honours the nation's enormous linguistic diversity, the paper ends by suggesting future paths for research, policy, and technology.

Keywords: *Artificial Intelligence, Indian Languages, Dialectal Variation, Pāṇini, Yāska, NLP, ASR, NEP 2020, Gurukula, Language Pedagogy, Educational Technology, Multilingualism*

1. INTRODUCTION

Artificial Intelligence is bringing tremendous changes to language teaching in India by enhancing personalized learning, supporting multilingual education, and bridging linguistic diversity, aligning with the goals of inclusive and effective pedagogy in classrooms. The linguistic diversity of India presents both a pedagogical challenge and a cultural asset, especially when addressing dialectal variations in language instruction (Annamalai, 2001; Mohanty, 2010). Languages with significant internal diversity—including many regional and social dialects—such as Sanskrit, Hindi, and popular South Indian languages like Tamil, Telugu, Kannada, and Malayalam, pose unique instructional challenges (Krishnamurti, 2003; Bhattacharya, 2016). The effectiveness of standardized language teaching techniques is greatly influenced by this dialectal complexity.

The absence of standardised teaching materials that take dialectal diversity into account is one of the main obstacles. Most of the time, textbooks and curricula focus on standardised forms, like Standard Hindi or Classical Sanskrit, while ignoring learners' natural speech patterns, which are frequently based on regional or local dialects (Agnihotri & Khanna, 1997).

Dialectal variations in pronunciation also make learning more difficult. For example, just as Awadhi or Bhojpuri differ from Standard Hindi, so do the phonetic subtleties in spoken Tamil between Madurai and Chennai (Masica, 1991). Students who have only been taught the standardised accent may find it difficult to understand and communicate verbally in everyday situations.

Another obstacle is vocabulary divergence; many dialect-specific terms are absent from official educational resources, creating gaps in functional usage. Similar to this, learners who are not familiar with alternative syntactic constructions may become confused by subtle grammatical variations (Saxena, 1994).

The readiness of teachers is a significant constraint. The ability of many educators to bridge the gap between standard and regional forms is diminished by their lack of formal training in dialectal variation (NCERT, 2020). This results in an authenticity gap, where students struggle in real-world situations where non-standard dialects predominate, even though they perform well in the classroom.

Context-sensitive methods are necessary to overcome these obstacles, particularly when teaching living languages like Hindi and South Indian dialects as well as classical languages like Sanskrit, where custom and regional usage must coexist peacefully (NEP, 2020).

2. AI-POWERED DIALECT DECORATION SOLUTIONS: A PANINIAN VIEW

The pedagogical challenges posed by dialectal variation in Indian languages can be creatively addressed by AI technologies. Notably, the fundamental understandings of Pāṇini's Ashtādhyāyī, with its unmatched accuracy in describing phonetics and grammar, offer an intellectual blueprint for AI developers working on processing Indian languages (Subramanian, 2019; Kulkarni & Chaitanya, 2021). Pāṇini's Sūtras on sound production (śikṣā and vyākaraṇa) provide an organised system of articulation, from the place of articulation (sthāna) to the manner of articulation (prayatna), which aligns with the reasoning behind contemporary Automatic Speech Recognition (ASR) systems (Sharma, 2020; Rao & Kumar, 2022).

2.1. Speech Recognition for Diverse Dialects

Contemporary AI-based ASR systems can mimic Pāṇinian logic by identifying subtle phonetic and phonological patterns across regions when trained on large datasets representing different Indian language dialects (Jha, 2010; Sproat, 2021). These kinds of systems can give:

Similarly to how Pāṇini systematised spoken Sanskrit using pratyāhāras and sandhi rules, AI models are able to transcribe dialectal speech with greater accuracy, producing textual outputs that accurately reflect pronunciation (Deshpande, 1993).

- Pronunciation Feedback: AI can give learners real-time corrective feedback to help them master dialect-specific pronunciation, just as Sanskrit phonology clearly and scientifically accounts for aspirated, unaspirated, voiced, and voiceless sounds (Hu et al., 2020).
- Dialect identification: AI systems can identify dialects by using lexical, phonological, and syntactic cues. This helps learners find resources specific to a given region, which helps close the authenticity gap (Sinha & Thakur, 2022).

By using AI applications that are based on linguistic traditions like Pāṇini's, we can hope for a time when language learning is both contextually sensitive to India's dialectal diversity and rigorously scientific.

3. YĀSKA-INSPIRED FRAMEWORK FOR NLP IN DIALECTAL ANALYSIS AND NUANCE DETECTION

Indian dialectal variations can be effectively analysed and decoded using Natural Language Processing (NLP). Its contemporary computing power echoes the ideas of Yāska's Nirukta, which focused on the study of semantic nuances, contextual meanings, and word origins (etymology) (Sarup, 1927; Sharma, 2018). According to Sarup (1927), Yāska proposed that comprehending a word necessitates examining both its usage within a particular cultural and linguistic context as well as its root (dhātu). Likewise, by combining the lexical, syntactic, and semantic layers of language, NLP methods can be developed to more accurately interpret dialectal nuances (Joshi & Singh, 2020; Patel et al., 2021).

- Dialectal Lexicon Creation: Natural language processing (NLP) can automatically extract and categorise dialect-specific vocabulary, assisting in the creation of regionally relevant glossaries and dictionaries, much like the Nirukta did when it identified and categorised words based on their meaning and function (Kumar & Rao, 2019).
- Grammar Variation Mapping: Yāska highlighted how usage changes grammatical constructions. By mapping syntactic and morphological differences between dialects, NLP can give students comparative grammatical frameworks (Iyengar, 2017; Rao & Srinivas, 2022).
- Sentiment and Contextual Analysis: According to Yāska, context and root determine meaning. For precise understanding and communication, NLP-based sentiment analysis can identify changes in meaning depending on regional context and usage (Chatterjee & Banerjee, 2019).

- Machine Translation with Dialectal Awareness: Drawing inspiration from Yāska's contextual fidelity principle, natural language processing (NLP) tools can be developed to translate not only literal meaning but also pragmatic and emotional tones, providing translations that are culturally and dialect-appropriate (Sharma & Verma, 2023).

Consequently, dialect-sensitive language learning and translation are advanced by contemporary NLP, which confirms the ancient Indian linguistic wisdom contained in Nirukta.

4. INDIVIDUALISED PATHWAYS: REFLECTIONS OF THE GURUKULA HISTORY

The ancient Indian Gurukula system placed a strong emphasis on individualised instruction, with the teacher (guru) tailoring the curriculum, pace, and method to the needs and cultural background of each student (śiṣya). This approach respected regional linguistic diversity and promoted holistic learning (Mukherjee, 2019). In today's multilingual educational contexts, it established a learning environment that is sensitive to linguistic and cultural identities.

This Gurukula philosophy is reflected in AI-powered adaptive learning systems in the digital age. These systems optimise student engagement and mastery by customising instructional delivery using learner analytics and real-time behavioural data (Luckin et al., 2016; Biswas & Roy, 2022).

- Tailored Content Delivery: AI systems can gradually introduce dialectal variation or deliver lessons in a learner's native tongue, improving comprehension and relevance. The Gurukula tradition of starting with well-known idioms and progressively increasing a learner's exposure is comparable to this (Saxena, 2020).
- Adaptive Assessment: AI is capable of assessing a learner's dialectal proficiency through non-standardized, dynamic methods, much like a guru would through dialogic interaction (Schmid et al., 2021).
- Intelligent Tutoring Systems (ITS): These online tutors provide immediate feedback and direction based on each student's unique learning path. It has been demonstrated that ITS systems greatly enhance student performance, especially in situations with a variety of linguistic backgrounds (VanLehn, 2011; Singh & Prasad, 2023).
- Immersive Learning Environments: Like Gurukula settings, which emphasise community-based learning, conversational AI agents can mimic dialect-rich interactions (Rai & Sharma, 2022).

In a nation with more than 1,600 dialects, AI-driven learning systems can support inclusive and successful Indian language education by adopting the individualised, culturally based values of the Gurukula tradition. This is particularly important (Census of India, 2011).

5. ILLUSTRATIVE CASE STUDIES/EXAMPLES: ANCIENT ROOTS AND MODERN INNOVATIONS

The concept of dialect-sensitive language instruction has deep roots in India. Ancient Sanskrit linguists like Pāṇini and Yāska, through concepts like deśa-bhāṣā (regional speech), recognized and incorporated regional linguistic diversity into structured grammatical systems. The Aṣṭādhyāyī and Nirukta systematized standard Sanskrit grammar while also acknowledging regional phonological and lexical variation, thus establishing a foundation for dialectal analysis within the Indian linguistic tradition. Modern AI projects now translate this historical awareness into digital tools and educational technologies.

- Newer dialects like Braj Bhasha, Awadhi, and regional Bengali variations are now included in the Indicate Language Corpora, which are being developed by major tech companies like Google and Microsoft. To train dialect-sensitive NLP models, these enriched corpora are essential (Ghosh et al., 2020; Shrivastava et al., 2021).
- By using extensive, region-specific user interaction data, localised voice assistants—like Google Assistant and Amazon Alexa—have greatly improved their capacity to comprehend Indian dialects. Real-time dialect detection and response are made possible by machine learning on multilingual datasets, which enables these advancements (Sitaram et al., 2021).
- Educational apps such as Bhasha.Dialectal features have begun to be incorporated into AI and region-specific language learning tools; one such example is a Tamil learning app that uses ASR and TTS systems to provide Madurai-style pronunciation guides (Raghavan & Vaidyanathan, 2022).
- NLP models for code-switching are addressing the widespread blending of Tamil-English or Hindi-English (Hinglish) in spoken and written discourse, especially in Indian academic labs like IIT Bombay and IIIT Hyderabad. These models, which are being used in chatbots, educational apps, and digital tutors, represent real-world multilingual usage (Jamatia et al., 2015; Srivastava et al., 2020).

These advancements blend cutting-edge AI with insights from ancient Indian languages, enabling inclusive, dialect-sensitive language learning that honors India's multilingual heritage.

6. Benefits of AI in Indian Language Pedagogy: Evidences from Contemporary Indian Research

Artificial intelligence has the potential to revolutionise understanding, engagement, heritage preservation, and inclusivity when incorporated into Indian language instruction. This influence is highlighted by recent developments made by Indian linguists and scientists, which provide verifiable proof of AI's contribution to innovative education.

- Improved Comprehension: To better understand the phonetic, lexical, and syntactic variation in Indian languages, researchers at IIT Madras, IIIT Hyderabad, and AI4Bharat have created speech recognition and text analysis models trained on regional dialect corpora (Kunchukuttan et al., 2020; Sitaram et al., 2021). In multilingual classrooms, these resources support comprehension by exposing students to linguistic diversity in everyday situations.
- Enhanced Engagement: EdTech companies that use AI to provide adaptive language learning experiences include Multibhashi (Bengaluru) and Entri App (Kochi). These platforms enhance engagement, motivation, and retention by customising content based on learners' dialectal preferences and performance data (Joseph & Kurup, 2022; KPMG & Google, 2017).
- Preservation of Linguistic Heritage: As a component of the Ministry of Electronics and Information Technology's (MeitY) National Language Translation Mission (NLTM), the Bhashini initiative seeks to use artificial intelligence (AI) to digitise and standardise Indian languages and dialects. Its main goal is to create open-source language tools and datasets for low-resource languages and dialects (MeitY, 2023).
- Accessibility and Inclusivity: Dialect-sensitive content has been made available in underserved and rural areas through AI-based applications created by C-DAC and the TDIL (Technology Development for Indian Languages) project. To promote inclusive language access, regional schools are implementing voice-enabled learning modules (Saxena et al., 2021).
- Effective Learning Paths & Empowered Teachers: NCERT and CBSE pilot projects, in partnership with AI research labs, have launched AI dashboards that assist teachers in monitoring students' progress in dialect-sensitive modules, enabling tailored educational interventions (CBSE, 2022; NITI Aayog, 2021).

Collectively, these advancements demonstrate India's dedication to balancing its rich linguistic diversity with cutting-edge educational technologies, paving the way for inclusive, culturally relevant, and AI-enhanced learning in the future.

7. Challenges of AI Implementation: Addressing India's Socio-Educational Realities

Even though artificial intelligence has the potential to completely transform Indian language instruction, there are many obstacles in the way of its successful application, particularly given the country's socioeconomic backwardness and rural-urban divide.

- One major obstacle is still the lack of data. The majority of annotated datasets used to train AI models come from standard and urban language varieties, excluding dialects used in marginalised, rural, or tribal communities. Existing disparities in access and representation are strengthened by this omission.
- In under-resourced schools, particularly in rural India, where access to modern hardware, high-speed internet, and even reliable electricity is still inconsistent, computational resource limitations are especially severe. Those who stand to gain the most from AI tools are unable to access them without this infrastructure.
- Cultural and ethical sensitivity are essential. Already marginalised linguistic groups may be further disadvantaged by algorithmic biases, in which AI systems favour urban speech patterns or dominant dialects. To prevent reinforcing linguistic hierarchies, diverse data inputs and culturally appropriate design are crucial.
- It is difficult to integrate systemsically into the current educational framework. Particularly in underdeveloped areas, government-run schools frequently use antiquated practices and struggle with bureaucratic sluggishness that prevents them from implementing cutting-edge technology.

One of the biggest obstacles is the lack of teacher training. Many educators in public schools are not digitally literate enough to use AI tools efficiently, particularly when teaching languages that require complex comprehension. Without consistent public funding or strong policy backing, scalable deployment is challenging due to costs associated with both initial development and ongoing maintenance. Inclusive

planning, regional data collection, and capacity-building programs, especially for India's underprivileged educational segments are necessary to address these issues.

8. Future Scope and Research Directions: Aligning with NEP 2020

AI holds promise for Indian language instruction, especially within the framework of the National Education Policy (NEP) 2020. The NEP prioritizes multilingualism, inclusive education, and technology integration to improve learning outcomes across diverse linguistic contexts. AI can be instrumental in achieving these goals through targeted research and development. Specifically, crowdsourcing dialectal data, where communities contribute speech and text samples, aligns with the NEP's mission to preserve local languages and ensures broader linguistic representation in AI databases.

Cross-lingual AI models can enhance machine learning in low-resource dialects by transferring linguistic patterns, fostering deeper understanding. Generative AI can create dialect-specific educational materials, aiding NEP's contextualized learning objective. Explainable AI (XAI) systems, by openly describing dialectal judgements, can build educator trust and acceptance. Furthermore, adaptive AI-powered linguistic mentors can personalize instruction based on a learner's dialect, supporting differentiated instruction, a key NEP 2020 goal.

Standardisation and policy initiatives are crucial, requiring collaboration between linguists, educators, government agencies, and AI developers to establish ethical, pedagogical, and technological standards for AI in dialect-sensitive language instruction across India.

Conclusion

An important turning point in balancing tradition and technology is the incorporation of AI into Indian language instruction. AI-driven educational tools can address longstanding issues in dialectal diversity, accessibility, and instructional equity by leveraging the linguistic depth of Yāska (Sarup, 1927), Pāṇini (Bhartrhari, 1965), and the experiential ethos of the Gurukula system (Sharma, 1996). The recent initiatives of Indian scientists, linguists, and ed-tech innovators reflect growing momentum towards inclusive, dialect-sensitive learning models (IITM, 2023; MeitY, 2022), despite the fact that urban-rural disparities and infrastructure gaps impede uniform implementation (Mehta & Kumar, 2021).

A suitable policy framework for this change is offered by the National Education Policy 2020 (MoE, 2020). Adaptive AI systems that respect regional languages and dialects are directly supported by its emphasis on multilingualism, contextual learning, and technology integration. India can guarantee that technology does not supplant pedagogy but rather increases its cultural and contextual relevance by making investments in explainable AI, community-based data collection, and personalised learning environments.

Technology, educators, linguists, and legislators must work together to address ethical issues, cost structures, and teacher preparation if they are to have a significant impact. Future work should concentrate on developing dialectal corpora that cover India's diverse linguistic landscape, cross-lingual AI models, and generative tools with a cultural foundation.

Ultimately, using AI to decode dialects is not just a technical undertaking; it is also a cultural necessity and a teaching opportunity to respect India's linguistic diversity and get the next generation ready for fair participation in a knowledge-based society.

REFERENCES

1. Agnihotri, R. K., & Khanna, A. L. (1997). Problematising English in India. Sage.
2. Annamalai, E. (2001). Managing Multilingualism in India: Political and Linguistic Manifestations. Sage.
3. Bhattacharya, T. (2016). The Politics of Language Contact in India. Cambridge Scholars Publishing.
4. Cardona, G. (1997). Pāṇini: A Survey of Research. Motilal Banarsidass.
5. CBSE. (2022). Handbook on Artificial Intelligence for Schools. Central Board of Secondary Education.
6. Chatterjee, A., & Banerjee, S. (2019). NLP for Indian Languages: Dialectal Sentiment Analysis. Journal of Linguistic Technology, 12(1), 45–61.
7. Deshpande, M. (1993). Transforming Tradition: Studies in Sanskrit and Tamil. Motilal Banarsidass.
8. Ghosh, S., Goyal, P., & Shrivastava, M. (2020). Building Linguistic Resources for Dialects: The Case of Indian Languages. Proceedings of LREC.
9. Houben, J. E. M. (2011). The Sanskrit Tradition and India's Linguistic Diversity. EFEO.
10. Iyengar, R. (2017). Grammar in Practice: Insights from Indian Dialects. Indian Linguistics, 78(3), 122–134.

11. Joseph, A., & Kurup, B. (2022). Personalized Learning in Indian Languages: A Study of Multibhashi and Entri. Journal of EdTech in India, 3(2), 22–34.
12. Jha, G. N. (2010). Developing Annotated Corpora for Indian Languages. Linguistic Research Bulletin, 45(2), 90–102.
13. Joshi, A., & Singh, M. (2020). NLP and the Indian Linguistic Landscape. AI and Society, 35(4), 477–488.
14. KPMG & Google. (2017). Indian Languages—Defining India's Internet. KPMG India.
15. Krishnamurti, B. (2003). The Dravidian Languages. Cambridge University Press.
16. Kunchukuttan, A., Mishra, P., & Bhattacharyya, P. (2020). AI4Bharat: NLP Tools for Indian Languages. ICON Proceedings.
17. Kulkarni, M., & Chaitanya, V. (2021). Pāṇini and Computational Linguistics. Indic Studies Journal, 5(1), 33–48.
18. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. Pearson.
19. Masica, C. P. (1991). The Indo-Aryan Languages. Cambridge University Press.
20. Mehta, S., & Kumar, R. (2021). AI Implementation in Rural India: Barriers and Pathways. Technology and Society Review, 10(1), 67–79.
21. MeitY. (2022). National Language Translation Mission: Bhashini Initiative. Ministry of Electronics and Information Technology.
22. Mohanty, A. K. (2010). Languages, Inequality and Marginalization: Implications for Mother-Tongue Based Education. UNESCO.
23. Mukherjee, S. (2019). Education in Ancient India: The Gurukula System. Aryan Books.
24. NCERT. (2019). Position Paper on Teaching of Indian Languages. National Council of Educational Research and Training.
25. NCERT. (2020). Teacher Competency Framework for Multilingual Education. NCERT.
26. NITI Aayog. (2021). Responsible AI for All: Strategy Document. Government of India.
27. Patel, R., Mehta, S., & Rao, A. (2021). Computational Semantics in Indian Contexts. Journal of Asian Linguistics, 14(1), 29–46.
28. Rai, R., & Sharma, A. (2022). AI for Immersive Language Learning. International Journal of EdTech, 8(4), 85–97.
29. Rao, M., & Kumar, S. (2022). Phonetics Meets AI: Revisiting Pāṇini. Indian Journal of Computational Linguistics, 3(2), 19–35.
30. Rao, P., & Srinivas, G. (2022). Dialect Morphology and Syntax: NLP Approaches. Language Technology Journal of India, 9(2), 111–127.
31. Saxena, S. (1994). Languages and Language Politics in India. South Asia Books.
32. Saxena, V. (2020). Culturally Responsive Learning Platforms: Bridging Tradition and Tech. Journal of Educational Innovation, 5(2), 102–119.
33. Saxena, A., Singh, M., & Goyal, N. (2021). Language Technology and Inclusion. South Asian Technology Journal, 8(1), 57–70.
34. Schmid, U., et al. (2021). AI and Education: A European Perspective. Springer.
35. Sharma, R. (1996). Ancient Indian Pedagogy and the Gurukula Model. Vikas Publishing.
36. Sinha, R., & Thakur, P. (2022). Dialect Classification with Deep Learning. Journal of Speech Technology in India, 4(1), 77–93.
37. Sitaram, S., Choudhury, M., & Bali, K. (2021). Speech Technology for Code-Switched Languages: A Survey. Computer Speech & Language, 68, 101199.
38. Srivastava, R., et al. (2020). Code-Mixed Language Modeling for Hinglish. ACL Anthology, 34(1), 345–356.
39. Subramanian, R. (2019). Pāṇini for the Digital Age. Indian Linguistic Tradition Review, 2(1), 11–26.
40. VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4), 197–221.