STRUCTURAL AND PHONOLOGICAL DIFFERENCES BETWEEN DRAVIDIAN AND INDO-ARYAN LANGUAGES: A CRITICAL STUDY

G. Saidulu
Research Scholar
Arni University
Himachal Pradesh

Dr. P. Suresh
Research Supervisor
Arni University
Himachal Pradesh

Dr. D. Sucharitha
Research Co-Supervisor
Arni University
Himachal Pradesh

ABSTRACT

The present paper is mainly concerned with the phenomenon of phonological uniformity among Indic languages. The paper begins with a brief overview of the critical notions in the field of 'linguistic area', introduced by M B Emeneau in 1956. It then takes up the cases of segmental and prosodic features that fill in the phonological spaces of the languages of six prominent language groups- Austro-Asiatic, Dravidian, Indo-Aryan, Tibeto-Burman, Tai Kadai and Great Andamanese. The segmental features are instantiated by the following- laryngeal types, retroflex sounds, nasal vowels, and hiatus. The prosodic features include syllable structures, stress, tone, and rhythm. While features of diversity are assumed as given in the languages of India, the focus here is on features of commonness. The general approach and goal of the present study is in consonance with the refrain of unity among the languages of India in the work of Professor Suniti Kumar Chatterjee.

Keywords: *Linguistic Area, Segmental Phonology, Hiatus, Prosodic Phonology, Rhythm*

INTRODUCTION

The Indian subcontinent is renowned for its exceptional language diversity and high population density, which justifies its classification as a "subcontinent." the region encompasses a population of almost 2 billion individuals residing in nations. these individuals communicate in over 100 prominent languages, as well as numerous minority languages and dialects, which belong to four major language groups: Indo-Aryan, Dravidian, Austroasiatic, and Sino-Tibetan. the Indo-Aryan languages, spoken by 78.05% of the population, and the Dravidian languages, spoken by 19.64%, dominate the linguistic landscape. Together, these are sometimes collectively referred to as Indic languages. Additionally, smaller fractions of the population speak Austroasiatic, Sino-Tibetan, Tai-Kadai, and other minor languages or isolates, accounting for 2.31% of the population. Over the course of time, these groups of languages have had considerable interactions, resulting in the creation of a linguistic region or sprach bund in India. Within this region, languages from diverse families share numerous.

The Indo-Aryan group is one of the few groups of Indo-European languages, if not the only one, for which no classification based on rigorous genetic criteria has been suggested thus far. The cause of such a situation is neither lack of data, nor even the low level of its historical interpretation, but rather the existence of certain prejudices which are widespread among Indologists. One of them is the belief that real genetic relations between the Indo-Aryan languages cannot be clarified because these languages form a dialect continuum. Such an argument can hardly seem convincing to a comparative linguist, since dialect continuum is by no means a unique phenomenon: it is characteristic of many regions, including those where Indo European languages are spoken, e.g. the Slavic and Romance-speaking areas. Since

genealogical classifications of Slavic and Romance languages do exist, there is no reason to believe that the taxonomy of Indo-Aryan languages cannot be established.

Living in a society, we humans communicate with each other to share our thoughts, feelings, ideas, and different types of information. We use different communication mediums, like text messages, emails, audio, and videos, to express ourselves to others. Besides this, people nowadays use a variety of emojis along with text messages to represent their feelings more precisely. However, without any doubt, among all the communication forms, speech is the most natural and easiest one to express ourselves.

In recent years, contact with computing devices has become more chatty. Dialogue systems like Siri, Alexa, Cortana, and many more have more widely penetrated the consumer market than before. Thus, to make them more like a human conversational partner, it is important to recognize human emotions from the user's voice signals. Understanding the emotional state of a speaker is important for perceiving the exact meaning of what he or she says. Therefore, research studies on automatic speech emotion recognition, which is the task of predicting the emotional state of humans from speech signals have emerged in recent years as it enhances human-computer interaction (HCI) systems and makes them more natural. Moreover, with the world being digitized day by day, speech emotion recognition has found increasing applications in our daily lives. Call centers, e-tutoring, surveillance systems, psychological treatments, robotics, and online marketing are just some of them.

It has been seen from cross-lingual speech emotion recognition studies that models trained with a language corpus do not perform well when tested on a different language corpus compared to the monolingual recognition rate. However, it will be interesting to find out whether those models perform better for different languages from the same groups. To carry out this study, the first thing is to find out the available resources of the target language group. Hence, in this study, we aimed to investigate recent advancements in SER for the Indo-Aryan and Dravidian language families. Indo-Aryan and Dravidian languages are spoken by 800 million and 250 million people, all over the world, respectively. The speakers of Indo-Aryan languages are mostly from Bangladesh, India, Nepal, Sri Lanka, and Pakistan, and speakers of Dravidian languages are mainly from southern India. Having a large number of speakers, yet most of them are low-resource languages. So far, there is no review work that highlights the SER experiments for the Indo-Aryan or Dravidian language groups. Therefore, this study presents a brief review of some work done for the development of SER for languages of the Indo-Aryan and Dravidian families.

LITERATURE REVIEW

John Peterson et.al (2010) The present study takes a closer look at language convergence in Jharkhand in eastern-central India, concentrating on Indo-Aryan and Munda languages. Although it is well-known that the Indo-Aryan languages which function as linguae francae in the region – such as Sadri, Bengali and Oriya – have had an enormous impact on the morphosyntax and lexicon of the Munda languages, in this study I call attention to a number of convergences which to my knowledge have so far gone unnoticed, many of which appear to originate in Munda, while others are of uncertain origin. These include, among others, the emergence of inalienable possession as a morphological category and incipient dual marking

in the pronominal paradigm in Sadri, similarities in categories denoting 'from' and 'to' or 'begin' and 'keep on', as well as a number of interesting areal developments of the genitive, including 3rd person marking, focus marking, or becoming part of the copular stem in several languages of the region.

Maria Casadei et.al (2023) Dakhni Urdu is an Indo-Aryan language spoken in the Deccan region of India, especially in the states of Andhra Pradesh, Telangana, Karnataka, Kerala and Tamil Nadu. This variety of Urdu, which developed in the Deccan from the 13th century, is the result of the contact between Urdu and the Dravidian tongues spoken in South India. It flourished as a literary vehicle during the 14th and 15th centuries and, after the conquest of Deccan by the Mughals in 1687, it saw a rapid decline that restricted it to the oral form. The proposed paper seeks to trace the history and development of Dakhni from its birth to its decline, paying closer attention to the social, political and linguistic choices that influenced the use of Dakhni in South India. In particular, the research looks at its origin, its literary production and, finally, at the social and political factors which led to its decline in the late 17th century. Moreover, the paper also gives a brief outline of Dakhni and its main linguistic features, stressing the Indo-Aryan and Dravidian influence on the language.

Kuwali Talukdar et.al (2023) Automatic Parts of Speech (PoS) Tagging is a sequence labeling problem. PoS Tagging research has undergone an evolutionary journey starting with Dictionary Lookup PoS Tagger model, and then using rule based and statistical schemes, and later on adopting hybrid methodology for enhanced performance. Emergence of Machine Learning (ML) has boosted the activities adding newer dimensions to look at the problem with deeper linguistics and computational perspectives, and in recent times this has shifted to completely self-learning models with incorporation of Deep Learning (DL) tools. Here we have recorded and analyzed this trajectory for the PoS tagger development and experimentation for the Indo Aryan languages. Rule based and statistical models performed to an acceptable level, but are not robust and dynamic. ML and DL based models outperformed all other models, and started giving higher performances, with reported accuracy to the tune of upto 97% in few cases.

Maimuna Hasmun Nahar et.al (2021) This study aims at studying various languages' impact on Assamese language over the period of history. The language families viz. Indo-Aryan, Tibeto-Burman, Austro-Asiatic, Dravidian and Tai-Kadai prevailing in Assam make the land a multilingual one. Other ethnic groups found on the state are Tai-Phake, Tai-Aiton and Tai-Khamti. Different tribes (plain and hill tribes) in Assam bear examples of different languages and cultures. Similarities are found between Assamese and other non-Aryan customs. A language develops from interactions that take place between certain speech communities. Assamese food habits, festivals and other cultural aspects have been greatly influenced. As a result, a lot of lexical items as well as other linguistic features are amalgamated with Assamese language. Assam is a great example of diverse lingua-cultural state where each language has made an impact on one another's vocabularies.

Miriam Butt et.al (2010) The original case system found in Sanskrit (Old Indo-Aryan) was lost in Middle Indo-Aryan and then reinvented in most of the modern New Indo-Aryan (NIA) languages. This paper suggests that: (1) a large factor in the redevelopment of the NIA case

systems is the expression of systematic semantic contrasts; (2) the precise distribution of the newly innovated case markers can only be understood by taking their original spatial semantics into account and how this originally spatial semantics came to be used primarily for marking the core participants of a sentence (e.g., agents, patients, experiencers, recipients). Furthermore, given that case markers were not innovated all at once, but successively, we suggest a model in which already existing case markers block or compete with newer ones, thus giving rise to differing particular instantiations of one and the same originally spatial postposition across closely related languages.

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.2Recent 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.

METHODOLOGY

Data transcription, tabulation, and classification based on morphological and syntactic variables are integral to our analysis procedures, which align with established linguistic analysis frameworks. Findings will be presented using data visualisation methods, focusing on identifying morphological and syntactic errors and shedding light on areas of linguistic divergence and convergence between Dravidian and Indo-Aryan speakers in the selected regions.

Universe: Uttar Pradesh, West Bengal, Tamil Nadu and Andhra Pradesh

- Uttar Pradesh

  Situated in the northern expanse of the Indian subcontinent, Uttar Pradesh occupies a multi dimensional role both geographically and socio-culturally. Enclosed by Bihar to the east, Madhya Pradesh to the south, Rajasthan to the west, Haryana and the National Capital Territory of Delhi to the northwest, and Uttarakhand and the international border with Nepal to the north, the state showcases a vast range of geographical and cultural peculiarities (Government of Uttar Pradesh, 2021). Its capital, Lucknow, is a salient cultural and administrative hub with historical landmarks and traditional art forms that complicate its identity. According to Census data 2011, Uttar Pradesh encompasses an area of about 243,290 square kilometers, making it the fourth-largest state within the Indian union (Census of India, 2011). This expanse harbors a population of approximately 199.8 million, presenting a multifaceted demographic composition representing various ethnicities, religions, and social strata (Census of India, 2011). Literacy rates, a pivotal indicator of development, showcase a marked gender disparity, with 77.28% of males and 57.18% of females reported as literate, according to the 2011 Census (Census of India, 2011). According to the Census of India 2011, most of Uttar Pradesh's population resides in rural areas. Approximately 77.73% of the population was classified as rural, while 22.27% identified as urban. The linguistic landscape of Uttar Pradesh is predominantly characterised by Hindi, the language spoken by the majority of its inhabitants and used extensively in administration and daily life (Office of the Registrar General & Census Commissioner, India, 2011).

CONCLUSION

The study provides a detailed quantitative overview of the prevalent difficulties encountered by speakers of Hindi, Bangla, Tamil, and Telugu in mastering English morpho-syntactic constructs. For instance, errors in the use of the superlative (-est) and the third-person singular present (-s) morphemes were notably high across all linguistic groups, pointing to

specific areas where targeted instruction could significantly improve learners' proficiency in English. Similarly, the pattern of errors in the past tense (-ed), plural (-s), and past participle (-en) morphemes suggests a widespread struggle with verb forms and noun number agreements, which are crucial for syntactic coherence and morphological accuracy in English. Moreover, the analysis reveals how educational and socio-psychological factors, such as the medium of instruction, gender, and motivational orientations, contribute to these errors. For example, learners instructed in their native languages showed different error patterns compared to those educated in English medium, indicating how the language of instruction influences the learning of English grammar. Gender differences in error frequencies further highlight the nuanced ways male and female learners navigate the challenges of English language learning, with certain error types more prevalent among one gender over the other.

REFERENCES

1. *John Peterson et.al (2010), "Language contact in Jharkhand Linguistic convergence between Munda and Indo-Aryan in eastern-central India", Himalayan Linguistics,ISSN:1544-7502,Vol.9,issue.(2),pages.56-86*
2. *Kuwali Talukdar et.al (2023), "Parts of Speech Taggers for Indo Aryan Languages: A critical Review of Approaches and Performances", IEEE access, ISSN: 2169-3536, vol. 11, DOI: 10.1109/I3CS58314.2023.10127336.*
3. *Maimuna Hasmun Nahar et.al (2021), "The Diverse Linguistic Impact on Assamese: An Indo -Aryan Language", International journal of creative research thoughts, ISSN: 2320-2882,Volume.9,Issue.3*
4. *Maria Casadei et.al (2023), "From a Literary Language to an Oral Tongue: AL inguistic Overview of the Dakhni Language", Journal of Language and Linguistics in Society, ISSN: 2815-0961,Vol.03,issue.03*
5. *Miriam Butt et.al (2010), "The redevelopment of Indo-Aryan case systems from a lexical semantic perspective", Morphology, ISSN: 1097-4687, Volume.21, pages. 545–572*
6. *Mohammad Ali et.al (2014), "Characterizing the genetic differences between two distinct migrant groups from Indo-European and Dravidian speaking populations in India", BMC genetics, ISSN 1471-2156,vol.15, doi: 10.1186/ 1471-2156-15-86*
7. *Saartje Verbeke et.al (2020), "Shaping Modern Indo-Aryan isoglosses", Poznan Studies in Contemporary Linguistics,vol.56,issue.(3), ISSN: 1897-7499, DOI: 10.1515/psicl-2020-0017*
8. *Saydali Mukhidinov et.al (2018), "Ancestral Home of Indo-Aryan Peoples and Migration of Iranian Tribes to South eastern Europe", SHS Web of Conferences, ISSN:2261-2424,vol.50, https://doi.org/10.1051/shsconf/20185001237*
9. *Sumana Mallick et.al (2018), "Early Indian Languages: An Evolution Perspective", Asian Review of Social Sciences,ISSN: 2249-6319,Vol.7, issue. 2,pp. 57-61*
10. *Syeda Tamanna Alam Monisha et.al (2022), "A Review of the Advancement in Speech Emotion Recognition for Indo-Aryan and Dravidian Languages", Advances in Human-Computer Interaction, ISSN:1687-5907,vol.5, https:// doi.org/10.1155/2022/9602429*
11. *Terisha Basumatary et.al (2021), "Phonology of Lexical borrowing of Indo-Aryan Languages used in Bodo",International Journal of Emerging Technologies and Innovative Research,ISSN: 2349-5162,Volume.8,Issue.3*
12. *Vishnupriya Kolipakam et.al (2018), "A Bayesian phylogenetic study of the Dravidian language family", Royal Society open science, ISSN 2054-5703, vol.5,http://dx.doi.org/10.1098/rsos.171504*
13. *Wiqas Ghai et.al (2012), "Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi a Case Study", International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-2 Issue-1*

14. *Yu-Chun Li et.al (2018), "Cultural diffusion of Indo-Aryan languages into Bangladesh: A perspective from mitochondrial DNA", Mitochondrion, ISSN: 1872-8278,Volume.38,Pages.23-30,https://doi.org/10.1016/j.mito.2017.07.010*
15. *Zygmunt Vetulani et.al (2021), "Development of real size IT systems with language competence as a challenge for a Less-Resourced Language: a methodological proposal for Indo-Aryan languages", Journal of Information and Tele communication, ISSN:2475-1847,Volume.5, Issue.4, https://d oi. org/10.108 0/ 2 4751839.2021.1966236*