Journal Title: *Mankind: Adam to Me*
Website: *https://juria.in/mankind*
ISSN: 3048-7668 (Online)
Vol. 3(2), Nov 2026, pp. 11-22

Published by: Juria Publisher, India

Review Article

Continuity, Change, and Language Contact in Indo-Aryan: A Historical-Comparative Review

Mohammad Qaseem Kashaf [1*], *Azizullah Jabarkhail* [2]

[1]*PhD Scholar, Department of Pashto, Faculty of Languages and Literature, Nangarhar University, Jalalabad, Afghanistan*
[2]*Research Supervisor, Department of Pashto, Faculty of Languages and Literature, Nangarhar University, Jalalabad, Afghanistan*

ARTICLE INFO

Keywords:
Indo-Aryan, Indo-European, Historical linguistics, Language contact, Linguistic change, South Asian languages

Article History:
Received: 11-10-2026
Revised: 15-11-2026
Accepted: 17-11-2026
Published: 22-11-2026

ABSTRACT

Indo-Aryan is one of the oldest branches of Indo-European languages. It is an extremely rich history of the language's development from Old Indo-Aryan to the Middle Indo-Aryan and the wide variety of languages of the New Indo-Aryan period, and as such offers a unique window on the ways in which languages "preserve inherited structures" and "undergo substantial internal and contact-related changes. The present article reviews the historical development of Indo-Aryan from three interrelated angles: genealogical inheritance, structural restructuring and language contact. It reflects on historical-comparative scholarship and research on South Asian linguistic convergence and looks at selected developments within the major historical periods in phonology, morphology, syntax and vocabulary. The progressive reorganization of inherited grammatical systems and the impact of ongoing contact with Austroasiatic, Tibeto-Burman, Iranian, Persian and Arabic traditions are given special attention. Instead of focusing exclusively on internal processes of change and contact, the review concentrates on the potential interaction between the processes in driving linguistic changes. It is the case that it is the historical record of Indo-Aryan in particular that is so important for the study of the link between genealogical continuity and areal convergence. The review also highlights those aspects of the particular Indo-Aryan developments, which are still subject to a dispute, and that further comparative and corpus-based research would allow the existing account of linguistic changes in South Asia to be improved.

Cite this article:
Kashaf, M. Q., & Jabarkhail, A. (2026). Continuity, Change, and Language Contact in Indo-Aryan: A Historical-Comparative Review. *Mankind: Adam to Me*, 3(2), 11-22. https://doi.org/10.55559/mankind.v3i2.33

Introduction

The Indo Aryan languages have significant importance in the history of Indo-European languages. The importance of their role stems from the fact that they were only first recorded in writing in ancient times, but also from the remarkable linguistic diversity that emerged over the following centuries. The Rigveda is the most prominent source of the Vedic Sanskrit language, which is one of the earliest fully documented Indo-European languages, and which has therefore been a major source for comparative reconstruction. The record of Classical Sanskrit, the different Middle Indo-Aryan traditions, and the modern Indo-Aryan languages is unusually long, in which it is possible to trace the processes of continuity and change in the language (Beekes, 2011; Clackson, 2007; Fortson, 2010). Indo-Aryan is an Indo-Iranian Indo-European language, having a significant amount of phonological, morphological and lexical features in common with Iranian. Comparative evidence indicates a common Proto-Indo-Iranian stage from which Indo-Aryan and Iranian later parted ways. However, the earliest Indo-Aryan proof is already sufficiently different to constitute an historical branch. Meanwhile, Vedic Sanskrit also contains a number of archaic grammatical and phonological traits that have been extremely useful in the reconstruction of the earlier Indo-Iranian and Indo-European structures (Beekes, 2011; Fortson, 2010; Hock & Bashir, 2016).

Traditionally the history of Indo-Aryan is broken down into three general periods: Old Indo-Aryan, Middle Indo-Aryan and New Indo-Aryan. The Old Indo-Aryan language is best attested in its literary forms, namely the Vedic and Classical Sanskrit. The latter is a standardized literary language, but the connection between the written and spoken varieties of language must be approached tentatively. Middle Indo-Aryan is a linguistically diverse category that groups Pāli, inscriptional and literary Prakrits, as well as Apabrahṃśa forms of later period. There are also a large and internally diverse group of modern languages of the New Indo-Aryan group, including Hindi, Urdu, Bengali, Punjabi, Marathi, Gujarati, Nepali, Odia, Assamese, Sindhi, Sinhala, and many regional languages and varieties (Jain & Cardona, 2007; Masica, 1991). The historical stages need not be interpreted as distinct linguistic periods however, nor as a linear sequence from Sanskrit through the modern languages. The

**Corresponding Author*:
qaseem_kashaf@nu.edu.af *(M. Q. Kashaf)*
© 2026 The Authors. Published by Juria Publisher, India. This is an open access article published under the CC-BY license
https://creativecommons.org/licenses/by/4.0

emergence of Indo-Aryan was characterized by overlapping varieties, regional differentiation, literary standardization, vernacular change, and ongoing contact between communities. In particular, Middle Indo-Aryan is more complex, and not just a simplified or corrupted version of Sanskrit. It is this variety diversity that enables the recognition of language changes within the Indo-Aryan vernaculars that took place during the process of its development, and this is true not only for the most significant phonological shifts, but also for the reorganization and change of inherited morphology. Innovative case-marking systems, postpositional constructions, auxiliary-based verbal constructions and new patterns of grammatical organization are among the further grammatical rejuvenations that followed the emergence of the New Indo-Aryan (Butt, 2022; Jain & Cardona, 2007; Masica, 1991). Another major aspect of the history of Indo-Aryans is that the changes took place in a multilingual context. South Asia is a region of longstanding contact between languages of several genealogically unrelated families. The Indo-Aryan languages have coexisted with the Dravidian, Austroasiatic, and Tibeto-Burman languages, and in certain periods, there has been considerable contact with Iranian languages as well as with Persian and Arabic. The interaction has affected vocabulary and in some instances phonological and grammatical structures as well. Emeneau (1956), in particular, was famous for describing India as a "linguistic area" and emphasising the structural similarities between different families of languages, irrespective of the genealogical distance between them. This view has been informed by subsequent scholarship and shown to be more complex and historically and geographically variable than the idea of a single uniform contact process, as is the case with South Asian linguistic convergence (Hock & Bashir, 2016; Southworth, 2005). This relationship of transmission and contact has placed Indo-Aryan in a special place for historical linguistics. A number of features can be reasonably connected with Indo-Iranian or Indo-European sources, while others are innovations that originated in later stages of the branch. Still other characteristics have been associated with prolonged regional contact. The explanations do not necessarily have to be mutually exclusive. A formerly constructed structure might be restructured inside, or an internal innovation might proliferate or be strengthened in a multilingual context. These processes must be differentiated by considering the chronology, comparative evidence, geographical distribution, and the structure of neighboring languages, not by assuming that similarity in structure implies direct borrowing (Hock & Bashir, 2016; Southworth, 2005). Indo-Aryan is therefore well documented historically, and can provide a record of the development of more than one linguistic branch. It allows us to look at the ways in which genealogical inheritance, internal restructuring, and contact have manifested over a long span of language history that has been documented. Significant innovations as well as continuity with the previous stages can be seen in the development of vocabulary, syntactic organization, nominal and verbal morphology, and sound systems. The study of these processes together can help to create a more balanced perspective of Indo-Aryan history, as well as to contribute in wider discussions in historical-comparative and contact linguistic studies.

The present study therefore examines the historical and structural evolution of Indo-Aryan, beginning in the early Indo-Iranian period, extending through the Old Indo-Aryan, Middle Indo-Aryan and the New Indo-Aryan period. It focuses especially on the nature of phonological, morphological, syntactic and lexical change, and how this is affected by multilingual contact. The study does not aim at the complete description of all Indo-Aryan languages, but focuses on representative developments that bring to light the general outlines of continuity, restructuring, diversification and convergence.

Review of Literature

The study of Indo-Aryan languages has grown out of a number of interrelated traditions, such as Indo-European comparative linguistics, Sanskrit and Prakrit studies, historical grammar, language typology and contact-induced language change in the Indian subcontinent. The comparative scholarship that followed early comparative studies firmly placed Sanskrit as a significant source to reconstruct Indo-European linguistic history and later studies began to focus on the processes of the changes, diversification, and interaction of Indo-Aryan languages with other languages. Contemporary scholarship, therefore, has taken a different view to the Indo-Aryan than just one that sees it as a branch of the PIE; it views Indo-Aryan as a historically layered branch that has developed through inheritance, internal restructuring, regionally distinct processes, and continued multilingual contact.

The position of Indo-Aryan in the wider context of Indo-European and Indo-Iranian has been the subject of much research. Clackson (2007), Beekes (2011) and Fortson (2010) show how systematic phonological, morphological and lexical correspondences can be used to reconstruct the relationship between the Indo-European languages. In this context, the intimate connection of early Indo-Aryan and Iranian is the main motivation for the reconstruction of Proto-Indo-Iranian. Vedic Sanskrit is significant especially for the fact that it has retained the older morphological categories, verbal forms, accentual patterns, and lexical material which can be compared with Avestan and other early Indo-European languages. Comparative scholarship, however, warns against the use of Vedic Sanskrit as an unaltered reflection of Proto-Indo-Iranian or Proto-Indo-European. It is a historically created language, with a fusion of inherited and innovations that are distinctive to Indo-Aryan (Beekes, 2011; Fortson, 2010). There has been equally extensive study into the internal history of Indo-Aryan. One of the most important broad descriptions of the New Indo-Aryan languages is that of Masica (1991), which is especially significant in the way that it takes account of geographical variation and historical development as well as structural comparison. Instead of regarding modern Indo-Aryan languages as simple lineal descendants of Classical Sanskrit, Masica stresses its complex relationship between Old, Middle and New Indo-Aryan and regional differences. Two detailed treatments of individual Indo-Aryan languages and linguistic traditions, by Jain and Cardona (2007), are also provided to illustrate the structural diversity of the branch. These papers demonstrate that the history of Indo-Aryan is not well captured by the simple paradigm of an unbroken lineage of Sanskrit to modern languages.

A particular importance for the study of structural change is the transition from Old Indo-Aryan to Middle Indo-Aryan. The situation with respect to the phonology and morphology of Middle Indo-Aryan is quite different from that of Old Indo-Aryan. Examples of these developments are consonant cluster shifts, restructuring of inherited inflectional paradigms, loss or reduction of case distinctions, and greater use of analytic grammatical strategies. The Middle Indo-Aryan period thus offers much more than an intermediate step; it offers insights into the processes by which inherited grammatical systems were reorganized prior to the emergence of New Indo-Aryan systems (Jain & Cardona, 2007; Masica, 1991). The long term restructuring is particularly evident in recent studies on case marking in Indo-Aryan languages. Butt (2022) traces the steps from the inflectional case system of Old Indo-Aryan to the substantial loss of inherited

case system in Middle Indo-Aryan, and the development of innovative case-marking systems in New Indo-Aryan. This work shows that the loss of inherited morphology can be accompanied by grammatical renewal and indicates that morphological 'simplification' does not necessarily imply grammatical 'simplification'. The new Indo-Aryan languages acquired oblique forms, new case markers and maintained the systematic links between the case marking and semantic differences. The evidence is significant not only because it demonstrates processes of simultaneous loss, reanalysis and innovation in linguistic change, but also because it indicates that these can occur in the same time period. Yet another line of research is in the direction of the multilingual situation in which Indo-Aryan arose. The influential characterisation of India as a linguistic area by Emeneau (1956) shifted the interpretation of the similarities between South Asian languages. Emeneau contrasted the assumption of genealogical relationships between Indo-Aryan and non-Indo-Aryan languages, and emphasized the impact of long-term interfamiliar contact. This method led to the development of areal linguistics as a significant facet of the history of South Asian languages, and sparked the interest of later scholars in the study of phonological, morphological, syntactic and lexical convergence.

Southworth (2005) expanded this historical-contact framework, incorporating language and the issues of population change, regional interaction, and the South Asian prehistory of the language. His work shows how crucial the phenomenon of lexical borrowing and structural convergence is for the reconstruction of relationships between the Indo-Aryan, Dravidian and Munda-speaking populations. Hock and Bashir (2016), on the other hand, gives a more general modern overview of the South Asian linguistic diversity, in which the genetic classification and the contact relationships are treated together. They stress the fact that the modern linguistic structure of South Asia is not necessarily best understood in family-tree terms; instead, historical relationships need to be considered in the light of long-standing multilingualism. The issue of contact must be handled with great care. As for the neighbouring languages, there are not necessarily structures in all of them that would automatically make it clear that one language borrowed from another. All these aspects of direction of influence, chronology, as well as possible mutual convergence need to be taken into account. This topic is especially important with regard to the Indo-Aryan retroflex consonants. The retroflexion has been often linked with early contact between Indo-Aryan and non-Indo-European languages of South Asia, in particular Dravidian. Yet the exact historical processes at work have been subject to debate, and Hock's account of Indo-Aryan retroflexion is instructive in pointing to the difficulty of outlining the distribution and origins of it. A responsible account should thus differentiate between general arguments of convergence and more specific statements about the exact origin or direction from which a specific development has come. A shift from South Asia as a region of a single language to localised studies on contact has occurred in research. For instance, Peterson (2017) looks at Indo-Aryan and Munda interaction in Jharkhand and considers the area a more definable contact zone language area within the larger contact zone of South Asia. This is important because the intensity of contacts and convergence in the structures is not geographically uniform. Regional linguistic histories can be either a mixture of Indo-Aryan, Munda, Dravidian or other languages, and can thus be explained in a few cases by a single subcontinental model.

Contacts between Indo-Aryan and Dravidian have also been interpreted as historically multiple and not one-way. The research on the South Asian convergence shows the material of extensive interaction over a long period of time, in which both Indo-Aryan and Dravidian languages engaged in structural and lexical exchange processes. Sridhar (2008), for example, speaks of 'convergence' as an ubiquitous aspect of South Asian linguistic history and points to the extent of Indo-Aryan influence on Dravidian and the features that have been read as evidence of Dravidian influence on Indo-Aryan. This multidirectional perspective is significant since it overcomes the dichotomy of donor–recipient models and presents a more historically realistic scenario of long-term multilingual interaction. This approach has been further developed in recent years for South Asian languages. However, instead of the contact hypothesis being a monolithic notion, current studies are able to make some differentiation between different times and places of contact and between lexical borrowing, phonological convergence, grammatical replication and other types of contact-induced variation. In the case of South Asian contact, for instance, Hock not only identifies phases and patterns, but also discusses post-Vedic contact as well as more structural and geographical issues (Hock & Bashir, 2016). Recent handbook treatments also place South Asian language contact in the broader theoretical context of multilingual convergence.

Traditional classification has also been complicated by study of the new Indo-Aryan. Modern Indo-Aryan languages exhibit considerable linguistic and geographical variation and there are frequently continuum regions between languages/groups. But as Masica (1991) shows, geographical, structural and historical factors do not always yield the same classification. It is also found that there is great diversity among individual Indo-Aryan languages as revealed by Jain and Cardona (2007). The use of terms like Northwestern, Central, Eastern, Western, or Southern, must therefore be interpreted as analytical categories, not as historically distinct genealogical entities. In particular, it is crucial when there are similarities in structures that are due to geographical contact rather than to common ancestry. Another recent advance in scholarship is the increasing reliance on computational techniques and on extensive linguistic databases. These approaches are not a substitute for the comparative method but do enable testing of historical hypotheses within a greater body of linguistic evidence. For instance, Arora, Farris, Basu and Kolichala (2022) highlight the potential of computational historical linguistics for the study of the exceptional linguistic diversity found in South Asia, and note that linguistic data which are dispersed is an important challenge for large-scale comparative research. Their work is part of a general trend towards the digital and computational turn in the field of traditional historical linguistics.

This development took a solid shape with the development of Jambu, a historical linguistic database for South Asian languages. Arora et al. (2023) compiled lexical data from hundreds of SA languages known as lects and attempted to classify their data into cognate sets for comparative research. The resource also provides computational investigation of historical relationships in Indo-Aryan. Such work shows that the future of the Indo-Aryan historical linguistics is likely to be a field in which philological skills and traditional historical comparative reconstruction will increasingly meet with the use of structured lexical databases and the modelling. As a whole, the literature shows the existence of a substantial shift in the perception of the history of the Indo-Aryan languages. A genealogical connection of Indo-Aryan with the broader Indo-European family had already been shown by earlier comparative scholarship, and Sanskrit proved to be of great importance in the reconstruction process. The research expanded later to include studies on the development of Middle and New Indo-Aryan, dialectal variation, grammatical reorganization, and South Asian multilingualism. Recent studies have also highlighted the importance of local contact, typology, grammatical renewal, and computation. Indo-Aryan scholarship has consequently shifted gradually from the purely genealogical interpretation to the multifaceted perspective of a linguistic history, with emphasis on the phenomenon of contact, on internal change, as well as on descent.

This scholarly tradition, however, is not necessarily connected to a cohesive explanation. Histories of Indo-Aryan focus on diachronic transformation, comparative Indo-European research privileges reconstruction and inheritance, historical language contact research puts more weight on convergence, and emerging computational research considers the processing of growing sets of data and quantitative patterns. There is therefore ample evidence in the literature on individual aspects of Indo-Aryan history, and space remains for a more comprehensive analysis of the interplay between inheritance, internal restructuring and contact throughout the Old, Middle and New Indo-Aryan spectrum.

Research Problem and Research Gap

Studies of Indo-Aryan languages have garnered significant knowledge on the origin of these languages, their historical evolution, their structures, internal variability, and contact with other languages of South Asia. Old Indo-Aryan has been important for the reconstruction of the IE system, and Middle and New Indo-Aryan have shown many phonological, morphological, syntactic and lexical changes (Beekes, 2011; Fortson, 2010; Jain & Cardona, 2007; Masica, 1991). At the same time, studies on the south Asian language area have shown that the history of Indo-Aryan cannot be explained without the persistent multilingual contact (Emeneau, 1956; Hock & Bashir, 2016; Southworth, 2005). All the same, there's still one major analytical difficulty. Historiography of Indo-Aryan history is dealt with on different bases. The most typical focus of Indo-European comparative studies is the genealogical and reconstructive, the most typical focus of Indo-Aryan is the diachronic and the most typical focus of contact-linguistic studies is that of convergence between Indo-Aryan and neighbouring language families. While each viewpoint has yielded valuable insights, they are each inadequate to explain the processes that turned the earlier documented stages of Indo-Aryan to the linguistic diversity observed today.

This is especially significant when separating features from the nature of the inheritance, restructuring from the inside and contact change from the outside. Some historical developments may fall into more than one of these classes. An inherited grammatical structure can undergo reanalysis, as can a development which is possible by internal means, if reinforced by contact with adjacent languages. However, the structural similarity of comparable structures from languages of different genetic origin is not evidence of direct borrowing, nor can it be used to determine the direction of influence. A framework is thus needed to account for all three types of evidence: genealogical, diachronic and areal, in the interpretation of Indo-Aryan (Hock & Bashir, 2016; Southworth, 2005). Another problem is the interrelationship between Old Indo-Aryan, Middle Indo-Aryan and New Indo-Aryan. Such labels are essential for historical description, but may provide a perception of distinct and internally homogeneous chronological phases. The linguistic record rather suggests the simultaneous occurrence of similar developments, regional differences, standardized literary traditions alongside vernacular speech, and differential rates of change of structure. The shift from the Old Indo-Aryan inherited inflectional system to Middle and New Indo-Aryan systems (e.g. Prakrit to Sanskrit) required a reanalysis of the grammatical system, and, to some degree, a loss of distinctions in earlier inflectional systems, as well as a development of new systems of case-marking (Butt, 2022). A satisfactory historical account therefore has to account for continuity, as well as change in terms of the loss or the simplification of Sanskrit structures. Recent scholarship also raises the questions regarding the synthesis of evidence about Indo-Aryan history. The traditional philological and historical-comparative methods continue to be crucial, but more detailed analyses of individual languages, of regional contact areas, of linguistic typology, and of computational historical linguistics have increased the amount of evidence available. These changes offer a chance to rethink general theories about Indo-Aryan evolution and also suggest that both genealogy and contact might be insufficient to explain the changes. The principal research gap the present study aims to fill is the lack of integration of these various explanations over the long history of Indo-Aryan. The study does not attempt a new classification or a reconstruction of the individual linguistic forms, but instead gathers the evidence for the historical continuity and restructuring in the Old, Middle and New Indo-Aryan continuum, as well as the evidence for contact in the same continuum. Through this, it aims to explain the overlap of these processes and to pinpoint where evidence available to the research suggests conflicting or ambiguous interpretations of the past.

Research Objectives

- To trace significant features of continuity and structural change between Old, Middle and New Indo-Aryan.
- To briefly look at certain phonological, morphological, syntactic and lexical changes in their historical setting.
- To evaluate the contribution of contact and areal processes to language development and diversification in the Indo-Aryan languages.
- To combine genealogical, historical-comparative, and contact-linguistic analysis to better understand the combined forces of these three approaches in understanding the evolution of Indo-Aryan language(s).

Research Questions

RQ1. What are the differences between linguistic inheritance and structural innovation across Old, Middle and New Indo-Aryan?

RQ2. What are the major phonological, morphological, syntactical, and lexical developments that show an Indo-Aryan continuity and restructuring?

RQ3. What role have long-term interactions with neighbouring traditions had in shaping Indo-Aryan languages?

RQ4. What can the historical development of Indo-Aryan tell us about the interplay between genealogical continuity, self-driven change and areal convergence?

Methodology

The design method used in this research is integrative review in a qualitative approach with historical comparative method. It is appropriate because Indo-Aryan's historical development has been recorded in different types of language; it is not recorded in one. History, comparative linguistics, grammars, text studies, epigraphic studies, individual Indo-Aryan languages and contact between languages and convergence studies in South Asia are relevant. An integrative review can include these multiple types of scholarship and provide a common analytical framework for them to be examined, and can compare the interpretations or theories of the past with newer ones in the field (Snyder, 2019; Torraco, 2016). The review is based on the background literature as well as on current research. Previous publications are preserved when these have contributed to significant developments in Indo-European reconstruction, Indo-Aryan historical linguistics, and theories about South Asia as a linguistic area. Among them are significant works on the historical development of the Indo-European and Indo-Aryan languages, as well as language contact (Emeneau 1956, Masica 1991, Southworth 2005). Latterly, more recent scholarship brings in to weigh how previous interpretations have been built upon, qualified, or rethought, with a particular focus on grammatical change, regional convergence, linguistic classification and computational methods for dealing with historical data. This is required because chronological recency is not a suitable standard for judging the quality of scholarship in historical linguistics: earlier works may continue to offer primary

description or theoretical explanation which are still at the core of contemporary scholarship.

Historical Development of Indo-Aryan Languages

Its historical evolution is spanning over 3000 years, and offers an unparalleled example of an Indo-European language that had undergone continuous changes across the ages. It is usually split into Old Indo-Aryan, Middle Indo-Aryan, and New Indo-Aryan, but these are only conventions, and the terms are not necessarily used to denote clear-cut linguistic eras. Changes in phonology, morphology, syntax, and vocabulary occurred gradually and were unevenly distributed across the region between them. Literary standardization also poses a challenge to this history because the language preserved in texts is not necessarily the full range of the contemporary spoken varieties (Jain & Cardona, 2007; Masica, 1991). The most prominent representatives of old Indo-Aryan are the Vedic Sanskrit and the classical Sanskrit. A comparatively rich inflectional system is found in the Rigveda, with distinctions of case, number, gender, person, tense-aspect and mood, and the language of the Rigveda shows an early stage of Indo-Aryan. It retains features that are of significance for the reconstruction of Proto-Indo-Iranian and Proto-Indo-European language. Vedic Sanskrit also had a pitch accent system as well as grammatical distinctions, which were lost or reshaped in later Indo-Aryan dialects (Beekes, 2011; Fortson, 2010).

The Vedic-Sanskrit transition should not be taken to mean a mere chronological replacement of the one language with the other. Classical Sanskrit is a very formalised literary and intellectual language whose grammatical system came to be tightly linked to the Sanskrit grammatical tradition, especially that developed by Pāṇini. Cardona (1997) illustrates how complex Pāṇinian grammar was and how crucial it was to understanding the structure of Sanskrit. Standardisation played a role in maintaining Sanskrit as a great medium of scholarly learning, religion, philosophy, literature and administration for many centuries, while spoken Indo-Aryan languages evolved in many directions, not fully reflected in the classical literary standard.

This is an important distinction between literary Sanskrit and vernacular development that is important for understanding Middle Indo-Aryan. The Middle Indo-Aryan stage is a non-homogeneous group of varieties, rather than a single language. The Prakrits as found in inscriptions and literary traditions (as they are called in Indian traditions) Pāli, Gāndhārī, and then Apabhraṃśa are all evidence for linguistic changes after the Old Indo-Aryan period. Their relationship with Sanskrit was complicated; some of the Literary Middle Indo-Aryan languages were standardized, and there was a considerable diversity in the vernacular languages on which they were based (Liao, 2000; Jain & Cardona 2007; Masica 1991). The restructuring of Old Indo-Aryan phonology was one of the easiest changes to notice during the shift to Middle Indo-Aryan languages. Complex consonant clusters were often simplified and/or assimilated; intervocalic consonants had several changes; and inherited vowel and consonant patterns were re-organized. These changes were not uniform amongst the Middle Indo-Aryan varieties, and it is unhelpful to suggest that all of them underwent the same change in the same way. However, the overall profile was more and more different from the Vedic and Classical Sanskrit (Jain & Cardona, 2007) profile.

These developments can be better presented through the history of individual lexical elements than in a general description alone. For example, Sanskrit sapta is the equivalent of Pāli satta and finally to form like Hindi sāt. Likewise, Sanskrit hasta 'hand' corresponds historically to the Middle Indo-Aryan reflexes hattha, and the New Indo-Aryan reflexes hāth. The examples show some of the effects of consonantal assimilation and the subsequent phonological development that took place at different stages. These should not be taken as a sign that each form of the branch is a direct descendant of a single standardized Middle Indo-Aryan form, but as easily accessible illustrations of historical processes that recurred throughout the branch of Indo-Iranian (Masica, 1991; Turner, 1966).

The significance of the morphological change was also great. Old Indo-Aryan had a complex system of case marking for the nouns, with eight cases and three numbers, as well as a very rich verbal system. In Middle Indo-Aryan, a number of inherited differences became less distinct, merged or vanished. Later Indo-Aryan languages were, however, not grammatically poor due to the wear and tear of the inherited inflectional morphology. Instead, new combinations of nominal forms, particles, postpositions, participles and auxiliary constructions increasingly came to be used to express grammatical relationships (Butt, 2022; Masica, 1991). Case marking is an example of this process particularly useful for its development. The history of Indo-Aryan case is not so simple a transition from complex to simpler, as is suggested by Butt (2022). Most of the Old Indo-Aryan case system has been lost in Middle Indo-Aryan, but in New Indo-Aryan new case systems and postpositional languages have emerged. The outcome was thus not just a loss of form but a renewal and reorganization. Modern languages like Hindi-Urdu employ direct and oblique nominal forms, postpositions, and case markers that in terms of structure are distinct from the older system of inflections in Sanskrit, but which nonetheless share the same long historical continuum (Butt, 2022). This is seen also in the verbal system. Old Indo-Aryan had a well-developed system of finite verbal categories, while later Indo-Aryan began to use participial and auxiliary-based constructions more. Several tense-aspect constructions of the New Indo-Aryan languages evolved from grammatical resources found in Middle Indo-Aryan. The change to New Indo-Aryan, therefore, can be only understood as reorganization of the grammatical functions in terms of new analytic-periphrastic structures (Jain & Cardona, 2007; Masica, 1991). After centuries of peripheral differentiation and reworking of grammar, new Indo-Aryan was formed. It includes many languages spoken in northern, western, central and eastern South Asia and elsewhere. Some of the major literary languages are Hindi-Urdu, Bengali, Punjabi, Marathi, Gujarati, Sindhi, Nepali, Assamese, Odia, Sinhala and more, with a much wider continuum of regional languages and dialects. Their phonology and grammar are not only different but also their history of innovation, contact, standardization and sociopolitical development has been different, as well (Jain & Cardona, 2007; Masica, 1991).

This continuity via restructuring can, therefore, be understood more clearly between Old Indo-Aryan and Middle Indo-Aryan as well as between the Middle Indo-Aryan and the New Indo-Aryan. Certain inherited features remained in a relatively recognisable phonological or grammatical form, while others were formally transformed, and others were eliminated, and their functions were realised by new grammaticalised forms. In contrast to the older approach to Indo-Aryan history that mostly focused on "simplification," this pattern poses a challenge. Butt (2022) reports a decrease in some areas of the grammatical system, such as inflectional morphology, and an innovation and renewal in others.

Language features and structural change in the Indo-Aryan languages

The structural history of Indo-Aryan displays a gradual process of interaction of inherited material and innovations which developed at later phases of the branch. Old, Middle and New Indo-Aryan are very distinct languages in both their phonological and grammatical structure, but they do not mark any sharp discontinuities between separate systems. They show a long time in which inherited forms have been maintained, changed, eliminated or superseded by newly grammaticalized forms. Finally, because the study of phonology, morphology, and syntax

together offers greater insight into the nature of Indo-Aryan change while maintaining a genealogical perspective (Jain & Cardona, 2007; Masica, 1991), the chapters that follow examine various components of this system.

Old Indo-Aryan had a fairly intricate consonantal system, retaining distinctions between voiced and voiceless stops, between aspirated and unaspirated stops and between a variety of place of articulation. Retroflex consonants are a particularly significant feature of Indo-Aryan and several other South Asian languages as they differentiate them from many other Indo-European languages in the South Asian context. This has therefore made retroflexion the prominent feature of the discussion of the language region of South Asia. Emeneau (1956) interpreted the retroflex consonants as part of the larger picture of convergence between the genetically separate South Asian language families, and more recent research has demonstrated that the historical development of the retroflex consonants calls for a more cautious approach to chronology, geographical distribution and possible contact mechanisms (Hock & Bashir, 2016; Southworth, 2005).

Of course the development of retroflexion is especially instructive, showing how hard it is to distinguish the influence of contact from internal phonological change. The suggestion that Dravidian and other South Asian languages have exerted a strong influence on Indo-Aryan, causing retroflex consonants to be used more often than they were in Indo-European, has been invoked numerous times to explain the Indo-Aryan features. However, this does not imply that the whole Indo-Aryan retroflex system was nothing but borrowed from one external source. There are some areal environments that can be associated with Indo-Aryan developments, and there is the extensive South Asian distribution which indicates that older areal interaction also held an important place. Therefore, the evidence should be interpreted with an understanding of historically layered development and convergence rather than through a donor–recipient model (Hock & Bashir, 2016).

Other phonological changes are especially noticeable in the change from Old to Middle Indo-Aryan. Assimilation and/or reduction of complex consonant clusters was common. For example, Sanskrit sapta 'seven' is equivalent to Pāli satta and eventually to New Indo-Aryan forms like Hindi sāt. Sanskrit hasta 'hand' is etymologically connected with Middle Indo-Aryan forms like hattha and Hindi hāth. The examples show recurrence of consonantal assimilation and simplification processes, as well as further developments in single New Indo-Aryan languages (Masica, 1991; Turner, 1966). These changes were not uniform, but taken together, they helped further phonological separation between Middle and Old Indo-Aryan.

The restructuring of morphology was also a major process. Vedic Sanskrit retained a complex case, number and gender nominal system based on inflection. In Old Indo-Aryan the paradigm of noun words is characterized by eight cases: nominative, accusative, instrumental, dative, ablative, genitive, locative and vocative, and of three numbers: singular, dual, plural. The verbal morphology, too, was rich, with distinctions marked by person, number, tense-aspect, mood and voice using extensive inflectional paradigms (Beekes, 2011; Fortson, 2010). These features are very significant clues to the historical connection between Indo-Aryan and other Indo-European languages.

In Middle Indo-Aryan, however, there were several inherited morphological differences that were eliminated, combined, or rearranged. Dual was lost, differences in the nominal endings diminished, and the older case forms merged. However, when summarising these developments solely as a "morphological simplification" it is easy to mislead the reader in that regard, as the reduction of inherited inflection was paralleled by new ways of establishing grammatical relations. In the course of the developement of the new Indo-Aryan languages oblique nominal forms began to be used together with postpositions and other grammatical devices, so that the resulting systems were not necessarily less systematic but structurally different from the Old Indo-Aryan languages (Butt 2022; Masica 1991).

This process can be best seen in the historical development of case marking. Butt (2022) demonstrates the erosion of case morphology inherited from its Sanskrit predecessor and rejuvenation of case structure in Indo-Aryan. The grammatical relations in Hindi-Urdu are manifested in direct and oblique nominal forms (and other nominal forms) as well as grammatical markers like ko, se, and ne. Although they serve some similar grammatical and semantic functions, these structures cannot be identified with the inherited Sanskrit case endings. The emergence of their shows how grammatical categories can retain their function while the mechanisms employed for their expression undergo significant reorganization.

The same kind of change took place in verbal morphology. In Old Indo-Aryan, there was a large number of verb forms, whereas in Middle and New Indo-Aryan constructions with participles and auxiliaries were becoming more and more common. In some New Indo-Aryan languages, tense and aspect is marked not merely with the type of synthetic inflection typical of Old Indo-Aryan, but also with combinations of lexical verbs, participial forms and auxiliaries. This is part of a trend toward periphrastic expression as a whole, but verb elements remained part of the new constructions in this case (Jain & Cardona, 2007; Masica, 1991).

Syntax is more complex historically because the syntactic features of a language are less easily derived from the properties of its lexicon than the phonological or morphological properties. The Indo-Aryan languages generally have head-final tendencies, with subject-object-verb being the predominant order of operations, both historically and currently, throughout the language. New Indo-Aryan languages typically have postpositions and many have flexible constituent ordering that is affected by information structure, emphasis, and discourse factors. To refer to Indo-Aryan as "SOV" is to describe an important typological tendency, but does not exhaust the possible variation of actual usage, as Jain and Cardona (2007) and Masica (1991) have pointed out.

The syntactic evolution of case marking-agreement-aspect is one of the most significant changes in New Indo-Aryan. Split ergative is found in several languages, in which tense-aspect- or construction-related alignment differs. A familiar case is Hindi-Urdu, in which other arguments determine the agreement of the verb, but the perfective transitive constructions have ne on the agent. This is an illustration of how far the New Indo-Aryan system of grammatical construction has departed from the older system of nominative-accusative inflection (Butt, 2022; Jain & Cardona, 2007).

Many New Indo-Aryan languages have another characteristic feature: complex predicates. The aspect, completion, intensity, and/or point of view distinctions are provided by verb-verb combinations, and by constructions in which the main lexical element is combined with a semantically lighter verb. The constructions have been well documented due to their importance for understanding process of grammaticalization and lexical-semantic restructuring in the south Asian languages. Their ubiquity in the region also poses questions on the links between internal development and areal convergence, and this further suggests the need to avoid either inheritance or contact explanations (Hock & Bashir, 2016).

The most obvious proof of the various layers in Indo-Aryan is lexical. Many of the basic words are in fact of Indo-Iranian and Indo-European origin. Meanwhile, Indo-Aryan languages have gained words from long use with other South Asian languages, and more recently, politically, economically, literarily and religiously. Thus a modern Indo-Aryan language can have inherited elements, as well as Sanskrit elements and regional borrowings, and contact elements from Persian, Arabic,

Portuguese, English, or other contact traditions (Masica, 1991; Southworth, 2005).

The case of Hindi and Urdu is particularly telling of how lexical differences can emerge without an accompanying genealogical split of the underlying grammatical system. The two languages, standard Hindi and standard Urdu, have a large shared core of grammatical structures and common vocabulary, but also have historically different lexical sets for formal and literary uses. Sanskrit loan words are also widely used in the process of standardisation in Hindi, while formal Urdu has borrowed from the Persian and Arabic vocabulary traditions. The differences in the data illustrate the potential for significant lexical variation between closely related language varieties, as a result of differences in their sociocultural histories, literary practices, and language planning (Jain & Cardona, 2007; Masica, 1991).

It is helpful to consider these developments two ways: structurally, and formally. Changes in the form of a grammatical function can leave the meaning of that function unchanged. On the other hand, a linguistic form that has been inherited may remain and be used in a new grammatical or semantic context. The historical relations of Indo-Aryan stages cannot, therefore, be evaluated solely on the basis of the forms of Sanskrit which are preserved. It can only be reconstructed from the system of correspondences and the historical processes that led to the reorganization of forms and functions.

Areal Convergence and Language Contact

Genealogical descent from Indo-Iranian does not completely account for the historical development of Indo-Aryan. Throughout Indo-Aryan's history, beginnin from its earliest documented stages, it was a multilingual community in which speakers interacted with communities who spoke languages from a number of genealogically distinct families. The interactions were highly variable from region to region and period to period, and most pronounced in the development of the vocabulary and, more remotely, in the phonology, morphology, and syntax. The interaction between inherited linguistic structures and mutual linguistic convergence due to long-term contact is therefore a crucial context in which to look at the role of South Asia (Emeneau, 1956; Hock & Bashir, 2016; Southworth, 2005).

The concept of India as a linguistic area was a major conceptual pillar for this approach. This argument was significant because structural similarity between the languages of South Asia did not necessarily have to be due to common genealogical ancestry. Languages of the Indo-Aryan, Dravidian, Munda and Tibeto-Burman traditions have coexisted for long periods, which have led to situations in which linguistic features were able to spread beyond the confines of the family. The concept of a linguistic area moved away from the family-tree model and thus required a consideration of historical impact of multilingual interaction.

The idea of South Asia as a linguistic region does not, however, mean that all of its elements are of a similar origin or geographical distribution. Subsequent research has focussed on the fact that linguistic convergence can take place at varying levels and at different times and degrees of contact. While some are found throughout the subcontinent, others are only found in specific areas or contact zones. Hock and Bashir (2016) then see the linguistic structure of the South Asian region as the result of both genealogical diversity and longstanding interaction, and Southworth (2005) investigates the deeper historical ties among language communities in the subcontinent based on linguistic evidence.

The retroflexion, for instance, is one of the most talked about phenomena. Retroflex consonants are common in Indo-Aryan and Dravidian, and also found in other south Asian languages. It is these which have always been associated with the linguistic context in which Indo-Aryan speakers came in contact with South Asian languages (Emeneau, 1956). Contact with Dravidian has repeatedly been a part of this explanation and there have been other suggestions of possible substrate connections, such as with the Mundas. However, the evolution of retroflexion is not so straightforward as the unconditional assimilation of a whole phonological category from one language family to another.

Careful interpretation is therefore required. The expansion and distribution of retroflexion in Indo-Aryan is consistent with an Indo-Aryan history of areal convergence, but can be equally driven by internal Indo-Aryan processes. Contact and internal phonological change do not have to be mutually exclusive: contact may reinforce distinctions which are structurally possible, or serve to broaden their distribution. The Indo-Aryan evidence therefore underlines the methodological challenges in determining a single source for a feature that has been developed over the course of centuries in a multilingual linguistic area (Hock & Bashir, 2016; Southworth, 2005).

There is also speculation about the influence of Dravidian on grammar. Typological features that are shared by Indo-Aryan and Dravidian languages are: the head-final way of word order, postpositional or suffixal grammatical relations, and some convergent patterns in complex constructions. However, no structural similarity is conclusive proof of borrowing. Some features can be traced to independent historical sources and others might be enhanced by extended bilingualism. Because of this, any claims of grammatical convergence need evidence of historical periodicity, geographic spread, and the grammatical structure of the languages involved – not typological similarity (Hock & Bashir, 2016).

Another important dimension of the contact history between the Indo-Aryan and the Munda languages is the relationship between the two. The Austroasiatic group of languages, the Munda languages have been in prolonged contact with Indo-Aryan varieties in Eastern and Central India. Lexical evidence is especially valuable in illuminating the historical relationships that can be identified in both local environments and agriculture, as well as in terms of material culture and subsistence (Southworth 2005). The importance of such lexical layers lies in the fact that they can help to document interactions that may not have been directly recorded in historical texts.

The concept of a single South Asian linguistic area has been further complicated by regional studies. For example, Peterson (2017) investigates the contact zone of Jharkhand in which the Indo-Aryan and Munda languages have had extensive interaction. This is a regional approach which demonstrates that areal convergence is not uniform within the subcontinent. Linguistic features can be spread in a dense manner within a smaller geographical space, and neighbouring areas will have different historical paths. South Asia can then be considered not as a single linguistic region, but rather a series of "contact zones" with different histories and levels of convergence (Peterson 2017).

The contact between Tibeto-Burman languages is particularly important at the northern and Himalayan and north-eastern boundaries of Indo-Aryan speaking area. In these areas Indo-Aryan dialects have formed in parallel with the Tibeto-Burman branch of the Sino-Tibetan family of languages and have thus provided opportunities for lexical and morphological convergence. The impact of such contact varies across the regions and languages, and cannot be translated into Indo-Aryan in general. The regional variation echoes the larger theme that context matters for contact-induced developments—historical and geographical (Hock & Bashir, 2016). A separate layer of contact contact was created through interaction with Iranian languages and even more so from the medieval period forward, Persian. Persian was used as an administrative, cultural and literary language in significant portions of South Asia. It is particularly evident in the vocabulary of Indo-Aryan languages which were involved in significant cultural and administrative

networks with Persia. Urdu is perhaps the clearest case as many Indo-Aryan languages borrowed the Persian derived terms. The Arabic vocabulary has also been transmitted extensively, via Persian as well as via religious, scholarly and commercial contact (Jain & Cardona, 2007; Masica, 1991). The effects of Persian and Arabic contact are especially significant since they illustrate how lexical borrowing can be influenced by social history. The vocabulary that is borrowed often groups itself around a field, like administration, law, literature, religion, commerce and material culture. Such borrowing does not affect the genealogical classification of a language. For instance, although Urdu has a large percentage of Persian and Arabic loanwords, genealogical classification is not based on the source of a large part of the borrowed words, but rather upon the historical development of the Urdu core grammatical system and systematic correspondences (Jain & Cardona, 2007; Masica, 1991).

The relationship between Hindi and Urdu is a good example to highlight this difference. The standardized forms vary widely in script and in the choice of lexical resources in the highly formal registers, with Standard Hindi being more heavily dependent on Sanskrit-based words and Standard Urdu on Persian and Arabic resources. Concurrently, they have a considerable amount of grammatical structure and common vocabulary in common. Their history provides a good illustration of the ways in which linguistic varieties can become differentiated without their being genealogically differentiated, as a result of literary traditions, political history, standardization and the sociocultural identity (Masica, 1991). The influence of language contact, therefore, has been felt at various stages in the history of Indo-Aryan. The most obvious one is lexical borrowing; it is also the one in which the historical source of particular words is often easily determined comparatively. The phonological convergence might be more complex to evaluate since similar sound patterns can occur internally and in contact. Problems of grammatical convergence are even more carefully to be handled as typological similarity may not show either its historical origin or the direction of influence. It is important to note these differences in order to prevent the use of the word "contact" as a blanket explanation for any Indo-Aryan cases where it deviates from the reconstructed Indo-European ones.

The internal Diversity, and Classification of Indo-Aryan Languages

The division of Indo-Aryan languages is not as easy as a regional classification, but rather, there is a more complicated aspect. On the largest historical scale, the difference between Old Indo-Aryan and Middle Indo-Aryan and New Indo-Aryan offers a helpful chronological framework for the study of the linguistic development. The internal classification of the New Indo-Aryan languages is far less easy, however. Geographically related language groups are not necessarily genealogically distinct and in longstanding contact between neighbouring varieties can lead to language similarity that makes it difficult to distinguish shared innovations. This classification must therefore be based on historical factors, structural similarity, and geographical considerations (Jain & Cardona, 2007; Masica, 1991).

The earlier attempts to classify New Indo-Aryan were based on a combination of the phonological, morphological, geographical, and historical migration features. The separation of so called "Inner" and "Outer" Indo-Aryan languages was one such debate that was particularly influential. Variants of the hypothesis suggested a separate historical stage of development for some peripheral Indo-Aryan languages based on a different geographical location from languages in the more central region. The proposal sparked much discussion due to the fact that linguistic traits that support the identifications of the subgroups do not necessarily align neatly, and that some of the traits that describe the apparent subgroups may also be "recent contact and diffusion" (Masica, 1991). The problem is compounded by the fact that many of the Indo-Aryan speaking areas are in a dialect continuum. Linguistic divisions do not always conform to political divisions; commonly, political, linguistic, and literary divisions do not align. Varieties of the same language can have significant differences between each other, and varieties of different standardized languages can share phonological, grammatical and lexical properties. Therefore, the categorization of Indo-Aryan should consider not only the historical linguistic evidence, but also the processes of the sociolinguistic character that have made certain varieties of the language standardized and accepted as distinct languages (Jain & Cardona 2007; Masica 1991).

A broad geographical orientation can still be helpful for the description of New Indo-Aryan, but geographical labels should not be taken as necessarily secure genealogical branches. The north west Indo-Aryan zone encompasses varieties of Punjabi and Sindhi and is geopolitically of significance in terms of interaction history with Iranian and other north west languages. The central Indo-Aryan contains Hindi and a number of regional varieties of Hindi; western Indo-Aryan contains languages like Gujarati and Rajasthani varieties. Eastern Indo-Aryan languages are Bengali, Assamese, Odia and more. While Marathi and Konkani have a significant presence in the south-west, Sinhala reflects the larger Indo-Aryan tradition of Sri Lanka. These are the large geographical divisions for which a descriptive map is provided, but do not constitute a universally accepted family tree (Jain & Cardona, 2007; Masica, 1991). Kashmiri shows some of the challenges of applying easy geographies. It is an Indo-Aryan language of the north-western part of the subcontinent, but is often talked about in relation to the Dardic languages. But the status "Dardic" needs special attention. It has frequently been employed as an umbrella term for a group of northwestern Indo-Aryan languages, but it should not be taken for granted that it is a well-defined independent group running alongside Indo-Aryan and Iranian. It is generally accepted that the languages classically categorized as Dardic are part of the Indo-Aryan language family, but that they have special historical and areal features (Jain & Cardona, 2007; Masica, 1991). In the case of Sinhala, too, geographical position does not reflect the direct genealogical subgrouping. It is the historical path and the extensive contact with Dravidian languages, particularly Tamil, that have occurred in Sri Lanka that has created this distinct historical trajectory. Its geographical isolation from the Indo-Aryan speaking belt does not exclude it from the Indo-Aryan genealogy. It demonstrates that extensive geographical separation and migration can lead to significant structural differences in a language family, while leaving traces of common ancestry (Jain & Cardona, 2007; Masica, 1991).

There is another classification problem with reference to Hindi and Urdu. From a structural and historical linguistic point of view, their standard forms are quite similar and have a good degree of grammatical and colloquial base. This difference is particularly evident in the use of writing, literature, and lexical items in formal contexts. Standard Hindi is written in Devanagari, and is widely lexified in Sanskrit, while Standard Urdu is written in a Perso-Arabic script and has been influenced by Persian and Arabic extensively in formal vocabulary. Their case illustrates that the social-political notion of separate languages does not necessarily reflect significant genealogic separation (Jain & Cardona, 2007; Masica, 1991). There have been some old questions raised again, which are now being tackled with a quantitative approach. For instance, Cathcart (2020) reexamines some of the tenets of the Inner–Outer hypothesis through computational modelling of sound changes between Old Indo-Aryan and modern Indo-Aryan varieties. The analysis showed statistical support for a general dialect trend that follows a broad pattern of core-periphery, but it did not make the history of New

Indo-Aryan a two-divided affair. Instead, they illustrate the use of computational tools to test historical hypotheses, which were previously tested mostly on a qualitative basis. The evidence used for subgrouping is also undergoing changes on a large scale due to digital resources. Much larger sets of data are possible to compare than in pre-existing linguistic history, thanks to historical linguistic databases of lexical forms and cognate relationships from a large number of South Asian varieties. These methods could be useful for detecting patterns of co-innovation, lexical memory, geographical diffusion, etc., but quantitative cluster analysis still needs historical interpretation. When features are found to be statistically similar, it is not possible to say if this is because they are common descent, contact, parallel development or later diffusion.

Traditions and evidence from the past

Indo-Aryan importance in historical linguistics is one of the main reasons because of its very long textual history. There is evidence of each stage in the religious texts, in the grammatical works, in inscriptions, in manuscripts, in literary compositions, in administrative documents and in later printed documents. Written evidence is available for comparison over long time periods but this must be used with caution. Forms may be preserved in ways not directly related to the way they are spoken in the modern era, due to orthographic conventions, literary standardization, and scribal practices and the social status of particular languages (Jain & Cardona, 2007; Masica, 1991; Salomon, 1998).

Vedic Sanskrit is the earliest old Indo-Aryan tradition. The Vedic value of the texts is that they are transmitted with care, and preserve phonological, morphological, accentual, and lexical information of great value for comparative reconstruction. In this regard the Rigveda is a central document, as it is a source of evidence for an early stage of Indo-Aryan, the language which can be compared with Avestan and other ancient Indo-European languages. The Vedic tradition was first handed down through song, however. The documentation of Indo-Aryan languages thus starts not only with the invention of writing but also with highly developed techniques of oral documentation with which the archaic language material has been preserved over the generations (Fortson, 2010; Witzel, 1997). It is another significant relationship between the history of the language and the tradition of texts, the Sanskritization of the language.The next standardization of Sanskrit is another important relationship between the history of language and textual tradition. Pāṇini's analysis of Sanskrit is conventionally believed to be from the middle of the first millennium BCE and served as a very systematic description of Sanskrit, and later was used as a foundation for the Sanskrit grammatical tradition. This process of standardisation helped in the continuity of Sanskrit as language of intellectual, religious and literary production in the long run. However, a methodological problem arises from this success for historical linguistics: the fact that Sanskrit is still written by means of a standard system does not mean that the vernacular Indo-Aryan language was not structurally changed (Cardona, 1997). Inscriptions are thus of special value as epigraphic evidence, since they contain linguistic material which is not subject to the highly codified Sanskrit literary tradition. The inscriptions of Aśoka of the third century BCE are one of the most significant early written Middle Indo-Aryan documents. The distribution of these is quite extensive both geographically and linguistically, and they have an important role in the study of the diffusion of writing and regional differences in the Indo-Aryan system. Salomon (1998) also stresses that inscriptions are important for the reconstruction of the linguistic and cultural history of ancient India because they can give information about forms and regional patterns that may not be seen in literary sources. Brahmi is in the middle of Indo-Aryan writing history. It is well attested in the Aśokan inscriptions, and was adopted as the ancestor of many subsequent scripts in South Asia. Over the centuries, regional variations led to the evolution of a variety of scripts based on Brahmi, which are used to write many Indo-Aryan and non-Indo-Aryan languages. This is a large-scale historical tradition that culminates in the Devanagari script, as well as Bengali-Assamese, Gujarati, Odia and other modern scripts (Salomon 1998). The wide-ranging variation in writing systems in turn mirrors aspects of the graphical changes and the histories and processes of language standardization in different regions.

The other important source of early Indo-Aryan evidence is Kharosthi, especially in the NW region. Kharosthi was used from right to left and is historically known from the northwest part of South Asia, as well as neighbouring regions, and it is particularly significant for the recording of Gāndhārī. The manuscript and inscriptional material has greatly enriched the knowledge of the transmission of northwestern Middle Indo-Aryan and early Buddhist texts. The importance of such evidence is that it shows linguistic traditions which simply cannot be reconstructed from Sanskrit or from later literary Prakrits (Salomon, 1998). The textual evidence of the Middle Indo-Aryan is also very varied. Pāli was subsequently closely linked to the Theravāda Buddhist textual tradition, and the various Prakrits had significant functions in Jain, Buddhist, dramatic and secular literature. Later, in the literary production of Middle Indo-Aryan cultures, apabhraṃśa was to become important. The traditions reflect structural changes that occurred outside of Old Indo-Aryan, do not necessarily reflect ordinary speech, and cannot be straightforwardly interpreted as such. Many other literary forms of Middle Indo-Aryan languages were grammatically described and developed into literary conventions, similar to Sanskrit. Thus, the Indo-Aryan history of textual language/vernacular relationship was still complex (Jain & Cardona, 2007; Masica, 1991). With the advent of new Indo-Aryan literary strands, the link between language and script further complexified. There are multiple writing systems used in Modern Indo-Aryan languages and in some cases, various linguistic varieties are written using the same or similar scripts. In India, it is primarily written in Devanagari script (Hindi), Bengali script (Bengali), Gujarati script (Gujarati), Gurmukhi script (Punjab) and a Perso-Arabic derived Nastaliq script (Urdu). Sindhi also exists in the more than one script depending upon the geographic and historical context. The relationships show that the use of a script does not imply a shift in the linguistic family of a language, that is, a language does not change its linguistic family simply by using a script, historically associated with another linguistic or cultural tradition (Jain & Cardona, 2007).

Hindi and Urdu are a great example that can be used to illustrate this. While they use different scripts, and use a different literary and formal lexical tradition, the difference in script does not imply a difference in genealogical origins. The Devanagari script is used for Standard Hindi and visually links it with the literary and Sanskrit oriented culture, while Urdu culture is linked with the Persianate culture through the script which is a mixture of Urdu and Arabic. Such orthographic deviations were significant markers of literary and sociocultural differentiation and the grammatical structure of the underlying varieties remained largely shared (Masica, 1991). Historical sound change is also influenced by writing systems. Orthography can maintain differences when they've become weak in the spoken language, may not consistently represent differences in phonetics, or may represent spelling conventions from earlier periods. Historical linguists are thus not able to deduce pronunciation mechanically from written sources. For modern periods direct linguistic documentation is also required; for other periods textual evidence must be considered in conjunction with the grammar, metrical evidence, comparative correspondences, and manuscript variation. This is of special importance when comparing

languages which are written in different scripts or texts in which the writing system is well-standardized. The importance of the written evidence to Indo-Aryan history is thus not just that some of these writings are old but also that there are several different kinds which can be compared across time and space. Oral-Textual Vedic evidence for Old Indo-Aryan, Inscriptions regional Middle Indo-Aryan evidence, Literate Prakrits and Pāli for further historical varieties, later manuscripts and modern traditions of textual evidence for emergence and standardization of New Indo-Aryan languages. Each of the type of sources gives a different aspect of the language record; there are limitations for each. Therefore, it is not appropriate to read the history of Indo-Aryan as a prechosen path of Sanskrit → Prakrit → modern languages. Historically, the situation of standardized and vernacular varieties and the simultaneous existence of overlapping literary traditions, regional differences and multilingualism, as well as changes in script and sociolinguistic prestige were all present. These factors help written evidence to be most useful, and, when considering textual forms, by comparison of the text in relation to others, rather than as a direct representation of spoken language.

Discussion

The historical development of Indo-Aryan shows that it is possible to sustain a language despite changing structures. In all three groups of Indo-Aryan languages, there is a striking genealogical resemblance in the phonological, morphological and lexical structure, but the grammatical structure underwent significant change. The material examined in this study, therefore, merits an interpretation of Indo-Aryan history that hinges upon three interactive modes: genealogical inheritance, internal restructuring, and language contact. None of these processes alone can be responsible for the present diversity of the branch. The first research question was related to the Old, Middle and New Indo-Aryan languages and the relationship of language inheritance and structural innovation. There is significant continuity however, especially in the core vocabulary and the historically traceable grammatical forms, while at the same time, there is a lot of restructuring. Vedic Sanskrit has inherited from Proto-Indo-Iranian and ultimately from Proto-Indo-European a wealth of inflectional morphology and grammar that is useful for comparative reconstruction (Beekes 2011; Fortson 2010). Later Indo-Aryan versions did not just retain these structures in increasingly modified forms. Some differences went away, others combined, and new grammatical devices emerged to carry out former functions of inherited morphology (Butt, 2022; Masica, 1991). This is especially true for nominal morphology. The loss of the case system of the Old Indo-Aryan language during Middle Indo-Aryan may seem to be a step towards grammatical simplification. An interpretation of this kind, however, only reflects a limited aspect of the historical process. In the New Indo-Aryan languages, new case markers, postpositions, and oblique forms were developed, which provided new means of expressing grammatical and semantic relationships. History of case is thus an example of a cycle of loss and renewal, or of re-distribution of grammatical functions over new formal structures, respectively (Butt, 2022). It implies that a description that is only analytically increasing or that is only reductionist in terms of morphological complexity can mask the creative aspect of the process of grammatical change.

The same goes for verbal development. Old Indo-Aryan had a very sophisticated finite verbal morphology while the later phases gradually started to use participles, auxiliaries and periphrastic constructions. Some of these changes were the disappearance of inherited categories, but also new constructions with tense-aspect forms and new ways of marking the arguments. This shift from synthetic to more analytic strategies is then to be understood not as erosion, but as restructuring. This is aligned with the historical-linguistic fact that a grammatical system can re-establish distinctions in the system via grammaticalization and reanalysis while older morphological material is being eliminated (Jain & Cardona, 2007; Masica, 1991).

The second research question was the view of the evidence for continuity and restructuring, namely in the areas of phonology, morphology, syntax, and lexicon. Some of the clearest evidences of regular historical transformation are phonological developments like consonant-cluster assimilation. Relationships like Sanskrit sapta/Middle Indo-Aryan satta, or Middle Indo-Aryan satta/Sanskrit sapta, and Sanskrit hasta/Middle Indo-Aryan hattha/Middle Indo-Aryan hāth are good examples of the possibility of significant formal change without masking genealogical connections, provided intermediate developments and systematic correspondences are considered (Masica, 1991; Turner, 1966). Such examples underscore the need for the comparative method: with historical relationships, there is a need for repetitive patterns of correspondence, rather than a superficial similarity between isolated words.

Syntax, on the other hand, is more complicated. However, there are important continuities across Indo-Aryan, as well as innovations in alignment, agreement, complex predicates, and in auxiliary-based constructions in New Indo-Aryan languages. It is important to consider split-ergative patterns in languages like Hindi-Urdu, which demonstrate some great reanalysis of syntactic and morphological organization. It is not the grammatical structure itself that is preserved over the centuries in Indo-Aryan, but the changes that are apparent in the inherited linguistic material, which are part of successive linguistic systems (Butt, 2022; Jain & Cardona, 2007).

The third research question focused on Indo-Aryan development as a result of contact with the language. The evidence makes it clear that multilingual interaction is important, but also suggests the need to be careful in invoking contact. Emeneau's idea of India as a linguistic area is still basic for two reasons: first, it provides a reason for the similarities among the structurally different languages of South Asia that are genealogically unrelated; and, second, it was the first to provide such an explanation. This overall conclusion has been reinforced by subsequent scholarship, but complicated by the difficulty of establishing a single source or direction for each common attribute (Hock & Bashir, 2016, Southworth, 2005).

A case in point is retroflexion. Its distribution is uniform over Indo-Aryan, Dravidian and other South Asian languages, and contact has thus been a key factor in historical accounts of its prominence. However, not all retroflex development in Indo-Aryan needs an external explanation; nor is there a single donor language because this type of consonant is found in many languages. The proffered explanation is that phonological internal changes happened in a linguistic environment in which retroflexion was already structurally prominent. Contact can thus maintain, prolong and consolidate progress without necessarily being the source of each separate instance (Emeneau, 1956; Hock & Bashir, 2016).

This separation of change due to contact from change due to contact reinforcement has other applications besides phonology. The similarities in word order, complex predicates, case marking systems and other grammatical features between the various South Asian language families have sometimes been mistakenly interpreted as borrowing. However, no inferences of historic direction can be made from structural similarity. To show the grammatical changes caused by contact it is necessary to pay attention to chronology, geographical distribution, earlier stages of language, bilingualism, and plausible mechanisms of language diffusion. The most responsible conclusion in some cases is then that contact led to convergence without being able to be reconstructed precisely in which direction the influence was.

The regionality of contact also helps to debunk the notion of a 'whole South Asian region' as a single linguistic entity. Smaller

contact zones, such as those in Jharkhand, where convergence emerged in a specific historical and sociolinguistic context, are also significant because they offer insights into the process and factors involved in the emergence of new languages (Peterson, 2017). In the same manner, Indo-Aryan contact with Tibeto-Burman languages is of particular interest in the Himalayan and north-eastern parts of the subcontinent, whereas Dravidian contact has different histories and intensities in the central, western and southern parts of the subcontinent. The idea of the South Asian linguistic region is thus most fruitful if viewed as a process of convergence of multiple histories, rather than as a one-off, uniform diffusion process.

In the case of Persian and Arabic influence, a somewhat different situation of contact is observed because a significant amount of the interaction is documented historically. For centuries Persian was an important administrative, literary and cultural language in South Asia and has made significant contributions to the vocabulary of several Indo-Aryan languages. The Indo-Aryan language also borrowed from Arabic, for religious and scholarly purposes, and in many cases via the mediation of the Persian language. These lexical levels illustrate the possibility of vocabulary change owing to sociopolitical and cultural history without a change in genealogical affiliation. The Urdu language has a historical base in Indo-Aryan and similar systematic linguistic correspondences place it in Indo-Aryan (Jain & Cardona, 2007; Masica, 1991), and thus it has been classified as Indo-Aryan.

The fourth research question focused on the contribution of Indo-Aryan to the general understanding of genealogical descent, internal change, and areal convergence. These should not be viewed as rival historical interpretations, the evidence indicates. Genealogical inheritance is the reason for Indo-Aryan languages having systematic features that are traceable to Indo-Iranian and Indo-European. Internal change is the root of many of the reorganisations of inherited structures across the different historical periods. Contact discusses other types of lexical transfer and structural convergence that occur when a language is used for a period of time. But in real linguistic history, all of these processes often overlap. Variants of an inherited form can be subject to contact processes of internal phonological change, and they may also be reanalysed by later varieties. The material borrowed, on the other hand, may get completely absorbed into an Indo-Aryan grammatical structure and then take part in internally-motivated developments. In this sense, then, "inherited" and "borrowed" refer only to the historical origins of a linguistic feature, not to the entire history of it. The same can be inferred from the internal diversity of New Indo-Aryan. The presence of dialect continua and regional convergence contributes to the challenges of building universally accepted subgroups, as do incomplete historical evidence. Varieties from geographically contiguous areas may have aspects in common, and historically related languages may differ significantly depending on the conditions of contact. In other words, geographical, typological and genealogical classifications do not necessarily overlap (Jain & Cardona, 2007; Masica, 1991). Large-scale patterns may be identified using recent quantitative methods, but before such patterns can be used as evidence for the presence of a subgrouping, they need to be interpreted historically in light of the computational similarity.

Indo-Aryan has another dimension to this problem in the form of its documentary history. Equivalent kinds of linguistic evidence are not available in the Vedic texts, the Classical Sanskrit, inscriptions, Pāli, Prakrit, Gāndhārī, Apabraṃśa or the later New Indo-Aryan traditions. Some are very uniform literatures, others more regional. The continued use of Sanskrit on the page does not necessarily indicate that there was no change in the vernacular language. On the other hand, the boundaries of the literary traditions are not always congruent with those of the spoken languages. Such evidence from epigraphy and manuscripts can only be understood and used in the context of the text and sociolinguistic situation (Cardona, 1997; Salomon, 1998). Put together, these observations illustrate a potential way to look at Indo-Aryan history that is more productive, one of continuity and restructuring rather than of preservation and/or simplification. Indo-Aryan was genealogically continuous, experiencing a great deal of variation in phonological, inflectional, case marking, verbal organization, syntax, word inventory, and geographic distribution. Some of the changes were related more to the internal developments, others were more deeply connected to multilingual interaction, while others are best explained by the interaction between the two. It is this interpretation that is of special importance to historical linguistics because of the Indo-Aryan remains. The branch offers unusually rich documentation from which linguistic structures can be compared over long time spans. It shows that even with significant structural changes and when language contact takes place, the genealogical identity of the language can be clearly recovered. Indo-Aryan is more broadly an illustration of the shortcomings of purely family-tree-based or areally based models for the origin of the language family. A more detailed picture of the inheritance, restructuring, diversification and contact aspects of language history becomes available when these aspects are viewed as interacting.

Conclusion

Indo-Aryan provides a very rich historical record of how languages maintain genealogical continuity while undergoing significant structural change. All of these linguistic changes, from the Old Indo-Aryan to the spectrum of the Middle Indo-Aryan and the languages of the New Indo-Aryan, show no continuity in the past or simple linguistic simplifications. Rather, Indo-Aryan history has been marked by continuity by restructuring, where linguistic material inherited previously was subjected to a number of changes at the phonological, morphological, grammatical, lexical, and regional levels (Jain & Cardona 2007, Masica 1991). This is well seen through the transition that takes place across the major historical stages. While the OIA kept a complex inflectional system, as well as many features from Indo-IrA and Indo-IE, the MIA experienced significant phonological and morphological changes. The same trends were observed in New Indo-Aryan, where the loss of certain older inflectional differences was paralleled by the introduction of new grammatical features for case marking, the development of postpositional constructions, the use of auxiliaries to form verbal structures, and others. It is particularly revealing that the case-marking system developed historically, because it shows that a reduction in linguistic complexity does not necessarily results in the loss of case marking but rather can be accompanied by the renewal of the grammatical system (Butt, 2022).

Meanwhile Indo-Aryan evolved in one of the world's longest-standing multilingual settings. The contact with Dravidian, Munda and Austroasiatic languages, Tibeto-Burman languages, Iranian varieties and later with Persian/Arabic languages have made different contributions over time and space. It is useful to consider an area of South Asia as a linguistic space for the purpose of understanding structural convergence across genealogical boundaries (Emeneau, 1956; Hock & Bashir, 2016). The evidence looked at in this study does, however, indicate that caution must be applied in attributing individual features to contact. The borrowing of a structure cannot be confirmed by structural similarity alone, and borrowing can be supplemented by areal influence or even stimulated by it. Contact is thus best seen as a part of a larger historical story than as a general explanation of innovation in Indo-Aryan. Ironically, the intralinguistic variability of New Indo-Aryan adds to the difficulty of classification.The intralinguistic variability of New Indo-Aryan further highlights classification problems. The modern Indo-Aryan languages are the result of regions, dialects, literature, and

social factors becoming intertwined in a complex series of patterns of descent, differentiation, dialect contact, literary standardisation, and sociolinguistic change. That close geographical proximity does not always imply close genealogical subgrouping, and that structural similarity can be due to diffusion, not to shared innovation. As the history of Indo-Aryan involves genealogical, geographical, typological and sociolinguistic evidence, they have to be taken together. This historical interpretation is reinforced by the long documentary tradition of Indo-Aryan, but is demanding methodological caution. Different types of linguistic evidence remain in the vedic texts, Classical Sanskrit, inscriptions, pānīpañhīya traditions, gāndhārī materials, apabhraṃśa and more recently New Indo-Aryan sources. Their usefulness is not in providing a smooth and seamless view of spoken language, but in enabling the comparison of linguistic changes in different contexts and at different times in the same language (Cardona, 1997; Salomon, 1998). This difference in the text traditions is hence crucial for reconstruction of Indo-Aryan linguistic history and the emergence of the vernacular development. Overall, this review suggests that the evolution of Indo-Aryan languages can be explained by the interplay of genealogical inheritance, internal reorganization and language contact. Such processes are analytically separable but in a historical sense linked. Later varieties emerged on top of inherited structures, internal mechanisms reshaped phonological and grammatical systems, while multilingual interactions brought further layers of convergence, borrowing and regional variation. The Indo-Aryan is an interesting case for historical linguistics, in that it illustrates the genealogical identification of a language family in the face of significant structural change over a long period of time.

The study could be further enriched by the use of additional historical and regional contact studies, larger historical corpora and computational methods. Special care must be taken with less well-documented Indo-Aryan varieties, and at contact areas where the direction and timing of the linguistic influence are unclear. Such work may help to extend the field away from the dichotomy of inheritance and borrowing and towards a more historically meaningful explanation of how linguistic systems develop in multi-lingual communities.

References

Arora, A., Farris, A., Basu, S., & Kolichala, S. (2023). Jambu: A historical linguistic database for South Asian languages. In *Proceedings of the 20th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 68–77). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.sigmorphon-1.8

Beekes, R. S. P. (2011). *Comparative Indo-European linguistics: An introduction* (2nd ed., M. de Vaan, Rev.). John Benjamins Publishing Company.

Butt, M. (2022). Case markers in Indo-Aryan. *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.292

Cardona, G. (1997). *Pāṇini: A survey of research*. Motilal Banarsidass.

Cathcart, C. (2019a). Gaussian process models of sound change in Indo-Aryan dialectology. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (pp. 254–264). Association for Computational Linguistics.

Cathcart, C. (2019b). Toward a deep dialectological representation of Indo-Aryan. In *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 110–119). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-1411

Clackson, J. (2007). *Indo-European linguistics: An introduction*. Cambridge University Press.

Emeneau, M. B. (1956). India as a linguistic area. *Language, 32*(1), 3–16. https://doi.org/10.2307/410649

Fortson, B. W., IV. (2010). *Indo-European language and culture: An introduction* (2nd ed.). Wiley-Blackwell.

Hock, H. H., & Bashir, E. (Eds.). (2016). *The languages and linguistics of South Asia: A comprehensive guide*. De Gruyter Mouton. https://doi.org/10.1515/9783110423303

Jain, D., & Cardona, G. (Eds.). (2007). *The Indo-Aryan languages*. Routledge.

Masica, C. P. (1991). *The Indo-Aryan languages*. Cambridge University Press.

Peterson, J. (2017). Jharkhand as a 'linguistic area': Language contact between Indo-Aryan and Munda in eastern-central South Asia. In R. Hickey (Ed.), *The Cambridge handbook of areal linguistics* (pp. 551–574). Cambridge University Press. https://doi.org/10.1017/9781107279872.021

Salomon, R. (1998). *Indian epigraphy: A guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages*. Oxford University Press.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Southworth, F. C. (2005). *Linguistic archaeology of South Asia*. Routledge.

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Turner, R. L. (1966). *A comparative dictionary of the Indo-Aryan languages*. Oxford University Press.

Witzel, M. (1997). The development of the Vedic canon and its schools: The social and political milieu. In M. Witzel (Ed.), *Inside the texts, beyond the texts: New approaches to the study of the Vedas* (pp. 257–345). Department of Sanskrit and Indian Studies, Harvard University.