



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

A Study of the Relationship between the Pashto Language and Indo-Aryan Languages

Mohammad Usman Solizay Prof. Dr. Azizullah Jabarkhil***

Abstract:

Pashto is a historic language of the northeastern subgroup of the Aryan branch within the Indo-European language family. Based on its phonological, morphological, and syntactic features, it is directly descended from its parent Aryan language and is coexistent in origin with Avesta and Sanskrit. Within the Indo-European family, Pashto has well-maintained grammatical features of both Aryan and Indo-Aryan languages; therefore, it has at times been considered either Aryan or Indo-Aryan. Due to the presence of certain shared linguistic features with Indo-Aryan languages, some orientalists, local historians, researchers, and even some Pashtuns themselves have at various times classified Pashto among Semitic, Indo-Aryan, or Aryan languages. This study demonstrates that Pashto, as an independent earliest Aryan language, has directly grown from the Aryan parental language. Owing to geographical proximity and contact with Indo-Aryan languages, it has incorporated certain phonological and morphological features of those languages into its structure and has also borrowed a number of lexical items from them. Since this research is limited to examining the position and relationship of Pashto with Indo-Aryan languages, the findings expose that Pashto has assumed some phonetic characteristics and linguistic principles of Indo-Aryan languages and combined them into its own system. As a qualitative study, the collected data have been analyzed and interpreted through a comparative approach.

Keywords:

Aryans, Pashto, Indo-Aryan, Relationships, Historic language, Avesta and Sanskrit

Introduction

Currently, more than seven thousand languages are spoken worldwide, belonging to various language families. The Indo-European language family is one of the most significant and extensive in the world, encompassing around 500 living languages. Pashto is also one of the important living languages of the Aryan (Indo-Iranian) branch of the Indo-European family, spoken by nearly seventy million people across different regions of the world (Dost, 2023).

Within the Indo-European language family, there is no full consensus among both Pashtun and non-Pashtun scholars regarding the classification of Pashto. Some have considered Pashto a daughter of Indic (Sanskrit), while others have regarded it as a sister or daughter of Avesta. A few have even gone further, classifying Pashto among Semitic languages or considering it a remnant of Hebrew.

The primary focus of this study is to clarify the position and relationships of the Pashto language within Indo-Aryan languages. Although some research has previously been conducted on this topic, Pashto still exhibits phonological and morphological features of both Aryan and Indo-Aryan branches. Despite preserving certain characteristics of Indo-Aryan languages within its structure, Pashto, in terms of antiquity and origin, is believed to have descended from a proto-language from which both Avesta and Sanskrit also originated. This hypothetical and

* Associate Professor, Department of Pashto Language and Literature, Faculty of Education, Bamyan University, Afghanistan, ORCID: 0009-0002-9835-9646, usman.solizay@gmail.com, u.solizay@bu.edu.af

** Faculty of Language & Literature, Nangarhar University, Nangarhar Afghanistan, ajabarkhail7@gmail.com, aziz.jabarkhail@nu.edu.af



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

reconstructed language is referred to as Aryan (Proto-Aryan). Although no direct records of this language exist, research suggests that Pashto, along with its sister languages (Avestan and Sanskrit), emerged simultaneously from this same Aryan source.

The study demonstrates that while Pashto shares and preserves phonological, morphological, lexical, and grammatical features with Indo-Aryan languages, it also possesses distinct characteristics that differentiate it from them. This research aims to fill an existing gap and resolve ambiguities surrounding the classification and position of Pashto. While Pashto retains certain grammatical features of Indo-Aryan languages, it directly evolved from the Aryan proto-language alongside both Aryan and Indo-Aryan languages. This study is qualitative in nature, and the collected data have been analyzed and interpreted through a comparative approach.

Research Questions

1. What is the relationship between Pashto and Indo-Aryan languages?
2. Is Pashto an Indo-Aryan language?

Research Objectives

- To evaluate the relationship between Pashto and Indo-Aryan languages.
- To examine Pashto as an Indo-Aryan language.

Research Methodology

This study is qualitative in nature, and the collected data have been examined through comparative, descriptive, and analytical approaches. Relevant and essential materials for the research were gathered from consistent scholarly sources, including academic books, journals (libraries), and online databases. After collection and evaluation, the data were systematically analyzed and interpreted.

Sources related to the selected research topic were identified from both printed and electronic works, carefully noted, and organized as research materials. Priority was given to works that directly addressed the subject matter of the study. The selected topics were comparatively examined and critically analyzed in line with the research focus. Both primary and secondary sources were utilized in the analysis of the research data.

Research Problem

This study focuses on examining the position of the Pashto language and its relationship with Indo-Aryan languages. It aims to clarify which linguistic features Pashto shares with the Indo-Aryan branch of the Indo-European language family. Previously, both domestic and international researchers have conducted studies on the relationship between Pashto and Aryan languages, leading to differing conclusions: some have classified Pashto as Semitic, others as Indo-Aryan, and still others as an independent language.

This research seeks to explain why certain orientalists, local scholars, and even some Pashtuns themselves have associated Pashto with Semitic, Hebrew, or Indo-Aryan languages. The issue is of particular importance for scholars, linguists, and specialists in Pashto language and literature, as it addresses the fundamental question of Pashto's classification within the Indo-European family, especially its position within the Indo-Aryan branch.



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

The study further aims to determine the extent of phonological, morphological, and lexical similarities between Pashto and Indo-Aryan languages, as well as to identify which grammatical features Pashto has preserved from its parent and sister languages.

Significance and Need of the Study

Investigating the position and relationships of Pashto within Indo-Aryan languages is important to clarify its connection with the Indo-Aryan branch of the Indo-European language family. This study highlights the areas in which Pashto shares similarities with Indo-Aryan languages and the aspects in which it differs.

It also provides clear and scholarly responses to earlier misconceptions and inaccurate assumptions about the classification of Pashto. The findings of this research will offer valuable guidance and support to scholars, researchers, and students of Pashto language and literature in understanding its linguistic position and in conducting further studies. Moreover, the research will help Pashtuns and the broader community understand the linguistic affiliation of Pashto within the Indo-European family and identify its true origin. It will also provide accessible research materials and references for future scholars working on related topics. Finally, the study contributes to resolving longstanding ambiguities and debates that have led to the misclassification of Pashto among different language groups.

Research Background

Based on the nature and scope of the topic, no comprehensive and independent study has previously been conducted specifically under the selected research title. However, related aspects have been addressed in a number of works under similar themes, such as the position of Pashto within the classification of Aryan languages and its status among Aryan languages. A brief overview of these works is presented below:

The Position of Pashto in Aryan Languages: This work is a research project by Basira Shuaa, written for the completion of her Master's degree. It was published in 2015 by the National Movement of Afghanistan, consisting of five chapters and 200 pages. A significant portion of this study focuses on examining several Aryan languages within the Indo-European language family and exploring their relationships with Pashto. Its primary objective is to identify and analyze the position of Pashto among Aryan languages. However, it should be noted that this work does not address the relationship between Pashto and Indo-Aryan languages.

The Position of Pashto in the Classification of Aryan Languages: Authored by Sharifullah Dost and published in 2023 by the Publications Department of the Afghanistan Academy of Sciences, this work consists of four chapters and 257 pages. The study is generally devoted to the analysis and classification of Aryan languages. However, it does not specifically examine the position or relationship of Pashto with Indo-Aryan languages.

A Comparative Study of Indo-European Languages: This comprehensive work by Rafiullah Niazi was published in 2019 in six chapters and 429 pages by the Afghanistan Academy of Sciences. It covers topics such as the classification of Indo-European languages, the geographical distribution of their speakers, the languages within the Indo-European family, and the division of linguistic elements within the family. It also addresses the position of Pashto within the Indo-



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

European language family. While the book provides a general discussion of Indo-Aryan languages, it does not specifically analyze their grammatical relationships, similarities, or differences with Pashto.

The Relationship of Pashto with Ancient Languages: Written by Wajihullah Shpoon and published in 2010 in Nangarhar by Momand Publishing House, this work consists of four chapters and 145 pages. It examines the relationship of Pashto with Aryan and Indo-Aryan languages, the classification of Aryan languages, Pashto dialects, and the phonological system of Pashto. Overall, it is a significant contribution, particularly in the study of phonological relationships between Pashto and Aryan/Indo-Aryan languages.

Aryans

Before addressing the main topic, it is necessary to briefly review the Aryans, Aryan languages, and their linguistic classification. This will provide a foundation for a comprehensive discussion of the core subject.

The Aryans lived in a region known as Ariana Vaejah before the Common Era, where they developed an early civilization. They spoke the Aryan (Proto-Aryan) language, from which many major and minor eastern and western languages later emerged (Ghorbandi, 2019). Around 2500 years ago, the Aryans referred to themselves as *Arya*; the term “Arya” is derived from Sanskrit, meaning “noble” or “honorable.” They called their homeland *Ariana Vaejah* (or Ariana Varsha) (Shinwari, 2018, pp. 47–48).

When the Aryans migrated eastward (toward India) and westward (toward Persia), a group remained in their original homeland (Ariana Vaejah), continuing to identify themselves as Aryans and their language as Aryan. Since these languages originated from a common parent language (Proto-Aryan), which over time diversified into various dialects and languages, the term “Aryan languages” broadly refers to those languages belonging to the Aryan branch of the Indo-European language family. Today, Aryan languages are spoken across regions such as Afghanistan, India, Pakistan, Iran, Central Asia, Turkey, and areas surrounding the Caucasus Mountains (Zyar, 2015).

According to historical and linguistic accounts, Aryan tribes migrated from their original homeland—believed to be around the Ural Sea and the Black Sea—approximately two millennia before the Common Era, moving toward Central and South Asia. While many migrated, some groups remained in Central Asia. At present, many languages spoken across Central and South Asia are believed to have originated from this common Proto-Aryan source. All peoples whose languages derive from this parent language and who migrated from Ariana Vaejah identify themselves as Aryans. Consequently, terms such as Aryan, Indo-Aryan, and Aryanic are used to describe these languages (Shpoon, 2010, pp. 16–17).

Linguists estimate that the original Aryan languages date back to approximately 2500–3000 BCE. Aryan groups that migrated as far as the Indus River composed their renowned sacred text, the Vedas. It is therefore suggested that by around 1400 BCE, the Aryan languages had begun to diverge. In general, Aryan peoples are classified into Central (Bactrian), Western, and Eastern Aryans (Shinwari, 2018, pp. 6–7).



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

Aryan Languages

Aryan languages refer to those languages spoken in regions such as Afghanistan, Pakistan, Iran, and Central Asia. In contrast, Indo-Aryan languages are those spoken not only in these regions but also widely in India. In general, both Aryan and Indo-Aryan languages are considered to have developed from the Proto-Aryan (Aryk) linguistic base. Aryan languages emerged from ancient Aryan languages, and together with Indo-Aryan languages, they constitute the eastern branch of the Indo-European language family.

The geographical distribution of Aryan languages includes Central Asia, Iraq, the Caucasus region, and areas extending from the Pamir regions of Afghanistan to the Indus River. Aryan (Indo-Aryan and Iranian) peoples separated from other Indo-European groups during the second millennium BCE and migrated from their original homeland—believed to be around the Ural Sea and the Black Sea—toward Central Asia. One group settled in Central Asia, occupying regions known as Ariana and its surroundings, while another group migrated through Afghanistan approximately 3,000 years ago and settled in present-day Sindh and Punjab, identifying themselves as Aryans.

Today, many of the populations living in Afghanistan, Iran, Pakistan, India, and Central Asia are considered Aryan, and their languages are classified as Aryan languages. These languages are believed to have originated from a common parent language (Proto-Aryan) around 2500 BCE and subsequently diversified. The populations of these regions are often referred to as Aryan, Arian, or Aryanic. The Proto-Aryan language later developed into Indo-Aryan, Dardic, and Iranian language groups (Dost, 2023).

Aryan languages—often also referred to as Iranian languages—constitute a group of languages whose origin can be traced back to a single source, namely the ancient Aryan language. This ancient Aryan language, together with the language of the ancient Indo-Aryans, forms the eastern branch (Aryk) of the Indo-European language family. These languages are spoken throughout Afghanistan, extending to the Indus region and the Pamir highlands.

Historical evidence suggests that the original homeland of the Aryan tribes was in Balkh and the regions on both sides of the Hindu Kush. From this area, the Indo-Aryan tribes migrated eastward earlier (around 2000 BCE), while the Iranian Aryan tribes remained in Bactria for several more centuries. As a result, Bactrian languages show closer affinity to Iranian languages (such as Persian and Median) than to Indo-Aryan languages. Consequently, these groups are sometimes collectively referred to as Aryan, while the three branches are also known under the broader term Indo-Aryan in certain classifications (Shuaa, 2015, p. 172).

Pashto and Aryan Languages

Few languages have attracted as many complex and divergent views regarding their origin and identity as Pashto. Both local and foreign scholars have long debated fundamental questions such as: Who are the Pashtuns? What type of language is Pashto? To which language family does it belong, and from which source did it originate? Based on its phonological, morphological, and syntactic structures, with which languages does it share affinity? These are among the key questions this study seeks to address.

Aryan languages—often referred to by some linguists as Iranian languages—comprise a group of languages which, together with Indo-Aryan languages, form the eastern branch of the



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

Indo-European language family (Proto-Aryan). Unfortunately, no direct records of the Proto-Aryan language have survived; it is therefore considered a hypothetical and reconstructed parent language. The strong connection between Pashto and Proto-Aryan is further evidenced by the continued use in Pashto of terms such as *ār*, *aryan*, *arya*, *aryana*, and related forms.

Pashto belongs to the northeastern (Bactrian) subgroup of the Aryan (Indo-Iranian) branch of the Indo-European language family, and is often associated with the Saka (Scythian) group. The Aryans in their original homeland (*Ariana Vaejeh*) spoke a common language (Proto-Aryan), from which Indo-Aryan, Iranian, and Bactrian languages later developed. This language is believed to have been spoken around 3000–2500 BCE, although no direct evidence of it survives. Scholars generally agree that Sanskrit, Avestan, and Old Persian all originated from this Proto-Aryan source, and that by around 1400 BCE, these languages had diverged from their common parent (Zyar, 2015).

On the basis of several arguments, Pashto appears to be one of the closest living languages to this original parent language from which Sanskrit and Avestan emerged. First, many archaic lexical items from Proto-Aryan remain preserved in Pashto. Second, the Bactrian Aryans divided into three groups during their migrations: one group moved toward India, giving rise to Sanskrit; another migrated westward, and their language was later identified by orientalist as Old Persian; while a third group remained in Bactria (Balkh) and later migrated (c. 1300–1000 BCE) across the Hindu Kush into eastern mountainous regions and along the Indus River. These groups were historically referred to as *Bakhdi* (linked etymologically to *Pakht*, *Pasht*, *Pashtun*), and both their name and language have survived to the present.

Given that Afghanistan lies geographically between the historical centers of Sanskrit and Avestan, Pashto has retained linguistic features of both. For this reason, some scholars classify it among Eastern Aryan languages, while others associate it with Western Aryan languages. Nevertheless, Pashto should be regarded as an independent Aryan language (Shuaa, 2015, p. 130).

Although Pashto shares certain phonetic features and vocabulary with Avestan and Sanskrit, it differs from both in its morphological system and word formation processes. The Pashtuns, as an Aryan group, preserved many archaic features of their original language (Proto-Aryan), including grammatical distinctions such as gender (masculine and feminine), adjective–noun agreement, numeral–noun distinctions, subject–verb agreement, pronominal concord, and in some cases object–verb agreement.

Aryan and Indo-Aryan languages have historically been spoken mainly in Iran and India; hence, they are often collectively referred to as Indo-Iranian. Since these are languages of the Aryan peoples, they are also called Aryan languages. In other words, “Indo-Iranian” and “Aryan” are two terms used for the same linguistic branch. Linguists generally believe that Indo-European speakers split into two major groups: one migrated toward Europe and Russia, while the other settled in Iran and India. The languages spoken in India became known as Indo-Aryan, while those in Iran were termed Iranian. Because both groups originally shared a common language and homeland, they are collectively referred to as Aryan or Indo-Aryan languages (Alakozai, 2020, pp. 8–9).

All Aryan languages and dialects ultimately derive from a single parent language (Proto-Aryan). This means that Aryan languages—including Indo-Aryan and Dardic languages—



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

originated from a common root. Proto-Aryan constitutes the eastern branch of the Indo-European language family and forms part of its broader linguistic foundation. Due to the absence of written records of the original Aryan language, most related languages lack early documentary evidence. Over time, these dialects evolved through long historical processes into distinct and independent languages, undergoing such extensive changes that their speakers are no longer mutually intelligible (Shpoon, 2010, pp. 61–63).

The classification of Aryan languages is a complex and challenging issue. These languages and dialects form a unified branch that diverged from a single source (Proto-Aryan). In general, Aryan and Indo-Aryan languages together constitute the eastern branch of the Indo-European language family. From a grammatical and historical perspective, Aryan languages have undergone three major stages:

1. Old Aryan Languages (c. 1200–331 BCE): This stage includes languages such as Avestan, Old Persian, Median, Saka, Scythian, and Alanian. Among these, only Avestan and Old Persian have surviving written records, mainly in the form of inscriptions.

Avestan: A major ancient Eastern Aryan language, used around 600 BCE, and the language of Zoroastrian religious texts. Its corpus consists of multiple layers composed in different periods and regions.

Old Persian: The ancestor of Modern Persian, preserved primarily in Achaemenid inscriptions written in cuneiform script, dating from approximately 640–425 BCE.

2. Middle Aryan Languages (331 BCE–Islamic period): This stage includes languages such as Parthian and Middle Persian (Pahlavi) in the western group, and Sogdian, Khwarazmian, Bactrian, Khotanese, and others in the eastern group. Pahlavi was one of the most prominent languages of this period.

- **Western (Persic) Group:** Parthian and Middle Persian (Sasanian Pahlavi).
- **Eastern (Bactrian) Group:** Sogdian, Khwarazmian, Bactrian, and related languages.

3. Modern Aryan Languages (from the 8th century CE to present): Modern Aryan languages are spoken across a vast area from the Caucasus to the Pamir and from Oman to Central Asia. Most are written in the Arabic script, while some in former Soviet regions use the Cyrillic script. Major modern Aryan languages include:

- **Persian (Dari/New Persian):** A continuation of Middle Persian, which expanded significantly after the early Islamic period.
- **Pashto:** A northeastern Aryan language spoken in Afghanistan, Pakistan, Iran, India, and parts of Xinjiang (China).
- **Pamir Languages:** Including Wakhi, Shughni, Rushani, and others spoken in Badakhshan and Tajikistan (Shpoon, 2010, pp. 52–60).

Pashto as an Indo-European Language

Historically and geographically, Pashto and its speakers are associated with northeastern regions such as Balkh and the Pamir. The ancestors of the Pashtuns were known by names such as Saka, Sakae, and Kassites, and Pashto is considered a descendant of the Saka language. These groups established various local polities across regions such as Ghor, Herat, Helmand, Arghistan,



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

and the Suleiman mountain ranges, eventually extending into India and Kashmir—thus establishing contact with Indo-Aryan languages (Garcia & Munir, 2016).

Pashto has maintained strong phonological and morphological ties with Aryan languages and is classified within the eastern branch of this group. Despite migrations, Pashtuns preserved their language, and references to *Arya*, *Ariana*, and related terms continue to reflect their historical identity. After separating from Saka groups and settling in regions such as Sistan, Pashtuns developed Pashto as a distinct language from earlier Saka varieties. They are considered descendants of ancient Bactrian Aryans, who retained their original language while other groups, through contact with Iranian, Turanian, Mongolic, or Indo-Aryan peoples, gradually lost theirs. Historical accounts suggest that these Bactrian Aryans, known as *Pakht* around 1400 BCE, migrated from their original homeland (Bactria) toward the Indus region and eastern mountainous areas between 1400 and 1000 BCE, preserving both their ethnic identity and their language—Pashto—which remains in use to this day (Shinwari, 2018, pp. 7–8).

Indo-Aryan Languages

Following the broader discussion of Aryans and Aryan languages, this section focuses specifically on Indo-Aryan languages. Based on linguistic research, Indo-Aryan languages are generally divided into three historical stages:

1. Old Indo-Aryan Languages (c. 2500–500 BCE)

This period extends up to the compilation of the *Rigveda* (c. 1400 BCE), with Vedic literature forming its core. The term *Veda* (from Sanskrit) means “knowledge.” The two principal languages of this stage are Vedic and Sanskrit, both of which are well-documented.

- **Vedic:** The earliest known Indo-Aryan language, brought by the Aryans into the Indian subcontinent. It is preserved in religious poetic texts known as the *Vedas*, which include 1,017 hymns organized into four collections: *Rigveda*, *Samaveda*, *Yajurveda*, and *Atharvaveda*.
- **Sanskrit:** One of the most important ancient Aryan languages, preserving many archaic linguistic features. Its earliest grammatical description was composed by Panini in the 4th century BCE. Sanskrit functioned as the formal and scholarly language of ancient India and later developed into various Prakrit dialects.

Structural Features of Old Indo-Aryan Languages:

- Highly inflectional (rich morphological systems)
- Presence of vocalic /r/ and /l/
- Diphthongs such as /ai/ and /au/
- Complex consonant clusters (e.g., *kt*, *kr*, *tr*, *sm*)
- Three grammatical genders, three numbers, and eight cases
- Both inflectional and compound verb systems
- Multiple verbal moods (indicative, imperative, optative, etc.)
- Distinction of active, middle, and passive voices

2. Middle Indo-Aryan Languages (c. 500 BCE–1000 CE)

This stage includes the Prakrit languages and represents a transitional phase from highly inflectional to more analytical structures.



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

- **Pali and Prakrits:** Pali is considered the earliest preserved Middle Indo-Aryan language, especially associated with Buddhist texts. Prakrits (meaning “natural” or “vernacular”) were widely used in inscriptions, literature, and religious traditions (e.g., under Emperor Ashoka).
- **Apabhramśa:** A later development representing the transition from Prakrits to modern Indo-Aryan languages, used in literary traditions from around the 5th century CE.

Key Linguistic Changes:

- Loss of vocalic /r/ and /l/
- Simplification of diphthongs (/ai/, /au/ → /e/, /o/)
- Reduction of consonant clusters
- Loss of final consonants
- Increased phonological simplification
- Merging of grammatical cases (especially genitive and accusative)
- Development of compound verbs
- Greater emphasis on word order and syntactic structure

3. New (Modern) Indo-Aryan Languages (c. 1st century CE–present)

Modern Indo-Aryan languages emerged gradually toward the end of the early centuries CE, with clearer development by the 3rd century CE. This stage reflects further simplification and standardization.

Characteristics:

- Continued weakening or loss of short final vowels
- Simplification of vowel systems and formation of new diphthongs
- Development of analytic grammatical structures
- Emergence of new pluralization patterns
- Expansion of compound verb constructions
- Standardization of SOV (Subject–Object–Verb) word order

Scholars such as George Grierson classified modern Indo-Aryan languages into outer, intermediate, and inner groups, while Suniti Kumar Chatterji divided them into five groups: Northern, Central, Western, Eastern, and Southern.

Phonological Relationship Between Pashto and Indo-Aryan Languages

Certain phonological features in Pashto show clear connections with Indo-Aryan languages:

1. Aspirated Consonants Simplification:

Indo-Aryan languages retain aspirated consonants such as *ph*, *th*, *dh*, *jh*, *chh*, *kh*, *gh*, whereas Pashto simplifies them into unaspirated forms:

- Indo-Aryan: *ph*, *th*, *dh*, *jh*, *chh*, *kh*, *gh*
- Pashto: *p*, *t*, *d*, *j*, *ch*, *k*, *g*

2. Retroflex (Cerebral) Sounds:

Indo-Aryan languages have retroflex consonants such as *ṭh*, *ḍh*, *ṛh*, *ṇh*, while Pashto retains simplified forms:

- Indo-Aryan: *ṭh*, *ḍh*, *ṛh*, *ṇh*
- Pashto: *t*, *d*, *r*, *ṇ*



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

There is scholarly debate regarding the origin of these sounds. Some argue they were borrowed from Indo-Aryan into Pashto, others suggest the reverse, while a third view traces them to Dravidian languages and then into Indo-Aryan and Pashto. However, most research supports the view that Pashto borrowed these sounds from Indo-Aryan languages.

3. Retention of Aryan-Specific Sounds:

Pashto has preserved unique Aryan phonemes such as /ts/ (څ) and /dz/ (ځ), which are not typical of Indo-Aryan languages.

4. Shared Phonemes Across Aryan Languages:

Earlier assumptions that certain sounds (e.g., *p*, *ch*, *zh*, *g*) were exclusive to Persian have been disproven. These phonemes are now recognized as common features across various Aryan languages, including Pashto and Pamir languages (Dost, 2023).

Overall, while Pashto shares certain phonological features with Indo-Aryan languages, especially through historical contact and borrowing, it maintains its own distinct structural identity within the broader Aryan branch of the Indo-European language family.

General Relationship of Pashto with Indo-Aryan Languages

Within the Indo-European language family, Pashto is among those languages whose classification and status have long been debated by both Eastern and Western scholars. Some have associated it with Hebrew and Semitic languages; others have regarded it as a sister of Persian or Indo-Aryan languages. Certain scholars have considered it closely related to Zend (Avestan), Pahlavi, or Sanskrit, while others have viewed it as a distinct branch of the Proto-Aryan language.

Debates regarding the origin of Pashtuns and the Pashto language began in the 18th century and continued into the mid-19th century. Figures such as Nematullah Harawi, Abul Fazl, Akhund Darweza, Afzal Khan Khattak, and even some Persian-language scholars—along with foreign orientalists—either intentionally or unintentionally classified Pashtuns as (Bani Israel). However, modern linguistic and scientific research, including comparative linguistics and DNA studies conducted in recent decades, has rejected such claims. These findings confirm that Pashtuns are neither of Israelite origin nor is Pashto directly related to Semitic languages.

From a linguistic perspective, Pashto is not derived from any Indo-Aryan language; rather, it belongs to the Aryan (Indo-Iranian) branch and shows close affinity with Avestan. In other words, Pashto is an independent language within the Aryan branch of the Proto-Aryan division of the Indo-European language family, sharing features with many languages of this group (Niazi, 2019).

The French linguist James Darmesteter considered Pashto a daughter of Zend (Avestan) and argued, based on linguistic principles, that Pashto belongs to the Aryan branch, not the Indo-Aryan group, and is not derived from Sanskrit. Similarly, Abdul Hai Habibi concluded that Pashto is neither directly Indo-Aryan nor purely Persian but serves as a connecting link between the two, having evolved independently from ancient Bactrian or Proto-Aryan. Meanwhile, **Georg Morgenstierne** classified Pashto among the northeastern Aryan languages and suggested that it may have developed from dialects of the Saka language in the northeastern regions of Ariana.

From these perspectives, it can be concluded that Pashto has been influenced by Indo-Aryan languages at phonological and grammatical levels, yet it remains fundamentally an



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

independent language that evolved from the Proto-Aryan source alongside both Aryan and Indo-Aryan languages (Shuaa, 2015, pp. 128–129).

Why Pashto Has Been Considered Semitic, Indo-Aryan, or Aryan?

The divergence of opinions among scholars—both local and foreign—has led to multiple and sometimes conflicting classifications of Pashto.

- **Semitic Theory:** Some scholars associated Pashto with Semitic languages due to factors such as the influx of Arabic vocabulary after the advent of Islam, the migration of Jewish communities into regions such as Balkh, Jowzjan, and Herat, and certain historical narratives promoted in courts and literary traditions.
- **Indo-Aryan Theory:** Pashto has been linked to Indo-Aryan languages due to geographical proximity to India, historical political control between the regions, Aryan migrations into the Indian subcontinent, and long-standing cultural, economic, and linguistic exchanges.
- **Aryan (Indo-Iranian) Classification:** Since Pashto, Sanskrit, and Avestan share a common Proto-Aryan origin and once existed within a unified linguistic environment, Pashto is often seen as a bridge between Indo-Aryan and Iranian languages.

According to Niazi (2019), Pashto possesses features of both Avestan and Sanskrit; however, these are not merely borrowed elements but reflect shared inheritance from a common parent language. Thus, Pashto did not originate from Indo-Aryan or Iranian languages but developed alongside them from Proto-Aryan.

Comparative Analysis of Pashto and Indo-Aryan Languages

Pashto contains certain retroflex (cerebral) sounds not originally present in early Aryan languages; these are believed to have entered Pashto through contact with Indo-Aryan languages. However, the core phonological system of Pashto remains fundamentally Aryan. Some key distinctions and similarities include:

- Pashto preserves gender distinctions (masculine and feminine) in nouns and adjectives, a feature also present in ancient Aryan languages.
- Verb morphology in Pashto reflects complex agreement patterns, unlike the relatively simplified systems of many Indo-Aryan languages.
- Pashto retains archaic grammatical features such as case distinctions and inflectional patterns closer to Old Aryan languages.
- Prefixes and suffixes in Pashto show continuity with older Aryan structures but differ from Indo-Aryan formations.

Phonological Similarities and Differences

- Pashto retains many original Aryan roots in its vocabulary.
- Both Pashto and Indo-Aryan languages show inflectional variation in nouns, verbs, and pronouns.
- Diphthongs such as /ai/ and /au/ are preserved in both.
- Consonant clusters (e.g., *kr*, *tr*, *sm*) remain in Pashto but were simplified in Indo-Aryan languages over time.
- Pashto includes fricative sounds that are limited in Indo-Aryan languages.



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

- Both language groups maintain grammatical gender.
- Pashto has fewer case forms compared to Indo-Aryan languages.
- Both follow an SOV (Subject–Object–Verb) syntactic structure.
- Indo-Aryan aspirated sounds appear in simplified form in Pashto.

Pashto: Plar 'father', wror 'brother', aam 'mango', kila 'banana', narandz 'orange', limbu 'limon', seb 'apple', alu 'potato', palak 'Spinach', mələi 'radish', mrəf 'paper', bendəi 'okra'.

Sanskrit: petar, bratar, amra, kadli, narandza, nimbuka, sewa, aluka, palaka, mulika, mriṭam, bindika

Lexical Borrowing

Pashto has borrowed a number of words from Indo-Aryan languages, often with little or no phonological change. Examples include:

Hindi: saoda 'trade', bazar 'market', hesab 'calculation', kaṛki 'window', ghaṛi 'clock', ṭap 'print', dzaḡa 'war', dzaṇḡa 'flag', talai 'plate', ṭapu 'island', ṭakar 'collision', ḡar 'fear', ḡaṇḡa 'stick', ḡal 'shield'.

Pashto: sawda, bazar, hesab, kəṛkəi, gaṛəi, ṭap, dzaḡa, dzaṇḡa, talay, ṭapu, ṭakar, ḡar, ḡaṇḡa, ḡal. (Shpoon, 2010).

These shared lexical items reflect historical interaction rather than linguistic dependence.

Since both Pashto and Indo-Aryan languages originated from a common Proto-Aryan source, their phonological and lexical similarities are natural. These similarities do not imply that Pashto is a derivative of Indo-Aryan languages. In conclusion, Pashto is not an Indo-Aryan language. Rather, it is an independent Aryan language that directly evolved from Proto-Aryan, while maintaining historical, geographical, and cultural connections with Indo-Aryan languages.

Fundamental Differences

No.	Feature	Language		
		Pashto	Indo-Aryan	
1	Language family	Indo-European	Indo-European	
2	Linguistic branch	Aryan	Aryan (Indo-Aryan)	
3	Verbal system	Complex	Simple	
4	Historical origin	Aryan	Sanskrit	
5	Phonological system	Independent	Heritage of Sanskrit	
6	Retroflex consonants (t, d, r, ṇ)	Simple retroflex	Sibilant retroflex	ṭh, ḡh, ṛh, ṇh



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

No.	Feature	Language		
7	Aspiration	Absent	Present	bh, ph, th, ṭh, jh...
8	Voiced and voiceless affricates (j, ch, c, ʒ)	All present	c and ʒ absent	
9	Morphological system	Inflectional and derivational	Inflectional and derivational	
10	Gender	Inflected	Inflected	Tense and mood
11	Number	Inflected	Not inflected	ek laṛkā (sg.), ek laṛke (pl.)
12	Person	Inflected	Inflected	hū, ho, hai
13	Nominal cases	Limited	Extensive	
14	Verbal system	Complex	Simple/Complex	
15	Retroflex sounds	Present	Present	ṭ, ḍ, ṛ, ṇ
16	Consonant clusters	Initial, medial, final	Initial, medial	kr, tr, sm...
17	Plural forms (masculine/feminine)	Present	Present	ek laṛkī, do laṛkiyā
18	Word order	SOV	SOV	maī kitāb paṛh rahā hū
19	Writing system	Right to left	Various	
20	Semivowels	w and y	r, l, w, y (various)	
21	Gemination	Rare	Extensive	jaṅgeḍal, taṅgawal, raṅgawal...
22	Palatalization	Rare	Extensive	chāṅd, bāṅd
23	Elision	Present	Present	šu'ā' (šu'ā), iṭilā' (iṭilā)



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

The above table briefly illustrates the similarities and differences between Pashto and Indo-Aryan languages.

Discussion

Prior to this research, regarding the specified topic, the work titled "The Position of the Pashto Language in the Classification of Aryan Languages" explored the relationships of Pashto with various sister Aryan languages. According to the researcher, Pashto, as a contemporary Aryan language, belongs to the northeastern group of the Aryan branch. The course of Pashto's development and evolution is complex because it possesses, alongside Aryan languages, features of Indo-Aryan languages as well. Pashto is counted among the contemporary Aryan languages because no written records from its ancient or medieval periods are available.

At the same time, another important work is "The Status of the Pashto Language among Aryan Languages." This research work also extensively discusses the subject. The researcher concludes that the hypothetical mother language of Indo-European languages (Aric) was spoken in the region of Ariana before the migration and division of the Aryans, and after the Aryan migration, languages such as Avestan, Sanskrit, and Median emerged. Since the Pashtuns settled in the region between these two languages (Avestan and Sanskrit), features of both languages are observable in Pashto. This is why some linguists include it among the Eastern Aryan languages, while others consider it part of the Indo-Aryan languages; this is despite the fact that Pashto is an independent language.

Another extensive and arguably most significant work is "A Comparative Study of Indo-European Languages," whose final chapter discusses the aforementioned topic. According to this researcher, Pashto is neither among the Indic languages nor among the Semitic languages; rather, based on certain features, it shares some connections with Indo-Aryan languages, but overall, in terms of phonology, morphology, verbal system, and case system, it is much closer to Aryan languages. Thus, Pashto is contemporaneous with Avestan and Sanskrit and has evolved from a common mother Aryan language (Aric) after its emergence.

In the course of the above research and findings, when the results of this study were comparatively analyzed and explored alongside the aforementioned works, the following conclusion was reached: Since Pashto and Pashtuns, by geographical location, inhabit a region situated between the western (Aryan) branch and the eastern (Indo-Aryan) branch of the Indo-European language family, it is evident that their people would have had interaction with languages from both sides. They have had and continue to have economic, political, cultural, and various other connections. All these factors cause the Pashto language to be influenced by these languages in phonological, morphological, and syntactic domains. However, despite all these traces and signs, a number of lexical items and grammatical features from its original mother language (Aric) remain alive in Pashto. This itself demonstrates that Pashto is neither a language of the Aryan branch nor of the Indo-Aryan branch, but rather, alongside and contemporaneously with Aryan and Indo-Aryan languages, it is a direct offspring of the Aric language.

Recommendations

At the conclusion of this research, I present the following recommendations for future research and broader investigations to other experienced and professional researchers:



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

1. Pashto should be comparatively studied with Indo-Aryan languages in terms of syntactic structure, so that their proximity, depth, and differences are made clear to readers.
2. For the purpose of comparing and tracing sounds and phonetic features, the ancient texts and manuscripts of Pashto and Indo-Aryan languages should be evaluated in specialized professional experiments, in order to establish their phonetic independence and authenticity.
3. Specialized, professional, and comparative research should be conducted in the domain of Pashto, Hindi, Sanskrit, and Urdu languages to identify grammatical proximity, similarities, differences, and their historical development.

Conclusion

This research was specifically dedicated to investigating the position and relationships of the Pashto language among Indo-Aryan languages. Consequently, the findings of the study revealed that Pashto is an ancient language belonging to the northeastern group of the Aric branch of the Indo-European language family. One objective of this research was to clarify whether Pashto is affiliated with the Aryan branch of the Indo-European family or with the Indo-Aryan branch. The analysis and interpretation of the data collected in the research contents demonstrated that some phonological and morphological features of Pashto are similar to Aryan languages, while in other respects, it aligns with Indo-Aryan languages. This means that Pashto has accommodated within its structure both phonetic and morphological characteristics of the two language groups. The study also revealed that Pashto possesses a number of sounds and phonetic features distinct from both aforementioned language groups, yet traces of these features are observable in the daughter languages descended from the mother and original language (Aric).

The results of this study will assist language and linguistics enthusiasts, specialists, and scholars in identifying and clarifying the position of Pashto within the Indo-European language family. Furthermore, the documentation and findings of this research will be presented, demonstrating that due to these factors and reasons, Pashto is not an Indo-Aryan language, but rather diverged directly from its mother language (Aric) contemporaneously with its sister languages (Avestan and Sanskrit). The conclusion of this study will especially help Pashtuns and, more broadly, enthusiasts and researchers of the Pashto language, so that in the future, they do not repeat previously settled and stereotypical conceptions in their writings and research. Instead, they will assert without hesitation that Pashto is a direct offspring of the Aric language and not a language reconstructed from the combination and arrangement of other languages. Although this study is neither the first nor the only one in this field—since some research has been conducted previously to establish Pashto's connection with Aryan languages—it represents the first comprehensive effort of its kind to investigate the relationships between Pashto and Indo-Aryan languages. Therefore, the definitive stance and distinction of this research from earlier studies is that, based on a range of grammatical (phonological and morphological) features, Pashto does not belong to any other Aryan or Indo-Aryan language. Rather, based on existing documents and evidence, it is considered a language born simultaneously with Aryan and Indo-Aryan languages and from a single mother language.



SRAI JOURNAL OF LINGUISTICS AND SOCIAL SCIENCES

In accordance with research principles, this study first discussed the Aryans, their original homeland, and their settlements. Subsequently, a general overview of Aryan languages was provided. A comprehensive discussion was held regarding Pashto's connection with Aryan languages, along with a brief examination of a set of features of ancient, medieval, and contemporary Aryan languages and their relationships with Pashto. Following that, the study discussed Indo-Aryan languages and shed light on the linguistic features of ancient, medieval, and contemporary Indo-Aryan periods, thereby highlighting the similarities and differences with the Pashto language. In this context, it became clear which sounds Pashto has borrowed from Indic, which sounds it shares with Indic, and how and which Indic sounds are used in Pashto. Subsequently, the position of Pashto within the Indo-European language family was briefly addressed, and the position and relationships of Pashto among Indo-Aryan languages were exhaustively discussed. At the end of the study, a brief overview of various perspectives of different domestic and foreign researchers regarding Pashto was provided.

References

- Alkozi, M. F. (2020). *Indo-Aryan Languages*. Jalalabad: Hashemi Publishing Association.
- Dost, Sharifullah. (2023). *The Position of the Pashto Language in the Classification of Aryan Languages*. Kabul: Academy of Sciences of Afghanistan, Department of Information and Public Relations.
- Ziar, M.A. (2015). *Pashto and Pashtuns in the Light of Linguistics (with a Historical Overview)*. Kabul: Danish Publishing Association.
- Ghurbandi, M. (2019). *The History of the Pashto Language*. Kabul: Jahan-e-Danish Publishing Association.
- Shapoon, W. (2007). *The Connections of the Pashto Language with Ancient Languages*. Nangarhar: Momand Publishing Association.
- Shua, B. (2015). *The Status of the Pashto Language among Aryan Languages*. Kabul: Cultural Department of the National Movement of Afghanistan.
- Shinwari, M. S. (2018). *The Historical Evolution of the Pashto Language*. Nangarhar: Classic Publishing Association.
- Niazi, R. (2019). *A Comparative Study of Indo-European Languages*. Kabul: Academy of Sciences of Afghanistan, Department of Information and Public Relations.
- Klein, J. Joseph, B. Fritz M. & Wenthe, M. (2017). *Handbook of Comparative and Historical Indo-European Linguistics*. Walter de Gruyter; Berlin.
- Garcia, M.I.M & Munir, B. (2016). *Origins of the Pashto Language and Phases of its Literary Evolution*. Journal of Research (Humanities).
- The Indo-European Family of Languages: Evolution and Locale Maps. By: *Jonathan Slocum*
- Zograph. G.A. (2022). *Languages of South Asia A Guide*. Second Edition. Routledge Tylor and Francis Group. New York.
- Cardona. G and Jain d. (2007). *The Indo-Aryan Languages*. Routledge Taylor & Francis Croup. New York.
- http://mnabievart.com/oldwebsite/old/info_mn/article_6.html

