

Proto Forms in Pashto Phonology and the Puzzle of Pashtun-Pathān

Farooq Babrakzai ¹

Abstract:

This paper examines three previous views about the origin of the word Pashtun and its relation to the Indic word Pathan. These include earlier oral traditions according to which the title Pathan was given to Qais Abdul Rashid, the putative ancestor of Pashuns; the derivation of Pashtun through folk etymology from Baxdi, the earlier name of Balkh, the historic city in northern Afghanistan; and through common developments in Persian and Waziri Pashto. However, these views are based on a single word and focus on the cultural and historical identity of Pashtuns, but do not reflect the linguistic realities of Pashto / Persian and the related Aryan languages. This paper introduces proto forms in Pashto, presents data from Persian and Pashto as evidence to systematically show how all the related words Pashtun, Pathān, Paktiā, Paktikā, and Pakhtā developed from the single source word paktā- which was a place name south of the Hindukush Mountain range.

Keywords: Pashtun, Pathan, Proto Forms in Pashto

Introduction

Scholars in Afghanistan, Pakistan, and elsewhere, have attempted for decades to explain the origin of the word paštun and the related word paṭhān in Indic languages, but so far have not provided any logical and satisfactory explanation. These scholars quote many historical sources in which these names have appeared and consider this as sufficient evidence for the existence of Pashun and Pashto in the history of the region.

Other related words are paktyā and paktikā, names of two adjacent provinces in southern Afghanistan, and pakhtā the name of an ethnic group mentioned in the list of Aryan tribes in Vedic literature. All these have historically originated from the same source and are related terms but have developed different meanings in the course of history.

One point that needs clarification is that toponyms, that is, place names recorded in history are older and more stable than the ethnic names of the people living there. Inhabitants living in a region are generally referred to by the name of the location, not the other way around. For example, Persian from the region of pārs- whose Arabized form is fārs-, and Russian from Russia. Thus, paktā was the original name of the region in south-eastern Afghanistan, where Pashtuns lived and where their historical identity first developed.

Broadly, there are three different lines of investigation into the origin of the word Pashtun:

1. Oral traditions and myths that Pashtuns have accepted as historical facts.

¹ Defense Language Institute

2. Folk etymology in the 20th Century as evidence of shared linguistic relations and history.
3. Earlier attempts by European linguists through comparative method and internal reconstruction.

This paper briefly examines each view and then presents data as evidence of the regular phonological developments in the history of Iranian languages, Pashto and Persian, and Indo-Aryan languages. Data is also presented to shed light on the origin of Pashtun and Pashto, and other related words, such as *pakt-ik-ā*.

Oral Traditions and Myths

The earliest recorded perspective on the origins of Pashtuns is through genealogies, based on oral traditions and myths, as the descendants of the lost tribes of the Israelites. This view was first documented in the early seventeenth century at the court of the Moghul Emperor Jahangir. The original author, Khwaja Neamatullah Haravi (1612), wrote the history in Persian because it was the language of the court and administration during the Moghul period. In other words, the view that Pashtuns are descendants of Israelites was not indigenous but was given to them by non-Pashtuns at the Moghul court in Delhi. By the 17th Century Pashtuns had been Muslims for more than a few hundred years. All traces of the Buddhist and Hindu traditions had been lost except for the relics and artifacts in historical ruins. Persian was the dominant language of scholarship and Arabic was the language of religious practices.

When the new religion, Islam, penetrated the lands of Pashtuns, the earlier oral genealogies were naturally changed and re-aligned with notable historical figures in Islamic history. According to this view, in the early Islamic period, Prophet Mohammad predicted that the offspring of the putative ancestor of Pashtuns, Qais Abdul Rashid, would be numerous, and in “strength, be like the keel when constructing a ship, which wood the seamen call Pathan: on this account he conferred upon Abd Ullashid the title of Pathan also. Among them arose derveeshes, devotees, and saints, excelling both in deed and speech” (Neamet Ullah, 1612 (1836), 38). This view is culturally popular as a legend, but does not provide any concrete evidence, as *pathan* is not an Arabic word, though many people in Afghanistan and Pakistan still believe it.

Tegey (2003, 4) writes that Prophet Mohammad bestowed the title of *batān* on Qais Abdul Rashid, which eventually became *pathān*. But ironically, the word *pathān* is common east of the Indus River, in Indic languages, and not used by those in the Pashtun heartland.

Hayat-Khan (1999/1874, 54) writes “the most probable explanation of ‘Pathan’ is that it is merely a Hindi corruption of ‘Pashtun.’” Hayat-Khan, like others, pointed to the common view, but did not explain how Pashtun developed into Pathan.

Folk Etymology

A second view emerged in the 20th Century, which claimed that Pashtuns were the indigenous Aryans of the Balkh region in northern Afghanistan. Inspiration for this came from Iran which changed its name from Pars/Fars to Iran in 1935. Iran was the historic name associated with the

Aryan tribes in the western region. Soon afterward, Afghan scholar, Habibi (2005 (1944)) used folk etymology as evidence to show that terms *paštun* and *pašto* were derived from *baχdi*, an earlier name of the city of Balkh in northern Afghanistan. He claimed that Pashto was a contemporary with Sanskrit and Avesta. He listed numerous cognates, copied from the works of European linguists, to show the genetic relations between Pashto and Sanskrit and between Pashto and Avesta as evidence. He quoted European sources which stated that the Central Asian Aryan people who migrated south-east toward India developed Sanskrit as their language and those who migrated south-west developed Old Persian and then Pahlavi language. But those who stayed in the Balkh region spoke Pashto, thus giving Pashtuns a unique identity. The assumption was that northern Afghanistan, particularly the Balkh region, was the original home of the Aryan people, and that the Avesta texts were composed in Balkh. Habibi used folk etymology to derive the term *paχtun* / *paštun* from *baχdi* → *paxt-i* → *paχto* and *paχtun* (Habibi 2005:191).

Kaka-Khel (1999:71) followed a similar line of argument. He stated that Vedic *pakht* and the Avesta *baχd* is one word. The phonetic change took place like this: *p* → *b*, *k* → *χ*, *t* → *d*. Thus, the Vedic *pakht* changed to Avesta *baχd*. However, at the same time, Kaka-Khel writes that, “Avesta was written between 1000 – 600 BC and there is no explicit mention of Pashtun in the text.” (ibid).

Linguistically, the word *baχd* is related to the word *Bākhtar*, Greek Bactria, name of the region north of the Hindukush range in history and has no connection with the word *paχtun* / *paštun*, which originated from *paktā* the name of a region south of the Hindukush range. The modern name of the region Balkh is the metathesized form of *baχd* → *badχ* → *balχ*. Here *badχ* is the proto form of *balχ*, where *d* → *l*.

Another view is that word Pashtun developed from the Persian *pusht-* or *pushta* ‘knoll, hill’, referring to Pashtuns as ‘people of the mountain’. Rond-Miakhel (2008, 190) writes that Pashtuns had migrated from the west to the region of Alburz Mountains in north-eastern Iran. When they migrated further east, their name developed from *pušt-* → *puštyā* → *paštīn* which later developed into *paštun* and *paštānə*. While this may be a make-believe theory, it defies the linguistic realities of the Iranian languages, particularly that of Pashto.

These views are popular in cultural contexts and self-assuring to Pashtuns that they are an ancient and important people who have been mentioned in the history of the region, but from the perspective of historical linguistics, these views do not constitute factual knowledge.

Earlier Linguistic Attempts

Morgenstierne (1975) attempted to derive the word *Pašto* from **parsawā*, following the phonetic change of *rs* > *št*, based on the Persian infinitive *pursidan* and the Waziri Pashto *puštəl* ‘to ask’, thus indirectly claiming that the earlier form of *Paštun* was **Pars(u)wāna*.

He states, “Phonetically *Pašto* would go back to **Pars(u)wāna*, a feminine form of the adjective ‘the Parsu’ (language). The primary stem **Parsu* is preserved in *Pušt*, and *Paštun* is derived from *Pars(u)wāna*. The change of *rs* > *št* must be relatively early, since a transitional from **Paštān*, has been introduced into Indian border-dialects of an archaic type in the form **Paštāna*, from which modern Indo-Aryan *Paṭhān* is derived” (1975:168-174).

This analysis sounds plausible but is rejected here because it is based on the phonetic change in two words, which cannot be generalized to a set of data to show shared developments in Persian and Pashto and then in Aryan languages.

Linguistic Reconstruction

The Pashtun – Pathan question is primarily a linguistic issue but historians have changed that into of historical and cultural identity. This paper shows the Pashtun – Pathan connection by providing evidence through phonological reconstruction. To solve the puzzle comprehensively, other historic terms like paktyā, paktikā and pakhtā are also explained, since all these have developed from a single source in different periods of history and contexts. Each change is supported by a set of data which sheds light on the internal phonological structure of Pashto, Pashto-Persian relations, and furthermore, the historical developments in Indo-Iranian languages.

- Similarities between Verb Stems of Pashto and Persian

Persian and Pashto are the two major languages in the Iranian group. Persian is classified as West Iranian and Pashto as East Iranian. Further to the south-east of Pashto, are the Aryan or Indic languages such as Nuristani, Pashai, and Hindko. Pashto exhibits many linguistic features of Indic languages, such as retroflex consonants, oblique forms of words, and ergativity in the past tense, which Persian does not.

The first similarity is in the verb stems of Pashto and Persian. Both languages show a phonetic contrast in the past and present stems of the verbs. The past stems are either augmented or show a phonetic change.

-- Persian	Present Stem	Past Stem	Gloss
	dār-	dāšt-	have
	rav-	raft-	go
	χor-	χord-	eat
	soz-	soxt-	burn
	band-	bast-	tie
	gir-	grift-	capture
	kon-	kard-	do
-- Pashto	lwəl-	lwəst-	read
	mɾ-	məɾ-	die
	kaw-	kɾ-	do
	χwɾ-	χwəɾ-	eat
	ğwāɾ-	ğwəxt-	want
	boz-	bot-	take away (animate)

The relevant point in the data is that if final consonant is a [stop], the preceding consonant is always a fricative, for example, in words like bast- and ġwəxt-. (See 2.3)

- Proto Forms in Pashto

A proto form is a hypothetical, an earlier intermediate form, of a modern word. Proto forms tell us what earlier shapes of words might have existed at some point and let us see the general historical trends in language. Proto forms also help generalize a rule to large set of data rather than a single word.

-- Paštān - Paṭhān

Indo-Iranian languages exhibit several shared developments, that is, systematic phonetic correspondences, on the one hand between the Iranian languages, Pashto and/or Persian, and on the other, the Aryan languages, Hindi and Punjabi. The development relevant here is the voiceless palatal fricative *š* in Iranian which corresponds to the voiceless retroflex alveolar stop *ṭ* or to its aspirated form *ṭh* in Indic languages.

Persian	Pashto	Indic	Gloss
hašt	atə	āṭh	eight
šutor	uš	ūṭ	camel
angušt	guta	anguṭhā	finger/thumb
χišť	χə št-	īṭ	brick
pušt		piṭh	back
ušt (Parachi)		uṭh	rise
mušt		muṭhi	fist, grip
	lašta	lāṭhi	stick
	paštān	paṭhān	Pashtun

Here paštān is the proto form of modern paštun. The vowel raising *ā* → *u* before a nasal consonant *n/m* is a later development in Iranian languages, explained in section 2.4 below.

The development of *št* → *ṭh* in Aryan languages occurred during the transition from Sanskrit to various Prakrits. In Sanskrit, in Panini's time (\pm 400 BC), the phonological split had not taken place. Panini's classical work on Sanskrit was named *aṣṭādhyāyī* 'eight chapters', when 'eight' was *aṣṭā*, Persian *hašt*. In the transitional period, the Sanskrit, *ṣṭ-* developed into Prakrit *-ṭh-* (Gray 1902, p. 7). Other examples of this change are (Gray 1902, 215):

Sanskrit	Prakrit	Gloss
uṣṭra	uṭha	camel
prṣṭha	paṭhi	back
oṣṭha	oṭha	lip

It is therefore safe to conclude that in Panini's time the word paṭhān was not in existence, and neither was the word paštun, both are later developments from paštān.

- Vowel Change in Pashto *ā* → *o*

Pashto and Persian share a large of number of cognates. It is assumed here that the proto forms of many Pashto words were closely identical with words in Avesta, Pahlavi and/or Classical Persian, but they show regular vowel shift in contemporary Pashto. The Avesta, Pahlavi / Classical Persian *ā* changes to *o* in modern Pashto:

Persian	Pashto	English
āb	obə	water
tār	tor	dark
zāl	zoṛ	old

Other proto forms go through an intermediate stage where a consonant or a segment of the original syllable is lost and as a result, the preceding vowel is lengthened, called compensatory lengthening. “The term compensatory lengthening (CL) refers to a set of phonological phenomena wherein the disappearance of one element of a representation is accompanied by a corresponding lengthening of another element” (Kavitskaya, 2001). When the vowel is lengthened, it then follows the rule $\bar{a} \rightarrow o$.

The Pashto proto forms posited below are hypothetical, that is, there is no recorded evidence that they existed at some period. However, when these are generalized, they correctly predict many modern Pashto words.

Avesta	Persian	Proto-Pashto	Modern Pashto	Gloss
(ātarš)	ātaš	ār	or	fire
	ārd	ārd	oṛə	flour
	χwāhar	χwār	χor	sister
	sard	sārd	soṛ	cold
	berādar	wrār	wror	brother
	mādar	mār	mor	mother

- ā → u / _ nasal consonant

The third type of vowel change is conditioned by the presence of a nasal consonant n / m. If the long back vowel ā is followed by a nasal consonant, it is raised to u in modern Pashto.

Proto Form	Pashto	Gloss
paštān	paštun	Pashtun
dām	lum	trap
nām	num	name
namāz	lmunz	prayer
χāna	χuna	house
chopān	špun	shepherd
-tān / dān	-tun	location (pohan-tun)

It is this vowel raising that resulted in paštun from its proto form paštān. The same changed is seen in other important names in Pashtun history. The two sons of Sarban, who was the son of Qais Abdul Rashid, the eponymous ancestor of Pashtuns, are mentioned in the genealogies as Kharshbun and Sharakhbun. According to the phonetic rule $\bar{a} \rightarrow u / _ _ n/m$, these names were originally Kharshbān and Sharakhbān, but by the time Pashtun genealogies were first recorded in the early 17th Century, the endings of these names had already changed from -bān to -bun, parallel to paštān → paštun.

In modern spoken Iranian Persian, the long *ā* regularly changes to *u* when followed by a nasal consonant *n* /*m*.

Standard	Spoken	Gloss
tehrān	tehrun	Tehran
irān	irun	Iran
nān	nun	bread
kodām	kodum	which?

- stop → fricative / _ stop#

The next relevant point here is how *paštān* developed from the root *pakt-*. This is explained through two phonetic rules in Pashto. The first is that the suffix *-ān* is used to designate a person's affiliation with a location, in words such as *pārsi-ān* 'persian' from the stem *pārs-*, the historic name of modern Iran, and *russi-an* from Russia.

The second rule states that the penultimate consonant becomes fricative if the word final consonant is a stop. This is a synchronic phenomenon in contemporary Persian and in Pashto. For example, the present stem of the verb changes to a fricative when the past tense marker is suffixed to it.

Present	Past	Gloss
yāb-	yāft-	find
tāb-	tāft-	shine
kub-	kuft-	grind

This process in Pashto resulted in the final cluster *-kt-* in *pakt-* to become *paxt-* / *pašt-* in Pashto and *pakht-* in Indic languages, depending on the dialect of the region. Similarly, words borrowed from other languages also undergo this change, for example, the Arabic *waqt* → *waχt* 'time'.

- pakt- → pakiā / paktiyā and paktikā

The place name *paktiā* or *paktiyā* from *paktā* is an internal development in Pashto phonology. In some contemporary Pashto dialects, particularly southern regions of Afghanistan, [y] is inserted as an epenthetic semi-vowel before the long [ā]. For example, the Persian word *tār* 'darkness' becomes *tyārə* in Pashto.

Standard/ Proto	Dialect	Gloss
lār	lyār	path, way
mə lā	mə lyā	back
wāla	wyāla	small stream
paktā-	paktyā	Paktia
dai wāyi	dai wyāyi	he says

The form *paktikā* has the diminutive suffix *-k-* which is common to Persian, Pashto, and in the earlier forms of Aryan languages. Pashtuns called the non-Pashtun speakers of Hindko as *hindki* who spoke *hindko*, which has *hind-* Indic, *-k-* diminutive marker, and *-o* as feminine / diminutive ending, parallel to *pašt-o*. Here it is sufficient to say that *paktikā* is an alternative name for *paktiā*

as both words preserve the original pakt- in them, just like jinəy and jinkəy ‘girl’ are both used in Pashto.

- Proto Forms in Contemporary Pashto

Another type of phonetic change is internal to Pashto and is morphophonemic in nature, that is, it is conditioned by the grammatical form of the word. When the word is pluralized, or when it is grammatically an oblique form, it reverts to its proto form; its back round vowel o / u changes to the original ā.

Singular	Plural	Gloss
paštun	paštānə	Pashtun
ɽund	ɽāndə	blind
špun	špānə	shepherd
soɽ	sāɽə	cold
wror	wrārə	nephew
moɽ	mārə	full (belly)

Vowel Shift in Southern Dialects

The dialects of Pashto spoken by the tribes in the central mountainous region on both sides of the Durand Line, have undergone another vowel shift, which has given a distinct flavor to the Pashto of these tribes. These include Mangal, Zadran, Khostwal on the Afghan side and Wazir and Mehsud on the Pakistan side. The dialects of other Pashtuns living at the periphery of these tribes exhibit various degrees of the vowel change, such as Wardak and Zazi in Afghanistan and Dawar and Banochi in Pakistan. This vowel shift is natural, the result of living sedentary, isolated life in the mountainous regions where there is no literacy in standard Pashto, and mobility has been rare in the region. The following examples show the vowel shift common in the Wazir / Zadran Pashto:

ā → u u → i
o → e a → ā

Standard Pashto	Southern Pashto	Gloss
lās	lus	hand
plār	plur	father
paštun	paštin	Pashtun
mug/muž	mig/miž	we
mahsud	masid	Mehsud
pašto	pašte	Pashto
χor	χer	sister
mor	mer	mother
ğar	ğār	mountain
motar	metār	motor

It is this vowel shift that is reflected in the name of the Pashtun activist Manzur Paštin, who is from Waziristan.

Notes:

1. The name afġān is used for Pashtuns by the Turkic people in the north and Persians in the west. The original work on Pashtun genealogies by Nematullah Haravi in the early 17th Century is titled maḡzan-e-afġān ‘Afghan Treasure’ because the ruling Pashtuns were called Afghans in India.
2. The word afġān in the Hazara dialect of Persian is afġu/awġu, following the rule ā → u / ___ n.
3. Words marked with [*] indicate proto or earlier forms of words in historical linguistics. Proto forms given in this paper are hypothetical, but they help generalize a single rule to explain a large set of data.
4. Persian [d] changes to [l] in many words Pashto.

dast	lās	hand
dār-	lar-	have
juda	jala	separate
sad	sal	hundred
dah	las	ten
dāman	laman	skirt

5. The final consonant cluster -rd in Persian changes to the retroflex -ṛ in Pashto.

mard	☞ meṛə	man
sard	☞ soṛ	cold
zard	☞ žaṛ	yellow
mord	☞ maṛ	dead

6. In Pashto n also changes to l when followed by m, for example,

namāz	☞ lmunz	prayer
nambar	☞ lambar	number
namak	☞ māлга	salt

References:

Encyclopedia Iranica. 1982. Volume 1: Afghanistan. London: Routledge & Kegan Paul.

Fox, Anthony. 1995. *Linguistic Reconstruction*. An Introduction to the Theory and Method. New York: Oxford University Press.

Gray, Louis H. 1965 (1902). *Indo-Iranian Phonology*, With Special Reference to the Middle and New Indo-Iranian Languages. New York. AMS Press.

Habibi, Abdul Hay. 2005 (1944). *Da Paxto Adabeyāto Tārix* ‘History of Pashto Writing’ Vol. I & II. Peshawar: Danish Publishers. (in Pashto)

Henderson, Michael Magnus. 1972. *Dari (Kabul Persian) Phonology*. Doctoral Dissertation. University of Wisconsin.

Kaka-Khel, Syed Bahadar Shah Zafar. 1999(1965). *Pashtāna da Tārikh pa Ranā ke*, 'Pashtuns in the Light of History'. Peshawar: University Book Agency. (in Pashto)

Kavitskaya, Darya. 2001. *Compensatory Lengthening: phonetics, phonology, diachrony*. Doctoral Dissertation. University of California, Berkeley.

https://escholarship.org/content/qt9mw088r1/qt9mw088r1_noSplash_e8237417d10fb38b6561fd7a5ca2120b.pdf

Khan, Mohammad Hayat. 1999 (1874). *Afghanistan and its Inhabitants*. Translated by Henry Priestley. Lahore: Sang-e-Meel Publications.

Kohzad, Ahmad Ali. 2008 (1946?) *Tarikh-e Afghanistan* 'History of Afghanistan'. Kabul: Maiwand Publications. (in Persian)

List of Indo-Aryan Tribes: https://en.wikipedia.org/wiki/Indo-Aryan_peoples

Mackenzie, D. N. 1971. *A Concise Pahlavi Dictionary*. London: Oxford University Press.

Morgenstierne, Georg. 1973. *Indo-Iranian Frontier Languages*. Vol. II. Parachi and Ormuri. Oslo: Universitetsforlaget.

Morgenstierne, Georg. 1975. 'Pashto', Pathan' and the treatment of *r* + sibilant in Pashto. *Irano-Dardica*. Wiesbaden.

Neamet Ullah. 1612. *History of the Afghans*. Translated by Bernard Dorn (1836). London: Blackwell Crescent.

Rond-Miakhel, Mohammad Omar. 2008. *Da Pashtano Shajare aw Mene* 'Pashtun Tribal Genealogies and Homelands'. Kabul: Maiwand Publications. (in Pashto)

Tegey, Habibullah. 2003. *Pashtānə*. Peshawar: Danish Books. (in Pashto)