Linguistic Architecture of Pakistani and Russian Languages: A Comparative Typological Analysis of Phonology, Morphology, Syntax and Writing Systems in Urdu, Punjabi, Sindhi, Pashto and Russian

Ghulam Mustafa[1*], Taufique Alam[2]

Foreign Languages and Intercultural Communication, HSE University, Moscow, Russia

[*]Corresponding Author

DOI: https://doi.org/10.47772/IJRISS.2026.101000654

Received: 24 October 2026; Accepted: 30 October 2026; Published: 14 November 2026

ABSTRACT

Pakistan and Russia are home to some of the most linguistically diverse populations in Eurasia, yet their languages are rarely compared within a single typological framework. This study offers a descriptive-comparative analysis of four major languages of Pakistan (Urdu, Punjabi, Sindhi and Pashto) and Russian, drawing primarily on research, census data and reports published since 2020. Using a qualitative design, it compares the languages across five dimensions: (a) sociolinguistic ecology, (b) phonology, (c) morphology, (d) clause-level syntax and alignment, and (e) writing systems. Linguistic architecture is defined as a layered, multivariate profile of interdependent structural choices, and the comparison parameters were selected using explicit criteria of cross-linguistic comparability, documentary coverage, balance across levels and diagnostic value. A ten-parameter feature matrix is then used to estimate structural proximity. The matrix shows that the four Pakistani languages cluster closely: Urdu matches Punjabi, Sindhi and Pashto on nine of ten parameters each. Russian matches Urdu on only four parameters, diverging on word order, adposition type, alignment, retroflexion, palatalisation and the richness of nominal case. Russian converges with the Pakistani languages in marking grammatical aspect and gender agreement but realises both differently: aspect through perfective–imperfective verb pairs rather than through alignment, and gender through agreement shaped by case syncretism. The Pakistani languages share a South Asian profile of retroflex consonants, postpositions, SOV order and aspect- or tense-conditioned split ergativity, on which Punjabi lexical tone and Sindhi–Saraiki implosives are local developments. Russian, by contrast, displays a Slavic profile of prepositions, case syncretism and a pervasive hard–soft consonant contrast. Pashto belongs to the South Asian cluster partly through Iranian inheritance and partly through contact with Indo-Aryan, and structural distance follows branch and areal affiliation rather than shared Indo-European descent. The findings have implications for language teaching, translation and natural language processing between the two language spheres.

Keywords: comparative typology; Pakistani languages; Russian; split ergativity; palatalisation

INTRODUCTION

The organised set of structural properties through which a language encodes sound, meaning and grammatical relations can be called its “linguistic architecture”: its phonological inventory and prosody, its morphological organisation, its syntactic alignment and word order, and the orthography in which it is written. Recent research on word-order universals indicates that these properties are not independent: there are recurrent correlations between them, such as the link between object–verb order and adposition type (Hahn et al., 2020), which appear to arise from pressure for efficient communication. Such correlations allow languages from different families to be compared on common parameters. Section 3.1 develops this notion into a working definition and relates it to established typological theory.

Pakistan offers an exceptional setting for such a comparison. The 2023 digital census recorded Punjabi as the most widely reported mother tongue (36.98%), followed by Pashto (18.15%), Sindhi (14.31%), Saraiki (12.00%), Urdu (9.25%) and Balochi (3.38%) (Pakistan Bureau of Statistics [PBS], 2023). Although Urdu is the first language of a minority, it functions as the national lingua franca, and the politics of Urdu, English and the regional languages continue to shape education and identity (Shah et al., 2022; Waseem, 2025).

Russian, in contrast, is the dominant language of the Russian Federation; as an East Slavic language it has a six-case nominal system, prepositions and a hard–soft consonant contrast (Ćavar & Lulich, 2021; Chernova et al., 2022).

Nevertheless, little research compares Pakistani languages with Russian. Recent studies tend to describe single languages or to compare one of them with English, for example Punjabi and English case valuation (Khan & Kausar, 2021) and Pashto and English verb structure (Adib & Hood, 2025). At the same time, demand for Russian is growing in Pakistan: a recent mixed-methods study found that learners are motivated chiefly by the prospect of higher education in Russia, while teachers face a shortage of bilingual Russian–Urdu resources (Pathan & Bakhsh, 2025). A contrastive architectural description is therefore of both practical and theoretical value.

Research problem and objectives

The problem addressed here is the absence of a unified architectural description that places the major Pakistani languages and Russian on common typological parameters. The study pursues the following objectives:

- To outline the sociolinguistic ecology of the languages using the 2023 census and recent reports.
- To compare their phonological systems, focusing on retroflexion, aspiration, implosives, lexical tone and palatalisation.
- To compare morphological organisation, case, gender and aspect.
- To compare clause structure, word order and alignment, with attention to split ergativity.
- To compare writing systems and their fit with the underlying phonology.
- To quantify structural proximity through a ten-parameter feature matrix and discuss implications for pedagogy, translation and natural language processing (NLP).

Research Questions

- RQ1. Which architectural features are shared across Urdu, Punjabi, Sindhi and Pashto, and which distinguish them?
- RQ2. On which parameters does Russian diverge most from the Pakistani languages, and on which does it converge?
- RQ3. How do contact and genealogy help explain the observed patterns?

Significance

The study serves as a concise comparative reference for linguists, language teachers and language-technology developers working with South Asian languages and Russian, and it shows how sharply different architectures can develop within one language family through the interplay of areal convergence and genealogical inheritance.

LITERATURE REVIEW

Sociolinguistic Context

According to census-based summaries, Punjabi's share fell from about 44% in 1998 to 37% in 2023, while Urdu's rose from about 8% to 9% (Waseem, 2025). Scholars argue that language policy can either hinder or support social cohesion depending on how it is implemented: when Urdu and English were elevated as official languages early on, regional languages such as Sindhi were marginalised in education and public life, with consequences for cohesion (Shah et al., 2022). Ethnologue estimates about 210 million Russian speakers (133 million L1 and 77 million L2) and 246 million Urdu speakers (78 million L1 and 168 million L2) (Ethnologue, 2026).

Phonological Studies

Recent typological research shows that Indo-Aryan languages have relatively large retroflex inventories, and that retroflex implosives are unusually frequent in the implosive inventories of South Asia, a pattern attributed to the general frequency of retroflexes in Indo-Aryan (Hussain & Mielke, 2023). Instrumental evidence on Saraiki distinguishes unaspirated, aspirated, breathy-voiced and implosive stops: aspirated stops show long voice onset time, and implosives show a gradual increase in amplitude during closure (Atta et al., 2022). Acoustic evidence of tone is available for Punjabi, which has a three-way High, Mid and Low contrast cued primarily by duration and final velocity (Malik & Khan, 2018), and for the neighbouring variety Pahari, where pitch is complemented by duration, intensity and breathy phonation (Khan et al., 2020).

A study of Urdu based on 500 dictionary words and 20 speakers identifies regular deviations between pronunciation and spelling, including assimilation, dissimilation, deletion, insertion and nasalisation (Khan, 2023). For Russian, ultrasound studies analyse the palatalisation contrast in terms of tongue-body configuration (Padgett et al., 2024) and describe vowel allophony next to hard and soft consonants in terms of advanced tongue root (Ćavar & Lulich, 2021). Experimental evidence further suggests that assimilatory palatalisation involves incomplete neutralisation rather than a categorical process (Oh et al., 2024).

Morphology, Syntax and Alignment

Khan and Kausar (2021) describe Punjabi as a split-ergative language whose case patterns depend on aspect and person: in perfective clauses the external argument is marked with ergative case, whereas non-perfective clauses pattern as nominative-accusative; Punjabi also uses postpositions rather than prepositions. Pashto likewise shows split ergativity, with the verb agreeing with the subject in the present tense and with the object in past-tense transitive clauses (Adib & Hood, 2025); agreement targets the nominative argument and falls back to a default form when no nominative argument is available (Khan & Iqbal, 2021).

A large-scale feature comparison of 16 Indo-Aryan and 11 Munda languages identifies ergativity as one of the features that distinguish western from eastern Indo-Aryan (Ivani et al., 2020). Recent psycholinguistic work on Russian has focused on case syncretism, where different case–number cells share one form: syncretism affects the processing of nouns after case-governing prepositions (Chernova et al., 2022) and increases gender agreement attraction (Slioussar et al., 2022). Second-language acquisition research shows that the perfective–imperfective contrast is among the most difficult points for learners, who tend to default to the imperfective as a past tense (Whitehead Martelle & Shirai, 2023).

Technology and Language-Contact Studies

An evaluation of English-to-Urdu machine translation (MT) found that models specialised for low-resource Indic languages outperform general-purpose models, with training-data splits accounting for much of the difference (Basit et al., 2024). Urdu's Perso-Arabic orthography also complicates text processing because it diverges from the spoken language (Khan, 2023).

Research Gap

Although recent studies describe individual languages or compare one language with English, no study was found that compares the major languages of Pakistan and Russian on a common set of parameters. This study addresses that gap with an explicit feature matrix and draws out its pedagogical and computational implications.

THEORETICAL FRAMEWORK

Defining Linguistic Architecture

This research employs a definition of linguistic architecture that is based on four interdependent sets of structural choices, which a language makes at the level of its segmental and prosodic inventory, the way the language packages up its grammatical information in words, the way the language marks the relations between a verb and its arguments, and the way the language's writing system maps onto its sound structure. It is better than 'language type' for two reasons. First, classical holistic typology tried to classify whole languages into a single morphological type (isolating, agglutinating or inflecting), which is now largely given up since the features at various levels do not correlate very strongly with each other (Croft, 2003). Second, current multivariate typology regards each language as a collection of language variables each of which can be measured independently and their co-occurrence is then explained by functional pressure and/or inheritance or contact (Bickel, 2007). For the purposes of this paper, "architecture" refers to the second, namely separately coded features, as well as the hypothesis that some of these features are interdependent, such as verb-final order and postpositions in Indo-Aryan and Iranian (Hahn et al., 2020) or perfective aspect and past tense (Haig, 2008) and ergative marking.

There are three aspects of the concept and each of these serves as a guide for the analysis. (a) Levels: It ranges from phonology all the way to orthography, whereas sociolinguistic ecology is considered as a context to the profile. (b) Linkage: features are expected to cluster and in the Discussion clusters as opposed to single features are the primary unit of comparison. (c) Source of similarity – common descent, areal convergence or independent functional pressure – similarity in the feature matrix description still needs to be explained (Section 6.3). Writing systems are included since they are the means by which the phonological architecture becomes visible to learners and to the processing system, but they are not part of the proximity matrix since a writing system is a cultural and not a grammatical choice.

Typological and Areal Framework

This is a functional-typological study that compares languages through cross-linguistically definable categories. Word-order parameters are treated as correlated, consistent with evidence that grammars balance ease of production against the avoidance of ambiguity and thereby reproduce the correlation between object–verb order and adposition type (Hahn et al., 2020). Following recent work on South Asia, areal convergence is understood as the geographical patterning of morphosyntactic features across genealogically related and unrelated languages (Ivani et al., 2020). Alignment is analysed using the established categories of nominative-accusative, ergative-absolutive and split systems, as applied in recent descriptions of Punjabi and Pashto (Khan & Kausar, 2021; Adib & Hood, 2025). The South Asian linguistic area, characterised among other traits by retroflex consonants and verb-final syntax (Emeneau, 1956), provides the areal baseline against which Pashto and Russian are compared.

Comparative Concepts and Their Limits

In order to make crosslinguistic comparison, categories need to be defined independently of the grammar of any particular language. As in Haspelmath (2010), the ten parameters are considered as comparative concepts, not as the categories of the individual languages. The agreement for gender, for example, is coded as present if the verb, adjective or other target agrees with the gender of its controller, although Russian has three genders and the languages of Pakistan two. This renders the languages similar but also implies that the same code can implement different mechanisms, a matter discussed in Section 6.5.

RESEARCH METHODOLOGY

Design

This is a qualitative, descriptive-comparative study based on secondary data. No fieldwork or corpus collection was undertaken. The comparison draws primarily on evidence published since 2020, so that it reflects current research; standard reference grammars and the World Atlas of Language Structures are used only for baseline features that the recent studies do not address (see Section 4.5).

Language Sample

Five languages were selected: Urdu (national language and lingua franca), Punjabi (the largest mother tongue), Sindhi and Pashto (the major provincial languages from the Indo-Aryan and Iranian branches respectively) and Russian (a Slavic reference language). Saraiki is discussed in the literature reviewed (Atta et al., 2022) but was not coded.

Data Sources

Sociolinguistic data were taken from the 2023 census of the Pakistan Bureau of Statistics (PBS, 2023) and the census-based summary by Waseem (2025). Structural data are drawn from the peer-reviewed studies cited in Section 2, supplemented for baseline features by standard reference descriptions (Masica, 1991; Schmidt, 1999; Tegey & Robson, 1996) and the World Atlas of Language Structures (Dryer & Haspelmath, 2013). Worldwide speaker figures for Urdu and Russian come from Ethnologue (2026). Because Ethnologue's full data tables are available only by subscription, the figures were taken from secondary tabulations of the 2026 edition. These agree on the totals (246 million for Urdu and 210 million for Russian) but differ on the first- and second-language split for Russian; this is noted where the split is reported.

Analytical Procedure

Each of the ten parameters was chosen with respect to four criteria (summarised in Table 1). (i) Comparability: Every parameter had to be amenable to a description in terms of a comparative concept (Haspelmath, 2010) and – as much as possible – to the comparison of a feature of the World Atlas of Language Structures, to ensure that the codes could be validated against an external source. (iii) Coverage: Each parameter needed to be described in the literature in the last five years, or in a standard reference description, in all five languages. (iii) Balance across levels: the parameters for each level (phonology, morphology, and syntax) were distributed so that none of the levels dominate the proximity score. (iv) Diagnostic value: Each parameter had to differentiate between at least one of the sample languages and the remaining languages, or have been demonstrated in the literature to differentiate between the South Asian profile and the Slavic profile. Parameters 8 and 10 do not discriminate within this sample, as they are defining parameters of both profiles and test the functional convergence or not (Section 6.5). Features which did not meet criterion (ii) (e.g., compound verbs, evidentiality, pronominal clitics) are referred to when they occur, but not coded.

Table 1. Parameters of the feature matrix and the criteria for their selection

Parameter	Level	WALS feature	Diagnostic contrast in the sample
1. Dominant word order	Syntax	81A	SOV (Pakistani languages) vs SVO (Russian)
2. Adposition position	Syntax	85A	Postpositions vs prepositions
3. Split ergativity	Syntax / morphosyntax	98A (partly)	Present in all four Pakistani languages; absent in Russian

4. Lexical tone	Phonology	13A	Punjabi only
5. Implosive consonants	Phonology	7A	Sindhi only
6. Secondary palatalisation	Phonology	— (reference descriptions)	Russian only
7. Retroflex series	Phonology	— (Hussain & Mielke, 2023)	Pakistani languages vs Russian
8. Overt grammatical aspect	Morphology	65A	Non-discriminating; tests formal vs functional convergence
9. Rich nominal case	Morphology	49A	Pashto and Russian vs Indo-Aryan languages
10. Gender agreement	Morphology	30A	Non-discriminating; number of genders differs (two vs three)

Source: Author's selection. WALS feature numbers refer to Dryer and Haspelmath (2013); "—" marks parameters without a directly corresponding WALS feature.

Ten parameters were defined: (1) dominant word order, (2) adposition position, (3) split ergativity, (4) lexical tone, (5) implosive consonants, (6) a secondary palatalisation contrast, (7) a retroflex series, (8) overt grammatical aspect, (9) richness of nominal case and (10) grammatical gender agreement. Each language was coded 1 (present), 0 (absent) or 0.5 (partial). The proximity of two languages is the number of parameters on which their codes are identical, so a partial code counts as a full mismatch against both 0 and 1. A graded alternative, which subtracts the absolute difference between codes, changes only the pairs involving parameter 9 (for example, Urdu–Pashto rises from 9 to 9.5 and Urdu–Russian from 4 to 4.5) and leaves the overall pattern unchanged. Because the coding is coarse and the feature set is small, the scores are descriptive heuristics, not inferential statistics. The full coding sheet is given in Appendix A.

Reliability and Limitations of Coding

Coding was based primarily on the recent studies cited in this paper. Where those studies did not address a parameter for a given language (for example, word order and adposition type in Urdu and Sindhi, or retroflexes in Pashto), the code follows standard reference descriptions: Masica (1991) for Indo-Aryan, Schmidt (1999) for Urdu, Tegey and Robson (1996) for Pashto, and the World Atlas of Language Structures (Dryer & Haspelmath, 2013). Pashto has prepositions and circumpositions as well as postpositions; it is coded as postpositional following the World Atlas of Language Structures. Parameter 3 records whether a split exists, not whether it is conditioned by aspect (Urdu, Punjabi and Sindhi) or by tense (Pashto); that difference is discussed in Section 5.4. Parameter 9 (nominal case) was coded 0.5 for the Indo-Aryan languages, because postpositions take over much of the case-marking function, while Pashto and Russian were coded 1.

Varieties Coded and Dialectal Scope

The codes were assigned to a single variety of each language, and do not represent statements about other varieties. For Urdu, the standard written form was coded for the standard Urdu of Pakistan; for Punjabi, the standard written form was coded for central (Majhi) Urdu; for Sindhi, the standard written form was coded for the standard written form of Sindhi used; for Pashto, the standard form of the language was coded for the variety described by Tegey and Robson (1996); and for Russian, the standard form of the language was coded for contemporary standard Russian. Table 2 presents the parameters known to influence the code in the context of dialectal/structural variation. The code in each case is the reference variety and the variation is reported not averaged.

Table 2. Dialectal and structural variation affecting the codes

Parameter	Language	Variation	Coding decision
4. Lexical tone	Punjabi	Tone arose from the loss of historical voiced aspirates and /h/; its realisation varies across dialects (Bhatia, 1993; Malik & Khan, 2018)	Coded 1 for the central standard
5. Implosives	Sindhi (also Saraiki)	Sindhi has implosives at four places of articulation (Nihalani, 1995); Saraiki shares the series (Atta et al., 2022)	Coded 1; Saraiki not coded
7. Retroflex series	Pashto	Retroflex stops and flap occur across dialects; the retroflex fricatives of southern varieties are realised as velars in northern varieties (Tegey & Robson, 1996)	Coded 1
1. Word order	Russian	Order is flexible and governed by information structure; SVO is the neutral order (Bailyn, 2012)	Coded 0 (SVO)
2. Adposition position	Pashto	Prepositions and circumpositions occur alongside postpositions (Tegey & Robson, 1996)	Coded 1, following WALS
3. Split ergativity	All Pakistani languages	Conditioned by perfective aspect in Indo-Aryan and by past tense in Pashto; Punjabi is also sensitive to person; Sindhi marks the ergative subject by oblique case and pronominal suffixes rather than a dedicated postposition (Masica, 1991)	Coded 1 for the presence of a split only
9. Rich nominal case	Urdu, Punjabi, Sindhi	Direct, oblique and vocative inflection, with postpositions carrying most case functions (Masica, 1991)	Coded 0.5
10. Gender agreement	All	Russian has three genders; the Pakistani languages have two (Corbett, 1991)	Coded 1 for presence

Source: Author's compilation from the works cited in the table.

Uneven Documentation across Languages

Inconsistent documentation in languages To make the differences in the available documentation clear we present in Table 3 whether the codes are based on recent empirical studies, on standard reference descriptions or on both.

Empirical studies have been conducted at most levels for Urdu and Punjabi, and for Russian, they have been conducted at all levels, while coding for Sindhi is based more on reference description, and neither the writing system of Punjabi nor Sindhi and Pashto have been studied recently.

Thus any differences in the depth of analysis shown in the languages in section 5 is not due to any differences in the way they are analytically handled but simply to the difference in the literature. The conclusions drawn on Sindhi are drawn with more modesty.

Table 3. Evidence base for each language and level

Language	Phonology	Morphology and syntax	Writing system
Urdu	Empirical (Khan, 2023; Hussain & Mielke, 2023)	Reference (Schmidt, 1999; Masica, 1991)	Empirical (Khan, 2023)
Punjabi	Empirical (Malik & Khan, 2018); reference (Bhatia, 1993)	Empirical (Khan & Kausar, 2021)	Reference only
Sindhi	Reference and typological (Nihalani, 1995; Hussain & Mielke, 2023)	Reference (Masica, 1991)	Reference only (Masica, 1991)
Pashto	Reference (Tegey & Robson, 1996)	Empirical (Adib & Hood, 2025; Khan & Iqbal, 2021); reference (Tegey & Robson, 1996)	Reference only
Russian	Empirical (Padgett et al., 2024; Ćavar & Lulich, 2021; Oh et al., 2024)	Empirical (Chernova et al., 2022; Slioussar et al., 2022; Whitehead Martelle & Shirai, 2023); reference (Bailyn, 2012)	Empirical (Padgett et al., 2024)

Source: Author's compilation. "Empirical" = acoustic, experimental, corpus or analytical study; "Reference" = reference grammar, phonetic illustration or typological database.

ANALYSIS AND FINDINGS

Sociolinguistic Ecology

Table 4 summarises the position of each language. The 2023 census figures are shown in Figure 1. Figure 2 compares the 1998 and 2023 shares of Punjabi and Urdu as reported by Waseem (2025): Punjabi declined while Urdu rose slightly.

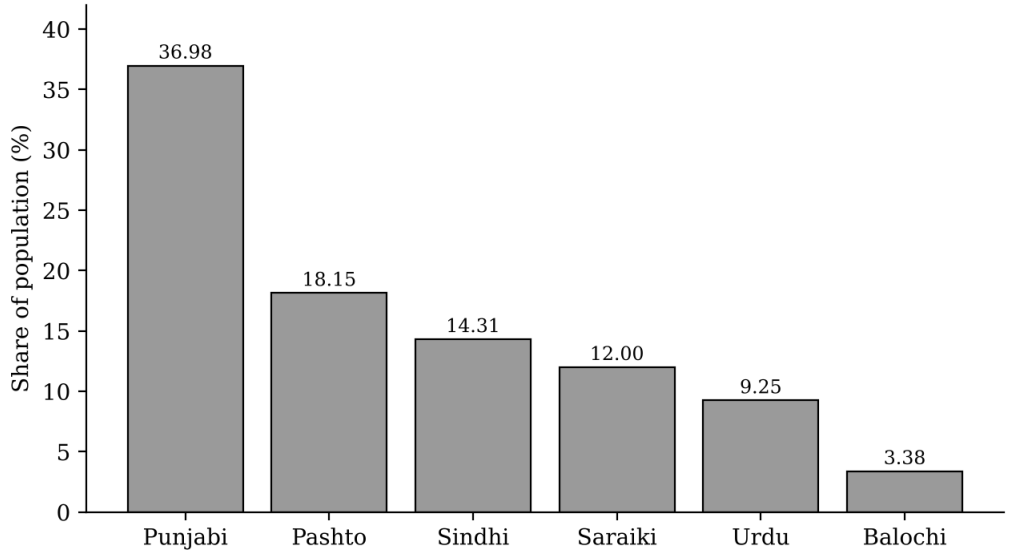
Table 4. Sociolinguistic profile of the languages under study

Language	Branch / family	Main region	Status and share
Urdu	Indo-Aryan (Indo-European)	Pakistan-wide; lingua franca	9.25% mother tongue in Pakistan (2023); national and co-official language; about 246 million speakers worldwide, 78 million of them first-language (Ethnologue, 2026)
Punjabi	Indo-Aryan (Indo-European)	Punjab province	36.98% (2023); largest mother tongue, down from about 44% in 1998
Sindhi	Indo-Aryan (Indo-European)	Sindh province	14.31% (2023); regional language

Pashto	Iranian (Indo-European)	Khyber Pakhtunkhwa, northern Balochistan	18.15% (2023)
Russian	East Slavic (Indo-European)	Russian Federation and wider post-Soviet space	About 210 million speakers worldwide: 133 million first-language and 77 million second-language (Ethnologue, 2026)

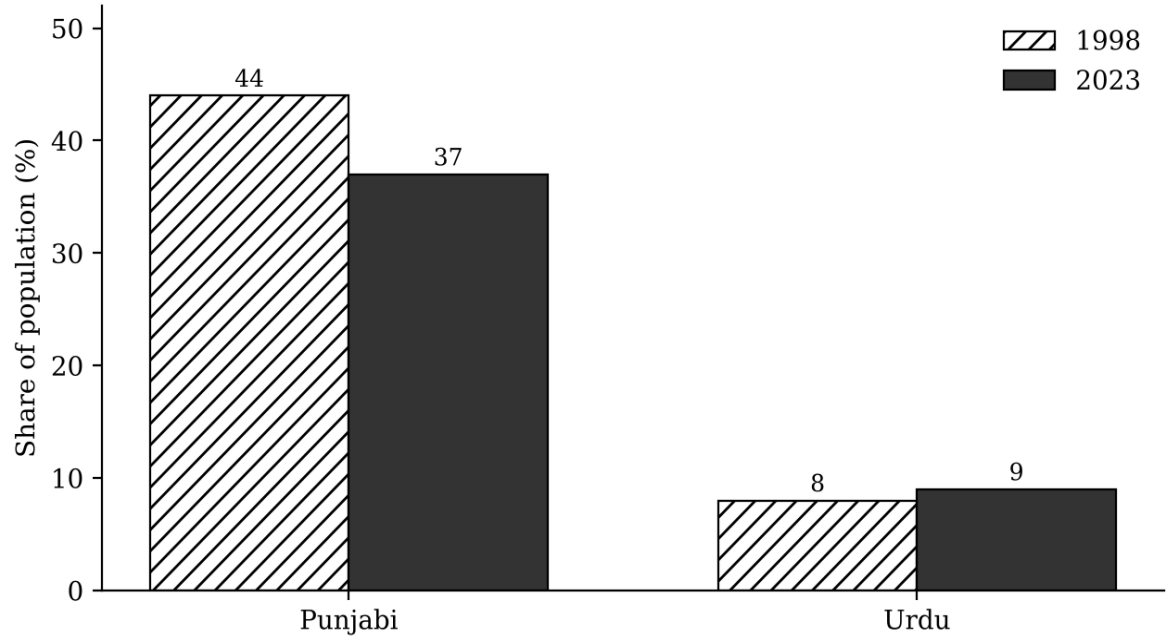
Sources: PBS (2023) and Waseem (2025) for Pakistani census shares; Ethnologue (2026) for worldwide speaker counts of Urdu and Russian.

Figure 1. Mother-tongue shares in Pakistan, 2023 census (%)



Source: Author's chart based on PBS (2023).

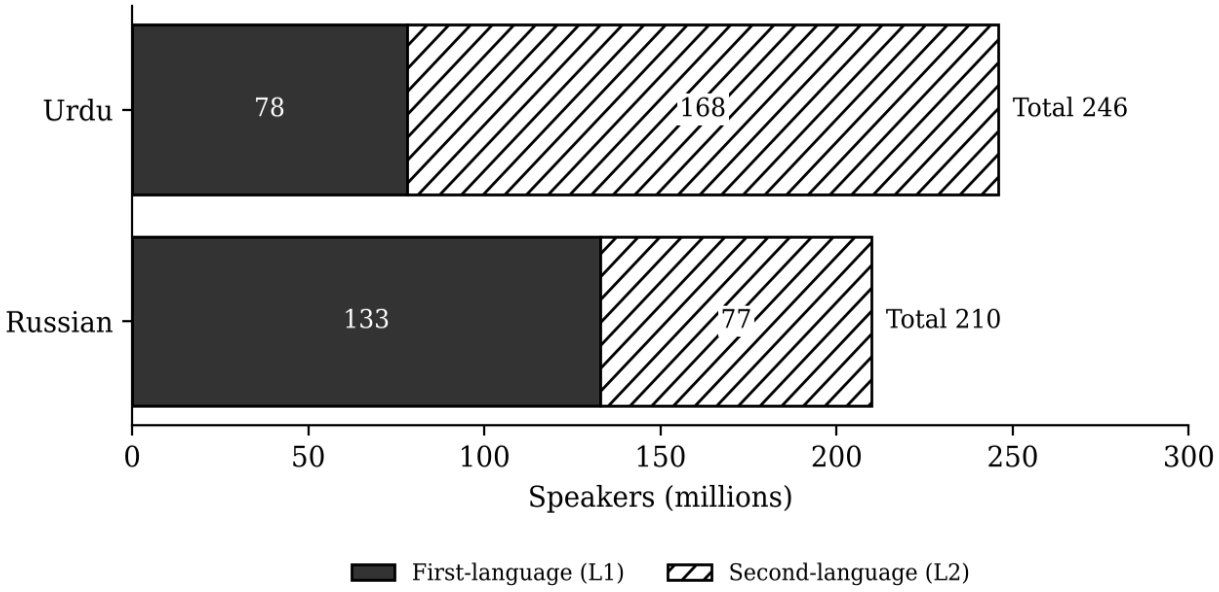
Figure 2. Change in the share of Punjabi and Urdu, 1998–2023 (%)



Source: Author's chart based on figures reported in Waseem (2025).

The two languages also differ in how their speaker populations are composed (Figure 3). Urdu has more second-language than first-language speakers (168 million against 78 million), which reflects its role as a lingua franca, whereas Russian has more first-language than second-language speakers (Ethnologue, 2026). The Russian split is reported as 133 million L1 and 77 million L2 in one tabulation and as about 145 million L1 and 65 million L2 in another; both give a total of 210 million and both show an L1 majority.

Figure 3. First-language (L1) and second-language (L2) speakers of Urdu and Russian worldwide, 2026 (millions)



Source: Author's chart based on Ethnologue (2026) figures: Urdu 78 L1 + 168 L2 = 246; Russian 133 L1 + 77 L2 = 210 (another tabulation gives Russian as about 145 L1 + 65 L2, with the same total). Urdu is counted separately from Hindi.

Phonology

Pakistani Languages

Retroflex consonants are pervasive in the consonant systems of the Indo-Aryan languages of Pakistan (Hussain & Mielke, 2023). Sindhi and Saraiki have implosives at several places of articulation, a textbook case of implosives in Indo-Aryan (Hussain & Mielke, 2023; Atta et al., 2022). Punjabi has a three-way lexical tone contrast (Malik & Khan, 2018), which is uncommon among Indo-European languages, and related varieties realise tone through several acoustic cues (Khan et al., 2020). Urdu pronunciation departs from spelling in regular, rule-governed ways, which has implications for lexicography and speech technology (Khan, 2023). The Punjabi tone description applies to the central (Majhi-based) standard; because tone is realised differently across Punjabi dialects (Bhatia, 1993), the code in Appendix A refers to the standard variety only. Sindhi has implosives at four places of articulation (Nihalani, 1995). Pashto has retroflex stops and a retroflex flap, while its retroflex fricatives vary by dialect: southern varieties retain them, whereas northern varieties realise them as velars (Tegey & Robson, 1996).

Russian

Russian lacks the retroflex, implosive and tonal contrasts found in the Pakistani languages. Its phonology is instead organised around the contrast between hard and soft (palatalised) consonants, which ultrasound studies interpret in terms of tongue-body and tongue-root configuration (Padgett et al., 2024; Ćavar & Lulich, 2021) and which shows incomplete neutralisation in assimilation contexts (Oh et al., 2024).

Table 5. Comparison of salient phonological features

Feature	Urdu	Punjabi	Sindhi	Pashto	Russian
Retroflex series	Yes	Yes	Yes	Yes	No
Implosives	No	No	Yes	No	No
Lexical tone	No	Yes (3 tones)	No	No	No
Secondary palatalisation contrast	No	No	No	No	Yes
Spelling–sound mismatch documented	Yes	No recent study found	No recent study found	No recent study found	No recent study found

Sources: Hussain & Mielke (2023); Atta et al. (2022); Malik & Khan (2018); Khan (2023); Padgett et al. (2024); Ćavar & Lulich (2021); Tegey & Robson (1996) for Pashto retroflexes; Nihalani (1995) for Sindhi implosives. "No recent study found" marks rows not addressed in the studies reviewed. Values refer to the reference varieties listed in Section 4.6; see Table 2 for dialectal variation.

Morphology

Russian morphology is organised around case, gender and aspect. Prepositions and syncretism across paradigms shape case processing and agreement (Chernova et al., 2022; Slioussar et al., 2022), and the perfective–imperfective contrast is central to verb use and L2 acquisition (Whitehead Martelle & Shirai, 2023). In Punjabi and Pashto, by contrast, case and agreement morphology is conditioned by perfective aspect or past tense, and in these contexts the verb may agree with the object (Khan & Kausar, 2021; Adib & Hood, 2025; Khan & Iqbal, 2021). Aspect is therefore grammatically encoded in both groups, but Russian expresses it mainly through verb pairs, whereas the Pakistani languages express it largely through its interaction with alignment. Unlike Sindhi, which marks the ergative subject by oblique case and pronominal suffixes in perfective transitive clauses, Urdu has the postposition ne, which is also used in perfective transitive clauses, to mark the ergative subject as in (2) below (Masica, 1991; Schmidt, 1999). Thus, in Table 6 the grouping of the three Indo-Aryan languages is based on similarity of alignment, not on morphology. Russian also has gender systems of varying sizes; it has masculine, feminine and neuter, whereas the four Pakistani languages have only two genders – masculine and feminine (Corbett 1991).

Table 6. Morphological architecture

Dimension	Punjabi / Urdu / Sindhi	Pashto	Russian
Case marking	Postpositions carry much of the load; ergative case in perfective clauses (Punjabi)	Split-ergative: ergative marker on past transitive subjects	Case governed by prepositions; syncretism across paradigms
Agreement	Subject or object depending on aspect (Punjabi)	Subject in present and intransitive clauses; object in past transitive	Gender agreement, subject to attraction effects
Aspect	Conditions case and agreement	Interacts with agreement	Perfective / imperfective contrast, a learner difficulty

Sources: Khan & Kausar (2021); Adib & Hood (2025); Khan & Iqbal (2021); Chernova et al. (2022);

Slioussar et al. (2022); Whitehead Martelle & Shirai (2023). Urdu and Sindhi are grouped with Punjabi following standard descriptions of Indo-Aryan morphosyntax (Masica, 1991; Schmidt, 1999). Sindhi differs from Urdu and Punjabi in lacking a dedicated ergative postposition (Masica, 1991).

Syntax and Alignment

Word order is where the syntactic contrast is clearest. Pashto has basic SOV order, which can be varied for emphasis (Adib & Hood, 2025), and Punjabi uses postpositions (Khan & Kausar, 2021). Russian is an SVO language with prepositions that govern specific cases (Chernova et al., 2022). Russian word order is flexible and sensitive to information structure, but SVO is its neutral order (Bailyn, 2012). The pairing of verb-final order with postpositions, and of verb-medial order with prepositions, is what the efficiency-based account predicts (Hahn et al., 2020). The second major difference is alignment. Punjabi and Pashto show aspect- or tense-conditioned split ergativity, whereas Russian is nominative-accusative throughout (Khan & Kausar, 2021; Adib & Hood, 2025). Large-scale comparison shows that ergativity is among the features that separate western from eastern Indo-Aryan (Ivani et al., 2020).

A few simple examples can be used to illustrate the alignment contrast. Unmarked in Urdu, the subject of an imperfective clause also determines the case of the verb (1); in the perfective, the subject is marked with the postposition ne and the verb is in agreement with the feminine object (2). Both tenses of Pashto are conditioned by tense (3) with verb agreeing with the subject in the present, and with the subject in the oblique case and the verb agreeing with the object in the past. The perfective–imperfective contrast is not signaled by the verb in Russian; the subject is nominative and agrees in both number and gender.The perfective–imperfective contrast is not marked on the verb in Russian; the subject is nominative, and it agrees in both number and gender (5a–b). The examples are all based on the descriptions from Schmidt (1999), Tegey and Robson (1996) and Whitehead Martelle and Shirai (2023); glosses are based on the Leipzig conventions.

Urdu

(1) *ali* *kitāb* *paṛh-t-ā* *hai*

Ali(M) book(F) read-IPFV-M.SG be.PRS.3SG

‘Ali reads a book.’

(2) *ali* *ne* *kitāb* *paṛh-ī*

Ali(M) ERG book(F) read.PFV-F.SG

‘Ali read the book.’

Pashto

(3) *zə* *ḍoḍəy* *xwr-əm*

1SG.DIR bread(F) eat-1SG

‘I eat bread.’

(4) *mā* *ḍoḍəy* *wə-xwaṛ-a*

1SG.OBL bread(F) PFV-eat.PST-F.SG

‘I ate bread.’

Russian

(5a) *ja* *čita-l* *knig-u*

1SG.NOM read.IPFV-PST.M book-ACC

'I was reading / used to read a book.'

(5b) *ja* *pro-čita-l* *knig-u*

1SG.NOM PFV-read-PST.M book-ACC

'I read (finished reading) the book.'

Table 7. Clause-level typology

Parameter	Urdu	Punjabi	Sindhi	Pashto	Russian
Basic order	SOV	SOV	SOV	SOV	SVO
Adposition type	Postpositions	Postpositions	Postpositions	Postpositions	Prepositions
Alignment	Split (aspect)	Split (aspect, person)	Split (aspect)	Split (tense)	Nominative-accusative

Sources: Khan & Kausar (2021); Adib & Hood (2025); Khan & Iqbal (2021); Chernova et al. (2022); Hahn et al. (2020). Values for Urdu and Sindhi, and the Pashto adposition value, follow Masica (1991), Schmidt (1999), Tegey and Robson (1996) and Dryer and Haspelmath (2013). Sindhi realises ergative alignment without a dedicated ergative postposition (see Section 5.3), and Russian SVO is the neutral rather than the only order (Bailyn, 2012).

5.5 Writing systems

Urdu, Punjabi (Shahmukhi), Sindhi and Pashto are written in modified Perso-Arabic scripts, while Russian is written in Cyrillic. Urdu orthography is broadly phonemic, but its systematic departures from pronunciation, together with properties of the script itself, make Urdu difficult to process computationally (Khan, 2023). In Russian, palatalisation is signalled through the choice of letters, although the phonetics of the hard–soft contrast are complex (Padgett et al., 2024).

Table 8. Writing systems

Language	Script	Direction	Documented issue
Urdu	Perso-Arabic	Right-to-left	Spelling–pronunciation divergence via sound-change rules (Khan, 2023)
Punjabi	Shahmukhi (Perso-Arabic) in Pakistan	Right-to-left	No recent study found
Sindhi	Extended Perso-Arabic	Right-to-left	No recent study found; implosives require additional letters

Pashto	Extended Perso-Arabic	Right-to-left	No recent study found
Russian	Cyrillic	Left-to-right	Encodes hard–soft contrast; relation to phonetics is complex (Padgett et al., 2024)

Sources as indicated. Entries marked "No recent study found" identify gaps in the recent literature. The Sindhi entry on additional letters for implosives follows Masica (1991).

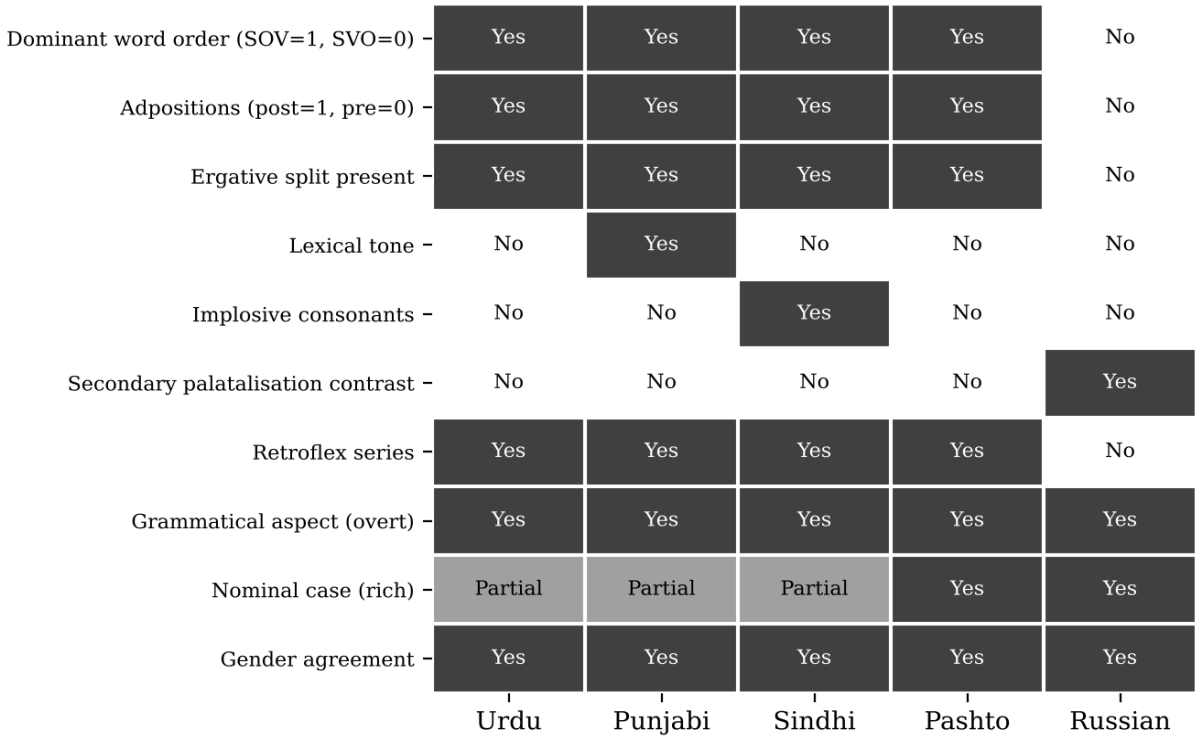
Structural proximity: the ten-parameter matrix

The coded feature matrix is presented in Figure 4 and the pairwise proximity scores in Figure 5. Urdu matches each of the other Pakistani languages on nine of ten parameters, differing from Punjabi only in lexical tone, from Sindhi only in implosives and from Pashto only in nominal case richness.

All four Pakistani languages share verb-final order, postpositions, split ergativity, a retroflex series, overt aspect and gender agreement; Pashto's place in this cluster reflects both Iranian inheritance and contact with its Indo-Aryan neighbours (Section 6.3). Russian matches Urdu on only four parameters (absence of tone, absence of implosives, overt aspect and gender agreement) and matches Punjabi and Sindhi on three each.

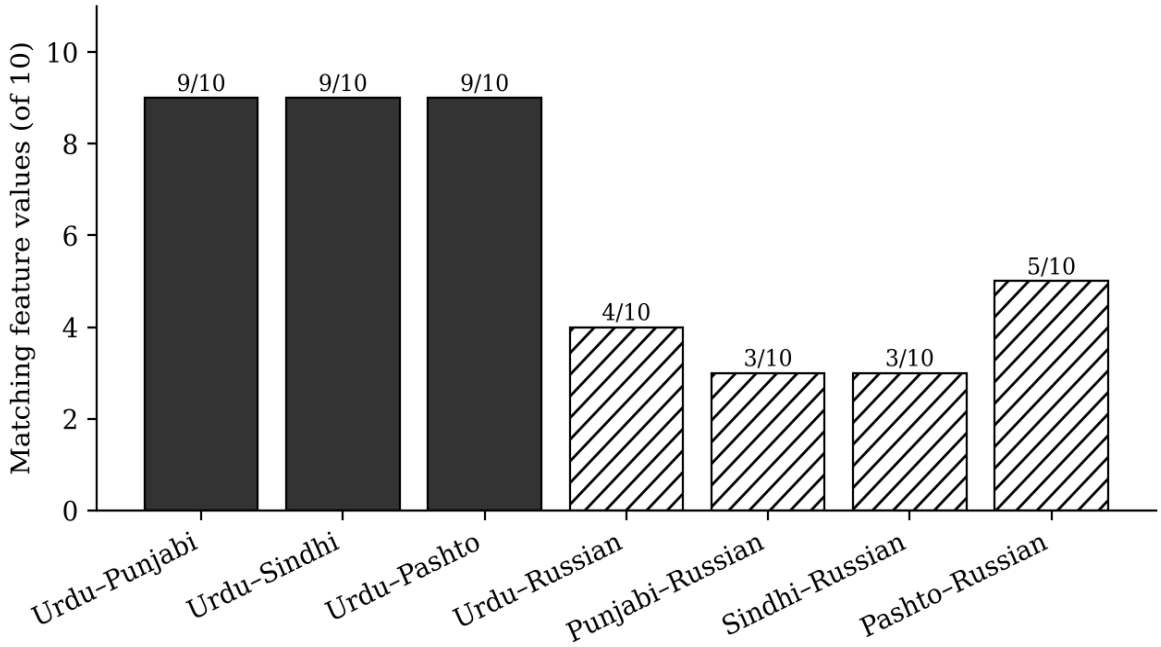
Its highest score is with Pashto (five of ten), because both are coded as having rich nominal case. Because Parameters 8 and 10 do not discriminate within the sample, the scores are driven by the other eight parameters: excluding them lowers every pairwise score by two (for example, Urdu–Punjabi to 7/8 and Urdu–Russian to 2/8) without changing the ranking.

Figure 4. Ten-parameter typological feature matrix



Source: Author's coding, based on the recent studies cited in Sections 2 and 5. 1 = present, 0 = absent, 0.5 = partial. Codes are heuristics, not measurements.

Figure 5. Pairwise structural proximity (number of matching parameters out of 10)



Source: Author's calculation from the matrix in Figure 4. Solid bars compare Urdu with other Pakistani languages; hatched bars compare Russian with Pakistani languages.

DISCUSSION

A South Asian Profile

In answer to RQ1, the Pakistani languages share a combination of retroflexion, postpositions, SOV word order and aspect- or tense-conditioned ergativity (Hussain & Mielke, 2023; Khan & Kausar, 2021; Adib & Hood, 2025). Distinctive features, such as Punjabi tone and Sindhi and Saraiki implosives, are layered on this common base (Malik & Khan, 2018; Atta et al., 2022). Pashto belongs to the Iranian branch, yet it shares this profile. Its verb-final order and past-tense ergativity are also found in other Iranian languages and are best seen as inherited (Haig, 2008), whereas its retroflex consonants point to contact with Indo-Aryan (Masica, 1991).

Russian as a Contrasting Architecture

In answer to RQ2, Russian differs on word order, adposition type, alignment, retroflexion, case richness and the hard–soft consonant contrast; the word-order and adposition differences cluster as efficiency-based typology would predict (Hahn et al., 2020). It converges with the Pakistani languages in having aspect and gender agreement, but realises them differently: aspect through perfective–imperfective verb pairs (Whitehead Martelle & Shirai, 2023) rather than through its interaction with alignment, and gender through agreement shaped by case syncretism (Slioussar et al., 2022).

Genealogy, Contact and Typology

For RQ3, the evidence implicates both genealogy and contact. Inheritance accounts for much of what Pashto shares with its Indo-Aryan neighbours, while contact best explains its retroflex consonants. The implosives of Sindhi and Saraiki, by contrast, are attributed to a local sound change from voiced geminates rather than to contact (Hussain & Mielke, 2023). At the regional level, a large-scale comparison shows that morphosyntactic features in South Asia are patterned geographically rather than genealogically (Ivani et al., 2020). The Russian profile is clearly a product of its Slavic inheritance and its own palatalisation system (Ćavar & Lulich, 2021), and shows minimal South Asian influence. Because the two language spheres have

had no sustained contact, the features Russian shares with the Pakistani languages are better explained otherwise: gender agreement reflects common Indo-European inheritance (Corbett, 1991), whereas the Slavic perfective–imperfective system and the aspect-conditioned alignment of the Pakistani languages developed independently.

6.4 Implications for pedagogy and language technology

Table 9 summarises the practical implications of the structural contrasts identified in Sections 5.2–5.5.

Table 9. Implications of structural contrasts for pedagogy and language technology

Contrast	Pedagogical implication	NLP / translation implication
SOV versus SVO order; postpositions versus prepositions	Learners need explicit practice with clause reordering and with adposition–case government	Reordering and case handling are needed in translation between Urdu-type and Russian-type languages (Basit et al., 2024)
Split ergativity versus nominative-accusative alignment	Russian-speaking learners of Urdu or Punjabi must learn aspect-conditioned subject marking and object agreement	Agreement and case tagging must be aspect-aware (Khan & Iqbal, 2021)
Hard–soft consonants versus retroflexes and implosives	Pronunciation training should target palatalisation for Pakistani learners of Russian and retroflex contrasts for Russian learners of Urdu	Speech and grapheme-to-phoneme systems need language-specific modelling (Khan, 2023)
Perfective–imperfective verb pairs	Aspect is a known learning difficulty and is often over-defaulted to the imperfective (Whitehead Martelle & Shirai, 2023)	Aspect-sensitive generation is needed in Russian output
Limited Russian–Urdu materials	Bilingual Russian–Urdu resources are scarce (Pathan & Bakhsh, 2025)	Parallel data are needed for low-resource pairs

Source: Author’s synthesis of Sections 5.2–5.5.

Theoretical Implications

There are three implications to the comparison, in addition to the description of the five languages. First, it provides a perspective on the architecture of language as multi-layered and not a single entity. Despite Pashto's Iranian language, the morphosyntactic layer (word order, adposition type and adposition alignment) shares characteristics with its Indo-Aryan neighbours, whereas the phonological layer includes innovations found exclusively in Pashto, like Punjabi tone, and in Sindhi, like implosives. The morphosyntactic layer is always enhanced by the inheritance layer and areal convergence, with the latter layer being more likely to preserve local development (Haig, 2008; Ivani et al., 2020). The concept of "typological distance" is therefore related to which layer is measured; it is therefore important to read a single proximity score as a summary of several more or less independent histories.

Second, it is in line with efficiency-based analyses of word-order correlations (Hahn et al., 2020). All five languages are harmonic: the four verb-final languages are mostly postpositional, as is the verb-medial language. The sample is insufficient for statistical testing of the account, but there is no counterexample found, and from the prepositions and circumpositions of Pashto, it is clear that harmony is not necessarily the rule.

Third, the two parameters along which the five languages agree, that is, overt aspect and gender agreement,

are the ones along which they differ most in mechanism, for typological method. In Russian case and aspect are independent of each other and aspect is largely lexicalized using a verb pair, while the Pakistani languages use case and agreement as a grammatical cue to aspect or tense. Similarly, in Russian, the gender system is three-gender and it is in interaction with case syncretism (Slioussar et al., 2022), while in the Pakistani languages, it is two gender system that is interacting with the alignment controller that shifts with case. Formal convergence is therefore captured in binary feature matrices, but these may overestimate functional similarity. This reinforces the idea that comparative concepts need to be separated from language-particular categories (Haspelmath 2010) and that a typological profile is in essence a multivariate rather than single score (Bickel 2007). These parameters could be further broken down in future architectural comparisons, such as lexical vs. grammaticalised aspect, and two gender vs. three gender.

LIMITATIONS AND FUTURE RESEARCH

This study has several limitations. First, the evidence base is restricted mainly to works published since 2020; standard reference grammars were consulted only for baseline features, so finer-grained differences among Urdu, Punjabi and Sindhi may be understated. Second, the feature matrix rests on a small number of coarsely coded parameters, and identical-code matching treats partial values as full mismatches. Third, other languages of Pakistan, including Saraiki, Balochi and Brahui, were not coded. Fourth, worldwide speaker numbers come from secondary tabulations of Ethnologue (2026), which differ on the Russian L1/L2 split, and Pakistan's census data are politically contested (Shah et al., 2022). Fifth, documentation is uneven across the languages (Table 3), particularly for Sindhi, and each language was coded for a single reference variety (Section 4.6), so dialectal variation is reported but not quantified. Future studies could draw on recent work on Urdu and Sindhi syntax, extend the sample, use corpus-based measures and test the pedagogical predictions in classroom or machine-translation studies.

CONCLUSION

This paper has compared the linguistic architecture of four Pakistani languages and Russian, drawing primarily on research, census data and reports published since 2020. The Pakistani languages share a South Asian profile of retroflexion, postpositions, SOV order and aspect- or tense-conditioned ergativity, with Punjabi tone and Sindhi–Saraiki implosives as local developments. Russian has a Slavic profile of prepositions, case syncretism, a perfective–imperfective verb system and a pervasive hard–soft consonant contrast. The ten-parameter matrix quantifies this contrast and shows that structural distance follows branch and areal affiliation rather than shared Indo-European membership: Pashto, an Iranian language, patterns with its Indo-Aryan neighbours, while Russian stands apart. The results provide a basis for teaching, translation and language technology across the two language spheres. In addition to description, the study introduces as a new concept the notion of linguistic architecture as a profile in multiple parameters, demonstrating that each such layer is subject to different patterns of change through areal and branch affiliation, local innovation, and apparent convergence on aspect and gender. The conclusions are tentative for Sindhi, which uses a lot of reference descriptions in its coding.

Ethical Considerations

Ethical Approval: Not applicable. This study did not involve human subjects or animals. It is based entirely on published secondary sources and publicly available census data, so ethical approval was not required.

Conflict of Interest: The authors declare that they have no conflict of interest.

Data Availability

Statement: All data analysed in this study are publicly available. Census data are available from the Pakistan Bureau of Statistics (https://www.pbs.gov.pk/sites/default/files/population/2023/material/pakistan_census_data.pdf); the census-based summary is available from ISSRA (https://issra.pk/insight/2025/unity-in-diversity-exploring-pakistans-linguistic-landscape/insight.html); speaker figures are available from Ethnologue

(https://www.ethnologue.com/insights/ethnologue200/); and baseline typological values are available from the World Atlas of Language Structures (https://wals.info). The coding sheet underlying the feature matrix (Figure 4) and the pairwise proximity scores (Figure 5) is provided in Appendix A.

REFERENCES

1. Adib, S. S., & Hood, H. M. (2025). A comparative analysis of the roles of subject and object in verb structure in English and Pashto. Balkh International Journal of Social Science, 1(2), 1–23. https://bjss.ba.edu.af/index.php/bjss/article/view/59
2. Atta, F., van de Weijer, J., & Zhu, L. (2022). Saraiki. Journal of the International Phonetic Association, 52(3), 541–561. https://doi.org/10.1017/S0025100320000328
3. Bailyn, J. F. (2012). The syntax of Russian. Cambridge University Press.
4. Basit, A., Azeemi, A. H., & Raza, A. A. (2024). Challenges in Urdu machine translation. In Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024) (pp. 44–49). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.loresmt-1.4
5. Bhatia, T. K. (1993). Punjabi: A cognitive-descriptive grammar. Routledge.
6. Bickel, B. (2007). Typology in the 21st century: Major current developments. Linguistic Typology, 11(1), 239–251.
7. Ćavar, M. E., & Lulich, S. M. (2021). Variation in the articulation of Russian stressed vowels and the mechanics of palatalization in consonants. Phonological Data and Analysis, 3(3), 1–44. https://doi.org/10.3765/pda.v3art3.31
8. Chernova, D., Slioussar, N., Alexeeva, S., & Bursov, K. (2022). Trans-paradigmatic syncretism in case form processing in Russian. Journal of Slavic Linguistics, 30(3), 1–16. https://doi.org/10.1353/jsl.2022.a923075
9. Corbett, G. G. (1991). Gender. Cambridge University Press.
10. Croft, W. (2003). Typology and universals (2nd ed.). Cambridge University Press.
11. Dryer, M. S., & Haspelmath, M. (Eds.). (2013). The world atlas of language structures online. Max Planck Institute for Evolutionary Anthropology. https://wals.info
12. Emeneau, M. B. (1956). India as a linguistic area. Language, 32(1), 3–16. https://doi.org/10.2307/410649
13. Ethnologue. (2026). What are the top 200 most spoken languages? SIL International. https://www.ethnologue.com/insights/ethnologue200/
14. Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. Proceedings of the National Academy of Sciences, 117(5), 2347–2353. https://doi.org/10.1073/pnas.1910923117
15. Haig, G. L. J. (2008). Alignment change in Iranian languages: A construction grammar approach. Mouton de Gruyter.
16. Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. Language, 86(3), 663–687. https://doi.org/10.1353/lan.2010.0021
17. Hussain, Q., & Mielke, J. (2023). Place typology and evolution of implosives in Indo-Aryan languages. Linguistic Typology, 27(2), 429–453. https://doi.org/10.1515/lingty-2022-0040
18. Ivani, J. K., Paudyal, N., & Peterson, J. (2020). Indo-Aryan – a house divided? Evidence for the east–west Indo-Aryan divide and its significance for the study of northern South Asia. Journal of South Asian Languages and Linguistics, 7(2), 287–326. https://doi.org/10.1515/jsall-2021-2029
19. Khan, A. Q., Xu, Y., & Sohail, A. (2020). Multidimensionality of tone in Pahari. Lingua, 245, Article 102923. https://doi.org/10.1016/j.lingua.2020.102923
20. Khan, A. R., & Kausar, G. (2021). Case valuation in transitive clauses: A comparative study of Punjabi and English syntax. Linguistics and Literature Review, 7(1), 1–15. https://doi.org/10.32350/llr.71.01
21. Khan, R. W., & Iqbal, M. (2021). On agreement and its interactions with case: A Lexical Functional Grammar perspective. Pashto, 50(661), 23–44. https://pashto.org.pk/index.php/path/article/view/167
22. Khan, T. A. (2023). An orthographic analysis of sound changing rules in the Urdu language. Linguistics and Literature Review, 9(2), 158–188. https://doi.org/10.32350/llr.92.07

23. Malik, T. G., & Khan, A. Q. (2018). Tone and its acoustic correlates in Punjabi. Kashmir Journal of Language Research, 21(2), 1–13. https://kjlr.pk/index.php/kjlr/article/view/134
24. Masica, C. P. (1991). The Indo-Aryan languages. Cambridge University Press.
25. Nihalani, P. (1995). Sindhi. Journal of the International Phonetic Association, 25(2), 95–98. https://doi.org/10.1017/S0025100300005235
26. Oh, S., Shaw, J. A., Durvasula, K., & Kochetov, A. (2024). Russian assimilatory palatalization is incomplete neutralization. Laboratory Phonology, 15(1). https://doi.org/10.16995/labphon.10185
27. Padgett, J., Bennett, R., McGuire, G., & Ní Chiosáin, M. (2024). Russian palatalization is a matter of the tongue body. Journal of Slavic Linguistics, 32(3), 1–25. https://doi.org/10.1353/jsl.2024.a950575
28. Pakistan Bureau of Statistics. (2023). 7th population and housing census 2023: Pakistan census data. https://www.pbs.gov.pk/sites/default/files/population/2023/material/pakistan_census_data.pdf
29. Pathan, H., & Bakhsh, I. (2025). Russian language in Pakistan: Course analysis and development pathways. Russian Language Studies, 23(4), 656–672. https://doi.org/10.22363/2618-8163-2025-23-4-656-672
30. Schmidt, R. L. (1999). Urdu: An essential grammar. Routledge.
31. Shah, A. A., Isani, U. H. S., & Shah, M. A. (2022). Sociolinguistic impacts on Sindh: A case study on the Sindhi language. Pakistan Social Sciences Review, 6(1), 197–206. https://doi.org/10.35484/pssr.2022(6-I)16
32. Slioussar, N., Magomedova, V., & Makarova, P. (2022). The role of case syncretism in agreement attraction: A comprehension study. Frontiers in Psychology, 13, Article 829112. https://doi.org/10.3389/fpsyg.2022.829112
33. Tegey, H., & Robson, B. (1996). A reference grammar of Pashto. Center for Applied Linguistics.
34. Waseem, S. (2025, April 24). Unity in diversity: Exploring Pakistan's linguistic landscape. Institute for Strategic Studies, Research and Analysis, National Defence University. https://issra.pk/insight/2025/unity-in-diversity-exploring-pakistans-linguistic-landscape/insight.html
35. Whitehead Martelle, W., & Shirai, Y. (2023). The aspect hypothesis and L2 Russian. Frontiers in Language Sciences, 2, Article 1119026. https://doi.org/10.3389/flang.2023.1119026

APPENDIX A. Coding Sheet For The Ten-Parameter Feature Matrix

Table A1. Feature codes by language (1 = present, 0 = absent, 0.5 = partial)

Parameter	Urdu	Punjabi	Sindhi	Pashto	Russian
1. Dominant word order (SOV = 1, SVO = 0)	1	1	1	1	0
2. Adposition position (post = 1, pre = 0)	1	1	1	1	0
3. Split ergativity	1	1	1	1	0
4. Lexical tone	0	1	0	0	0
5. Implosive consonants	0	0	1	0	0
6. Secondary palatalisation contrast	0	0	0	0	1
7. Retroflex series	1	1	1	1	0
8. Overt grammatical aspect	1	1	1	1	1
9. Rich nominal case	0.5	0.5	0.5	1	1
10. Gender agreement	1	1	1	1	1

Scoring: a pair of languages scores one point for each parameter on which their codes are identical. Resulting pairwise scores (Figure 5): Urdu–Punjabi 9/10; Urdu–Sindhi 9/10; Urdu–Pashto 9/10; Urdu–Russian 4/10; Punjabi–Russian 3/10; Sindhi–Russian 3/10; Pashto–Russian 5/10. Under a graded measure (10 minus the sum of absolute differences) the scores are 9, 9, 9.5, 4.5, 3.5, 3.5 and 5.