

## Prenominal relative clauses in Northern Tajik Persian: Analytic to synthetic morphology and a new contact perspective

Afsheen Sharifzadeh, Hossep Dolatian & Jonathan N. Washington\*

**Abstract.** We examine the evolution of synthetic relative clauses (RCs) in Northern Tajik Persian, driven by prolonged contact with Uzbek. While Standard Tajik Persian uses analytic postnominal RCs with the complementizer *ki*, Northern Tajik has synthetic prenominal RCs with non-finite verbs, with a high degree of structural parallelism with Uzbek RCs, suggesting they were acquired through contact. We explore the dynamics of metatypy in Northern Tajik and evaluate the sociolinguistic factors contributing to its grammatical ‘Turkification’. Initially, stable bilingualism in the vast rural contact zone favored the adoption of isomorphic structures for code-switching. Later, these varieties permeated Tajik-speaking cities like Samarkand and Bukhara, analogous to rural-to-urban shifts which diachronically brought Bedouin-type Arabic dialects into Baghdad. In modern times, these urban centers have become significant contact zones, further transforming Northern Tajik through Uzbek-medium education and media exposure.

**Keywords.** relative clauses; language contact; verb morphology; morphological typology; metatypy; sociolinguistics; historical linguistics

**1. Introduction.** Persian relative clauses (RC) are typically postnominal, use a finite verb, and have the complementizer *ki*. We treat this strategy as analytic because the marker of relativization is a separate word, *ki*. This is the only relativization strategy in Standard Tajik Persian (1).

- (1) Standard Tajik  
 odam-on-e ki me-rav-and  
 person-PL-DEF who IPFV-go-3PL  
 ‘People who go’

By contrast, Northern Tajik is a dialect of Persian spoken throughout Uzbekistan (~8 million people)<sup>1</sup>, including in the cities of Samarkand and Bukhara, and northern Tajikistan (Sughd province), including in the city of Khujand. Northern Tajik uses three kinds of RCs based on register. For the formal literary register, the inherited Persian formula is used (1), but it is also heard in colloquial speech that is invested with an intellectual or poetic quality. In the informal register, or Colloquial Northern Tajik, the RC is most commonly either (a) prenominal with a non-finite verb (participle or verbal adjective), constituting a ‘participial clause’, or is (b) postnominal with a non-finite verb and a linking suffix on the noun, or *ezafe*. We treat the former strategy (2a) as synthetic because the marker of relativization is the participle suffix on the verb, while the *ezafe* RC (2b) represents an analytic intermediary between (1) and (2a), since a separate linking

---

\* Authors: Afsheen Sharifzadeh, UMass Chan Medical School ([afsheen.sharifzadeh@gmail.com](mailto:afsheen.sharifzadeh@gmail.com)), Hossep Dolatian, Stony Brook University ([hossep.dolatian@alumni.stonybrook.edu](mailto:hossep.dolatian@alumni.stonybrook.edu)), & Jonathan N. Washington, Swarthmore College ([jonathan.washington@swarthmore.edu](mailto:jonathan.washington@swarthmore.edu)).

<sup>1</sup>Multiple experts, international commentators, as well as Tajiks within and outside of the republic suggest that there may be as many as eight to twelve million Tajiks in Uzbekistan, constituting between 20-30% of the republic’s 35 million population. See Foltz 2023.

particle is used between the noun and a non-finite verb.

- (2) Colloquial Northern Tajik
- a. me-raft-agi odam-o  
IPFV-go-PSPT person-PL
- b. odam-o-i me-raft-agi  
person-PL-EZ IPFV-go-PSPT  
'People who go'

Many speakers use both strategies with high frequency, although the *ezafe* RC may occur only in varieties east of Bukhara province. Notably, (2a) appears to be preferred among younger Uzbekistani Tajiks, most of whom have not received Tajik-language education (Ido 2007; Finke, & Sancak 2012). Self-stigmatization against its use exists on the grounds that it is the result of alien (un-Persian) influences. Conversely, speakers of Central and Southern Tajik dialects do not use either of the formulas illustrated in (2). While known to them from interactions with northerners, these structures are anecdotally perceived as ungrammatical or uninterpretable.

At first glance, (2b) appears more conservative since it uses an analytic *ezafe* construction inherited from Persian. But it, too, is unknown as a relativization strategy in any other Persian variety.<sup>2</sup> It further employs a series of augmented participles typical of Northern Tajik which function nearly identically to those of Uzbek rather than Classical Persian. This suggests that the *ezafe* strategy was possibly formulated later, or as a response to the synthetic strategy. Thus, (2b) may have emerged via hypercorrection of (2a) through prescriptive influences from the syntax of a more conservative dialect, or Classical Persian.<sup>3</sup> This occurred at a time when the augmented participles had already become indispensable to daily speech, but the prenominal position of the verb was perceived as an alien influence. This idea is supported by the observation that in Samarkand, the *ezafe* RC is frequently heard alongside the *ki* RC in contexts where effort is made to use 'pure' Tajik, to the exclusion of the synthetic strategy. In Khujand, the synthetic strategy appears to have been mostly leveled in favor of the *ezafe* strategy. A corresponding *ezafe* strategy does not exist in Uzbek, which only has the prenominal RC that uses a non-finite verb (3).

- (3) Uzbek
- ket-adigan odam-lar  
go-FTPT person-PL  
'People who go'

We hypothesize that contact-induced metatypy introduced synthetic RCs into Northern Tajik. This is because of the parallelism in structure between Uzbek and Northern Tajik synthetic RCs, the lack of this structure in any other Persian variety, and the alignment of Northern Tajik's use of participles with Uzbek. Data was acquired from public Telegram channels administered by Northern Tajik speakers based in Bukhara and Samarkand, Uzbekistan. We also double checked with native speakers. Uzbek is written in the orthography. Tajik data is rendered in transliteration (ALA-LC) but with <š, č, ø> instead of <sh, ch, ø>. We use Leipzig glossing

---

<sup>2</sup> While the formula NOUN + EZAFE MARKER – PAST PARTICIPLE exists in other Persian varieties, it does not function as a productive relativization strategy. Rather, the participle functions purely as an adjective describing the state of the noun; i.e. Iranian Persian: *sib-e leh šode* "a crushed apple", glossed as /apple-EZ crush be-PSPT/.

<sup>3</sup> Perhaps it was the prescriptive influence of a more conservative urban dialect—in all probability, that of Samarkand—which still had a productive *ezafe*. It thus may have emerged when 'Turkified' Tajik, with its prenominal RCs, penetrated the city, where monolingual urbanites likely resisted its alien word order.

rules in addition to-EZ (ezafe), PRPT (present participle) and PSPT (past participle).

## 2. Relative Clauses in Northern Tajik.

2.1 TYPOLOGICAL PROPERTIES OF SYNTHETIC RELATIVE CLAUSES. Northern Tajik participial clauses, like those in Turkic, are syntactically autonomous and function as adjectival phrases, inflecting like other nominal categories based on their syntactic roles. RCs are either headed by a nominal head (4a) or headless (4b). Headless RCs are permitted in Uzbek, but not in Standard Tajik.

- (4) Northern Tajik as spoken in Samarkand
- a. Azob kašt-agi odam  
 pain pull-PSPT person  
 ‘A person who suffered/has suffered’
- b. Tu-ba mexr dod-agi-ba du baravar ziyod mexr te  
 you-DAT love give-PSPT-DAT two times much love give.IMP  
 ‘To [the person] who gave you love, give twice as much love’

Participial clauses can serve as non-finite complement clauses (CC). Unlike standard participial clauses, non-finite CCs do not involve relativization (5). The participle agrees with the subject via a possessive suffix, as in Uzbek. While in RCs, the participle lacks subject agreement morphology.

- (5) Northern Tajik as spoken in Samarkand
- Vay odam aslašba man-ba dōst na-bud-ag-eš-a  
 That person in.reality me-DAT friend NEG-be-PSPT-3POSS-ACC  
 fam-id-am  
 understand-PST-1SG  
 ‘I understood that in reality that person has not been/was not my friend’

CCs introduced by *ki* are typical of Persian varieties, and they are also used in Northern Tajik (6), especially with cognitive verbs such as *donistan* ‘to know’. Colloquially, an Uzbek-style construction with a preposed subordinate clause is also used.

- (6) Northern Tajik
- Fam-id-am ki vay odam aslašba man-ba dōst ne  
 Understand-PST-1SG that that person in.reality me-DAT friend NEG  
 ‘I understood that in reality that person has not been/was not my friend’

In Northern Tajik, as in Uzbek, the subject constituent's case marking in RCs is categorized into two types. The first type involves a participle construction where the subject of the RC, like *pisar* ‘guy’, is in the nominative case (7). Compare with Uzbek (8).

- (7) Northern Tajik: Nominative relative clause
- Pisar dod-agi axirangi gul  
 Guy give-PSPT last flower  
 ‘The last flower that the boy gave/has given’
- (8) cf. Uzbek
- Yigit ber-gan oxirgi gul  
 Guy give-PSPT last flower  
 ‘The last flower that the boy gave/has given’

In the second type (9), the subject of the RC appears in the genitive case, as in *duxtara*

“girl’s” with a possessive agreement suffix attached to the head noun. These constructions are referred to as Genitive RCs and, to our knowledge, this is the first report of this type of RC in Northern Tajik, reflecting close structural alignment with Uzbek. These, like nominative RCs, can be based on any participle form.

(9) Northern Tajik: Genitive relative construction

Duxtar-a      navist-agi      kitob-aš  
 Girl-GEN      write-PSPT      book-3POSS  
 ‘The book that the girl wrote/has written’

(10) cf. Uzbek

Qiz-niņ      yoz-gan      kitob-i  
 Girl-GEN      write-PSPT      book-3POSS  
 ‘The book that the girl wrote/has written’

An alternative construction using the *ezafe* (as in 2b), which is unknown in Uzbek, exists in some varieties of Northern Tajik (11). In this formula, the subject of the RC is in the nominative case as in (7), but the object is proposed and linked with an *ezafe*. The subject may be specified explicitly or implicitly by a pronominal suffix on the participle.

(11) Northern Tajik: *Ezafe* relative construction

Kitob-i      navist-ag-eš  
 Book-EZ      write-PSPT-3POSS  
 ‘The book that [the girl] wrote/has written’

2.2. NOUN PHRASE ACCESSIBILITY HIERARCHY. Use of synthetic RCs is widespread in Colloquial Northern Tajik in a way that conforms with the Noun Phrase Accessibility Hierarchy (Keenan & Comrie 1977). The noun head of the RC can occupy different thematic roles in relation to the relativized verb. For example, (2a) has the noun interpreted as the subject of the RC, but a noun modified by a RC can also occupy an oblique role. Uzbek and Northern Tajik exhibit the same distribution regarding the syntactic status of relativizable nouns within the RC, which is also the same as described for Uyghur (Csató & Uchturpani 2010) but not Standard Tajik (Table 1).

	PRE-NOMINAL RC	HEADLESS ALLOWED	SUBJECT	DIRECT OBJECT	INDIRECT OBJECT	OBLIQUE ARGUMENT	POSSESSOR	OBJECT OF COMPARISON
UYGHUR	X	X	X	X	X	X	X	
UZBEK	X	X	X	X	X	X	X	
NORTHERN TAJIK	X	X	X	X	X	X	X	
STANDARD TAJIK			X	X	X	X	X	X

Table 1. Structural properties and possible thematic roles for the noun head

2.3. THE SEMANTICS OF PARTICIPIAL AFFIXES. The Northern Tajik augmented participles are in effect ‘replica functions’ for equivalent Uzbek morphemes, whereby the bilingual Tajik-Uzbek community gradually established an “interlingual equivalence of the morphemes or categories” without the borrowing of a single source language morpheme (Soper 1987:74). The Northern Tajik verb system has thus been modified to replicate nearly the entire Uzbek verb system (ibid.). This includes all the participial viewpoints and modalities found in Uzbek, the contact language.

RCs with *-agi* can be translated using various aspectual and temporal forms depending on context. Their interpretations correspond to a post-terminal viewpoint, like *-gan* in Uzbek (12).

- (12) Northern Tajik as spoken in Samarkand  
 Rəz-i            gandum-ba    man-a            patoft-a        raft-ag-o  
 day-EZ        bad.my-LOC   me-ACC        leave-CVB     go-PSPT-PL  
 ‘Those who abandoned me during my tough times’ (lit. ‘my bad day’)

The participle formed by *me*-[past stem]-*agi* conveys a non-focal intra-terminal meaning. This participle aligns in function with Uzbek *-(a)digan*. Depending on the context, it can be interpreted as referring to either an ongoing or a future event (13).

- (13) Northern Tajik as spoken in Samarkand  
 Vay    xudo-ba        na-me-forid-agi    koroy-a        mөл    me-kad-agi-s  
 He    God-DAT      NEG-IPFV-please-PSPT   actions-ACC    a.lot    IPFV-do-PSPT-EVID  
 ‘He apparently will do/does a lot of things that will displease/are displeasing to God’

The focal intra-terminal participle ending in *-(i)sudagi* (ultimately derived from *istod-agi* ‘stand-PSPT’) in Samarkand Tajik,<sup>4</sup> in contrast, primarily indicates situations or events that are currently in progress (14). This corresponds to the Uzbek participle ending in *-(a)yotgan*.

- (14) Northern Tajik as spoken in Samarkand  
 Raf-sudagi    yor-at-ba                    mon            na-gu  
 leave-PRPT   love-2POSS-DAT            stay.IMP        NEG-say  
 ‘Don’t tell your lover who is going (leaving) to stay’

There is rare and poorly studied participial viewpoint in Samarkand (including Judeo-Samarkandi) that has no Uzbek equivalent, termed ‘причастие настоящего определенного времени’ by Kerimova (1959). It is formed by *me*-[past stem]-*(i)sudagi*.<sup>5</sup> This adds an additional layer of imperfective aspect and/or intentionality to the focal intra-terminal participle.

The participle construction with *-ani šudagi* expresses a modal meaning that indicates the intention or will of the first actant. This construction has the same functions as the Uzbek *-moqchi bo’lgan*. Although *-ani šudagi* and *me*-[past stem]-*gi* can be used interchangeably for a future event, *-ani šudagi* is not used if the speaker does not wish to convey intentionality or will.

- (15) Northern Tajik as spoken in Samarkand  
 Raft-ani        šud-agi        odamo-ya        rah-aš-a            na-band-eton  
 leave-INT.PT   become-PSPT   people-GEN    way-3POSS-ACC    NEG-block-IMP.2PL  
 ‘Don’t block the way of people who want to go’

**3. Typological significance.** The use of isomorphic pronominal RCs in Uzbek and Northern Tajik represents a case of metatypy. Metatypy is when morphosyntactic change occurs because of multilingualism. Moreover, the relationship between Northern Tajik and Uzbek is broadly described as stable bilingualism, similar to a Sprachbund, due to their intense and prolonged contact which has resulted in a high degree of convergence. By contrast, the typical linguistic outcome of substratum/superstratum interference is limited lexical and phonological transference with the hallmark of simplification in morphology—approaching pidginization in extreme cases of interrupted transmission (McWhorter 2007). This clearly does not explain the extensive

<sup>4</sup> In Bukharan, the non-focal intraterminal participle is formed by the addition of the suffix *-(t)aštagi* (also ultimately derived from *istodagi*). See Ido 2007 for a full list of Bukharan participles.

<sup>5</sup> This appears analogous to the Bukharan “imperfective-progressive” participle mentioned by Ido (2007), which shares the same morphological elements as the Samarkandi participle.

structural borrowing and morphological complication seen in Northern Tajik (Thomason & Kaufman 1988:21;47; henceforth T&K). Such interference is well-documented in other rural contact scenarios, where the practical need for effective communication across different ethnic groups necessitates fluency in other languages while some ethnolinguistic boundaries are maintained (ibid.). Thus, Northern Tajik may be compared with Asia Minor Greek, another sub-branch of Indo-European that underwent agglutinative metatypy, including the acquisition of prenominal RCs, through L2 Turkic interference (ibid.:215). However, there is the important distinction that Asia Minor Greek generally held a subordinate position in its sociolinguistic context (ibid.:19) while Tajik varieties have been spoken in two large, prestigious administrative capitals, which is *prima facie* unexpected and demands more comprehensive investigation.

Moreover, the ancestor of Northern Tajik initially had only the analytic RC using the subordinating conjunction *ki*, inherited from Persian, but it later developed a synthetic strategy using non-finite verbs modeled on Uzbek RCs. We hypothesize that the use of synthetic RCs is due to two factors. One is the prolonged contact with Uzbek, a more agglutinative language, which took place in a large non-urban contact zone where stable bilingualism resulted in heavy structural borrowing into the more analytic language. The second is the reduction of linguistic ties with other Persian speakers following the conversion of the Safavid dynasty to Shi'ism at the start of the 16<sup>th</sup> century, which had the effect of removing many Northern Tajik speakers from 'correct' Persian input.<sup>6</sup> The distancing of the lects can further be observed in the retention of many archaisms in Tajik vis-à-vis Western Persian, although urbanites evidently continued composing music and literature in literary Persian. Further discussion on this topic extends beyond the scope of this article and will therefore not be elaborated upon here.

The typological distance between Persian and Uzbek was significant in conditioning the linguistic outcome in this contact scenario. While it is common for a mostly synthetic language to develop simplifactory changes because of internal drift or contact, particularly as a result of interrupted language transmission (McWhorter 2007; T&K 1988), it is relatively rare to observe the opposite process. Specifically, Northern Tajik exemplifies this uncommon phenomenon, having undergone widespread morphosyntactic 'complication' through contact, for example by incorporating a broad range of new tense, aspect, and modal categories into its verb morphology. We argue that the tendency for analytic languages to adopt L2 synthetic features in stable multilingualism is (1) due to the availability of 'free' or unbound morphemes that can be reorganized and (2) driven by ease of language switching, making speech processing and production more predictable and efficient in an enduring bilingual milieu. In other words, metatypy was driven—consciously or unconsciously—by reducing the cognitive load required for constantly switching between different grammatical systems. This was achieved by incorporating 'replica' structures from the contact language into Tajik (T&K 1988).<sup>7</sup> Crucially, convergence in stable multilingualism, like in a Sprachbund, tends to conserve complexity rather than drive simplification due to the robustness and resilience of the speech communities involved, in which languages are regularly transmitted from one generation to the next without interruption (ibid.:95).

---

<sup>6</sup> However, there are important exceptions to this, including the sizable Shi'a *Eroni* 'Iranian' community and the now-absorbed *Mavrigi* 'Mervian' community.

<sup>7</sup> The creation of 'replica' structures is seen in robust bilingual environments via L2 transference. For example, the analytic Western Persian dynamic modality formula CAPABILITY VERB + MAIN VERB (SUBJUNCTIVE) was, in all probability, borrowed into South Azerbaijani via bilingualism among L1 Turkophones, since bilingualism with Azerbaijani Turkish is hardly present among Persians. More importantly, few Persians live in the Iranian Azerbaijan region.

Beyond RCs and verbs, there are other areas of grammar where Northern Tajik has adopted structures from Uzbek (Doerfer 1991a). Standard Tajik employs prepositions to mark grammatical case. But due to contact with Uzbek, which uses suffixes, Northern Tajik varieties have developed suffixes from historical prepositions (see section 5.2). Additionally, Northern Tajik, like Uzbek, omits copula marking in the 3<sup>rd</sup> person present indicative tense, and the reporting of speech uses the postposed non-finite form *gufta/guftagi* ‘said’ (modeled on Uzbek *deb/degan*). Standard Tajik uses the traditional Persian genitive construction POSSESSED+EZAFE MARKER – POSSESSOR while the Central Asian Turkic pattern POSSESSOR+GEN – POSSESSED+POSSESSION AGREEMENT is widespread in Northern Tajik. These features are clearly not inherited from Persian and are unlikely to have arisen independently. In the next section, we examine the sociolinguistic dynamics that may have facilitated these grammatical transformations.

#### 4. Why Northern Tajik Looks So Un-Iranic

4.1 HISTORICAL CONTEXT OF PERSIAN AND KARLUK TURKIC IN CENTRAL ASIA. Tajik, as a dialect of Persian, is an Indo-European language that has been widely spoken in Central Asia since at least the 7th century AD. Previously, languages from the Eastern branch of Iranian, notably Sogdian and Chorasmian, were spoken in all the territory forming modern-day Uzbekistan. Persian, a Western Iranian language, appears to have taken hold in the major towns of Transoxiana at the expense of Eastern Iranian languages by the late Sasanian epoch and early Islamic period. It is thought to have been transplanted there by Sasanian aristocrats, bureaucrats, and soldiers from further west, presumably by way of Merv, and its status later solidified in Central Asia under the Samanid Dynasty in the 9<sup>th</sup> century (De la Vaissière n.d.). Of note, Central Asian (‘Bukharian’) Jews’ mother language since at least the beginning of the 9<sup>th</sup> century has also been Persian (Loy 2016).

In contrast, Uzbek<sup>8</sup> is a Karluk Turkic language descended from an Old Uighur dialect whose antecedents lie in the steppe lands to the north and east of its modern distribution (Doerfer, 1991a), making it genealogically distinct from Tajik and characterized by significantly different grammar and core vocabulary. Karluk Turkic has been spoken around Samarkand and Bukhara since approximately the end of the first millennium AD (ibid.). But Karluk Turkic (and other Turkic varieties that arrived in waves before and after it) has traditionally occupied different social and physical spaces from those dominated by Persian (Kontovas 2023). Although Turkic, in the form of the heavily Persianized Chaghatay language, gained some prominence in literature and administration during the Timurid era (mid-14<sup>th</sup> to early 16<sup>th</sup> centuries) in the capital city of Samarkand, Tajik has remained the prestige language and a key marker of identity for the Muslim and Jewish inhabitants of the Old City (ibid.), who continue to speak Tajik as a mother language. In Bukhara’s Old City, the population of which is also entirely Tajik-speaking, the historical written language was also generally Persian (Doerfer, 1991b).

4.2 FACTORS CONTRIBUTING TO STRUCTURAL BORROWING. Given their disparate origins and spheres of use, it is puzzling that the linguistic structure of Northern Tajik in these two cities closely resembles that of Uzbek rather than other Persian varieties. All Turkic languages, including Uzbek, are highly agglutinative (synthetic), i.e. words are modified using affixes and suffixes. Conversely, Persian is more analytic, using separate words for grammatical relationships and relying heavily on word order. Analytic languages are easier to acquire due to fewer morphological rules, aligning with cognitive load theories that simpler structures are easier to learn (Ellis, 2002). While synthetic languages often become more analytic through large-scale non-native

---

<sup>8</sup> Uzbek is the Turkic language that is spoken by the bulk of the Uzbek population. This does not include the Oghuz or Kipchak languages present in Uzbekistan.

language acquisition (e.g., Old English to Modern English), the reverse process—an analytic language becoming synthetic through contact—is rarer (McWhorter 2007). This process requires creating new inflectional and verb morphology and integrating them into everyday use, which demands a high level of linguistic proficiency and innovation which is less natural for learners. In most cases of substratum influence, then, shifting speakers adopt the target language's grammatical structures and vocabulary rather than causing it to converge morphologically and syntactically with their native language (T&K 1988). Generally, the acquired target language still reflects its genetic origins with minor to moderate traces from the shifting group (*ibid.*). In medieval Central Asia, however, linguistic shift primarily occurred from Iranian to Turkic, which does not conform to this model. Clearly then, while some grammatical structures may have been adopted through shift, something else happened to Northern Tajik.

Since neither language absorbed the other in rural Transoxiana, L1 Tajik speakers adapted to the long-term bilingual environment by restructuring their native vernacular to mirror the more complex morphology of Uzbek. This adaptation was achieved gradually, facilitating communication and easing the burden of employing two radically different grammars in mixed families, neighborhoods, and commercial settings. According to Thomason and Kaufman (1988), significant grammatical borrowing occurs when speakers learn another language intensively and use it regularly, incorporating L2 features into their native language. Thus, by imitating Uzbek morphology and syntax, the same sentence could be expressed in both languages by morpheme substitution. Interactions in the contact zones (such as agriculture, trade, intermarriage, etc.) fostered an environment of intense linguistic exchange in the countryside. The mutual borrowing of numerous kinship and nuptial terms between Uzbek and Northern Tajik highlights the role of intermarriage and thus bilingualism in this contact scenario.

To understand why Tajik necessarily received the bulk of structural borrowing, we must consider cognitive aspects of contact. The analyticity of Persian, characterized by the availability of 'free' or unbound morphemes, confers an underlying propensity toward agglutination even in the absence of external influences (McWhorter 2007). By contrast, Uzbek's greater morphemic boundedness, where morphemes are tightly integrated into the word structure, makes it inherently more resistant to reshaping. Further, speakers of the analytic language would have found it significantly more cumbersome to switch from a word-wise procession of their thoughts to a less intuitive morphological procession of the synthetic language present in their environment. In other words, an extended period of bilingualism catalyzed the analytic Tajik to undergo metatypy toward agglutination, modeled specifically on Uzbek rather than by random diachronic change, which aligns with accepted theories that borrowing in a Sprachbund is usually along lines that preserve complexity (T&K 1988). On the other hand, Uzbek spoken in proximity to Tajiks has undergone phonologic and lexical convergence with Tajik, traditionally attributed to substratum influence (shift), which is known to have occurred as consistent waves of incoming Turks made them numerically superior (*ibid.*; Soper 1987). However, the complete eradication of the Turkic vowel harmony rule in Uzbek and the adoption of a six-vowel inventory "almost identical down to phonetic detail to Tajik" (Comrie 1981:66), may also have been conditioned by the stable bilingual setting, since simplifying the vowel system to align more closely with Tajik undoubtedly made it easier to maintain proficiency in both languages. It is reasonable to conclude that vowel harmony, a highly systematic and pervasive phonological rule, is harder to integrate into a language that never had it. Consequently, the burden of convergence fell on Uzbek in the phonologic domain. The fact that no other living Turkic language, including Anatolian Turkish which spread gradually via shift, has entirely lost vowel harmony (rather than simply reducing it)

further suggests that the profound Tajik phonological influence on Uzbek involved processes beyond the substratum—namely, sustained Tajik input through stable bilingualism.

Thus, bilingual individuals recomposed Tajik grammar on the model of Uzbek out of convenience in light of enduring communications with Turkophones. This process likely began in the homes of Central Asia's agriculturalists and town folk, where social networks are dense and stable, and later spread to major cities. We discuss this process in the next section.

**5. Rural Dialects Supplanting Urban Dialects.** According to Soper (1987), the linguistic 'Turkification' of many Central Asian towns may be characterized as rural-centered diffusion, with the Iranian languages being displaced most rapidly in the countryside where the incoming Turkic tribes first settled. This attributes the massive rural shift to numerical superiority. After the Turks had acculturated to rural Iranian communities, it is reasonable to assume that neither language was particularly more prestigious in those settings. Hereto overlooked, we propose that this widely accepted diffusion model offers insight into the origin and fate of 'Turkified' Tajik dialects, too. We further attempt to explain how, without evidence of robust and sustained bilingualism in Samarkand and Bukhara cities until modern times, the Tajik varieties of these prestigious cultural and administrative capitals also underwent such extensive structural borrowing that they have been considered Turkic languages 'in embryo' (Doerfer 1967).

It will be argued that the 'Turkified' Tajik dialects of Samarkand and Bukhara cities likely emerged over two phases. First, a prolonged and intensive period of contact took place in the main contact zone, the countryside and small towns (Phase 1), which were historically home to the majority of Central Asia's population (Hugo 2019).<sup>9</sup> Later, these dialects permeated these two major cities in a restructured form that was typologically closer to Uzbek. These dialects thus possessed the capacity to mirror living Uzbek with high fidelity, particularly in modern times when urbanites became more proficiently bilingual (Phase 2). These interpretations are necessarily speculative, since the scarcity of historical materials in vernacular Tajik makes the endeavor of tracing individual borrowings with any degree of certainty effectively impossible. However, skepticism-based arguments, together with supporting evidence from the region and similar contact scenarios, allow us to outline this complex process within the established rural-centered language diffusion model. Furthermore, the rural-to-urban spread of 'Turkified' Tajik is reinforced by several factors: (1) historical settlement patterns of Turkic groups, (2) strong rural bilingualism and the absence of stable urban bilingualism, (3) possible linguistic remnants in urban dialects, and (4) sociolinguistic dynamics of the two urban centers.

5.1 HISTORICAL SETTLEMENT PATTERNS AND RURAL BILINGUALISM. First, it is known that beginning in the 10<sup>th</sup> century AD, bands of migrating Uighurs gradually settled in the vast Transoxianan and Khwarezmian plains where they interacted with sedentary Iranian speakers. There has been a prolonged age of contact through present times involving intensive agricultural and trade relations, shared social and cultural activities, and intermarriage. Because of this contact, many Persian-speaking communities shifted to Turkic, particularly in settings where Turkophones sufficiently outnumbered them. Thus, despite similar linguistic outcomes, bilingualism in this scenario differs from that of Asia Minor Greek, where it resulted in part from cultural pressure exerted by Turkish (T&K 1988:19). In contrast, Tajik-Uzbek bilingualism in rural areas is driven by practical necessity rather than the sociocultural dominance of one group over the other, since together they share a single religion, material culture, diet, and kinship structure (Finke &

---

<sup>9</sup> Indeed, even after massive urban growth in the late 20<sup>th</sup> century, the United Nations considered 36.3% of Uzbekistan urban by economy in 2010 (Hugo 2019).

Sancak 2012). This contrasts with the earlier shift to Persian among Eastern Iranian speakers in the region, which took place initially in the towns and only later in rural areas. Indeed, Kuzentsov noted in 1912 that the largest urban centers of Samarkand and Bukhara have remained “islands of Tajik-speaking inhabitants” through present times, around which the population in the rural areas is widely Uzbek-speaking (Soper 1987:58). Moreover, although shift occurred more slowly in the towns, Turkic eventually came to predominate in some of them, apparently due to elite dominance or their overwhelming numerical advantage over Tajik speakers (Ibid.).

Ethnographic research on the Bukhara Oasis by Finke and Sancak (2012) reveals robust multigenerational bilingualism among rural inhabitants. Rural areas of Bukhara, such as the municipality of Chilongu, exhibit a high degree of fluency in both Tajik and Uzbek among the population, so much that it is considered the social norm and those who are not bilingual are seen as ‘somewhat inadequate’ (ibid.:55). This bilingualism is evident in the frequent use of both languages in everyday interactions among both Uzbeks and Tajiks, with intensive intermarriage across linguistic lines having created a single speech community to the exclusion of other groups, such as Kazakhs and Eronis. For example, in the village of Hokimullomir, every household is bilingual, and code switching between Tajik and Uzbek on a daily basis is a fact of life among the residents (ibid.). This is quite remarkable, given that Tajik holds no official status in the country, and is evidence that Tajik-Uzbek bilingualism has long been a part of the local culture in these communities. As we shall see, the evidence from the cities is starkly different.

Thus, rural areas are well-established and expansive zones of bilingualism where language shift demonstrably occurred on a large scale. It stands to reason, therefore, that a multitude of Tajik-speaking enclaves—only a fraction of which remain—maintained their Tajik language within the bilingual environment rather than experiencing complete language shift. The simultaneous existence of various ‘Turkified’ Tajik varieties across different regions, employing unique postpositional suffixes (i.e. *-anda* “in” in the Ferghana valley for Samarkand *-ba* “in; to, for”) and distinct ‘replica structures’ of Turkic aspects, moods, and participles (see Rasturguyeva 1964)<sup>10</sup>, suggests an initially decentralized process with multiple loci of contact in rural settings. Rural communities are more isolated from centralized linguistic standards resulting in localized adaptations and innovations. Indeed, in the Middle Ages, Persian was still considerably more widespread than it is today, specifically in the Ferghana region, the Tashkent area, and the valleys of the Zarafshan, Kashkadarya, Surkhandarya, and many other areas, although sizable Tajik communities still exist in all these regions (Oranskiy 1960:238). Thus, while Turkic grew incrementally at the expense of Tajik, vast swathes of both urban and rural peoples retained their Iranian idiom, including those living in proximity to Turkophones.

5.2 SOCIOLINGUISTIC CONTEXTS AND EVIDENCE. Despite political transformations over the centuries that saw dynasties founded by both Tajiks and Turks in these cities, Samarkand and Bukhara maintained a cultural and social identity distinctly tied to the Persian language and Perso-Islamic culture. For example, in Bukhara province, the local Uzbek perception of Tajik is both as being attractive (described as *shirin* ‘sweet’) and the language of ‘urban culture’, and they are proud to speak it (Ido 2007). This continuity, which endures in both of the Old Cities where the inhabitants are almost exclusively Tajiks (ibid.; Kontovas 2023), must have provided resilience against complete language replacement. This reflects a complex identity where ethnic boundaries are

---

<sup>10</sup> E.g. the Tajik varieties of Chust and Kosonsoy in the Ferghana Valley construct the progressive aspect using the auxiliary verb *xoraftan* ‘to go to sleep, lie down’, modeled on Uzbek constructions using *yotmoq* ‘to lie down’, rather than constructions with *istodan* ‘to stand’ in Samarkand and Bukhara, modeled on Uzbek *turmoq* ‘to stand’.

maintained even as linguistic features are exchanged with other dialects. Moreover, the presence of a limited number of Turkic features in Iranian-speaking areas isolated from Turks, such as the Southern Tajik dialect of Kulob and the Pamiri languages (Borjian, 2014), suggests diffusion through Iranian intermediaries that had integrated Turkic elements rather than direct contact with Uzbek. This pattern is similar to the Balkan Sprachbund, where even monolingual communities adopted alien structures through bi- or multilingual intermediaries, despite not being in direct contact with the source languages. In addition, the Samarkandi and Bukharan Tajik “imperfect-progressive” participle is another piece of evidence that cities were not historically important bilingual zones. This participle is likely an internal innovation rather than L2 transference since it is morphologically akin to the other augmented participles, but no Uzbek equivalent exists. This unique verbal aspect probably arose among urbanites who spoke a ‘Turkified’ Tajik dialect but were not in contact with Uzbek, since it likely would have been subject to paradigm leveling.

On the other hand, major urban centers often maintain linguistic prestige and conservatism, making them less likely to adopt extensive structural changes from another language without significant external pressure or demographic shifts. But, despite being prestigious cultural hubs, cities can also be influenced by the dialectical patterns of incoming rural populations, especially if these migrants were sufficiently numerous and integrated into the socioeconomic fabric of the city. We cannot know when and how this happened in a sufficiently overwhelming surge to tip the scales in their favor. It was likely a gradual process, although larger population movements did occur as a result of political upheavals beginning in the early 13<sup>th</sup> century. Chief among these was the devastation of the Mongol invasions, which resulted in the destruction and depopulation of major cities including Gurganj, Merv, Samarkand, Bukhara, and Termez. Lamentably, according to ‘Atā-Malik Juvaynī, the inhabitants of Samarkand and Bukhara, like the other cities, “were all slain, none being spared”, thereby creating a demographic vacuum (Croner 2016). Economic opportunities and rebuilding efforts in these two decimated cities must have attracted rural populations. Later, massive Timurid and Shaybanid-era construction projects and then the Russian conquest and Soviet policies, too, led to significant urbanization of rural peoples. Thus, not unlike the case of Muslim Baghdadi Arabic, which experienced diachronic introduction of Bedouin-type features via rural influx (Palva 2009), the Tajik dialects of Samarkand and Bukhara, too, transformed with the input of people hailing from outside the confines of the city.

Moreover, the presence of a numerically limited group of allolingual urban elites (meaning an elite class that speaks a different language than the non-elites) has never been sufficient in any documented contact scenario to result in what happened to Northern Tajik. Superstratum influence typically brings about surface-level changes, such as loanwords and occasional phonological influences, yet substantial morphological and syntactic interference are unlikely to arise without extended and intensive bilingualism (T&K 1988). While modern Northern Tajik has acquired numerous Uzbek and Russian loanwords, the presence of thousands of Persian loanwords spanning all domains of Uzbek lexicon<sup>11</sup> is an indication that Tajiks were, at least in some spheres, socio-culturally dominant over their Turkic rulers (ibid.:116). Moreover, it is known that in Persian-Turkish symbiosis, the Persians were at first the giving part and the Turks the receiving one (Doerfer, 1991a). This is evident in the Turks’ official use of Persian for literary and administrative purposes (14<sup>th</sup>–20<sup>th</sup> centuries) and the extensive incorporation of Persian lexical, morphological, prosodic, and syntactic features into Chaghatay, which many authors described as ‘a Turkic-like language’ or a ‘tempered form of Persian’ (ibid.). By contrast, nearly no

---

<sup>11</sup> The ratio of Tajik Persian words in Uzbek was approximately 2:3 in the 19<sup>th</sup> century (Doerfer, 1991a).

traceable Turkic influences entered literary Persian in the same era (Doerfer, 1991b), underscoring the sociolinguistic reality that Chaghatay functioned as a literary language of the Turkic elite within a Persianate framework and alongside Persian, while Persian remained dominant in urban life, scholarship, and religion. This superstratum scenario may be compared with the example of the Germanic conquests of southern France and Spain, where the Germanic conquerors, eventually abandoning their native languages, left limited lexical borrowings in the Romance TLs, mostly confined to military terms and the like (T&K 1988:121). Even more relevant to our discussion are the cases of Safavid Isfahan and Qajar Tehran, where the presence of Persianized Turkic-speaking elites, as well as communities of Turkic peasantry, left behind clusters of lexical items but no hint of Turkic morphological or syntactic influence on the local Persian. The elite Turks' superstrate influence on Tajik then, like that of the Normans on English, would have been rather superficial compared to the more widespread and consistent bilingualism in rural areas where the two languages held, and continue to hold, more equal sociolinguistic footing.

Beyond the lack positive historical evidence, the hypothesis that transient (since they were eventually absorbed) Turkophone minorities in Samarkand and Bukhara cities introduced a plethora of complicating rather than simplifying changes to the target language (TL)—and that this group's shared learning errors were adopted by all L1 Tajik speakers—is improbable on both cognitive and sociolinguistic grounds. First, as discussed, the advanced proficiency required to replace the grammatical system of a TL using only TL morphemes would imply high-fidelity language acquisition by learners at a young age. This would theoretically protect against radical morphosyntactic restructuring of the TL, and thus the scale of the observed changes clearly cannot be attributed to shifting speakers' imperfect learning. There is further no reason to believe that urban Tajiks would have felt particularly motivated to emulate the Turkophones' rendition of their mother tongue and, less likely yet, a version that had been mutated on such a drastic scale. As discussed, one reason for this stability may be the economic importance of these Tajik-speaking cities but also the prestige of Classical Persian as the principal medium of science, literature, and administration within them under both Tajik and Turkic dynasties (Soper 1987:58).

Second, several analytic elements present in the dialects of these two cities must be relics from an earlier time when there were no Turkic-influenced case suffixes. These include the dative/locative preposition *da* in Bukharan, the privative prefix *be*, the typical Persian spatial preposition *a(z)* (which has no postposed variant but is being replaced by a new suffix, *-ban*, in Bukhara), and the causal preposition (*az*)*baroyi*. The comitative *qati*, now a circumposition, curiously retains the *ezafe* particle *-i* as a postposition, suggesting that it became transposed at a time when the *ezafe* was no longer very productive. As we have also seen, the possible vestige of an early leveling event lies in the form of the *ezafe* RCs (2b and 11), which may have arisen through hypercorrection of the 'Turkified' prenominal RC. Other clues, such as the common Samarkandi phrases *az aft(aš)*, lit. 'from (its) face', *ba fikram* 'to my thought', and *ba xayolam* 'to my imagination' used broadly in expressing perception, appearance, possibility, or opinion, may descend from the older analytic layer before the epistemic modality and evidentiality became embedded within verbs—another synthetic transformation modeled on the complex Uzbek system. That these phrases contain lexicalized prepositions further reveals their antiquity.

Third, and quite crucially, reports of poor Uzbek proficiency among older and female city residents casts serious doubt on the idea that Uzbek bilingualism was a historical reality in these cities (Kerimova 1959:4). All Samarkandi and Bukharan Tajik informants reported that their grandparents' generation, born in the early 20<sup>th</sup> century, had poor knowledge of Uzbek, but many had been educated in Tajik with Russian as a second language. Russian further does not appear

to have replaced Uzbek, since when the Turkologist Radlov visited Samarkand as early as 1868, he remarked that the people spoke Tajik exclusively (Bergne 2007:15). Of note, in Finke and Sancak's case study, it is stated of an individual born in 1937 in Hokimullomir village in Bukhara province that his "mother was from the city of Bukhara and—as many urban dwellers at that time—never spoke a single sentence of Uzbek in her life" (2012:59). In fact, she seems to have held a dislike for the language, as she always urged others to "not speak Turki" in her house (ibid.). All of this implies that, with exceptions, urbanites were monolingual in the urban prestige language: Tajik. The currency of Tajik throughout both oases, where it endures as a marker of status, instead imposed the necessity of bilingualism on Uzbeks (Ido 2014). In Bukhara in particular, Uzbeks from surrounding towns are largely bilingual in Tajik—quite startling for a language with no official status. Finke reports of incoming Uzbeks from other regions needing to learn Tajik to live comfortably in Bukhara city (2014:82). This, coupled with the reported stigma against speaking Uzbek<sup>12</sup> among some urbanites despite its new role as the official language, refutes the perception of Uzbek as a historically dominant first language or urban *lingua franca*.

This dynamic, whereby Turkic immigrants to the city were compelled to adopt Tajik for social recognition and status, is hardly suitable for the type of intensive and stable urban bilingualism that could upend and replace the entire verbal, morphological and syntactic proto-systems of a prestige language (T&K 1988). Instead, our analysis highlights relevant language acquisition theories and a sociopolitical context where Tajik Persian has held significant cultural and linguistic sway within these cities throughout the ages (Finke & Sancak 2012:55), and was therefore unlikely to be transformed on such a drastic scale as a result of shifting Turkophone elites or peasantry. The subsequent implication of this idea—that 'Turkified' Tajik was born in cities and then spread to the environs—is likewise untenable since the predominance of Uzbek-speaking villages in the vicinity of Bukhara and Samarkand would be wholly unexpected.

Taken together, this challenges the unsupported notion that urban environments were historically important sites for the 'Turkification' of Tajik. Instead, the argument by incredulity is that, like the diachronic Bedouinization of Muslim Baghdadi Arabic, significant demographic shifts transplanted suburban peoples into Samarkand and Bukhara, leading to the displacement and transformation of older lects. This hypothesis aligns nicely with language distributions and the well-known phenomenon that Turkic came to dominate in some Central Asian towns via rural influx, and is therefore grounded in theory and historicity. It further avoids an urge to force the facts to fit in any remotely satisfying way to see these traditionally Tajik-speaking cities as important stages for the 'Turkification' of Northern Tajik grammar.

5.3. SECOND PHASE OF TURKIFICATION IN CITIES. The 'second phase' of Turkification began in the Soviet era and continues today. During the Soviet period, Turkic nationalism in the Uzbek SSR was promoted to align with Soviet ideological goals, encouraging the use of Uzbek in education, administration, and public life, which diminished the status and usage of Tajik Persian despite its historic significance (Foltz 2023). By the end of the Soviet period, there were no Tajik language schools or publications in the Uzbek SSR (ibid.). Tajik speakers increasingly adopted Uzbek, the only official language since 1991, for socio-political and economic advantages. Under President Karimov's long reign (1991-2016), Tajik language and identity were routinely repressed. Today, Uzbek's dominance is further reinforced through media such as music, movies, and TV

---

<sup>12</sup> One informant from an Uzbek-speaking village in the Samarkand oasis reports a lamentable episode where his grandfather was harassed as a teenager for speaking Uzbek in the streets of the city. Another Samarkandi Uzbek informant, fluent in Tajik, criticized 'Tajik nationalism', which he defined as "a mentality against Uzbek".

programs. Thus, while rural-to-urban migration historically introduced Turkified elements into these cities, the post-Soviet shift towards Uzbek-medium education, digital media, and increasing Turkic immigration have made Bukhara and Samarkand robust contact zones.

Metatypy continues in this setting, as shown by the ongoing process of suffix formation. For instance, the now ubiquitous Bukharan ablative suffix *-ban*, not mentioned by Kerimova (1959) or Rastorguyeva (1964), must have emerged after their research. This suffix likely derives from the merged dative/locative suffix, *-ba*, augmented by a final coronal consonant, and is used frequently in lieu of the inherited preposition *a(z)*. The final coronal is the only phonological distinction in the corresponding Uzbek suffixes (cf. Uzbek *-dan* and *-da*), making *-ban* a convenient match for code-switching. Additionally, the Tajik dative/benefactive preposition *baroi* is now often heard as a postposition among the youth (*baruy* in Bukhara and *baro(ya)š* in Samarkand), mirroring Uzbek *uchun*. These examples illustrate the ongoing process of metatypy in Northern Tajik, of which there are many more (Sharifzadeh forthcoming).

**6. Conclusion.** The grammatical transformation of Northern Tajik under the influence of Uzbek exemplifies a case of heavy structural borrowing and contact-induced metatypy through L2 interference. This process, driven by stable bilingualism in rural Transoxiana, similar to the contact between Asian Minor Greek varieties and Turkish, led to the integration of Uzbek's agglutinative structures into the analytic framework of Tajik. Moreover, our study underscores the typological significance of Turkic RCs in Northern Tajik, illustrating how stable bilingualism rather than substrate/superstrate resulted in the adoption of 'replica structures' for cognitive efficiency and the ease of code switching. Focusing on rural areas as the primary setting for the 'first phase' of metatypy challenges traditional metrocentric views and offers new insights into the significant role of rural-urban dynamics in Central Asian linguistic history. This perspective proposes that rural-centered diffusion, not unlike the diachronic Bedouinization of Muslim Baghdadi Arabic through massive rural influx (Palva 2009), introduced elements of rural Tajik speech into the vernaculars of these two prestigious urban centers. In the modern era, Samarkand and Bukhara have experienced a 'second phase' of Turkification, driven by the decline of Tajik-medium schools, the influence of Uzbek-language media, and increased domestic migration.

Despite these transformations, Persian (Tajik) remains a key cultural and social identity marker in these two cities (Samarkand and Bukhara). The future of Persian in Uzbekistan appears bright, where younger generations use it in digital communication. Recent softening of language restrictions from the government has allowed a resurgence of Tajik-language programming and increased collaborations with Tajikistan, Iran, and Afghanistan. This has fostered an environment where most nationally acclaimed Uzbekistani artists, both ethnic Uzbeks and Tajiks, have multiple Persian songs in their repertoire. In Bukhara and Samarkand, there is growing interest among youth in speaking *adabi*, or Standard Persian, which they attempt to learn independently. However, without official status for Tajik or the revival of Tajik-medium schools, synthetic features such as the RC will likely persist in the living language.

## References

- Bergne Paul. 2007. *The birth of Tajikistan: National identity and the origins of the republic*. London and New York: I.B. Tauris.
- Borjian, Habib. 2014. Kulābi dialect. In *Encyclopædia Iranica*, online edition. <http://www.iranicaonline.org/articles/kulabi-dialect>.
- Comrie Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.

- Croner, Don. 2016. *Chingis Khan Rides West: The Mongol invasion of Bukhara, Samarkand, and other great cities of the Silk Road, 1215-1221*. Ulaanbaatar: Polar Star Books.
- Csató, Éva Ágnes & Muzappar Abdurusul Uchturpani. 2010. On Uyghur relative clauses. *Turkic languages* 14. 69-93.
- De la Vaissière, Etienne. n.d. The origin of New Persian.
- Doerfer, Gerhard. 1967. *Türkische Lehnwörter im Tadschikischen*. Wiesbaden: Steiner.
- Doerfer, Gerhard. 1991a. Central xiv. Turkish-Iranian language contacts. In *Encyclopædia Iranica*, vol. 3, 226-235. <http://www.iranicaonline.org/articles/central-asia-xiv>.
- Doerfer, Gerhard. 1991b. Chaghatay Language and Literature. In *Encyclopædia Iranica*, vol. 4, 339 – 343. <https://iranicaonline.org/articles/chaghatay-language-and-literature>.
- Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2). 143-188. . <https://doi.org/10.1017/S0272263102002024>.
- Finke, Peter. 2014. *Variations on Uzbek identity: Strategic choices, cognitive schemas and political constraints in identification processes*. 1st ed., Berghahn Books.
- Finke, Peter & Meltem Sancak. 2012. To be an Uzbek or not to be a Tajik? Ethnicity and locality in the Bukhara Oasis. *Zeitschrift Für Ethnologie* 137(1). 47–70.
- Foltz, Richard. 2023. *A history of the Tajiks: Iranians of the East*. I.B. Tauris.
- Hugo, Graeme. 2019. Patterns and trends of urbanization and urban growth in Asia. In Kankesu Jayanthakumaran, Reetu Verma, Guanghua Wan & Edgar Wilson (eds.), *Internal migration, urbanization and poverty in Asia: Dynamics and interrelationships*, 13-45. Singapore: Springer. [https://doi.org/10.1007/978-981-13-1537-4\\_2](https://doi.org/10.1007/978-981-13-1537-4_2).
- Ido, Shinji. 2007. *Bukharan Tajik*. München: Lincom.
- Ido, Shinji. 2014. Bukharan Tajik. *Journal of the International Phonetic Association*, 44(1), 87-102. <https://doi.org/10.1017/S002510031300011X>.
- Keenan, Edward & Comrie, Bernard 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.
- Kerimova, A. A. 1959. *Govor tadjikov buxary*. Moskva: Izdatel'stvo Vostonoy Literatury.
- Kontovas, Nicholas. 2023. Language and transformation in Samarqand. <https://www.ias.asia/the-newsletter/article/language-and-transformation-samarqand>.
- Loy, Thomas. 2016. *Bukharan Jews in the Soviet Union: Autobiographical narrations of mobility, continuity and change*. WiesbadenReichert Verlag.
- McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford University Press.
- Oranskiy, Iosif. 1960. *Vvedeniye v iranskuyu filologiyu*. Moskva: Izdatel'stvo vostochnoy literatury.
- Palva, Heikki. 2009. From Qəltu To Gələt: Diachronic notes on linguistic adaptation in Muslim Baghdad Arabic. In Enam Al-Wer and Rudolf de (eds.), *Arabic dialectology*, 17-40. Leiden: Brill. <http://dx.doi.org/10.1163/ej.9789004172128.i-298.12>
- Rastorgueva, V. S. 1964. *Opyt sravnitel'nogo izucheniya tadjikskix govorov*. Moskva: Izdatel'stvo Nauka.
- Soper, John David. 1987. *Loan syntax in Turkic and Iranian: The verb systems of Tajik, Uzbek, and Qashqai*. University of California at Los Angeles dissertation.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics* University of California Press. <https://doi.org/10.2307/jj.8501341>.