



TAJIK II. TAJIK PERSIAN

TAJIK

ii. TAJIKI PERSIAN

Tajik Persian (self-designation [*zabon-i tojik*, *tojikī*, *forsī-i tojikī*; also called Tajiki, Tojiki, Tadjhik) is the variety of New Persian used in Central Asia. From the 1920s it was officially fostered in the USSR as the national literary language of the Tajik SSR (since 1991, the Republic of Tajikistan). It is also spoken in parts of Uzbekistan (notably in the cities of [Bukhara](#) and [Samarqand](#)), and is the vernacular of the Bukharan Jews. Tajik is the common written language and contact vernacular (usually called *fārsi*) in the province of Mountain [Badakṣān](#), where most people speak one of several Eastern Iranian languages (Pamir languages). It is widely spoken natively in southern Badakṣān and in northeastern parts of Afghanistan. The so-called Tajiks of southwestern Xinjiang in China are speakers of Sarikoli or Wakhi, and not of Tajik. Tajik speakers number at least 5 million in Tajikistan, and several million elsewhere in the region and abroad (chiefly Russia, western Europe, and the United States).

The literary language, written in a modified Cyrillic script, was to a great extent standardized under Soviet tutelage, as was the educated speech in the capital city of [Dushanbe](#). These varieties are characterized by a large Russian component in the vocabulary, and a number of Uzbekisms in the syntax. Spoken Tajik in other regions of Tajikistan is represented by several dialect groups, divided roughly into northern and southern varieties. Northern



dialects are all more or less strongly influenced by Uzbek, while southern dialects are less so, being generally closer to the spoken Persian of adjacent areas of Afghanistan, including Kāboli (see [AFGHANISTAN v. LANGUAGES](#); Rastorgueva, 1964, pp. 146-48, 151-53). Tajik preserves some archaic elements, particularly of phonology and the lexicon, which have disappeared from the Persian of Iran, and displays a number of innovations in syntax and lexical derivation and composition.

History. Early New Persian (*pārsi-e dari*), a continuation of spoken Middle Persian, spread to Central Asia during the 8th century CE as the language of Iranian converts to Islam who were attached to the invading Arab armies. The Samanid rulers of Bukhara (9th-10th centuries) patronized it as the literary language, in which form it soon spread throughout Iran. In the region of Samarqand it displaced Sogdian, the indigenous Iranian language, whose descendant (Yaghnobi) still survives in the mountains of western Tajikistan. As a written language, Persian of Central Asia was hardly distinguishable from Classical Persian of Iran, Afghanistan, and India up until the early 20th century. From Timurid times (15th century), Indo-Persian was modeled on the writing of Central Asia; this influence can still be seen in some vocabulary common to Tajik and Indo-Persian/Urdu, which is not usual in Persian of Iran. Invasions and settlement of Turkic peoples in the Oxus basin and its foothills during the past one thousand years (most recently, the Uzbeks) interrupted the dialect continuum. Spoken Persian of Central Asia evolved independently of Persian of Iran, and northern dialects in particular were strongly influenced by Turkic speech. Persian speakers of the region came to be called Tajiks, in contradistinction to Turks, but their language was still called *fārsi* ‘Persian’ until the Soviet period.

The Uzbek Emirate of Bukhara, which ruled most of the Persian-speaking regions of the Oxus basin and the Pamirs since the middle of the 18th century, was reduced to a dependency of imperial Russia in 1868. After a Bolshevik-aided revolution at Bukhara in 1920, in accordance with Soviet nationalities policy, an ethnic Tajik Soviet republic was established, and a literary language called “Tajik” was engineered on a vernacular base close to the Uzbekized spoken Persian of Bukhara and Samarqand (these Tajik cultural centers, ironically, were incorporated into the Uzbek SSR). Under the guidance of writers who were mostly of Bukharan or other northern origin, such as Sadriddin Ayni (Şadr-al-Din ‘Ayni, 1878–1954; see [‘AYNI, ŞADR-AL-DIN](#)), this language became the vehicle of a considerable native literature and a lively

periodical press. From 1926 to 1939 a modified Latin alphabet was in use, and a concerted educational campaign produced impressive gains in adult literacy. In 1939-40 the writing system was switched to Cyrillic, as was the case in the Turkic-language republics of Soviet Asia (see Rzehak, esp. pp. 222-48, 312-33). The Bukharan Jews, with a Tajik literature using Hebrew script, developed a slightly different Latin alphabet and used it until 1935, when they adopted the common Tajik Latin alphabet until 1940 (TABLE 1).

During the period of ca. 1948-88, Tajik lost much of its prestige, vocabulary, and domain of use to Russian. With *perestroyka* and *glasnost'* in the 1980s came a sudden revival and re-Persianization of the national language, which continues (at a slower rate) in post-Soviet Tajikistan. Policies legislated by the Language Laws of 1989 and 1992 included the official use of Tajik in government and public domains, replacement of Russian vocabulary by Persian (both native coinages and copies from Persian of Iran), and teaching of the Perso-Arabic writing system in schools (Perry, 1996).

Phonology and Orthography. The Tajik sound system is shared almost entirely with that of Uzbek. The Cyrillic orthography is based on Russian usage. Modified characters, representing consonants not found in Russian, are: Cyrillic letter “ghe” with stroke (Ғ) for [ɣ] (*ḡeyn*), Cyrillic letter “ka” with descender (қ) for [q] (*qāf*; note that [ɣ] and [q] are distinct phonemes in Tajik while they have merged into one phoneme in standard Persian); Cyrillic letter “ha” with descender (ҳ) for [h] (both *h* and *ḥ*), and Cyrillic letter “che” with descender (ҷ) for [dʒ] (*j*). The Cyrillic “hard sign” (Ђ) represents the glottal stop corresponding to written *ʿeyn* or *hamza*, and the Cyrillic “soft sign” (Ь), which was used in accordance with Russian rather than Tajik orthography, has been dropped from current Tajik. Also abolished in the spelling reform of 1998 was the Cyrillic letter “tse” (Ц), which has been replaced by *s*.

Among the vowels, Cyrillic letters “u” and “i” have their variants with macrons (transliterated as *ū* and *ī* respectively). These accents have two different functions: *ū* represents a different quality from *u*, namely the vowel *u* (see below, and FIGURE 2), while *ī* is used for *i* in word-final position to distinguish a stressed morphological syllable from the unstressed enclitic of *izofat* (Pers. *eẓāfa*). Words followed by the *izofat* and other enclitics, prefixes, suffixes, and all compound words are conventionally written as single words (in this article, the morpheme divisions are shown in examples by hyphenation).

Cyrillic letter “o” represents the Persian back vowel written in Perso-Arabic



with *alef*, which in Tajik is more rounded than in standard Persian (thus Tajik *kitob* is Persian *ketāb*). The yotated vowels of Russian, “yo,” “yu,” and “ya,” represent the Tajik syllables *yo*, *yu*, and *ya* (*bayon*, Pers. *bayān* ‘declaration’); Cyrillic letter “e” stands for *e* after a consonant, but also for *ye* initially or after a vowel (*oed*, Pers. *āyid* ‘come!’); Cyrillic letter “è” (ə) stands for initial *e*.

Tajik consonants correspond to those of Persian (but note the distinct *ġ* and *q*). The liquid [r], in Tajik a flap as in Persian, tends to be dropped before a dental in many dialects (as in Afghan Persian): *kadam*, *kadum* ‘I did,’ Pers. *kardam*. Labiodental [v] tends toward bilabial [β] or [w] in the environment of rounded vowels: *gov*, *gow* ‘cow,’ Pers. *gāv*. Both the diphthongs are fully preserved in Tajik: *mayda* ‘small,’ *qawm* ‘clan.’ In written Tajik, the Cyrillic letter “ve” (v) is used for the second component of the diphthong *aw/au*, but this is often realized as *w*, especially in final position or before a labial: *qavl* ‘promise’ and *rav* ‘go!’ are generally heard as *qawl* and *raw*, but the plural of the latter is *raved*.

Also preserved is final *a*: Taj. *šuda* ‘having become,’ Pers. *šode* (the Tajik vowel is not raised to *e*, except in the names of the days of the week ending in *šanbe*). Historically, *e* and *ū* continue the so-called *majhul* vowels of early New Persian (*ē* and *ō*), which in Persian of Iran merged with ‘long’ *i* and *u* (see [FIGURE 1](#)). They also occur as raised allophones of *i* and *u* before *h* or the glottal stop, as in Taj. *ehtimol* ‘probability, probably’ and *šū’la* ‘flame.’ In addition, *ū*, sounding between *u* and *ü* (phonemic *u* in [FIGURE 1](#)), is shared with Uzbek, in which it corresponds to Turkic *ü* or *ö*.

Tajik has thus reduced the inventory of eight vowels in Middle and early New Persian to six, but in a quite different way from Persian of Iran ([FIGURE 1](#), [FIGURE 2](#)). Length has been neutralized in most dialects (including literary Tajik), and it is replaced by a contrast between ‘stable’ vowels [e, o], which retain their quality in stressed and unstressed positions, and ‘unstable’ vowels [i, u, a], which tend to be reduced in duration and/or quality in unstressed positions (Sokolova, pp. 76-77; Lazard, pp. 127-29). In written Tajik, *i* and *u* thus stand for both the ‘long’ vowels of Persian (as written in Arabic script with characters *yā* and *vāv*, respectively) and the ‘short’ vowels (unwritten in Perso-Arabic). A consequence of the non-distinction of vowel length in the phonographic alphabet (one character to one phoneme), devised by Soviet scholars, is that the classical Persian verse, where in accordance with the system called ‘*aruḏ*’ the meter conventionally depends on alternation of long and short syllables, cannot be appreciated or composed by anyone unfamiliar



with the Perso-Arabic writing system.

Nominal Morphology and the Noun Phrase. Tajik retains final *-y* (as seen in classical Persian) in nouns with a final vowel *ū* or *o*, though sometimes as an optional variant: *mūy* ‘hair’, *poy* ‘foot, leg.’ Some personal and demonstrative pronouns are distributed differently from Persian, as follows:

vay (general), *ū* (literary), *in* (dialect) ‘he, she;’

on (literary), *in*, *[h]amin* (colloq.), *vay* (dialect) ‘it;’

on[h]o (general), *in[h]o* (colloq.), *vay[h]o* (dialect) ‘they’ (all classes).

The deferential pronoun *ešon* for 3rd person plural (‘he, she,’ lit. ‘they;’ cf. Pers. *išān*) evolved into an honorific title for religious notables, and has been replaced in Tajik by *in kas* ‘this person.’ The other plural pronouns, which may refer deprecatingly or deferentially to a singular person, can add plural suffixes as ‘explicit plurals’: *mo* ‘we/I’ becomes *mo-yon*, *mo-ho*, *mo-hon* ‘we;’ *šumo* ‘you’ (sg. or pl.) becomes *šumo-yon*, *šumo-ho* ‘you (pl.);’ cf. verb endings, below. In addition to *in* ‘this’ and *on* ‘that,’ *vay* can also function as a demonstrative pronoun or adjective, with a rhetorical nuance, e.g., *qin budagi-st*, *vay kor* ‘it must be tough, that job’ (see Lazard, p. 137; for the Conjectural Tense, see below). The pronominal enclitics also have extended idiomatic uses, e.g., to bind a complex noun phrase such as a reduced relative clause: *mard-i mūy-aš safed* ‘the white-haired man,’ lit. ‘the man whose hair is white.’ Pronouns *yagon*, *kadom*, *kadom yak* ‘some– (or other),’ ‘any– (at all)’ are indefinite qualifiers peculiar to Tajik: *yagon rūz ba kōna-i mo marhamat kuned* ‘please come and visit us some day;’ *vay bo kadom sabab-e javob nadod* ‘for some reason he didn’t answer.’ Similar is a series of Uzbek-Tajik hybrids formed on Uz. *kim* ‘who?’: *kim-kī* ‘someone or other, anyone,’ *kim-kujo* ‘somewhere, anywhere,’ *kim-kadom* ‘some– or other,’ etc.: *vay az kim-čī norozī ast* ‘she/he is unhappy about something.’

The postposition *-ro* fulfills the basic Persian function of marking the definite direct object. In addition to this, it (or rather its colloquial and dialect reflexes *-a* and *-ya*) can be found in several other uses. It may occur after nouns governed by prepositions: *baroi kī-ro?—baroi man-a* ‘for whom?—for me;’ *a [az] kōndidan-a murdem* ‘we died laughing.’ In northern dialects, a construction using *-ro* widely replaces the Persian-style possessive *izofat*: *man-a pīsar-am* ‘my son,’ *muallim-a kitob-aš* ‘the teacher’s book.’ The word order



here is that of the equivalent Uzbek phrase, *muallim-ning kitob-i*, lit. ‘of-the-teacher his-book.’ The predicative possessive pronouns ‘mine, yours, his,’ etc. are expressed by the prepositional phrase *az on* ‘from that (of)’ in *izofat* construction with an independent pronoun: *in pūl az oni kī ast?—az oni mo-st* ‘whose money is this?—(it is) ours.’

Prepositions include *bar* ‘upon’ and *be* (Persian *bi*, ‘without’), which in standard Persian are no longer used except as prefixes. Tajik also employs postpositions, such as *barin* (‘like’): *man-barin odam* ‘a person like me;’ and *qatī* ‘with,’ also found as a preposition in southern dialects and in Afghanistan: *tu qatī* ‘with you,’ *bo qošuq qatī* ‘with a spoon,’ here as a circumposition with the synonymous preposition *bo*; and *da* (< *dar* ‘in, at, to’): *ow-da raft* ‘she’s gone to (fetch) water.’ There is likewise a range of compound post- and circumpositions: e.g., *az avvali moh-i ramazon in-jonib/in-taraf rūza me-došta-ast* ‘he has been fasting since the first of Ramadan’ (lit. ‘from– to this side,’ see Perry, 2005, pp. 101-6).

Comparison of adjectives shows some differences from standard Persian. The preposition *az* ‘from, than’ may be expanded, especially in colloquial speech, to a circumpositional phrase, *az– dida* ‘as compared with,’ lit. ‘seen from,’ that is, with the standard in view: *in darakt az on dida baland(tar) ast* ‘this tree is taller than that one’ (see Perry, 2005, p. 140). Note also that the comparative suffix *-tar* may be omitted. An attributive superlative adjective may be preposed to the related noun, as it is in Persian, or postposed with *izofat*: *kalontarin pīsar* or *pīsar-i kalontarin* ‘the biggest boy.’

Verbal Morphology. The simple tenses of Tajik verbs (Present Indicative and Subjunctive, based on the Present Stem; Past, based on the Past Stem) are the same as in Persian, except for the *majhul* vowels in the Present/Imperfect prefix, and in the personal endings of 1st person plural and 2nd person plural (see TABLE 2).

The form for 2nd person plural may also add an ‘explicit plural’ supplement (cf. Pronouns, above) derived from the pronominal enclitic *–ton*: *šin-eton, rafīqon* ‘sit down, friends’ (*šined + ton*).

Neither the Imperative nor the Subjunctive forms normally take the stressed prefix *bi-* (Pers. *be-*): *navis, navised* ‘write!’; *boyad ravam* ‘I must go.’ The prefix *bi-* may, however, be used to mark the subjunctive in elevated style or poetry: *agar bigūyad...* ‘if he says....’ Optionally it occurs with the Imperative in some



common verbs: *bidon* ‘know,’ and before a stem beginning with *b-* its vowel is modified to *u*: *bu-baḵšed* ‘excuse me,’ *bu-bined* ‘see!’.

It is noteworthy that *bi-* occurs commonly as an inseparable suppletive in Present tenses of the two common verbs *omadan* ‘to come’ and *ovardan* ‘to bring’: *me-biyoyam* ‘I come, am coming,’ *me-biyorad* ‘she/he will bring (it).’ Similarly, in some complex verbs, the preverbs *dar* and *bar* are attached to the stem of these same verbs: *me-dar-oyad* ‘he comes in,’ *na-bar-ovardand* ‘they did not produce (it);’ cf. Pers. *dar-miāyad*, *bar-nayāvardand* in the same meanings.

In compound tenses and moods, Tajik verbal morphology has expanded beyond that of Persian. Two progressive tenses are formed on the Perfect tense of a desemanticized verb *istodan* ‘to stand,’ in which the main verb is represented by a non-finite ‘past participle’: *bačaho ovoz konda istoda-and* ‘the children are singing;’ *rūz ba oḵir rasida istoda bud* ‘the day was drawing to an end.’ An alternate use of *gaštan* for the auxiliary usually gives a Perfect progressive sense: *kor karda gašta-ast* ‘he has been working.’

Present tenses of *istodan/ist-* may be used, in place of the Perfect, to indicate that an action once begun will continue: *to omadan-i šumo man kor karda meistam* ‘until you arrive, I shall keep on working’ (see Perry, 2005, pp. 223-27; the standard Persian construction with *dāštan* is not used).

An epistemic mode of the Indicative (called Non-witnessed, or Evidential) has three basic tenses, one of which is identical with the regular Perfect tense functioning as an evidential present: *vay sayohat-ba rafta-ast* ‘he went/has gone on a trip (—so I surmise/am told)’; note also the use of Persian preposition as a postposition. The Non-witnessed durative combines *me-* with the Perfect tense form: *Šūro-hukūmat ba yatimon galla medoda-ast* ‘the Soviet government is/has been (reportedly) supplying grain to the farmhands.’ This tense is frequently used in reportage, when the narrator cannot, or does not, wish to vouch for his statement; depending on context, it may designate habitual or iterative actions in past, present, or future time. The Non-witnessed Past corresponds to the Past Perfect: *az suḵanon-i modar-aš mo fahmidem ki ū kayho ba šahr kūčida buda-ast* ‘from what his mother said, we realized that he had moved to the city long ago.’ This mode, which implies hearsay, inference, or speculation as the source of the statement, also includes progressive tenses: *šumo yak asar-i nav navišta istoda-buda-ed* ‘you have been writing a new work (—so I gather/see).’



Another Tajik innovation, the Conjectural mood (misleadingly called “Conditional” in translations from Russian grammars; in Tajik, *siġa-i ehtimolī*), uses an augmented form of the past participle in *-agī* to form three tenses expressing a probable situation or event: *yagon kor-i ganda karda-gi-st* ‘he must have done something bad’ (Past); *dast-u rū mešusta-gi-st-ed* ‘(I imagine) you will want to wash your hands and face’ (Present/Future); *holo as kob bar-košta va čoy nūšida istoda-gi-st* ‘by now (I suppose) he would have got up and be drinking tea’ (Present Progressive; see Perry, 2005, pp. 243-47).

In addition to the standard Persian participles—*kunanda* (Present), *karda* (Past), and *kardanī* (Future)—Tajik has an augmented Past participle *kardagī* and a Present/Future participle *me-kardagī*. These two tensed adjectives help to form the Conjectural tenses (see above), and, like the simple Past participle, commonly construct adjectival phrases that are nominalized relative clauses: *ana kitob-i ovardagi-am* ‘here is the book that I brought,’ lit. ‘brought-of-me;’ *noma-i ba Bukoro me-firistodagi-aton* ‘the letter that you are sending to Bukhara;’ *gurekta-istodagiho* ‘those who are/were fleeing, the fugitives’ (see Lorentz; Perry, 2005, pp. 271-78).

The Future participle is more productive and versatile in Tajik than in Persian. It forms a quasi-future tense with present forms of the verb “to be,” generally expressing an intention: *kay omadanī-and?* ‘when will they be coming?’; *heč hujum kardanī nestand* ‘they are not going to attack.’ Colloquially, the copula may be omitted: *šumo ozmoiš kardanī-mī?* ‘are you going to try it out?’ (for *-mī* see below). Depending on context, the participle may be active or passive in sense, even from a transitive verb: *guftanī nabuda-ast* ‘it seems she/he is not about to tell’ (Evidential); *joho-i noguftanī* ‘locations not to be divulged’ (see Perry, 2005, pp. 264-67).

The infinitive is extensively used in nominalized sentential complements: *mo kujo raftan-i kud-ro medonem* ‘we know where we are going,’ lit. ‘we know our going-where.’ A speech string may also be reported indirectly as an infinitive phrase object of the verb of saying: *vay ba šahr naraftan-aš-ro guft* ‘she said that she was not going to town,’ lit. ‘told her not-going;’ for other ways of reporting speech, see below.

The Passive Voice in Tajik is more frequently used in modern prose than formerly, and is formed as in standard Persian. A few differences, however, stand out. Whereas Persian usually prefers the ‘short passive’ of composite verbs (that is, replacement of transitive *kardan* by intransitive *šudan*), Tajik



tends to favor the ‘long passive’ (explicit passivization of *kardan* by addition of *šudan*): cf. Persian *zamin taqsim šod* vs. Tajik *zamin taqsim karda šud* ‘the land was divided up.’ Colloquially (frequently in northern dialects), the *karda-gī* type of Past participle may replace the usual *karda* type, especially with stative verbs: *dar yak taraf gahvora monda šudagī bud* ‘on one side the cradle was standing/had been placed’ (Persian would have it *gozāšta šoda bud*; Taj. *mondan* here is the transitive verb ‘to put, let,’ typical of Afghan Persian and Tajik).

As an Active Voice Conjunct verb auxiliary (see below), *šudan* adds the nuance ‘to finish (doing), do completely’ to the main verb, which precedes it in the form of a Past participle: *Zulayko šartnoma-ro navišta šud-u ba man nigoh kard* ‘Zulaikha wrote up the contract and looked at me.’ Here the presence of a direct object, that is, the fact that *navišta šud* is transitive, sufficiently distinguishes it from a Passive Voice form.

Sentence syntax. Simple sentences, and dependent clauses after conjunctions, are constructed similarly to standard Persian, with a finite verb in final position. In volitional sentences, a dependent subjunctive may be used, or simply an infinitive: *bačaho bozī kardan namekōstand* ‘the children did not want to play.’ Peculiar to Tajik is the concessive clause using the adverb *ham*, usually in the form *bošad ham* ‘be it even’: *havo kunuk nabošad ham, barf meborid* ‘although the weather was not cold, it was snowing.’

There are often alternative constructions (to be heard in northern dialects, and read in many examples of Soviet literary Tajik), in which Uzbek influence is directly or indirectly at work. The Turkic interrogative particle *mi* is used in final position, or as an enclitic on the component questioned: *muallim-a pisar-aš raft-mi?* ‘has the teacher’s son left?’ The Persian conditional conjunction *agar* may be placed after the verb: *yagon čil-ta kadu girad agar, nağz* ‘if he gets forty or so pumpkins, that is good’ (the enclitic *-ta* or *-to* is the post-numeral noun classifier; cf. Persian *čeheltā*).

The Conjunct construction—the use of a non-finite form identical with the Past participle in conjunction with another, finite verb—is a mainstay of several Tajik idioms. This structure is analogous to—and in most cases derives from—the Turkic predilection for a single finite verb in a sentence, with one or more dependent non-finite participles or gerunds. It is the preferred construction with the modal verb *tavonistan* ‘to be able’: *man rafta (na)metavonam* ‘I can(not) go;’ also found in the Tajik of Afghanistan. In more



complex forms, the reporting of speech centers on *gufta* (occasionally *gūyon*), a non-finite form of *guftan* ‘to say,’ which is followed by an inflected form of the specific speech verb (with the speech string preceding): *kud-i ū kist?—gufta man az Rahim Qand porsidam* ‘“who’s he?” I asked Rahim Qand;’ lit. ‘... saying, I asked...’. This construction has further evolved as an idealized quotation to explain the cause or purpose of the action in the main clause: *kūrjin-ro ham ba šumo mukofot gūyon dihad* ‘he will give you the saddlebag too, as a reward;’ lit.; ‘... reward, saying...;’ on the analogy of a typically Turkic construction, using *dep* ‘saying’ in Uzbek (Perry, 2005, pp. 321-26). A grammaticalized use of the Conjunct construction has been mentioned as forming Progressive tenses, and a lexical use of Conjunct verbs will be illustrated below.

The verbal adverbs ending with *-on* are used similarly (and more freely than in Persian, e.g., with composite verbs) to denote an action carried on simultaneously with that of the main verb (cf. *gūyon* above): *modar tabassum-kunon ba bača-aš nigoh mekard* ‘the mother looked at her child with a smile;’ *ū dar-ro taraqqos-zanon pūšid* ‘he slammed the door shut’ (*taraqqos zadan* ‘to bang, to make a loud noise’).

The Lexicon. Tajik derivational morphology is very rich. A great number of composite adjectives, and their derived substantives, adverbs, and quality nouns, are formed with prefixes and suffixes, some of them different from (or more productive than) their Persian counterparts. Thus, among adjective-forming suffixes, *-nok* is highly productive, denoting something having the quality of the base noun: *foida-nok* ‘beneficial, profitable’ (*foida* ‘use, profit’); *sado-nok* ‘vowel’ (*sado* ‘sound, voice’); *qarakter-nok* ‘characteristic, specific.’ The relative *-ī* may take the forms *-vī*, *-gī*, or *-ngī* after a vowel, e.g., *bilet-i partiyavī* ‘Party card,’ *qabilagī* or *qabilavī* ‘tribal,’ *rūznoma-idinangī* ‘yesterday’s paper’ (*dina* or *dina-rūz* ‘yesterday;’ cf. Pers. *dī[ruz]*). Among prefixes, *ser-* ‘sated, full’ indicates an abundance of the base noun: *ser-gap* ‘garrulous’ (*gap* ‘talk’); *ser-odam* ‘populous, crowded.’ The prefix *to-* ‘up to, until’ produces words with the meaning ‘pre-’, e.g., *to-maktabī* ‘pre-school,’ *to-inqilobī* ‘pre-revolutionary;’ this use of the preposition *to* (unknown with Persian *tā*) is probably calqued on similar use of Russian *do-* ‘pre-.’

Formation of compound nouns generally resembles the Persian types. During the Soviet period, the productivity of present-stem activity nouns of the type *roh-barī* ‘leadership’ was particularly encouraged, especially from putative composite verbs, e.g., *mablaḡ-judokunī* ‘appropriation, disbursement of funds,’ lit. ‘sum-separate-making;’ *avtomobil’-kor-karda-bar-orī* ‘automobile



production;’ lit. ‘auto-work-done-out-bringing.’

Transitivizing denominal verbs (obtained by infixing *-on-*) are more productive than in Persian, e.g., *gūr-on-idan* ‘to bury, inter’ (*gūr* ‘grave, tomb’); *mukofot-on-idan* ‘to recompense;’ *kollektiv-on-idan* ‘to collectivize’ (the three samples are formed from a Persian, an Arabic, and a Russian noun respectively). Causatives are similarly constructed, from verbal present stems: *šiştan/šin-* ‘to sit’ (intr.) gives *šin-on-dan* ‘to set, seat,’ and derived nominals such as *darakt-šironī* ‘tree-planting, forestation;’ they may be formed not only from simple verbs, but also from complex and composite verbs: *papirus dar-me-gir-on-ad* ‘she/he lights a cigarette’ (cf. *dar-me-girad* ‘it catches fire’).

Characteristic of Tajik are Conjunct verbs (serial verbs), of which the Progressive tenses (see above) are grammaticalized instances. There are some eighteen lexically established conjunct auxiliaries (corresponding to models in Uzbek), which in regularly conjugated tenses furnish adverbial ‘modes of action’ for the non-finite participle—which is, semantically speaking, the main verb (see [CENTRAL ASIA xiv. TURKISH-IRANIAN LANGUAGE CONTACTS](#), pp. 231-33; Perry, 2005, pp. 467-77). Some are fairly literal in sense: *kitob-mitob karida mebarad* ‘he buys (up) books and stationery (and takes them away with him),’ to highly metaphorical: *in kurta-ro pūšida bin* ‘try on this tunic’ (*didan/bin-* ‘to see,’ tentative mode; cf. Eng. ‘see if it fits’). Other typical conjunct auxiliaries are *giriftan* ‘to take’ (self-benefactive): *dars-i nav-ro navišta giriftem* ‘we copied down the new lesson;’ *dodan* ‘to give’ (other-benefactive): *nom-i kud-ro navišta mediham* ‘I shall jot down my name (for you);’ *partoftan* ‘to throw (away)’ (complete or thorough action): *berunho-ya toza karda rūfta parto!* ‘sweep all the outside nice and clean!’ This last illustrates a double conjunct construction, the auxiliary governing both of the non-finite forms of *rūftan* ‘to sweep’ and *toza kardan* ‘to clean’ (a typical Persian-type composite verb).

Applicability of the auxiliary often depends on the nature of the main verb: *omadan* ‘to come,’ for example, combined with other verbs of motion expresses motion toward the speaker: *davida omad* ‘he ran up (to us).’ With a range of other verbs, it denotes the successful completion of an action or the culmination of a process: *onho suporišho-i kud-ro 160-foizī ijro karda meoyand* ‘they are fulfilling their quotas 160 percent.’ In colloquial and dialect usage, the participial component is sometimes postposed: *kord biyor rafta* ‘go fetch a knife.’



Foreign vocabulary and lexical distribution. Russian loanwords flooded into Tajik only in recent decades (ca. 1925-55; for a selection, see Bashiri). Most derivatives from Russian loans freely use Tajik Persian formatives, as in the adjectives *bolševik-ī*, *bolševik-ona* ‘Bolshevik, Bolshevik.’ Outright loanwords are widely supplemented by calques; e.g., Taj. *sar* ‘head,’ by analogy with the Rus. *glav-*, forms *sar-duktur* ‘chief physician’ (Rus. *glavvrach*), *sar-muhandis* ‘chief engineer’ (Rus. *glavnyi inzhener*; cf. also *to-inqilobī* above). Stalin ruled that Russian words incorporated into the newly Cyrillicized languages should retain their Russian spelling, and many earlier assimilated loans were re-Russianized, e.g., *istansa* ‘station’ became *stantsiya*; the borrowed Russian *krovat* ‘bed,’ however, retains its assimilated pseudo-Arabic plural form *karavot* ‘(raised) Western-style bed.’

Though individual loans are targeted for replacement by Persian terms in post-Soviet Tajikistan, certain turns of phrase and paralinguistic habits introduced through Russian are likely to remain. Among these are bureaucratic acronyms and abbreviations, such as *VABK* (*Viloyat-i avtonom-i Badaḡšon-i kūhī* ‘the Gorno-Badakhshan Autonomous Region,’ now renamed *Viloyat-i muḡtor-i kūhiston-i Badaḡšon*).

Uzbek loanwords are of course plentiful in the everyday language, e.g., *boy* ‘rich,’ *qišloq* ‘village,’ *tūy* ‘wedding, circumcision celebration,’ *yaroq* ‘weapon,’ *yordam* ‘help,’ and several terms for kinship. Like Russian loans, they readily adopt Persian derivational morphology: *boi-gar-ī* ‘wealth, riches,’ *yaroq-parto-ī* ‘disarmament’ (calqued from Arabic- Persian *kal‘-e selāḥ*). Such Uzbek and other Turkic vocabulary, which infiltrated into Tajik over a period of centuries, remains too basic to be endangered; however, some of the syntactic Uzbekisms introduced into literary Tajik during the 1920s and 1930s (such as those involving reported speech; see under *Syntax* above) are already fading from post-Soviet Tajik writing. On the other hand, the spoken Tajik of Samarqand and other northern regions, where speakers are generally bilingual in Uzbek, is still permeated with Uzbekisms, to the extent that Doerfer characterized it as a Turkic language in embryo (Dorfer, pp. 52-53).

Arabic borrowings, as in all varieties of Persian, comprise the earliest and most tenacious stratum of foreign vocabulary in Tajik. While Tajik of Afghanistan is almost entirely free of Russian loans, and has many fewer Uzbekisms than Tajik of Tajikistan, it still shares most of the features of Arabic vocabulary noted here. Patterns of distribution, however, differ from those of standard Persian. For instance, Tajik uses *peš* and *pas* rather than *qabl* and



ba'd for 'before' and 'after,' *pok* 'clean' instead of *tamiz*, but *oid ba/oid-i* (< Ar. 'ā'ed) 'concerning, relating to' rather than *rāje' be* or *darbāra-ye* as in Persian; *iflos* 'dirty' corresponds to Persian *kaṭif*; *madaniyat* 'civilization' to Persian *tamaddon* or *farhang*; *tayyor* (Pers. *ṭayyār*) 'ready' to *ḥāẓer* or *āmāda*; *hozir* (Pers. *ḥāẓer*) 'now' to *ḥālā*; and *ittifoq* (Pers. *ettefāq*) '(labor) union' to *etteḥād*.

Several Arabic plural forms, both suffixal and mutated ('broken'), have been lexicalized with collective or singular meanings: *hašarot* 'insect' with regular plural *hašarotho* 'insects'; *talaba* 'student,' pl. *talabagon*; *a'zo* 'member (of an institution),' pl. *a'zoyon*, but *a'zo-i badan* 'parts/limbs/members of the body' (also *uzvho-i badan*, using the original singular in its literal meaning); *marotiba* 'time, occasion' (sg. *martaba* is used in Persian); and *šaroit* 'condition, stipulation.'

Lexical distribution also varies between native Persian vocables; sometimes the same words have evolved surprisingly disparate forms and meanings, sometimes different words have been favored; quite often Tajik Persian vocabulary corresponds to that of Afghanistan (and of Indo-Persian). Below is a small list of samples (the Persian equivalent is in parentheses; see also Lazard, p. 180): *pagoh* (*fardā*) 'tomorrow,' *begoh* (*diruz*, 'ašr) 'yesterday, evening,' *tiramoh* (*pā'iz*) 'autumn,' *daryo* (*rud*, *rudkāna*) 'river,' *bahr* (*daryā*) 'sea,' *paḵta* (*panba*) 'cotton' (but Taj. *panba-dona* 'cottonseed'); *tireza* (*panjara*) 'window,' *kel* (*jur*) 'sort, kind,' *vaznin* (*sangin*) 'heavy,' *sangin* (*sangi*) '(of) stone,' *kalon* (*bozorg*) 'big, great, old,' *mayda* (*kord*) 'tiny, small change,' *kurd* (*kuček*) 'small, little, young,' *kampir* (*pir-e zan*) 'old woman,' *nağz*, *nek* (*kub*, *qašang*) 'good, nice,' *ganda* (*bad*, *ḡarāb*) 'bad,' *kasal* (*bimār*, *mariz*) 'sick,' *monda* (*kaste*) 'tired,' *aftidan* (*oftādan*) 'to fall,' *šištān* (*nešastan*) 'to sit,' *mondan* (*godāštan*) 'to let, put,' *gundoštan* (*jam' kardan*) 'to gather.'

BIBLIOGRAPHY

Azim Baizoyev and John Hayward, *A Beginner's Guide to Tajiki*, London, 2004.

Iraj Bashiri, "Russian Loanwords in Persian and Tajiki Languages," in *Persian Studies in North America: Studies in Honor of Mohammad Ali Jazayery*, ed. M.



Marashi, Bethesda, Md., 1994, pp. 109-39.

Ḥabib Borjiān, “Sāktemān-e kaṭṭ -e tājiki,” *Nāma-ye Farhangestān* 4/1, 1998, pp. 103-16.

Ravān Farhādi, *Zabān-e tājiki-e Mā-warā’-al-nahr*, Kabul, 1981.

Gerhard Doerfer, “Türkische Lehnwörter im Tadschikischen,” *AKM* 37/3, Wiesbaden, 1967.

Gilbert Lazard, “Caractères distinctifs de la langue tadjik,” *Bulletin de la Société de Linguistique de Paris* 52/1, 1956, pp. 117-86.

Manfred Lorenz, “Die partizipialen Wendungen—ein Unterscheidungsmerkmal des Ta^aikischen gegenüber dem Persischen,” *Mitteilungen des Instituts für Orientforschung* 10, 1964, pp. 133-39.

Randall B. Olson and Rochelle A. Olson, *Standard Tajik-English Dictionary*, Dushanbe, 2000.

John R. Perry, “From Persian to Tajik to Persian: culture, politics and law reshape a Central Asian language,” in *Non-Slavic Languages 8. Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, ed. H. I. Aronson, Chicago, 1996, pp. 279-305.

Idem, “Script and Scripture: the Three Alphabets of Tajik Persian, 1927-1997,” *Journal of Central Asian Studies* 2/1, 1997, pp. 2-18.

Idem, “Epistemic Verb Forms in Persian of Iran, Afghanistan and Tajikistan,” in *Evidentials: Turkic, Iranian and Neighboring Languages*, ed. L. Johanson and Bo Utas, Berlin and New York, 2000, pp. 229-57.

Idem, *A Tajik Persian Reference Grammar*, Leiden, 2005.

Aḥmad-‘Ali Rajā’i, *Yāddāšt-i darbāra-ye lahja-ye Boḵārā’i* (French title page: *Remarques sur le dialecte de Bukhārā*), Mashad, 1964.

V. S. Rastorgueva, *A Short Sketch of Tajik Grammar*, tr. and ed. Herbert H. Paper, Bloomington and The Hague, 1963; repr. Bloomington, 1992.

Idem, *Opyt sravnitel’nogo izucheniya tadzhikskikh govorov* (An essay in the comparative study of Tajik dialects), Moscow, 1964.



Lutz Rzehak, *Vom Persischen zum Tadschikischen. Sprachliches Handeln und Sprachplanung in Transoxanien zwischen Tradition, Moderne und Sowjetmacht (1900-1956)*, Wiesbaden, 2001.

V. S. Sokolova, *Fonetika tadzhikskogo yazyka* (Phonetics of the Tajik language), Moscow and Leningrad, 1949.

Idem, “Novye svedeniya po fonetike iranskikh yazykov. I. Yugovostochnye govory tadzhikskogo yazyka” (New research on the phonetics of the Iranian languages. I. Southeastern Tajik dialects), *Trudy Instituta yazykoznaniya* 1, 1952, pp. 154-72.

John D. Soper, *Loan Syntax in Turkic and Iranian: the Verb Systems of Tajik, Uzbek and Qashqay*, rev. and ed. A. J. E. Bodrogligeti, Bloomington, Ind., 1996.