PERSIAN, DARI AND TAJIK IN CENTRAL ASIA

William O. Beeman
Brown University

NCEEER

Executive Summary

There have been a number of informal debates among Persian language specialists concerning the status of Tajik and Dari vs. standard Persian. All linguists know that speech communities utilize a continuum of varieties of speech, and that the term "language" is more a political appellation than a scientifically accurate descriptor. The range of variation in Persian, Dari and Tajik communities is quite extensive, embodying regionalisms and borrowings from other language families. The term "register" has a special status in describing languages in that it represents a speech variety that is marked for particular specific occasions. Whereas Modern Persian and Dari are very close in form, Tajik has more divergent discourse structures. Based on fieldwork carried out in Tajikistan, I theorize that standard Persian as spoken in Iran has become a special register of Tajik marked for formal occasions such as political speech making, wedding orations, news broadcasts, and elevated scientific discourse. In this way the opposition between all the varieties of colloquial Tajik and standard Persian in Tajikistan resemble the diglossic opposition between dhimotiki and katherevusa in modern Greek. In this paper I will provide several examples, and speculate on the concretization and meaning of such diglossic vocal speech registers.

Introduction

The purpose of this paper is not to carry out a controlled comparison of "Persian," "Dari" and "Tajik." It is rather to begin to explore the social and cultural relationships between these language varieties, and to provide a sketch of the development of these varieties in recent years.

Let us be clear from the start that many of the base terms in linguistics are imprecise, leaving us in a struggle for good descriptors for the modes of communication used by particular communities of speakers. The term "language," referring to a particular variety of coded human communication has little meaning in the world of actual speakers. The term is best seen as a label for a set of social institutions serving to standardize communication codes. The old saw, "a language is a dialect with an army" reflects a basic truth. Any given community of speakers— hereafter a *speech community*--embodies communication practices that vary from a given standardized language to some degree.

In most speech communities individuals control a range of speech variables. Linguists often refer to specific clusters of these variables as *registers*, although this term can be deceptive, since they are rarely as unified as the term implies. In descriptive terms it is most often the case that speakers freely manipulate these variables to modify and shade their speech for specific social purposes. The connotational subtleties of the creative play of speech variables in interaction are infinite. For the purposes of this discussion I will focus on three broad categories of purpose: self-identification, context identification, and strategic action. They answer three fundamental questions: Who am I? What is going on? and What do I hope to accomplish?[1] The answers to all of these questions must be culturally defined for each of these language varieties.

When used for self-identification purposes speech variables can be used to identify one's self and others as belonging to a specific community, to indicate membership in a particular social class, or to reinforce one's gender identity. When used to identify context, speech variables distinguish between literary and conversational genres. They also mark particular culturally defined situations, such as public, private, academic, legal, formal, informal, and many others. When used strategically they can be used to indicate relative personal relations, such as status, formality and intimacy. They can also be used to indicate attitudes such as humor, sarcasm, irony, subordination, superordination, admiration, flattery, and others. Variables are *polysemic* in the sense that they can be used to indicate more than one thing. For example, a particular variable may indicate at the same time that one is an upper class male in a formal situation showing admiration toward one's companions in interaction.

In general speakers' use of speech variables exhibits a continuum of use ranging from one variable to the other. A good example from English might be the difference between the two morphemic variables *–in'* and *–ing* in the progressive tense (*walkin'* vs. *walking*, etc.). Almost every English speaker uses both variables freely, and an inventory of usage for any given speaker reveals a general pattern where the first variable is used in informal, intimate settings and the second in formal, non-intimate settings. However, the variable is not like a light switch—either on or off. Conversational analysis reveals that speakers mix this variable freely. One can't know exactly what is going on in speakers' minds, nor can they usually articulate precisely what they are doing (since usage patterns are largely unconscious), but the data suggests that they attempt to "fine tune" the speech situation by using proportionately more of one variable than the other as the social situation shifts and changes over the course of the event. I have documented some of these dynamics for standard Persian in other publications (1986, 1988).

Tajik, Dari and Persian

Tajik, Dari and Persian are "languages" in the sense that they have concretized canonical forms that are transmitted through institutionalized schooling and reference works, however structurally they are all varieties of Persian.

The history of all three varieties may be surprising to speakers of Modern Persian in Iran. In fact, Modern Persian in its literary form emerged first in Bukhara (present day Uzbekistan) during the Samanid Empire (9th-10th Centuries, C.E.). The term Dari derives from the phrase *Fārsi-ye Darbāri,* or "Court Persian" The term also dates from the Samanid Empire, although today it refers both to the variety of Persian spoken in Afghanistan, and to the variety spoken by Zoroastrians in Yazd and Kerman in Iran (also known as Gabri). Although its speakers have been active for millennia, Tajik, with its present name and in its present form is a 20th century creation—an artifact of the Soviet Union and its cultural policies, and some divergence between the two varieties is attributable to this political process.

Persian, Dari and Tajik encompass the kinds of variation referred to in the previous section, and there is much overlap in particular variable features. Some speakers of "Persian" in Khorasan communicate colloquially in a variety that is virtually identical with speakers in Heart, Samarkand or Dushanbeh. If we take Persian and Tajik as antipodes on a scale of variability, with Dari as an intermediate form, we can see some important dynamic relationships between the varieties.

There is a directionality in the relationship between the two varieties. Persian is seen by all speech communities as a prestige standard, and Tajik and Dari as colloquial forms. Dari, as spoken in Afghanistan, is seen as a stigmatized variety for many of its speakers when they find themselves in a primarily Persian speaking setting. Afghan residents in Iran will often resort to using a foreign language such as English rather than speak Dari. To reinforce this notion of hierarchy it is worth noting that speakers of Persian varieties rarely learn Tajik or Dari forms, whereas educated Tajik and Dari speakers all acquire some command of Persian forms.

It is important to note that Persian, Tajik and Dari are mutually intelligible. This is in stark contrast to their intelligibility with some other Iranian "languages" such as Kurdish or

Baluchi, and some varieties that are commonly referred to as "dialects" such as Tati or Kashi. The Pamir "languages" of the Gorno-Badakhshan region of Tajikistan are also unintelligible to Persian/Tajik/Dari speakers, despite the fact that these varieties have borrowed large amounts of standard Persian vocabulary. Shugni, or Shugnani, for example, no longer maintains any numbers above 10, the higher numbers being borrowed from Persian/Tajik.

Orthographic systems contribute to perceptions of intelligibility between the "languages." Tajik is written in Cyrillic characters despite some attempts to introduce Arabic script since independence. This leads many people to believe that the languages are less mutually intelligible than they actually are. This phenomenon is not uncommon elsewhere. Hindi/Urdu and Serbian/Croatian are examples of mutually intelligible varieties that differ primarily in their orthographic systems.

The presence of literature in the languages in question also contributes to the sense of difference. The extensive literature in Persian compared to the other two contributes to its prestige. However, vernacular published literature in Tajik—particularly in 20th century poetry— serves to concretize the idea of Tajik as a separate language.

The following table illustrates some of the differences between the formal languages discussed above:

	Persian	**Tajik**	**Dari**
Orthography	Arabic	Cyrillic	Arabic
Literature	Extensive	Moderate	Scant
Relative Prestige	High	Moderate	Low

2.1. Relations between Persian, Tajik and Dari

Markers of Persian and Tajik

Certain linguistic variables tend to mark Persian and Tajik. It is not possible to specify every difference in this brief presentation, but they fall into several broad categories roughly corresponding to standard linguistic descriptive categories.

Phonology

A simpler phonological structure tends to characterize varieties identified as Tajik as opposed to those identified as Persian. In theory both varieties have the same vowel and consonant structure as described in standard Persian grammatical literature. However, Tajik in general has a tendency to centralize vowels, particularly in unstressed syllables, and in grammatical prefixes ({mi-} and {be-}) and in personal suffixes (i.e. {-æm}). The phoneme /o/ in Persian seems quite unstable in Tajik, and is frequently realized as either [u] or , [ə][2]. There is a tendency for the prominent /a/ in Persian varieties to be realized as [o] or [ɔ] in Tajik varieties. Some of the same tendencies are seen in Dari, but Dari is generally closer in pronunciation to standard Persian. One generalizable difference is that Dari nearly universally realizes Persian /v/ as [w].

Morphology

Speakers of varieties identified as Persian generally see Tajik and Dari varieties as embodying completely recognizable, albeit occasionally archaic forms. In general Indo-European root forms are favored over Arabic forms in Tajik and Dari varieties, although many

transmitted Arabic vocabulary items are found. The third person singular pronoun /vai/ predominates over /u/ ("he, she") in Tajik, /besyor/ over /xeili/ ("very") and other similar preferences. Tajik differs from Persian an Dari in its increased number of Russian borrowings; and Arabic and Western European borrowings in Persian varieties add to the color of language use in Iran, but the high degree of overlap in the vocabulary of all three varieties is nearly complete.

Syntax

Two very distinct constructions differentiate Tajik varieties from Persian and Dari varieties in spoken language. The first involves the question construction. Tajik uses a terminal question particle {mi}, probably as a result of *Sprachbund* influence from Turkish varieties in the region as in the following

3.1 šəmo zæn dorid mi ? "Do you have a wife"

Persian and Dari varieties would eliminate the question particle.[3]

The second involves the use of the verb istadæn "to stand" in many Tajik constructions foreign to Persian varieties. In particular, with the truncated infinitive in Tajik progressive verb constructions where daštæn with the present tense would be used in Persian constructions. There are both literary and colloquial constructions, and even further regional variations on these colloquial constructions. In the examples below one widely used set of colloquial forms is provided. .

English	Tajik	Persian
We are eating (now)	Mo xorda istadæ-im (lit.)	Ma darim mixorim
	Mo istadæ-im xur (colloq.)	
We were eating	Mo xorda istadæ budim (lit.)	Ma daštim mixordim
We had been eating	Mo xorda istadæ budæ-im	Ma daštim mixordim

3.2 "We are eating" in Tajik and Persian

Finally, there is an unusual use of a gerund construcion with the suffix {-gi} in Tajik conditional constructions that rarely if ever occurs in Persian constructions where conditional forms collapse with normal indicative forms[4].

English	Tajik	Persian
We would eat	Mo mikhordagistim (lit.)	Ma mixorim
	Mo mikhordagim (colloq.)	
We would be eating	Mo xorda istadægistim (lit.)	Ma mixordim

	Mo xorda istadægim (colloq.)	
We would have eaten	Mo xordagistim (lit.)	Ma daštim mixordim
	Mo xordagim (colloq.)	
We would have been eating	Mo khorda istadægi budæ-im	Ma daštim mixordim

3.3 "We would eat," etc. in Tajik and Persian constructions

It should be noted that the past participle with the {-gi} suffix is widely used as a kind of impersonal construction in Tajik forms.

3.4 Vai ketobo xondægi, ræft. "Having read, he left"

Note that in the above, the translation of the tense of the first clause depends on the tense of the verb. Viz.

3.5 Vai ketobo xondægi, miravæd. "Reading the book, he goes."

Colloquially, this construction is also used as a simple past tense:

3.6 Shoma ketobo xondægi? "Did you read the book?"

This {-gi} construction is seen in Persian forms, but is fully nominalized in most cases (e.g. zendegi "living, life"), having presumably lost its function in verb constructions.

Contexts for Persian in Tajik

Looking at the previous section we can see that the primary areas where differences in Tajik and Persian varieties exist is in phonology and syntax. Morphology seems not to be a dimension of particular attention for speakers who posses both varieties.

Given that Tajik speakers all acquire some command of Persian forms, it is important to note where and under what conditions the tendency to use Persian forms is exercised.

In general, the Tajik situation tends toward *diglossia* as described by Ferguson (1959) in his classic article of the same title. Tajik speakers will demonstrate pronunciation and syntactic structures that tend toward Persian in literature, and in formal, public situations. They will tend toward Tajik constructions in face-to-face conversation and in informal, private situations. This resembles the diglossic opposition between *dhimotiki* and *katherevusa* in Modern Greek[5].

Curiously, and perhaps because the two varieties are so very close, Dari speakers do not generally command standard Persian pronunciation or intonation in spoken forms. Written Dari approximates standard literary Persian.

Literary usages in Tajik include journalistic writing, official government documents as well as some fiction, academic writing, non-fiction and poetry that emulates classic styles. Formal usages include political speeches, public addresses and formal social occasions, such as weddings. Toasting at banquets can also involve highly Persianized speech, especially on the part of the "toastmaster" who must introduce each individual making a toast. It must also be

noted that individuals wishing to appear erudite to others will adopt Persianized forms in their speech, at times to absurd degrees, indulging in a kind of *hypercorrection* (cf. Labov 1972) that can create an effect precisely opposite to that which they aspire.

Tajik forms dominate in personal contact situations. An individual using Persianized elements in speech risks alienating his or her intimate friends. It is certainly not a register that is designed to create intimacy. Colloquialized Tajik forms are also used in playwriting and colloquial literature as well as in comic strips, the most popular forms of journalism, and the lyrics of popular songs.

The divergence of Persianized registers from colloquial speech can be very great. Television is an important form of information and entertainment for most citizens in Tajikistan, but many rely on news broadcasts in Russian because they can not understand the Persianized register of the Tajik news broadcasts. President Rakhmanov is actually quite a skilled political speaker (in my opinion) because he manages to use a variety of speech in his public addresses that hits a medium between the use of Tajik and Persianized forms.

Other varieties of Tajik are found in Uzbekistan, notably in Samarqand, Bukhara and the Boysun region in the Surkhandarya region of the country. Unfortunately for Tajik speakers, the government of Uzbekistan is engaged in a systematic eradication of the language by closing schools, university faculties, publications and media outlets. The Bukhara, Samarqand and Boysun varieties of Tajik differ from each other in pronunciation and in some morphological respects, however, historically, the people of these regions all had knowledge of classical Persian to serve as a touchstone for the mutual interpretation of these regional differences. In a field trip to these regions in 2003, I discovered that young people, having lost formal Persian/Tajik

instruction in schools, and exposure to the language in the media were losing intelligibility for Tajik speakers outside of their own region.

As Tajikistan becomes more accessible to scholars, it is clear that much more research needs to be undertaken on the interrelationship between Tajik and Persian varieties of speech. The historical and genetic relationships are in need of clarification and further investigation. Since there is relatively little in terms of formal structure separating the two varieties most of the differences lie in the social realm. I hope that this small preliminary set of observations will start scholars thinking about the sociolinguistic dimensions of the relationship between the two speech communities.

Notes

---

[1] This is a greatly simplified version of schemata developed by Jakobson (1950) and Hymes (1970)
[2] Rastorgueva describes this phenomenon extensively (1963:4).
[3] Note however, that the {-mi} particle is eliminated if the initial question particle /oyo/ is used in Tajik constructions. In the Badakhshan region of Afghanistan, Tajik varieties take precedence over Dari in many areas.

---

[4] Cf. Rastorgueva 1963: 76-77 for a more complete analysis.

[5] Rzehak in his Tajik Grammar (1999) makes a clear distinction between spoken and written forms. Written Tajik, aside from its use of Cyrillic characters and some vocabulary differences is virtually indistinguishable from Modern Persian (Modern Persian in general uses more European loan words).

Bibliography

Beeman, William O.

    1986   *Language, Status and Power in Iran*. Bloomington, IN: Indiana University

            Press.

    1988   Affectivity in Persian Language Usage. In Good, Byron, Mary Jo Good

            and Michael M.J. Fischer, eds., *Affect and Healing in Middle Eastern*

            Cultures. Special Issue, Culture Medicine and Psychiatry 11 (4) 403-24

Ferguson, Charles A.

    1959   Diglossia. *Word, 15:* 325-340.

Lazard, Gilbert

    1956   Charactères distinctifes de la langue tadjik. In: Bulletin de la Société de

            linguistique de Paris. Paris 52 : 117-186

    1970   Persian and Tajik. in Thomas Sebeok, ed., *Current trends in Linguistics,*

            Vol. 6: 64-96. Paris: Mouton.

Labov, William

    1972   *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania

            Press.

Rastorgueva, V.S. (edited by Herbert Paper)

    1963   A Short Sketch of Tajik Grammar. *International Journal of American*

            Linguistics Vol. 29, No. 4. [Publication Twenty-eight of the Indiana

            University Research Center in Anthropology, Folklore and Linguistics.]

Rzehak, Lutz

    1999   Tadschikische Studiengrammatik. Wiesbaden: Reichert.

    2001   Vom Persischen zum Tadschikischen: Sprachliches Handeln und
Sprachplannung in Transoxanien zwischen Tradition, Modene und
Sowjetmacht (1900-1956). Wiesbaden: Reichert.

Persian, Dari and Tajik

William O Beeman

Department of Anthropology

Brown University

Tables and Examples

	Persian	**Tajik**	**Dari**
Orthography	Arabic	Cyrillic	Arabic
Literature	Extensive	Moderate	Scant
Relative Prestige	High	Moderate	Low

2.1. Relations between Persian, Tajik and Dari

3.1 šəmo zæn dorid mi ? "Do you have a wife"

English	Tajik	Persian
We are eating (now)	Mo xorda istadæ-im (lit.)	Ma darim mixorim
	Mo istadæ-im xur (colloq.)	
We were eating	Mo xorda istadæ budim (lit.)	Ma daštim mixordim
We had been eating	Mo xorda istadæ budæ-im	Ma daštim mixordim

3.2 "We are eating" in Tajik and Persian

English	**Tajik**	**Persian**
We would eat	Mo mikhordagistim (lit.)	Ma mixorim
	Mo mikhordagim (colloq.)	
We would be eating	Mo xorda istadægistim (lit.)	Ma mixordim
	Mo xorda istadægim (colloq.)	
We would have eaten	Mo xordagistim (lit.)	Ma daštim mixordim
	Mo xordagim (colloq.)	
We would have been eating	Mo khorda istadægi budæ-im	Ma daštim mixorde budim

3.3 "We would eat," etc. in Tajik and Persian constructions

3.4 Vai ketobo xondægi, ræft. "Having read, he left"

17

3.5 Vai ketobo xondægi, miravæd. "Reading the book, he goes."

3.6 Shoma ketobo xondægi? "Did you read the book?"

Figure 1 Persian Speaking Areas in Asia (Persian, Dari, Tajik)