

## SOME REMARKS ON THE ORIGIN OF THE IRANIAN STRATUM OF CAUCASIAN PERSIAN (TATI) LEXICON\*

*Artyom Tonoyan*

### **Abstract**

*The Caucasian Persian (Tat), being the only Southwest Iranian language, which is descended from Early New Persian and is spoken in the eastern part of Transcaucasus, is of great interest in the context of the study of the problems related to areal and historical-comparative linguistics, as well as in concern of the revision of issues related to historical contacts and interactions between Iranian languages and peoples in Transcaucasus.*

**Keywords:** *Caucasian Persian, Iranian stratum, Middle Persian archaisms, Northwest Iranian borrowings, Northwest Iranian substratum, Northern Talysh, Southwest Iranian loans, Caucasian Persian inherited lexicon.*

The study of the Caucasian Persian vocabulary shows that this language has had active contacts with various non-Iranian and Iranian languages in different historical periods.

1. **Southwest Iranian** stratum, which includes:

- Inherited lexicon of Caucasian Persian, which is almost identical to that of Early New Persian,
- Middle Persian archaisms,
- Direct or indirect borrowings from New and Modern Persian.

2. **Northwest Iranian** stratum, which includes:

- Northwest Iranian substratum,
- Lexemes of Northwest Iranian origin inherited from the Early New Persian,
- Borrowings from the Talyshi.

---

\* The article was submitted on June 27, 2021: The article was reviewed on July 21, 2021.

### Introductory notes

Caucasian Persian (Tati)<sup>1</sup>, which in different historical periods has been in direct and indirect contact with both Iranian and non-Iranian, including Arabic, Russian, Armenian, Caucasian and Turkic languages, is influenced by all these languages, especially in the part of lexicon<sup>2</sup>.

The issue of the stages of the formation of the Caucasian Persian cannot be considered completely resolved yet. However, both linguistic facts and historical evidence confirm that, although the core of the language is Early New Persian<sup>3</sup>, its formation and development must have taken place on the fertile ground of other Western Iranian language or languages<sup>4</sup>.

In other words, on the one hand, there is a tangible presence of Middle Persian, which is proved by a number of lexical archaisms preserved in various, especially Central and Quba dialects<sup>5</sup>, on the other hand, there is a substratum of a Northwest Iranian (NWI) dialect, the traces of which can be seen in the dialects of this language, spoken in the regions of Ismayli and Shamakhi. There is a strong linguistic basis for such an assumption, and the relevant historical data and toponyms of historical Shirvan also confirm that approach.<sup>6</sup> In addition, the southern border of historic Shirvan coincided with the northern border of the spread of Northern Talysh, which suggests

---

<sup>1</sup> The Caucasian Persian language, also known as the Tat or Caucasian Tat language, at present lacks a literary tradition and belongs to the south-western subgroup of Iranian languages. Historically, it was widespread in the area, which approximately coincided with the lands of the medieval Shirvanshah kingdom or the territories of Baku, Shirvan and Quba khanates in the 19th century. Nowadays the Caucasian Persian is the native language of the ethnic group of Iranian descent that is settled in the Apsheron, Baku, Sumgayit, Khizi, Siazan, Shabran, Quba, Khachmaz, Ismaili and Shamakhi districts of the Republic Azerbaijan. In various parts of the above-mentioned regions the Iranian-speaking population has different endoethnonyms, such as Pars, Lohij, Dağli, and Tat (**Miller**: 1929).

<sup>2</sup> See, e.g., **Gryunberg** 1961b: 11-23 and **idem** 1963: 112-113, also **Voskanian** 2016: 240-248.

<sup>3</sup> **Gryunberg** 1961a: 106-114.

<sup>4</sup> See, e.g., **Tonoyan** 2014: 193-197 and **idem** 2015a: 147-148.

<sup>5</sup> There is a widespread opinion, according to which the spread of Middle Persian towards the north of Araks river was due to the Iranians, who as soldiers of the Sasanian army, appeared in the northern parts of Transcaucasus in the 6th century AD to protect the borders of Iran from the attacks and invasions of the northern tribes (**Minorsky** 1934: 697-700).

<sup>6</sup> **Minorsky** 1958: 13-20.

that there must have been active contacts between Caucasian Persian and Northern Talyshi languages in above-mentioned region. In this regard, Veliyev-Baharli writes: “*We know from history about the movement of the Tats to the Caspian gates, the northern coastal region of Azerbaijan, where they came into direct contacts with the Talysh people. However, this connection was broken by the Turkic tribes, who wedged themselves like a wedge between these two neighboring people, forcing them to leave inaccessible mountains or continuously become Turks*”<sup>7</sup>.

Taking into account the presence of Middle Persian lexical archaisms in various dialects, as well as the possible traces left by the Northwest Iranian languages, including the Northern Talysh, on the Caucasian Persian, the examination of the Iranian lexical stratum of this language should be based on the following points:

#### **Southwest Iranian (SWI) stratum**

1. The SWI stratum of Early New Persian lexicon should be considered as a native lexical stratum in Caucasian Persian language. Moreover, only words that bypass the phonetic laws specific to Caucasian Persian can be considered borrowed from New and Modern Persian.

2. In all cases, if the lexeme existing in Caucasian Persian is Southwest Iranian in phonetic form, but does not have such evidence in Early New and Modern Persian, should be considered as Middle Persian archaism, which has been preserved from some Middle Persian dialect in Caucasian Persian.

#### **Northwest Iranian (NWI) stratum**

If the Caucasian Persian word in its phonetic form is Northwest Iranian, it could have appeared in that language in three following ways:

1. From the Early New Persian, where some of the Northwest Iranian words were inherited from Middle Persian and are considered to be Parthian borrowings, while others are directly derived from various Northwest dialects. In this case, the proposed law is valid when the

---

<sup>7</sup> Veliyev-Baxarly 1921: 46.

corresponding word is widely used and abundantly attested in classical Persian.

2. From the Talyshi, if the corresponding word is defined as of Northwest Iranian origin, is presented in the Tat dialects (Apsheroni and Central) which are spoken relatively close to the spreadzone of Talyshi. In addition, words of such type must exist in the corresponding or similar phonetic form in Talyshi.

3. From any other language of the Northwest Iranian branch, on which the Caucasian Persian or any of its dialects could be layered. In this case, it is possible to assume and prove, based on the data of early Arab historiography<sup>8</sup>, that some Iranian peoples, in particular Kurds and Deylamits, appeared in Shirvan earlier than the spread of the Early New Persian in the eastern part of Transcaucasus. A prerequisite for this point is that the corresponding word should first be defined as of Northwestern Iranian origin and be used in dialects far from the Talysh-speaking area, as well as be absent in Talyshi.

### **SWI inherited lexicon in Caucasian Persian**

Like other languages, the inherited lexicon of Caucasian Persian includes mainly the native stock of words, which consists of:

- More than 200 simple verbs like *bardan* < Early NP *burdan* “carry”, *daran* < Early NP *dādan* “give”, *čiran* < Early NP *čīdan* “to gather, reap”, *āmāran* < Early NP *āmadan* “to come”, *suxtan* < Early NP *suxtan* “to burn”, *zaran* < Early NP *zadan* “to hit”, *biran* < Early NP *budan* “to be”, *guftiran*//*kutiran* < Early NP *guftīdan* “to say”, *güftan*//*giftan* < Early NP *giriftan* “to take” etc.

- A large number of animals, birds and plant names, such as *xug* “pig” < Early NP *xūk*, *gürg*//*girg*<sup>9</sup> < Early NP *gurg* “wolf”, *xarguš* < Early NP

<sup>8</sup> See **Jakubovskij 1926**: 65, also **Tonoyan, Sargsyan 2016**: 262-272.

<sup>9</sup> *í//ü* and *ü//i* vowel changes typical for the Caucasian Persian dialects (cf. *dil*//*dül* “heart”, *giftan*//*güftan* “to get”, *birīstan*//*bürüştan* “to roast, fry, bake”, *düzd*//*dizd* “robber, thief”, *püšt*//*pišt* “back”), is one of the phonological features of the eastern dialect group of Azerbaijani (e.g., *bibi* > *bübü* “paternal aunt”, *şəkil* > *şəkül* “picture, form”, *siz* > *süz* “you”,

*xargūš* “rabbit”, *ruvi* < Early NP *rōbāh* “fox”, *kalāq* < Early NP *kalāy* “crow”, *šir* < Early NP “lion”, *sag//say* < Early NP *sag* “dog”, *as* < Early NP *asb//asp* “horse”, *giov* < Early NP *gāh* “plant”, *dār* “tree” < Early NP *dār* “wood”, *talū* < Early NP *talū* “thorn” etc.

- Names of some geographical and natural phenomena, celestial bodies, seasons, times such as *xāk* < Early NP *xāk* “earth”, *kuh // küh* < Early NP *kōh* “mountain”, *dāhār* “cave” < Early NP *dahār* “rock, scar”, *māh* < Early NP *māh* “moon”, *dara* < Early NP *darra* “canyon”, *rāh* < Early NP *rāh* “way”, *šāhāngum* < Early NP *šabhangām* “evening”, *zimistun* < Early NP *zimistān* “winter”, *pariz* < Early NP *pādiz* “autumn”, *zimilarz* “earthquake” (cf. Modern Pers. *būmlarz*), *tagarg* < Early NP *tagarg* “hail”, *ruz* < Early NP *rōz* “day”, *sāl* < Early NP *sāl* “year” etc.

- Kinship lexemes, such as *māy//moy//mö* < Early NP *mādar* “mother”, *piyar* < Early NP *pidar* “father”, *zan* < Early NP *zan* “woman”, *duxtār* < Early NP *duxtār* “girl”, *birār* < Early NP *barādar* “brother”, *xuvār* < Early NP *xāhar* “sister”, *birārzara* < Early NP *barādarzāda* “nephew, son of a brother”, *xuvārzara* < Early NP *xāharzāda* “nephew, son of a sister” etc.

- Somatic lexemes, such as *sar* < Early NP *sar* “head”, *gardun* < Early NP *gardān* “neck”, *dūš* < Early NP *dōš* “shoulder”, *išqam* < Early NP *šikam* “belly”, *dül* < Early NP *dil* “heart”, *ruru* < Early NP *rūda* “intestine”, *čüm* < Early NP *čašm* “eye”, *angūšt* < Early NP *angušt* “finger”, *pūšt* < Early NP *pušt* “back” etc.

- Numerals, pronouns, qualitative adjectives etc.

### SWI loans

Modern Persian and Caucasian Persian, as descendants of early New Persian, naturally have a huge number of words with the same phonetic

*sifət* > *sūfət* “face” etc.). The easternmost border of this vowel change is Türkan in the Apsheron peninsula, westernmost is Birinci Ərəbcəbirli in Göyçay district, southernmost is Xol Qarabucaq in Neftçala, and northernmost is Xudat in Xachmaz district, that coincide almost exactly with the borders of the area where Caucasian Persian was spoken. Consequently, it could be regarded as a result of Caucasian Persian phonological influence on the Baku dialect of Azerbaijani (Tonoyan 2019: 370-371).

form that cannot be considered Persian borrowings in Caucasian Persian. However, there are certain phonetic laws that are specific only to Caucasian Persian, and therefore the exceptions to these laws should certainly be considered Persian borrowings in Caucasian Persian.

The above-mentioned laws are as follows:

1. Old Iranian dental \*-t- shifts to -r- in intervocalic position, while in Early NP it shifts to -ḍ- (>-d-),

2. Old Iranian labio-dental \*v- is preserved in anlaut, while in Early NP initial \*v- shifts to b-.

If these historical phonetic laws are taken as a basis, then the following words of Caucasian Persian should be considered as borrowings from New Persian directly or in some cases indirectly through Azerbaijani. The following are examples of this group:

Lah. *haftod* “seventy” (< Early NP *haftād* < OIr. \**haftāti-*),

Lah. *haštod* “eighty” (< Early NP *haštād* < OIr. \**aštāti-*),

Lah. *nāvād* “ninety” (< Early NP *navad* < OIr. \**navatī-*),

Lah. *xudo* “God” (< Early NP *xudā* < OIr. \**xwatāya-*),

Abš. *bahar* “spring” (< Early NP *bahār* < OIr. \**vāhara-*),

Abš. *šad* (in other dialects *šar*) “happy” (< Early NP *šād* < OIr. \**šyāta-*),

Abš. *šudan* “to be, to become” (< Early NP *šudan* < OP. *šuta-* (past participle),

Abš. *piyada* “on foot” (< Early NP *piyāda*),

*bist* (in all dialects) “twenty” (< Early NP *bīst* < OIr. \**visat-*)

*banöšā* (in all dialects) “violet” (< Early NP *banāšā* < OIr. \**vana-vaxša-<sup>10</sup>*),

*burun* (in all dialects) “rain” (< Early NP *bārān* < MP *vārān*).

### MP archaisms

There are a number of lexemes used in different dialects of Caucasian Persian, which have phonetic appearance typical of SWI languages, but have no evidence in New Persian. At the same time, it should be noted that some of these words are attested in Middle Persian. Therefore, taking into account

---

<sup>10</sup> Müller 1893: 368.

the above-mentioned two points, they can be considered as Middle Persian archaisms in Caucasian Persian.

The following are two examples of Middle Persian archaisms in Caucasian Persian:

*zafru* “down, under” (cf. MP *zurf*<sup>11</sup>, Parth. *žafra*<sup>12</sup> and NP *žarf*, which is a NWI loan in NP) < OIr. \* *ǰafra-* “deep”

*hāmi* “summer” (Qub. *homi*, Lah. *homin*, Abš., Šm̄. and Central *hāmi*) < MP *hāmīn*<sup>13</sup> < OIr. \* *hāmīna-*<sup>14</sup>.

### **NWI substratum and borrowings**

As was mentioned, words of NWI origin could appear in Caucasian Persian in three main ways:

1. Northwest Iranian, probably Deylamit substratum
2. Early New Persian
3. Talyshi language

In regard to the classification of the Northwest Iranian lexical substratum and borrowings according to the mentioned three groups which are left for further detailed examination, it should be mentioned that any word in Caucasian Persian can be placed under this group if it complies with the following historical phonetic laws:

- The preservation of Old Iranian labio-dental \**v-* in anlaut
- The shift of the Old Iranian predental-prepalatal consonants \**ǰ-(-ǰ-)* and \**č-(-č-)* to predorsal-prepalatal *ž* in anlaut and intervocalic position.

In addition to the words corresponding to these two points, this group may include those words which do not exist in Classical and Modern Persian but are present in the neighboring Talyshi or other Northwest Iranian languages and have other phonetic features specific to the languages of the Northwest Iranian branch.

---

<sup>11</sup> Nyberg 1974: 232.

<sup>12</sup> Durkin-Meistererents 2004: 198.

<sup>13</sup> Mackenzie 1986: 41.

<sup>14</sup> Rastorgueva, Edelman 2007: 351.

The following examples of this group are:

*varf* “snow” (cf. Gil. *varf*, Tal. *voə*, Kaš. *vafr* // *fafr*, Early NP *barf*) < Olr. \**vafra-*,

*vār* (Lah. *vor*) “wind” (cf. Tal. *vo*, Gil. *va*, Maz. *vā*//*vo*, Early NP *bād*) < Olr. \**vāta-*,

*valg* “leaf” (cf. Tal. *vəlg*, Gil. *vəlg*, Maz. *varg*//*valg* //*valk*, Early NP *barg*) < Olr. \**varka-*

*virixtan* (Abš. *gurixtan*) “to flee” (cf. Tal. *vite*, Early NP *gurixtan*) < Olr. \**vi-raēča-*

*vini* “nose” (cf. Gil. *vini*, Tal. *vini*, Maz. *veni*) < Olr. \**vaina-ka* (cf. Av. *vaena-*)

*žuži* “ hedgehog” (cf. Tal. *žaže*, Early NP *žūž(a)*, MP *zuzak*) < Olr. \**ǰū/uǰaka-*

*tiž* “sharp” (cf. Tal. *tiž*, Parth. *tēž*, Early NP *tīz* (<MP *tēz*) <Olr. \**taiǰa-*

*giž* “madman” (cf. Tal. *giž*, Kd. *giž*, Early NP *gēǰ*) < Olr. \**gaiza-*

The words phonetically of Northwest Iranian origin are numerous in Caucasian Persian and are not limited to the given list. There are some words of NWI origin not included in the list, which also are borrowings from Talyshi, e.g. *kila* “girl” found in the Lahiji and Shamakhi dialects and *delna* “woman” used in the central dialect<sup>15</sup>.

As a summary, it should be noted that although Caucasian Persian is a SWI language, but in different historical periods has had active contacts with various NWI languages, traces of which are preserved in the lexicon of this language. The proposed methodology for classifying NWI strata can be revised after the publication of new dialect data. A more detailed study of existing dialects and the publication of new material in the future could make it possible to review the methodology proposed in this article for the classification of NWI strata of Caucasian Persian.

---

<sup>15</sup> Tonoyan 2015b: 76.



### ***Abbreviations***

Apsh.	Apšeron dialect (Caucasian Persian)
Av.	Avestan
Gil.	Gilaki
Kash.	Kashani dialect
Kd.	Kurdish
Lah.	Lahīj dialect (Caucasian Persian)
Maz.	Mazandarani
MP	Middle Persian
NP	New Persian
NWI	Northwest Iranian
OP	Old Persian
Old Ir.	Old Iranian
Parth.	Parthian
Qub.	Quba dialect (Caucasian Persian)
Shmkh	Shamakha dialect of Caucasian Persian
SWI	Southwest Iranian
Tal.	Talyshi

### **BIBLIOGRAPHY**

- Durkin-Meistererents D. 2004:** Dictionary of Manichaean Middle Persian and Partian, Belgium.
- Grunberg A. 1961a:** O meste tatskogo sredi iranskih yazykov, Voprosy yazykoznanija (On the place of Tat language among the Iranian languages, Questions of linguistics), № 1, 106-114 (in Russian).
- Grunberg A. 1961b:** K voprosu o yazykovom vzaimodeystvii (na materiale yazyka severoazerbajdžanskikh tatov, Kratkie soobščeniya Instituta narodov Azii (On the issue of linguistic interaction (based on the language of the North Azerbaijani Tats, Brief reports of the Institute of the Peoples of Asia), vol. XL, 11-23 (in Russian)

- Grunberg A. 1963:** Jazyk severoazerbajdžanskix tatov (The language of the North Azerbaijani Tats,), Leningrad (in Russ.).
- Voskanian V. 2016:** Some Armenisms in Caucasian Persian language, Arevelagitut'yan harc'er (Problems of Oriental Studies), vol. 12, 240-248 (in Armenian).
- MacKenzie D.N. 1986:** A concise Pahlavi dictionary, London.
- Miller B. 1929:** Taty, ix rasselenie i govory, In Izvestija Obščestva obsedovanija i izučenija Azerbajdžana (Tats, their settlements and dialects, v News of the Society for the Survey and Study of Azerbaijan), Baku, 1-44 (in Russian).
- Minorsky V. 1934:** Tāt, Enzyklopaedie des Islām, Bd. IV, Leiden–Leipzig, 697-700.
- Minorsky V. 1958:** A History of Sharvān and Darband in the 10th-11th centuries, Cambridge.
- Müller F. 1893:** Pahlavi-, neupersische und armenische Etymologien, Wiener Zeitschrift für die Kunde des Morgenlandes, vol. 7, 366-382.
- Nyberg H. 1974:** A Manual of Pahlavi, II: Glossary, Wiesbaden.
- Rastorgueva V., Edelman D. 2007:** Etimologičeskij slovar' iranskix jazykov (Etymological dictionary of Iranian languages), t. 3, Moskva (in Russian).
- Tonoyan A. 2014:** Das Problem Des Phonems ž in Den Nordtatischen Dialekten von Berg-Schirwan, Wiener Zeitschrift für die Kunde des Morgenlandes, vol. 104, 193-197.
- Tonoyan A. 2015a:** Caucasian Persian: comparative historical research (unpublished Ph.D. dissertation), Institute of Language after Hr. Acharyan NAS RA, Yerevan (in Armenian).
- Tonoyan A. 2015b:** Etymological study of some words in Caucasian Persian, Banber Erevani Hamalsarani. Banasirut'yun (Bulletin of Yerevan university. Philology), № 1 (16), 72-79.
- Tonoyan A., Sargsyan H. 2016:** The Problem of some Arabic mentions of a word “Kurd” in Post-marzpanate Albania, Arevelagitut'yan harc'er (Problems of Oriental Studies), vol. 12, 262-272 (in Armenian).
- Tonoyan A. 2019:** On the Caucasian Persian (Tat) Lexical Substratum in the Baku Dialect of Azerbaijani. Preliminary Notes, Zeitschrift Der Deutschen Morgenländischen Gesellschaft, vol. 169, no. 2, 367–378.
- Veliev-Baxarly M. 1921:** Azerbaydz'an: fiziko-geografičeskij i ètnografičeskij očerk (Azerbaijan: physical-geographical and ethnographic sketch), Baku (in Russian).

**Jakubovskiy A. 1926:** “Ibn-Miskavey o pokhode Rusov v Berdaa v 332g. = 943/4 g. (Ibn-Miskavey about the campaign of Ruses to Berdaa in 332/943/4), Vizantijskij vremennik, t. XXIV, 63 -92 (in Russian).

**Artyom Tonoyan**

Institute of Oriental Studies of NAS RA, Department of Iran,  
Yerevan State University, Republic of Armenia, Department of Iranian Studies,  
artyom.tonoyan@ysu.am

## **ՄԻ ՔԱՆԻ ԴԻՏԱՐԿՈՒՄ ԿՈՎԿԱՍՅԱՆ ՊԱՐՍԿԵՐԵՆԻ ՔԱՌԱՊԱՇԱՐԻ ԻՐԱՆԱԿԱՆ ՇԵՐՏԻ (ԹԱԹԵՐԵՆ) ԾԱԳՄԱՆ ՇՈՒՐՋ**

*Արտյոմ Տոնոյան*

**Քանալի բառեր՝** Կովկասյան պարսկերեն, իրանական շերտ, միջին պարսկական հնաբանություններ, հյուսիսարևմտահրանական փոխառություններ, հյուսիսարևմտահրանական ենթաշերտ, հյուսիսային թալիշերեն, հարավարևմտահրանական փոխառություններ, կովկասյան պարսկական բնիկ բառազանծ:

Կովկասյան պարսկերենը (թաթերեն), լինելով Այսրկովկասի արևելյան հատվածում խոսվող իրանական լեզուներից միակը, որ պատկանում է դրանց հարավ-արևմտյան ճյուղին և սերում է վաղ նոր պարսկերենից, մեծ հետաքրքրություն է ներկայացնում և՛ արեալային լեզվաբանության, և՛ հատկապես պատմահամեմատական լեզվաբանության, իրանական լեզուների և դրանց կրող ժողովուրդների միջև պատմական շփումների և փոխազդեցությունների ուսումնասիրության համատեքստում:

Կովկասյան պարսկերենի բառապաշարի ուսումնասիրությունը բացահայտում է վերջինիս՝ ինչպես ոչ իրանական, այնպես էլ իրանական լեզուների հետ ունեցած առնչակցությունների և կրած ազդեցությունների տարբեր շերտեր, որոնք արդյունք են ինչպես նորագույն, այնպես էլ պատմական ավելի վաղ շրջաններում ընթացած շփումների և փոխազդեցությունների:

Հոդվածում ներկայացված են մի քանի դիտարկումներ կովկասյան պարսկերենի և դրա բարբառների իրանական շերտի տարբերակման

շուրջ: Այս լեզվի բառապաշարի ուսումնասիրությունը թույլ է տալիս իրանական ծագման բառերի համար առաջարկել դասակարգման հետևյալ մոդելը՝

1. **Հարավարևմտաիրանական շերտ**, որի մեջ մտնում են՝

- կովկասյան պարսկերենի բնիկ բառազանձը, որը գրեթե նույնական է վաղ նոր պարսկերենի հետ,

- միջին պարսկերեն հնաբանությունները,

- նոր և ժամանակակից պարսկերենից արված ուղիղ կամ միջնորդավորված փոխառությունները:

2. **Հյուսիսարևմտաիրանական շերտ**, որի մեջ մտնում են՝

- Մինչև լեզվի ձևավորումը տվյալ տարածքում խոսված որևէ հյուսիսարևմտիրանական լեզվից կամ լեզուներից պահպանված և հետագայում կովկասյան պարսկերենին անցած բառային ենթաշերտ:

- Լեզվի ձևավորման վաղ շրջանում վաղ նոր պարսկերենից ժառանգված հյուսիսարևմտաիրանական բառեր:

Թալիշերենից կատարված փոխառություններ: