Univerzita Karlova v Praze

Filozofická fakulta

Ústav obecné lingvistiky

Obecná jazykověda – Indoevropská srovnávací jazykověda

Ľubomír Novák

Problem of Archaism and Innovation in the Eastern Iranian Languages

K problému archaismu a inovace ve východoíránských jazycích

Teze

vedoucí práce – Doc. PhDr. Petr Vavroušek, CSc.

The Eastern Iranian languages form an independent group within the Iranian branch of the Indo-European languages. The presented thesis aims to present an outline of development of the Eastern Iranian languages — as languages develop, they usually start to differ from its relatives by development of various innovations and/or by preservation of archaisms. The spread of innovations and preservation of archaisms may vary in individual languages or dialects and study of sets of common innovations and/or archaisms may characterize grouping of languages of a given branch. To see the Eastern Iranian archaisms and innovations I have decided to focus on three fields of study — 1) an outline of the Eastern Iranian languages, 2) historical grammar of Sogdian and Yaghnōbī and 3) lexical study.

The first part will be dedicated to the description of the attested Eastern Iranian languages and dialects – each language (or a subgroup) will be briefly described with focus on common data about the individual language(s), with an overview of main phonetic changes and grammar outline. For the overviews I will mark only some archaic and innovative features of the individual languages as for each language can be written separate book on its historical grammar and phonology. I would also like to (re)examine commonly accepted grouping of the Eastern Iranian languages into the Northern and Southern branches as it seems to me that this grouping needs a new revision.

The second part will present comparation of development of Sogdian and Yaghnōbī – i.e. two languages that are considered closely related by many scholars (e.g. BOGOLYUBOV 1956; KLIMCHITSKIY 1935), but none of them has ever presented thorough study of their differences – Yaghnōbī was in common just considered as a dialect quite different from literary Sogdian. By comparation of phonology and morphology of both languages I would like to show main differences between them and if possible I would like to try to define interrelationship of Sogdian and Yaghnōbī. The comparative study of Yaghnōbī and Sogdian has been taken intentionally – as both languages are comparable from diachronic point of view, their comparison may answer more questions than just their "dialectal" relationship. Historical development of Sogdian and Yaghnōbī will be compared with the other Eastern Iranian languages with focus on the Pāmīr group.

The third part will present a study of Sogdian and Yaghnōbī lexicon. I have compared Yaghnōbī and Sogdian lexicon according to the "Swadesh List" of 207 words combined with "Standard Word List Items" presented in *Sociolinguistic Survey of Northern Pakistan*. The lexical items of both languages will be supplemented by their etymology. The choice of the Swadesh List was not motivated by attempts of glottochronological study of both languages — I just wanted to exploit an accepted list of basic vocabulary. Both lists try to present unbiased choice of basic vocabulary so in this issue I have also to study eventual loans (mainly in case of Yaghnōbī).

As can be seen from outlines of all three parts, my study of the Eastern Iranian archaisms and innovations aims to present new classification of the Eastern Iranian branch with focus on position of Sogdian and Yaghnōbī within this language branch.

* * *

The Iranian languages form a group of genetically related languages and dialects that developed from the Indo-Iranian branch of the Indo-European languages. By use of methods of historical and comparative linguistics we can explain the origin of the Iranian languages as a split of the Indo-Iranian branch of *Proto-Indo-European language. The original *Proto-Indo-Iranian language broke up into the four main branches: Iranian, Nūristānī (or Kāfir), Dardic and Indo-Aryan. Particular prehistoric dialects of Indo-Iranian share with *Proto-Indo-European (and also with many other Indo-European languages) many common features – so called archaisms as well with series of innovations that set them apart from the proto-language. Some of the innovations can be observed in more branches of the Indo-European languages, but are not phenomena proper to the original system of reconstructed *Proto-Indo-European.

The Iranian languages are divided into two main branches – Western and Eastern. Their division is based on agreed conventional brake up of two Old Iranian dialects according to their geographical location to the East and West respectively from the deserts of Central Iran (ÈDEL'MAN 1986, 3). Present geographical spread of the Eastern and Western Iranian languages and their speakers has changed due to historical migrations of the Iranian peoples (e.g. Western Iranian Balōchī is nowadays located in Eastern Iran and Western Pakistan or the Eastern Iranian Ossetic is to be found on the Caucasus), the contemporary location of the Iranian languages is not relevant for their classification. The Iranian languages can be thus considered as an offspring of the Indo-European proto-language with which they are connected by genetic relationship and a preservation of some (*Proto-)Indo-European archaisms, on the other hand they differ from *Proto-Indo-European by several innovations which define this language family from historical point of view.

We are informed about the history of the Old Iranian languages by means of indirect sources. Herodotus for example mentions several Scythian words, in one case he even presents an etymology (HERODOTUS IV, 110; HINGE 2006). He also mentioned that the Sauromatians speak the language of Scythia, but they do not speak it well because the Amazons did not learn properly the Scythian language - Herodotus mentioned that the Amazons married some Scythians and by this the Sauromatian nation came into being (HERODOTUS IV, 117). Herodotus also writes about an older poem, Arimaspea, written by Aristeas of Proconnesus (HERODOTUS IV, 13). It is said that Aristeas described the habits and the language of Scythian Issedonians (Issedones) and Arimaspians (Arimaspi) who dwelled in regions to the North-East of the Pontic or Black Sea (ALEMANY I VILAMAJÓ 1999). Unfortunately, Aristeas' Arimaspea has not came down up to these days, it is only mentioned in the Histories of Herodotus and also in Пะค์ ปัง by Longinus and in Chiliades (or Book of Histories) by John (Ioannes) Tzetzes (TZETZES, Chil. VII, 686-692). In the Anabasis of Arrian there are mentioned several local tribal and personal names of Central Asia, but we miss any reference to the languages of the region, the only relevant information is that the river Ἰαξάρτης (Sīr Daryā) was called Ὁρξάντης in a language of barbarians of Sogdiana (ARRIAN III, 30.13). In Strabo's Geography is mentioned, that the northern part of Åρειανή (i.e. approximately area of modern Afghanistan, Eastern Iran, Tajikistan, Turkmenistan, and North-western Pakistan) is inhabited by Bactrian and Sogdian peoples who do speak similar languages (STRABO, Geography, XXV, 2:8). The city of Kūrkat in present northern Tajikistan is known from the antiquity – it is spelled either as Κῦρούπολις or as Κῦρούπολις is probably a calque of Iranian dialect by the analysis of both Greek names: Κῦρούπολις is probably a calque of Iranian appellative *Κūruš-kąβā- 'city of Cyrus'. What is even more interesting is the form Κῦρέσχατα, it can be an attempt to render the local name *Κūruš-kąβā- (cf. Tjk. and Pers. Κūrkát); the Greek name is probably contaminated by another Greek word ἐσχάτη 'the farthest' (probably by an influence by the name of the city of Alexandria the Farthest – Ἀλεξάνδρεια Ἐσχάτη, present Khujand, in the Soviet period known as Leninabad). City of 'Ρωξανάκη mentioned by Ctesias of Cnidus can be connected with city of Rōshān (Rōshānī Rixūn, Tajik Rūšōn) in Tajik Badakhshān (ABAEV 1949, 178).

The Iranian language family is conventionally divided into two basic groups – Eastern and Western Iranian. Differences between these two groups begun to appear probably in the Old Iranian period and became more distinctive in the Middle Iranian period. Each of these groups later split into two subgroups – South and North subgroup. Among the North Eastern Iranian are classified Scythian dialects and *Sauromatian (in the Old Iranian period), Sarmatian, Alanic, Sogdian (Middle Iranian period) and Ossetic and Yaghnōbī (New Iranian period). South Eastern Iranian languages are represented by dialects of the Saka (mainly Khōtanese and Tumshuqese), Bactrian (Middle Iranian period), the Pāmīr languages (Shughnī-Rōshānī group¹, Yazghulāmī, Wanjī, Wakhī, Ishkāshmī-Sanglēchī, Munjī-Yidghā and probably Sarghulāmī), Pashtō and Waṇetsī (New Iranian period). Questionable is classification of the Avestan language – it is probably one of the South Eastern Iranian, Khwārezmian is variously classified as North or South Eastern Iranian; the most complicated is classification of Parāchī and Ōrmuṛī – some scholars claim them as North Western Iranian but some other hive off new – Southeast branch within Eastern Iranian.

The South Western Iranian languages and dialects differ from other Iranian languages by significant isogloss Ir. *t, *dz, *tu > *t, *tu > *t > *t | *t |

However, distinctive features cannot be found only on phonological level. There were not many phonological differences between the Eastern and Western Iranian in the Middle Iranian period, one of the essential features was development of word-initial voiced stops *b-, *d-, *g- and development of clusters *ft and *xt. To establish a border between the Eastern and Western

¹ Id est Shughnī, Shākhdaraī, Bajūī, Khūfī, Rōshānī, Bartangī, Rāshārvī and Sarīqōlī.

² But in Wakhī *tu > s and in Khōtanese *tu > ss [ʃ].

Iranian, lexical (e.g. in many works presented example *gari- \times *kaufa- 'mountain' and *kapă- \times *mătia- 'fish'; cf. SIMS-WILLIAMS 1989a, 168-169) and grammatical differences should be also taken into account. There can be mentioned some other words from lexicon that can be considered typical for the Eastern Iranian area:

*abi-ar- 'to find, to obtain' > Sogd. s B M $\sqrt{\beta}yr$ C $\sqrt{b}yr$ / $\sqrt{\beta}$ ir/, Khwār. βyr -, Bactr. M $^2\beta yr$ -, Yazgh. vir-, Yagh. $v\bar{i}r$ -;

*(h)ánda- 'blind' > Khōt. hana-, Sogd. B ²nt M ²nd /aṁd/, Munj. yāndəy, Pasht. ṛūnd, Ōrm. hōnd (but cf. Parth. hand);

*aua-súxta-(ka-) 'clean, purified' > Khōt. Tumshuq. vasuta-, Sogd. B 'wswγty, 'ws(')wγt'k M 'wswγtyy /osuγdĕ/, Bactr. ωσογδο /osuγd/, Oss. (without prefix) sыγdæg || suγdæg, Khwār. (with other prefix) (')fsγd;

*dráua- 'hair' > Khōt. drau-, Sogd. B zw-y /žəwi/, Yagh. daráu | d¹ráu, Oss. ærdu | ærdo, Shugh. cīw, Rōsh. cōw, Yazgh. ců Ōrm. drī × Pers. mōj < *mauda-;

* $k\acute{a}ta$ - 'house' > Sogd. B kt'y, kt?k M qt, qty(y), $ktyy \in q\underline{t}y$ / $k \Rightarrow t\acute{e}$ /, Bactr. $\varkappa \alpha \delta(\alpha) \gamma \sigma$ /kad(a)g/, Yagh. kat, Shugh. $\check{c}id$, Rōsh. Khūf. $\check{c}od$, Bart. $\check{c}\ddot{o}d$, Rāshrv. $\check{c}\ddot{u}d$, Sarīq. $\check{c}ed$, Yazgh. $k\mathring{u}d$, Munj. $\acute{k}ay$, Yidgh. k' εi , Ave. kata- (+ Parth. Pahl. kdg) × Pers. $\varkappa an\acute{a} < *\varkappa an\^{a}$ -ka- (but Sogd. S B γ 'n'k(h) M χ 'n' / $\chi an \check{a}$ /, Wakh. χun , Ishk. χun , Sangl. χan);

*kúta-, *kutī- 'dog' > Sogd. s Β 'kwt-y M kwt-y, qwt-y / 'skwətí/, Bactr. κοδο /kud/, Yagh. kut, Oss. k^w ы $z \parallel kuy$, Shugh-Rōsh. kud, Sarīq. kыd, Yazgh. k^o od, Ishk. kь $d \times$ Pers. $sag < *tu\acute{q}-ka-$, Med. σπάκα (but Khōt. $\acute{s}ve$, Wakh. $\acute{s}a\check{c}$, Pasht. spay (f. spai), Waṇ. spa (f. $sp\bar{i}$), Ōrm. $\acute{s}spuk$, Parāch. $\acute{s}sp\ddot{o}$);

* $m\acute{a}i\Im a$ - 'day' > Sogd. s $my\eth$ B $m(^2)y\eth$ M $my\eth$, $my(y)\eth\eth$ C $my\Im$, $my\varOmega$, $my\varOmega$ /mē \Im /, Khwār. $my\Im$ /mē \Im /, Yagh. $m\bar{e}_{\bar{s}}\parallel m\bar{e}_{\bar{t}}$, Shugh. $m\bar{e}_{\bar{s}}$, Rōsh. Khūf. Bart. Rāshrv. $m\bar{i}\Im$, Sarīq. $ma\Im$, Yazgh. $mi\Im$, Ishk may, Sangl. $m\bar{e}_{\bar{i}}$, Munj. Yidgh. $m\bar{i}\check{x}$ × Pers. $r\bar{o}z$ < * $r\acute{a}u\check{c}a$ - (but Pasht. wraz, rwaz, Waṇ. wrez, Ōrm. wriez, wrioz);

*pati-gadz- 'to accept' > Khōt. pajāys-, Sogd. B √ρcγ²(²)z /√pəčγáz/, Khwār. pcγ²z-;

*tána- 'enemy' > Khōt. Tumshuq. sāna-, Sogd. s в м с s²n /sān/, Oss. son × Pers. dušmán < *duš-mana- (cf. SIMS-WILLIAMS 1989a, 169; SIMS-WILLIAMS 1996b [online]).

Division of Eastern Iranian languages into Northern and Southern branch (and eventually South-eastern branch if we will consider Ōrmuṛī and Parāchī as members of the Eastern Iranian languages) is often used by many scholars, only few of them explain the criteria of such classification, so it seems that this division was more based on (modern) geographical distribution of the Eastern Iranian languages. Some criteria for dividing the Eastern Iranian languages were presented by Vera Sergeevna RASTORGUEVA (1966, 198), but only a few of them fit the classification of the Northern and Southern branches of East Iranian. Comparation of selected sound changes and other features can characterize some isoglosses in the Eastern Iranian languages, but some changes are common for many of these languages regardless to their ranking to the Northern or Southern branch. Instead of classification of the Northern and Southern branch, there can be better postulated a dialect continuum than two different branches; the only (?) branch that seems to show more distinctive features is the South-eastern

branch which continues in the Ōrmuṛī-Parāchī subgroup. As distinctive features of the South Eastern Iranian branch can be considered 1) preservation of archaic formation of plural (i.e. absence of innovation of plural form by adding an abstract suffix *-tã); 2) sonorization of intervocalic *-ṣˇ-; 3) change of Ir. *rd, *rt; 4) change of Ir. *rdz, *rt and 5) emergence of innovated form of the second person plural personal pronoun from combination of forms of the second person singular and first person plural. All the above mentioned changes have not emerged in all South Eastern Iranian area: feature 1) have not took place in Yazghulāmī (and except some non-productive forms in Ishkāshmī); intervocalic *-ṣˇ- has not been sonorized in Bactrian and probably also in Sarghulāmī; changes under the point 3) have not taken place in Bactrian and Wakhī; in Munjī, Yidghā and Wakhī (and probably also in Bactrian) has not taken place change point 4); innovated forms of plural the second person plural (point 5)) are present in all South Eastern Iranian languages, but in Parāchī, Saka dialects and in Khwārezmian they come from a different source than from the above mentioned.

The Modern Eastern Iranian languages can be divided into five branches: I Northern (Sogdo-Scythian) group; II North-eastern (Saka) group, III Central (Pāmīr) group, IV Southern (Paṭhān) group and V South-eastern (Hindūkush) group. Group I can be defined by innovated plural ending *-tǎ- (comparable to Yazghulāmī), preservation of intervocalic *-š- (shared with Bactrian and Wakhī but excluding Ossetic). Groups III, IV, V have undergone common change of form of the second person personal pronoun, in languages of these groups there are innovated forms of plural, they may be influenced by Indo-Aryan or Dardic pronouns. Innovated forms of the second person plural often comes from combination of personal pronoun of the second person singular with form of the first person plural *talu-*ahmá-(k/xam-), or *talu-s,ma- copied from Indo-Aryan (cf. Maiyā tus; Ṣiṇā tsalo; Lahndā tus) different form is just in Parāchī. Groups II and IV share sonorization of word-initial *fr-, *\$r-, *xr-.

Some isoglosses can be demonstrated on following examples:

```
*čášman- 'eye'

I Sogd. s c(š)m-y m cm-y(y), cšm-y c c(y)m-y, cšm-y /či(š)mi/; Oss. cæst, casm || cans 'window-opening'

II Khōt. tse'iman-
III Ishk. com, Sangl. cām, Zēb. cōm, Munj. čōm, Yidgh. čam, Shugh. Baj. cēm, Rōsh. Khūf. cām, Bart. cēm, Rāshrv. cīm, Sarīq. cem, Yazgh. čăm, Wakh. čə(ž)m V Ōrm. cimī, čīm, cōm
? Khwār. cm-, cm-/camma/, Ave. čašman-

*Эrāja- 'three'

I Sogd. s ¬ðry mg ¬ðryw в (²)ðry м ¬ðry(y) с šy /¬ṣai/, Yagh. ṣaráy || tiráy, Oss. ærtæ
II Khōt. drai, Tumshuq. dre
```

III Yidgh. xiray, xuroy, Munj. xiray, Shugh. aray, Baj. Bart. Rosh. aray, Sariq.

```
aroy, Ishk. růy, Sangl. rōy, Yazgh. cůy, Wakh. trū(y) {Bactr. υαρηιο /hərēy/}
       IV Pasht. drē, Wan. dre
       V Ōrm. šō, řī, Parāch. šī, šu
       ? Khwār. šy /šē/, Ave. $rāiiō
*¡ušmā́xam 'you'
       I Sogd. S B (?)\check{s}m^2\gamma w, \check{s}m^2\gamma h M \check{s}m^2x(w), \check{s}m^2x C \check{s}m^2x /\check{s}m\dot{a}x(w)/, Yagh. \check{s}^um\dot{a}x,
       Oss. sыmax || sumax
       II Khōt. uhu, umă, umä, LKhōt. ama
       III Wakh. sá(y)išt, Ishk. tьтьх, Sangl. təтəх, Munj. mōf, Yidgh. măf, mŏf, Shugh.
       Rōsh. Khūf. tama, Bart. Rāshrv. tamāš, Sarīq. tamaš {Bactr. τωμαχο, τομαχο,
       ταμαχο, /tōmāx, tumāx, tamāx/}
       IV Pasht. tásē, tásō, Waņ. tās
       V Parāch. wā, Ōrm. tōs, tyūs
       ? Ave. yūžām, Khwār. hBy
*gauša- 'ear'
       I Sogd. S B M C γwš /γōš/, Yagh. γūš, Oss. qus || γos, Scyth. <sup>()</sup>γωσος
       II Khōt. gguv'a-, ggū'
       III Wakh. yǐiš, Ishk. yů, Sangl. yōl, Shugh. yūy, Rōsh. yōw, Sarīq. yawl,
       Yazgh. yəvon, Munj. yūy, Yidgh. yū(ĭ)
       IV Pasht. ywać, yważ
       V Ōrm. gōī, gōy, Parāch. gū
       ? Khwār. \( \gamma wx / \gamma \opix/\), Ave. \( gaoša-\)
```

The issue of reclassification of the Eastern Iranian languages was only outlined in this thesis, the question still waits for its thorough examination. Valentina Stepanovna Sokolova studied genetic relations of Yazghulāmī and the Shughnī-Rōshānī group (SOKOLOVA 1967) and later relations of the Shughnī-Yazghulāmī group with Munjī (SOKOLOVA 1973). Studies of genetic relations of Munjī and Yidghā with Bactrian and also interrelations of Bactrian with the Paṭhān languages can answer the question of position of Bactrian within the Eastern Iranian group. In a similar way can be studied relationship of Wakhī and the Saka languages – Wakhī appears to share several isoglosses with the Saka languages, but the language shows probable adstrate or substrate phenomena that link it closer to the languages of Pāmīr. Classification of the language of Khwārezm remains to be rather complicated – Khwārezmian shares several isoglosses with Alano-Ossetic languages and with the languages of Pāmīr on one hand, on the other hand there are some similarities with North-Western Iranian Sangesārī (cf. AʿZAMī – WINDFUHR 1972), there are also some isoglosses shared with Sogdian (cf. SIMS-WILLIAMS 1989a, 170); summary of possible connections of Khwārezmian with Avestan have been presented by David Neil MACKENZIE (1988) and by Vladimir Aronovich Livshits (1962, 140).

Majority of the Eastern Iranian languages can be studied mainly from synchronous point of view – these languages and dialects are attested as individual stages of the Eastern Iranian branch but with some exceptions we do not know their older development stages. There is exception within the North Eastern Iranian branch – in this case both Yaghnōbī and Ossetic can be compared with their closely related ancestors. The development of Ossetic can be continuously observed from the Old Iranian period – there are many similar features in the Scytho-Sarmatian dialects and in Alanic that can be compared with Ossetic and we can even suppose that Ossetic is a modern descendent of one of Alanic (or Sarmatian or even Sauromatian) dialects. Similar situation applies for Sogdian and Yaghnōbī – these two languages are very similar from many points of view, Yaghnōbī has been even labelled 'Neo Sogdian' by some authors (BOGOLYUBOV 1956; KLIMCHITSKIY 1935; SKJÆRVØ 1989, 375-376), nowadays many scholars are inclined to believe that Yaghnōbī may come from some non-attested non-literary dialect of Sogdian (BIELMEIER 1989, 480; SIMS-WILLIAMS 1989b, 173), Al'bert Leonidovich Khromov expressed an opinion that Yaghnōbī could have originate in a non-attested Sogdian dialect of Ustrōshana (KHROMOV 1987, 645), unfortunately there is no relevant data to confirm this hypothesis.

On the basis of the above mentioned data we can declare that a thorough diachronic and synchronic study of the Eastern Iranian languages is possible in its Northern branch – but in the case of Ossetic comparable material lies mainly in lexicon, development of grammar and syntax is blurred (cf. ABAEV 1949). It is of course possible to outline historical development of other (New) Eastern Iranian languages, but in these cases it is necessary to deal only with methods of historical and comparative linguistics because there are not attested direct ancestors of these languages.

Based on the above mentioned facts the main theme of this thesis is the comparison of Sogdian and Yaghnōbī – information on Sogdian are available in a large text corpora from which we can learn about Sogdian grammar, lexicon and syntax; Yaghnōbī as a living language is so far undrawn repository of knowledge – to linguists Yaghnōbī is known a little bit more than hundred years, within that period of time some texts, grammars and lexicons have been published, at the present time a research on the Yaghnōbī language and ethnography is under patronage of the Academy of Sciences of the Republic of Tajikistan, where under the Rūdakî Institute of Language and Literature falls the *Department of Yaghnōbī Studies*. In case of Yaghnōbī (and the other Modern Eastern Iranian languages except Ossetic) it is also necessary to follow development of Modern Persian, mainly its variety in Tajikistan.

A comparison of the Sogdian and Yaghnōbī material can solve the issue of the relationship of both of these languages. It can be supposed that both languages developed from one common North Eastern Iranian proto-language or proto-dialect, such proto-language will be labelled *Proto-Sogdic (i.e. a Central Asiatic variety of "Scythian") here. Later *Proto-Sogdic split into two (or even more) main dialects – *Proto-Sogdian and *Proto-Yaghnōbī. Both *Proto-Sogdian and *Proto-Yaghnōbī are reconstructed as predecessors of the attested languages – Sogdian and Yaghnōbī, besides those two languages there may have been Sogdian dialects of Bukhārā,

Ustrōshana and Zhetisu – *Bukhāran Sogdian is attested by several short texts, *Zhetisu Sogdian is attested on several inscriptions and from historical sources while *Ustrōshanian remains to be a hypothetical Early Mediaeval ancestor of Yaghnōbī, *Ustrōshanian is also thought to be an ancestor of hypothetic *Zarafshānī language/dialect which remained as substrate in Tajik dialects of Mastchōh, Falghar and Fōn.

The *Proto-Sogdic language split into two reconstructible dialects – *Proto-Sogdian and *Proto-Yaghnōbī. For description of the historical phonology of Sogdian it is necessary to outline several stages of development of the Sogdian language, see following scheme:

	*Proto-Sogdic	
	*Proto-Sogdian / *Proto-Yaghnōbī	
	*Old Sogdian	language of Sogdian translation of Aṣəm vohū
4 th -5 th cent.	Preclassical Sogdian	the Ancient Letters
	Early Classical Sogdian	Christian document C 2
7 th -9 th cent.	Classical Sogdian (& Bukhāran dialect)	majority of texts
	Postclassical Sogdian (& Zhetisu dialect)	Brāhmī documents, Christian document C 5
half of the 11 th (?) cent.	(death of Sogdian)	
(middle ages)	*Zarafshānī	preserved only in central Tajik dialects
up to cca. 1900	Early Modern Yaghnōbī	preservation of "majhūl" ō and ⋽
from cca. 1900	Contemporary Yaghnōbī	

Yaghnōbī appears in some aspects more archaic in comparison to Sogdian – it preserves archaic position of stress, augment (though the augment has been innovated in Yaghnōbī), and also better preserves inherited vowels (i.e. there is no reduction of unstressed vowels to *Schwa* as there were no phenomena related to the Sogdian *Rhythmic Law*) and Yaghnōbī dialects show that origins of both dialects can be of an old date. Archaic is also formation of ergative construction in Yaghnōbī and another archaism shared with Avestan, Khōtanese and Khwārezmian is preservation of archaic preterite ending of the third person plural *-ār. On contrary, Sogdian shows archaic features mainly in morphology – the operation of the Sogdian *Rhythmic Law* preserved archaic inflectional system for *light stem* words, and also verbal morphology – Sogdian preserves more inherited verbal forms then does Yaghnōbī.

Both languages share some innovations – main similarity is development of nominal inflection in Yaghnōbī and in case of the *heavy stems* in Sogdian – development of direct and oblique cases is comparable, moreover, Yaghnōbī lost vocative case. Another shared innovation (typical also for other North Eastern Iranian languages) is formation of plural with the abstract suffix *-t(u)ă-. Sogdian innovated ergative construction as it replaced copula by the verb * $d\bar{a}r$ - 'to hold' for transitive verbs (cf. similar development in Khwārezmian), another innovations can be seen in new suffixed forms of verbal inflection. The most important innovation in Sogdian was a stress-shift ("Stress III") and subsequent operation of the Rhythmic Law – in this case originally phonetic change strongly influenced morphology and phonology of the language. Yaghnōbī innovations show spread of prefixed augment by analogy to all verbal forms regardless

of their original prefixes and also reanalysis of verbal endings – original durative ending -išt serves to form simple present and future tenses or as durative marker for the imperfect.

Development of stress in the *Proto-Sogdic language is essential to understand phonology of Sogdian and Yaghnōbī and also to discover differences between both languages. It is not necessary to focus on position of stress in *Proto-Iranian because there was a stress shift in *Proto-Sogdic from which both languages developed. The reconstruction of *Proto-Iranian stress is complex – it can be supposed that the *Proto-Iranian stress was mobile and its position was similar to Vedic. For the reconstruction of Old Iranian stress is essential to study stress in Pashtō (GRYUNBERG – ÈDEL'MAN 1987, 38-39). Position of stress changed also in the other Eastern Iranian languages, mainly in the Pāmīr languages where stress shifts caused either syncopation of unstressed vowels or changes of stressed vowels under operation of \bar{a} - or i-Umlaut; nowadays all Pāmīr languages of Badakhshān have stress on the last syllable.

It seems that predecessors of both Yaghnōbī and Sogdian underwent the same or very similar stress shifts, the results of operations of stress slightly differ in both languages. Some Sogdian words point to original *Proto-(Eastern-)Iranian stress, the place of this stress (*Stress I*) can be reconstructed after operation of *i*-Umlaut, e.g. Sogd. s M C zyrn /zein/ < *dzárania- 'gold' (SIMS-WILLIAMS 1989b, 181). Stress later shifted to another position (*Stress II*): the stress fell on penultimate or antepenultimate syllable. Words with penultimate stress were either disyllabic words or words with a penultima containing long syllable i.e. syllable containing either long vowel (long either naturally or rhythmically) or a diphthong (diphthong could have been formed also by a nasal or (*)r) in a closed syllable; in other positions the stress shifts on antepenultima. Position of stress in Yaghnōbī comes from the results of operation of the *Stress II*, this stress can be observed in Sogdian in results of operation of *i*-Umlaut of several words. Such stress shift is also probably related with change of its strength – many unstressed vowels (in Yaghnōbī often all syllables) were reduced or even syncopated, mainly short vowels directly preceding or following a stressed syllable.

Other stress shift (Stress III) took place only in Sogdian, and this change is related operation of the Sogdian Rhythmic Law; but no such shift has taken place in Yaghnōbī. The Rhythmic Law, which was originally only a phonological feature caused many other changes in Sogdian morphology. The Rhythmic Law divides Sogdian words into two groups – in so-called light and heavy stems (cf. SIMS-WILLIAMS 1984; GMS §484-530). As the heavy stems we can classify words with stressed root syllable, in fact stress falls on the first possible rhythmically long syllable (i.e. either on a long vowel or on a diphthong – in this case diphthongs are considered groups V_i , V_i , V_i , V_i , V_i , V_i in closed syllable), the heavy stems end with a consonant in majority of words. In the light stems stress shifted to the ending – the light stem words do not have rhythmically long root syllables and the stress shifted towards the end of the word, and thus *Proto-Sogdian endings have been preserved. Emergence of the Rhythmic Law also influenced reduction of vowels in unstressed syllables, mainly when they followed stress – in the heavy stems the original endings

disappeared but they remained in the *light stem* forms. Subsequently the last stress shift (*Stress IV*) appears – this stress shifts to the ultimate syllable (Nicolas Sims-Williams suggests this development after an analysis of Sogdian documents in the Brāhmī script, some evidence of this feature can be found in several vocalized documents in the Syriac script; SIMS-WILLIAMS 1996a, 312-313)

As indicated above, mere shifts in stress position presented a significant feature which resulted in further sound changes in Sogdian and in Yaghnōbī. Both languages probably shared similar changes of stress for quite a long period of time during their common development. Yaghnōbī retained original stress on (ante)penultima (i.e. *Stress II*) Sogdian, however, was more progressive and there developed another innovation in stress (*Stress III*), this shift was motivated by rhythmical weight of a syllable – the operation of *Stress III* and the Sogdian *Rhythmic Law* is one of the most important distinctive features distinguishing Yaghnōbī from Sogdian.

We can distinguish three development stages of stress changes: Stress I, Stress II and Stress III – the first two stages can be observed in both languages (there are sources for position of the Stress I mainly in Sogdian, but they can be suggested in Yaghnōbī), Stress III is just Sogdian development – in the scientific literature the Stress III is labelled as the Sogdian Rhythmic Law. In the presented thesis I will use the term "Rhythmic Law" just for the outcome of the operation of the Stress III in all its complexity, mainly as a feature influencing Sogdian grammar; the label Stress III means only phonological shift of stress. In Late Sogdian Stress IV followed. A good example of all stress shifts can be seen in the following example: Stress I *adzám 'I' (Pasht. zə; Waṇ. ze; Munj. za; Yidgh. zo, zə; cf. Ave. azəm, Ved. ahám; Ide. *h¹eĝʰóm) > Stress II *ázam (Proto-Sogdic *ázu; Yagh. (*)az; Wakh. wuz; Ishk. az(i); Sangl. azə; azi; Yazgh. az; Shugh. (w)uz; Rōsh. az; Khūf. Rāshrv. Sarīq. waz; Bart. āz) > Stress III Sogd. S B M 'zw, /əzú/ > Stress IV Sogd. c zw /zu/ (?).

There are also several phonetic differences in development of Sogdian and Yaghnōbī – these features can be considered dialectal and probably they originally led to the assumption that Yaghnōbī may be a dialect of Sogdian. According to the analysis of stress shifts in languages derived from *Proto-Sogdic it can be suggested, that phonological development was also influenced by stress, namely in *(Proto-)Sogdian, where original short unstressed vowels changed to *Schwa* (\mathfrak{p} or its allophone \mathfrak{p}), but remained unchanged in Yaghnōbī.

In morphology the differences between Yaghnōbī and Sogdian arise, mainly due to the operation of the *Rhythmic Law*, but there are also other phenomena that have not been influenced by stress. Fundamental is development of augment in Sogdian and Yaghnōbī – in Sogdian augment has been lost for all non-prefixed verbs, but it has been preserved as so-called internal augment for prefixed verbs (i.e. reflects of augment can be seen after a verbal prefix, in this case prefix usually changes its phonetic form when followed by augment), but in Yaghnōbī augment remained as a distinctive feature of imperfect and was reanalysed by analogy for all verbs as a prefix even for those containing historical verbal prefixes. Other essential

morphological features are two archaisms preserved only in Yaghnōbī – preservation (and reanalysis) of peripheral preterite ending $-\bar{o}r < *-\bar{a}r < \text{Ide.} *-(o)ro / -(o)ror$ and preservation of imperfect ending of the first person plural $-\bar{o}m < *-\bar{a}ma$ in Western Yaghnōbī (in Eastern Yaghnōbī and in Sogdian the imperfect ending of the first person plural has been replaced by original optative ending $*-a\underline{i}ma > \text{Yagh. E } -\bar{i}m$, Sogd. $-\bar{e}m$). The fact that Yaghnōbī dialects developed two different imperfect endings of the first person plural may indicate an early split of *Proto-Yaghnōbī and *Proto-Sogdian, and subsequent innovation of imperfect endings in (*Proto-)Sogdian and *Proto-Eastern Yaghnōbī.

During the development of the Sogdian language, Sogdian nominal morphology gradually simplified inflectional cases and light stem nouns changed their case endings and analogically switched to agglutinative inflection as is attested for heavy stems - the light stems formed minority of nominal roots and as there was double system of nominal inflection in Sogdian the language tended to avoid such dichotomy. As the light stem inflection switched by analogy towards the heavy stem inflection, there remained system of three cases - direct, oblique and vocative, i.e. case system similar to *Proto-Yaghnōbī. This reduced inflectional system is attested in late Sogdian Christian document C 5 (cf. SIMS-WILLIAMS 1982). Also verbal endings tended to be unified for both light and heavy stems. Similarity in "agglutinative" system of late Sogdian inflectional system with Yaghnōbī is striking, but only formally (or say on synchronic level), but diachronically the development in both languages differ. The late Sogdian (or "C 5-Sogdian") system of nominal inflection cannot be considered as a source for development of Yaghnōbī inflectional system as there are still different patterns of stress development in both languages - diachronically Yaghnōbī still preserves stress on its position as it was in *Proto-Sogdic (i.e. Stress II), but (*Proto-)Sogdian certainly developed later stress shift -Stress III that influenced also morphology of the language (i.e. so-called Rhythmic Law), and probably later on another stress shift appeared in (late) Sogdian - Stress IV. The shift towards the Stress IV can be probably connected with the above mentioned simplification of nominal inflectional cases as attested in the document C 5 - the tendency to equalize the three-case system of the heavy stems and the six-case system of the light stems led towards a heavy stem-like agglutinative system. There was probable opposite tendency in stress - it tended to shift towards the end of a word, such tendency can be seen in analysis of Sogdian versification by Elio PROVASI (2009, 351-353) whereas the final state of the Stress IV shift can be seen in the Sogdian documents written in the Brāhmī script (SIMS-WILLIAMS 1996a, 312-313).

Lexicon of both Sogdian and Yaghnōbī differs. This fact can be caused by two facts – I) Sogdian is attested in various documents, but majority of texts are religious texts so the vocabulary often does not describe "basic" vocabulary connected with everyday life of peasants and other common people in Sogdiana, but such vocabulary is well attested in Yaghnōbī as the Yaghnōbīs are semi-nomadic pastoralists and their language preserves many "indigenous" terminology connected with animal husbandry and life in the mountains; and 2) there is

approximately a thousand years long gap between Sogdian and (Modern) Yaghnōbī, during this period the "world of the Sogdians" changed considerably and this development may be observed in development of Yaghnōbī lexicon.

After the fall of Sogdiana and gradual disuse of the Sogdian language (Arabic and) Persian became the *lingua franca* of Central Asia and Persian strongly influenced not only (Pre-Modern) Yaghnōbī, but also many other languages such as the Pāmīr languages, Pashtō, Indo-Aryan Urdū, the Nūristānī and the Dardic languages or Turkic Uzbek, Kyrgyz etc. Modern Yaghnōbī preserves approximately 27% of indigenous vocabulary, other parts of lexicon are borrowings, calques, or Yaghnōbī-Persian (Yaghnōbī-Arabic etc.) compounds. Sogdian lexicon contains also number of borrowings, mainly from Sanskrit, Old Turkic and Aramaic (but excluding "Sogdian" words written with Aramaic ideograms).

Both languages also show similar patterns of word-formation, even Yaghnōbī calques from Tajik show some Sogdic patterns of word-formation. In Yaghnōbī there still remain many suffixes attested in Sogdian, unfortunately many of such suffixes are unproductive in the contemporary language (cf. GMS §935-1166; LIVSHITS – KHROMOV 1981, 434-449; KHROMOV 1987, 665-670).

Some Yaghnōbī words have no Sogdian responses, Sofya Petrovna Vinogradova quotes several of them: $\gamma \acute{u}rda$ 'eye', γayk 'daughter', rax 'mouth'³, $n\bar{o}s$ - 'to take' (VINOGRADOVA 2000, 310), there are many other words without Sogdian etymology, but some of those words have etymology in the Pāmīr languages, e.g. Yagh. γayk 'daughter, girl' may be connected with Yazgh. $\gamma a\check{c}a\acute{g}$, Shugh. $\gamma \bar{a}c$, Rōsh. γac , Sarīq. γoc ; Yagh. $\acute{o}d(^i)ma$ 'Saponaria Griffithiana Boiss. plant' ~ Khūf. $wu\eth{m}$; Yagh. $par\acute{a}m$ 'Cousina umbrosa Buge plant' ~ Khūf. piram, Yagh. piram, Yagh. piram 'Siwēna 'home-made paper-like thin cotton cloth' ~ Shugh. piram, Bart. piram, Khūf. piram, Khūf. piram, Yagh. piram, p

Some other Yaghnōbī words have been recorded in past years, but they are not used in the modern language: man 'apple', $k^i m \acute{e}r$ 'red', $z \acute{e}rta$ 'yellow', \check{sou} 'black', $s^i p \acute{e}ta$ 'white', $v^u r \acute{u}k$ 'eyebrow', $\bar{t}p \acute{o}ra$, $\gamma albal\acute{a}$ 'much, many' and many other (cf. BOGOLYUBOV 1966, 359; KLIMCHITSKIY 1940; NOVÁK [in print]), some other *Early Modern Yaghnōbī words that were also similar in Sogdian were replaced by their Tajik similar-sounding counterparts: * $v\bar{o}\gamma$ (Sogd. $\beta\bar{a}\gamma$) 'garden' × Tjk. > Yagh. $b\bar{o}\gamma$, * $m\bar{o}x$ (Sogd. $m\bar{a}x$) 'moon, month' × Tjk. > Yagh. $m\bar{o}h$ 'month' (cf. BOGOLYUBOV 1966, 359).

From the above mentioned points it thus can be suggested, that Sogdian and Yaghnōbī are closely related languages, but there is no evidence that shows that Yaghnōbī developed directly

³ Yagh. rax has attested Sogdian form s ry k /rəxa/.

from Sogdian. If we assume that Yaghnōbī developed from a Sogdian dialect we have to define such dialect, but evidence of Sogdian dialects is quite deficient. It is certain that both Sogdian and Yaghnōbī developed from the same proto-language, but this proto-language equally differs from both languages in focus — I labelled the proto-language as *Proto-Sogdic which I find appropriate for explanation of development of both Sogdian and Yaghnōbī rather than *Proto-Sogdian as there has to be suggested a an intermediate development stage between *Proto-Sogdic and (literary) Sogdian.

Both Yaghnōbī and Sogdian show many differences, some of them are caused by approximately thousand years of discontinuity of development of both language as Sogdian has been replaced by Persian in the 10th and 11th centuries AD. After the Arabic conquest of Sogdiana both languages were gradually influenced by Persian, strong influence of Persian is visible mainly in Yaghnōbī. As both languages differ according to their attested forms, it can be said that from diachronic point of view they are two similar dialects/languages, both comparable in historical development as Sogdic dialects within the North Eastern Iranian language group.

Abbreviations:

Ave.	Avestan	Rōsh.	Rōshānī
Bactr.	Bactrian (in Greco-Bactrian	Sangl.	Sanglēchī
	alphabet)	Sarīq.	Sarīqōlī
M	Bactrian in Manichaean script	Scyth.	Scythian
Baj.	Bajūwī	Shugh.	Shughnī
Bart.	Bartangī	Shugh-Rōs	h. Shughnī-Rōshānī group
Ide.	(Proto-)Indo-European	Sogd.	Sogdian
Ir.	(Proto-)Iranian	В	Buddhist Sogdian
Ishk.	Ishkāshmī	С	Christian Sogdian
Khōt.	Khōtanese	M	Manichaean Sogdian
Khūf.	Khūfī	мд	Sogdian in Sogdian script -
Khwār.	Khwārezmian		Mount Mugh documents
Munj.	Munjī	S	Sogdian in Sogdian script
Ōrm.	Ōrmuṛī	Tjk.	Tajik
Oss.	Ossetic	Tumshuq.	Tumshuqese
Pahl.	Middle Persian, Pahlavī	Ved.	Vedic, Old Indic
Parāch.	Parāchī	Wakh.	Wakhī
Parth.	Parthian	Yagh.	Yaghnōbī
Pasht.	Pashtō	Yazgh.	Yazghulāmī
Pers.	(Classical) Persian	Yidgh.	Yidghā
Rāshrv.	Rāshārvī		

Bibliography:

ABAEV 1949:

Василий Иванович Абаев: Осетинский язык и фольклор. Москва – Ленинград, 1949.

ALEMANY I VILAMAJÓ 1999:

Augustí Alemany i Vilamajó: Els «Cants arimaspeus» d'Arísteas de Proconnès i la caiguda dels Zhou occidentals. Faventia 21/2, 1999, p. 45-55.

A^{ς} ZAMĪ – WINDFUHR 1972:

Cheragh Ali Azami – Gernot Windfuhr: A Dictionary of Sangesari, With a Grammatical Outline. Tehrān (: Franklin Book Company), 1972.

چراغعلی اعظمی – گرنت ل. ویندفوهر: واژهنامهٔ سنگسری، با مقدمهای از دستور آن زبان. تهران (: موسسه انتشارات فرانکلین)، ۱۳۵۱.

BIELMEIER 1989:

Roland Bielmeier: Yaghnōbī. In: Rüdiger Schmitt (ed.): Compendium Linguarum Iranicarum. Wiesbaden (: Dr. Ludwig Reichert Verlag), 1989, p. 480-488.

BOGOLYUBOV 1956:

Михаил Николаевич Боголюбов: Ягнобский (новосогдийский) язык. Исследование и материалы. Автореферат на соискание учёной степени доктора филологических наук. Ленинград 1956.

BOGOLYUBOV 1966:

Михаил Николаевич Боголюбов: Ягнобский язык. In: В. В. Виноградов (ed.): Языки народов СССР. Том первый: Индоевропейские языки. Москва (: Наука), 1966, р. 342-361.

ÈDEL'MAN 1986:

GMS:

Ilya Gershevitch: A Grammar of Manichaean Sogdian. Oxford, 1954.

GRYUNBERG – ÈDEL'MAN 1987:

Александр Леонович Грюнберг – Джой Иосифович Эдельман: Афганский язык. In: Вера Сергеевна Расторгуева (ed.): Основы иранского языкознания. Новоиранские языки II. – Восточная группа. Москва (: Наука), 1987, р. 6-154.

HINGE 2006:

George Hinge: Herodot zur skythischen Sprache. In: Glotta 81 2005[2006], p. 86-115.

KHROMOV 1987:

Альберт Леонидович Хромов: Ягнобский язык. In: Вера Сергеевна Расторгуева (ed.): Основы иранского языкознания. Новоиранские языки II. – Восточная группа. Москва (: Hayka), 1987, p. 644-701.

KLIMCHITSKIY 1935:

C.~~H.~~ Климчицкий: Ягнобско-согдийские соответствия. In: Записки института востоковедения академии наук \cdot VI. Ленинград, 1935. 15-25.

KLIMCHITSKIY 1940:

С. И. Климчицкий: Секретный язык у ягнобцев и язгулёмцев. In: Академия наук СССР – Труды Таджикистанской базы, т. IX – 1938 – История – язык – литература. Akademijaji Fanho

SSSR: Asarhoji ваzаji Toçikiston, çildi IX - Tarix - zавоп - adaвijot. Москва — Ленинград (: Издательство Академии наук СССР), 1940. 104-117.

LIVSHITS 1962:

Владимир Аронович Лившиц: Хорезмийский язык. In: С. Π . Толстов – T. A. Жданко – C. M. Абрамзон – H. A. Кисляков (eds.): Народы Средней Азии и Казахстана I. Москва (: Издательство Академии наук СССР), 1962, 138-140.

LIVSHITS - KHROMOV 1981:

Владимир Аронович Лившиц — Альберт Леонидович Хромов: Согдийский язык. In: Вера Сергеевна Расторгуева (ed.): Основы иранского языкознания. Среднеиранские языки I. Москва (: Наука), 1981, р. 347-514.

MACKENZIE 1988:

David Neil MacKenzie: Khwarezmian and Avestan. In: East and West, Vol. 38, No. 1/4 (: Instituto Italiano per l'Africa e l'Oriente), 1988, p. 81-92.

NOVÁK [in print]:

Ľubomír Novák: Yaghnobi: an Example of a Language in Contact. In: Chatreššar 2011. Praha (: Filozofická fakulta Univerzity Karlovy v Praze), 2011, p. XX-YY.

PAYNE 1989:

John Payne: Pāmir Languages. In: Rüdiger Schmitt (ed.): Compendium Linguarum Iranicarum. Wiesbaden (: Dr. Ludwig Reichert Verlag), 1989, p. 418-444.

PROVASI 2009:

Elio Provasi: Versification in Sogdian. In: Werner Sundermann – Almut Hintze – François de Blois (eds.): Exegisti monumenta. Festschrift in honour of Nicholas Sims-Williams. Wiesbaden (: Harrassowitz), 2009, p. 347-368.

RASTORGUEVA 1966:

Вера Сергеевна Расторгуева: Иранские языки. Введение. In: В. В. Виноградов (ed.): Языки народов СССР. Том первый: Индоевропейские языки. Москва (: Наука), 1966, р. 194-211.

SIMS-WILLIAMS 1982:

Nicolas Sims-Williams: The double system of nominal inflection in Sogdian. Transactions of the Philological Society 1982, Oxford, 1982, s. 67-76.

SIMS-WILLIAMS 1984:

Nicolas Sims-Williams: The Sogdian "Rhytmic Law". In: Wojciech Skalmowski, Alois van Tongerloo (eds.): Middle Iranian Studies: Proceedings of the International Symposium organized by the Katholieke Universiteit Leuven from the 17th to the 20th of May 1982. Leuven (: Katholieke Universiteit te Leuven), 1984, p. 203-215.

SIMS-WILLIAMS 1989a:

Nicolas Sims-Williams: Eastern Middle Iranian. In: Rüdiger Schmitt (ed.): Compendium Linguarum Iranicarum. Wiesbaden (: Dr. Ludwig Reichert Verlag), 1989, p. 165-172.

SIMS-WILLIAMS 1989b:

Nicolas Sims-Williams: Sogdian. In: Rüdiger Schmitt (ed.): Compendium Linguarum Iranicarum. Wiesbaden (: Dr. Ludwig Reichert Verlag), 1989, p. 174–192.

SIMS-WILLIAMS 1996a:

Nicolas Sims-Williams: The Sogdian manuscripts in Brāhmī script as evidence for Sogdian

phonology. In: *Ronald Eric* Emmerick *et ali*. (eds.): Turfan, Khotan und Dunhuang: Vorträge der Tagung "Annemarie von Gabain und die Turfanforschung" veranstaltet von der Berlin-Brandenburgischen Akademie der Wissenschaften in Berlin (9.–12. 12, 1994). Berlin, 1996, p. 307–315.

SIMS-WILLIAMS 1996b [online]:

Nicolas Sims-Williams: Eastern Iranian Languages. In: *Eḥsān* Yārshāṭer (ed.): Encyclopædia Iranica [online edition]. Costa Mesa, California.

URL: http://www.iranicaonline.org/articles/eastern-iranian-languages> [quot. 05. 12. 2012, 12:34]

SKJÆRVØ 1989:

Prods Oktor Skjærvø: Modern Eastern Iranian. In: Rüdiger Schmitt (ed.): Compendium Linguarum Iranicarum. Wiesbaden (: Dr. Ludwig Reichert Verlag), 1989, p. 370-383.

SOKOLOVA 1967:

Валетина Степановна Соколова: Генетические отношения язгулямского языка и шугнанской языковой группы. Ленинград (: *Наука*), 1967.

SOKOLOVA 1973:

Валетина Степановна Соколова: Генетические отношения мунджанского языка и шугнаноязгулямской языковой группы. Ленинград, 1973.

VINOGRADOVA 2000:

Софья Петровна Виноградова: Ягнобский язык. In: Языки мира. Иранские языки III. – Восточноиранские языки. Москва (: ИНДРИК), 2000, р. 290-310.