



What Status of the Vasyugan Khanty Vernacular Do Calculations on the LingvoDoc Platform Support from the Point of View of Systemic Morphological Characteristics?

^{1,2} V.V. Vorobeva, ORCID: 0000-0002-8729-0375 <victoriavorobeva@mail.ru>

³ I.V. Novitskaya, ORCID: 0000-0003-1559-8810 <irno2012@yandex.ru>

¹ Ivannikov Institute for System Programming of the Russian Academy of Sciences,
25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.

² National Research Tomsk Polytechnic University,
30, Lenin ave., Tomsk, 634050, Russia.

³ National Research Tomsk State University,
36, Lenin ave., Tomsk, 634050, Russia.

Abstract. In this study the object of analysis is Vasyugan Khanty. Its status raises controversial opinions in Khanty studies. To clarify the status of the Vasyugan idiom as a separate dialect or as an accent of the Vakh-Vasyugan dialect, we employed modern methods of language data analysis. We used corpus data of the two varieties of the Khanty language, namely, Vakh Khanty and Vasyugan Khanty, available at the LingvoDoc platform to calculate their morphological proximity by means of the online virtual laboratory tool. The analysis results point to the fact that the morphological systems of the Vakh and Vasyugan Khanty vernaculars coincide by 98%, which confirms their morphological unity and affiliation with one and the same dialectal continuum. The machine analysis of the morphological dictionaries, cognate groups and transcriptions identified only three autonomous affixes in each idiom. Due to the fact that the volumes of the corpus data of the two varieties are unbalanced, the unique autonomous morphological affixes in each idiom can be regarded as a tentative argument subject to correction.

Keywords: Khanty language Vasyugan dialect; field data; text corpora; LingvoDoc; data analysis; language documentation.

For citation: Vorobeva V.V., Novitskaya I.V. What status of the Vasyugan Khanty vernacular do calculations on the LingvoDoc platform support from the point of view of systemic morphological characteristics? Trudy ISP RAN/Proc. ISP RAS, vol. 37, issue 2, 2025, pp. 237-246. DOI: 10.15514/ISPRAS-2025-37(2)-17.

Acknowledgements. The authors are grateful to the anonymous reviewers for their insightful comments to the article. The research was funded by the Russian Science Foundation, project no. 25-78-20002 “Capabilities of Artificial Intelligence for Comparative-Historical Study of Low-Resource Languages of the Peoples of the Russian Federation”.

Какой статус васюганского варианта хантыйского языка подтверждают вычисления на платформе ЛингвоДок с точки зрения его системных морфологических характеристик?

^{1,2} В.В. Воробьева, ORCID: 0000-0002-8729-0375 <victoriavorobeva@mail.ru>

³ И.В. Новицкая, ORCID: 0000-0003-1559-8810 <irno2012@yandex.ru>

¹ Институт системного программирования им. В.П. Иванникова РАН,
Россия, 109004, Москва, ул. Александра Солженицына, 25.

² Национальный Исследовательский Томский политехнический университет,
Россия, 634050, проспект Ленина, д. 30.

³ Национальный Исследовательский Томский государственный университет,
Россия, 634050, проспект Ленина, д. 36.

Аннотация. В данном исследовании объектом анализа выступает васюганский вариант хантыйского языка. Его статус вызывает противоречивые мнения в исследованиях хантологов. Для уточнения статуса васюганского идиома как отдельного диалекта или как говора вах-васюганского диалекта мы использовали современные методы анализа языковых данных. На платформе LingvoDoc были использованы корпусные данные двух вариантов хантыйского языка, а именно, ваховского хантыйского и васюганского хантыйского, для расчета их морфологической близости с помощью онлайн-инструмента виртуальной лаборатории. Результаты анализа указывают на то, что морфологические системы ваховского и васюганского вариантов хантыйского языка совпадают на 98%, что подтверждает их морфологическое единство и принадлежность к одному и тому же диалектному континууму. Машинный анализ морфологических словарей, родственных групп и транскрипций выявил только три автономных аффикса в каждом идиоме. В связи с тем, что объёмы корпусных данных двух диалектных разновидностей несбалансированы, уникальные автономные морфологические аффиксы в каждом идиоме можно рассматривать как аргумент, подлежащий дальнейшей корректировке.

Ключевые слова: васюганский хантыйский; полевые данные; текстовые корпуса; платформа LingvoDoc; анализ данных; языковое документирование.

Для цитирования: Воробьева В.В., Новицкая И.В. Какой статус васюганского варианта хантыйского языка подтверждают вычисления на платформе ЛингвоДок с точки зрения его системных морфологических характеристик? Труды ИСП РАН, том 37, вып. 2, 2025 г., стр. 237–246 (на английском языке). DOI: 10.15514/ISPRAS–2025–37(2)–17.

Благодарности. Авторы выражают искреннюю благодарность анонимным рецензентам за их ценные комментарии к статье. Исследование выполнено при финансовой поддержке РФН, грант 25-78-20002 «Возможности искусственного интеллекта для сравнительно-исторического изучения малоресурсных языков народов РФ».

1. Introduction

To date, linguists have reached no consensus on the status of the Vasyugan Khanty vernacular, which is considered to be a member of the eastern dialect continuum. According to one point of view, Vasyugan, together with Vakh, forms a single Vakh-Vasyugan dialect of the eastern Khanty [1-4]. Another viewpoint is that Vakh and Vasyugan belong to different Eastern Khanty dialects [5-7]. Russian researcher N.I. Tereshkin argues that the eastern group includes three vernaculars [8]:

- a) the Vakh-Vasyugan type, encompassing the dialects of the Vakh, Vasyugan and Aleksandrovsk Khanty;
- b) the Surgut type, encompassing the dialects of the Agan, Trom-Yugan, Pim and Yugan Khanty;
- c) the Salym type, encompassing two dialects [8], which are no longer in use.

According to N.I. Tereshkin's opinion, it is the dialects of the Vakh-Vasyugan subgroup alone that display some differences [8]. The Vakh dialect belongs to the l-type, the Vasyugan and Aleksandrovsky dialects belong to the j-type. The main difference is revealed in the use of the initial phoneme in word forms: cf. the form of the 3rd person singular pronoun with [l] in the Vakh *lõy* and the form with [j] in the Vasyugan *jõy*.

The same pattern is illustrated by the following examples: *lantjw – jantjw* 'ribbon, tape', *lalmjta – jalmjta* 'to steal', etc. [8]. In addition to phonetic differences, N.I. Tereshkin's dictionary also lists lexical differences between the two dialects, for example, Vakh. *wālā* – Vas. *wāla* 'bed, bunk' [8], Vakh. *ķāruyāl* – Vas. *qāruwāl* 'part of a rifle' [8], etc. Some lexical units, according to the dictionary materials, are specific to only one of the dialects, for example, Vakh. *ķarik* 'junk, rubbish', Vakh. *lāl* 'a gap between two spacers', [8], Vas. *kāyri* 'protection, hook', Vas. *kõt'kāmnaŋ* 'prey, trade' [8] and others.

In this study we draw on the opinion expressed by A.Yu. Filchenko [9], a researcher of the Vasyugan Khanty idiom, that the Vasyugan idiom is an autonomous dialect of the eastern group and is separate from the Vakh dialect [10-11]. The notion of dialect, however, is opposed to the notion of accent in the Russian linguistic tradition, so it is necessary to determine the status of the Vakh and Vasyugan varieties of the Khanty language either as dialects or as accents.

A dialect is understood as a regionally or socially distinctive variety of the language spoken by residents of a certain area, and identified by a particular set of words and grammatical structures [12]. An accent is defined as the cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially [12].

From the social and cultural perspective, the Vakh and Vasyugan Khanty peoples have a relatively common linguistic landscape, material and spiritual culture, traditions, literature, ethnic self-awareness, mutual intelligibility, and identity of speakers. From the linguistic perspective, there are differences between the Vakh and Vasyugan varieties of the language at the phonetic-phonological and lexical levels mentioned above. Differences of the phonetic type have also been dealt with in [9]. However, no discrepancies in the field of grammar have so far been described comprehensively in the scientific literature. This raises the question of the status of the Vasyugan Khanty vernacular. As it is known, dialects of one language display discrepancies in pronunciation, vocabulary and grammar of the language system. This poses the question of how legitimate it is to classify the Vasyugan Khanty as an autonomous dialect.

The purpose of this article is to clarify the status of the Vasyugan Khanty vernacular which is calculated on the corpus data by means of the tools on the LingvoDoc platform. It should be noted that it is the first attempt to calculate the degree of proximity between the Vakh and Vasyugan idioms of the Khanty language using online tools, although the effectiveness of the tools available on the LingvoDoc linguistic platform has already been tested with regard to the morphological and etymological analysis of a number of the Ural and Altai languages [13, 14].

The study was carried out on the corpora of the Vakh and Vasyugan varieties of the Khanty language of the eastern group. Both corpora are openly accessible at the LingvoDoc platform [15]. The platform uses Russian as an interface language.

The corpus of the Vakh dialect of the Khanty language, contains about 13 thousand word forms, is authored by V.V. Vorobeva [16] and S.V. Kovylin [17], and is represented by the materials from the expeditions to the village of Korliki in the Khanty-Mansiysk Autonomous Okrug–Yugra in 2017–2019. The corpus of the Vasyugan dialect of the Khanty language containing about 2,500 word forms, includes texts annotated by L.I. Kalinina (1966), a text by V. Steinitz (1950), texts obtained in the expeditions of A.Yu. Filchenko (2008) [18], and the Khanty written monument of St. Macarius (Nevskiy) (1887) [19].

2. Results of analysis and their interpretation

The calculation of the degree of proximity of the two eastern Khanty vernaculars was carried out in the virtual laboratory LingvoDoc [20].

LingvoDoc is a service designed with the focus on documenting languages by creating dictionaries of any structure, establishing etymological links between them and performing calculations based on language data [13, 21].

The calculation and analysis of the proximity of the two eastern idioms was carried out by means of the tool “Degree of morphological proximity of different dialects”. The program calculates mathematically a distance between different dialects/accents and presents results instantly in the form of diagrams and tables, which require further description and generalization.

At the initial stage, the corpora of the vernaculars under analysis are converted into morphological dictionaries to enable calculation of the morphological similarity between the Vakh and Vasyugan language varieties. The morphological dictionaries obtained from the corpus data represent a conglomerate of unordered affixes with their meanings, which are retrieved from the text corpora. Then comes the stage at which all cases of homonymy are manually removed. This results in creating dictionaries in which each affix is terminologically designated, is given one lexical entry. If it is necessary, a comment can be manually added that describes functional peculiarities of the affix. In addition, all instances of how this affix has been used are retrieved from the corpus and listed (see Fig. 1).

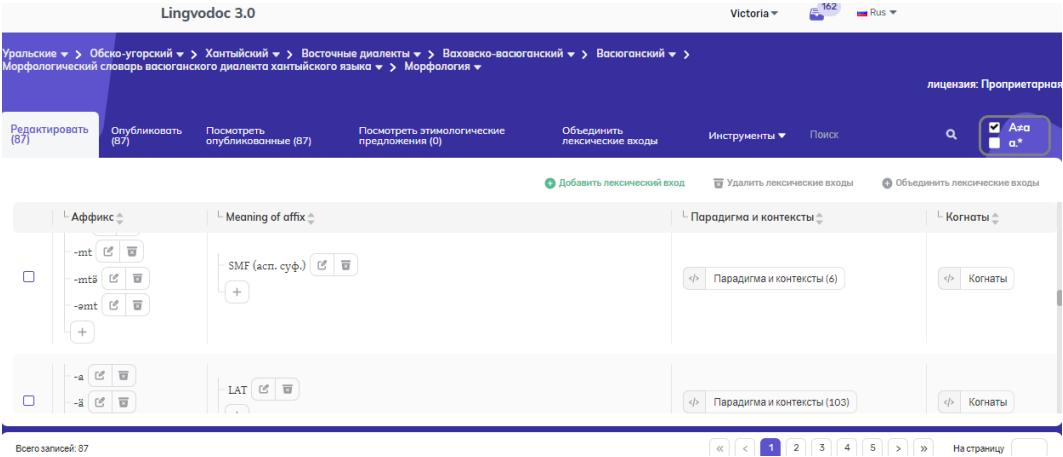


Fig. 1. Morphological dictionary on the Lingvodok platform.

After the manual removal of homonymy instances, the dictionaries under analysis can be connected with one another by invisible threads – cognates. To this end, one can add the “Cognates” column via the dictionary properties feature, which will offer etymologies for each affix obtained from the multiple dictionaries available on the LingvDoc. All one needs to do is to select the right one.

After all dictionaries have been prepared as described above, one can activate another virtual laboratory tool that mathematically calculates the proximity of the distance between the given vernaculars. This tool is called “Degree of morphological proximity between dialects”, which is available to all registered users of the laboratory. The tool enables one to obtain a portrait of the morphological proximity of the perspectives in the form of a report “from different angles”, which requires further interpretation.

In our case, the machine analysed the data of 2 dictionaries, 75 cognate groups and 171 transcriptions (Fig. 2). None of the cognate groups was excluded, which speaks about the accuracy of the results. The results of the machine analysis can also be exported in the format .xlsx.

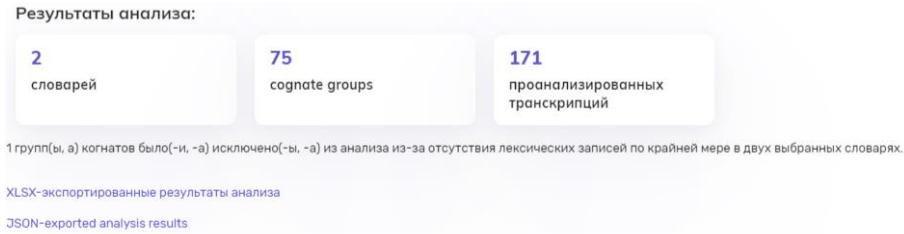


Fig.2. Summary of the data included in the calculation.

The software of the LingvoDoc platform shows the morphological proximity between dialects according to the processed dictionaries in the form of a graph that represents an etymological tree of distances between the two eastern varieties (see Fig. 3). In the diagram, the red dot represents the “Morphological Dictionary of the Vasyugan Idiom”, the green dot - the “Morphological Dictionary of the Vakh Idiom”. The diagram demonstrates the closest possible distance between perspectives 1 and 2, since the dots are located on the same line.

This result is further confirmed in terms of percentage in the table. One can see that the Vakh and Vasyugan Khanty idioms are 98% grammatically close (see Fig. 4).

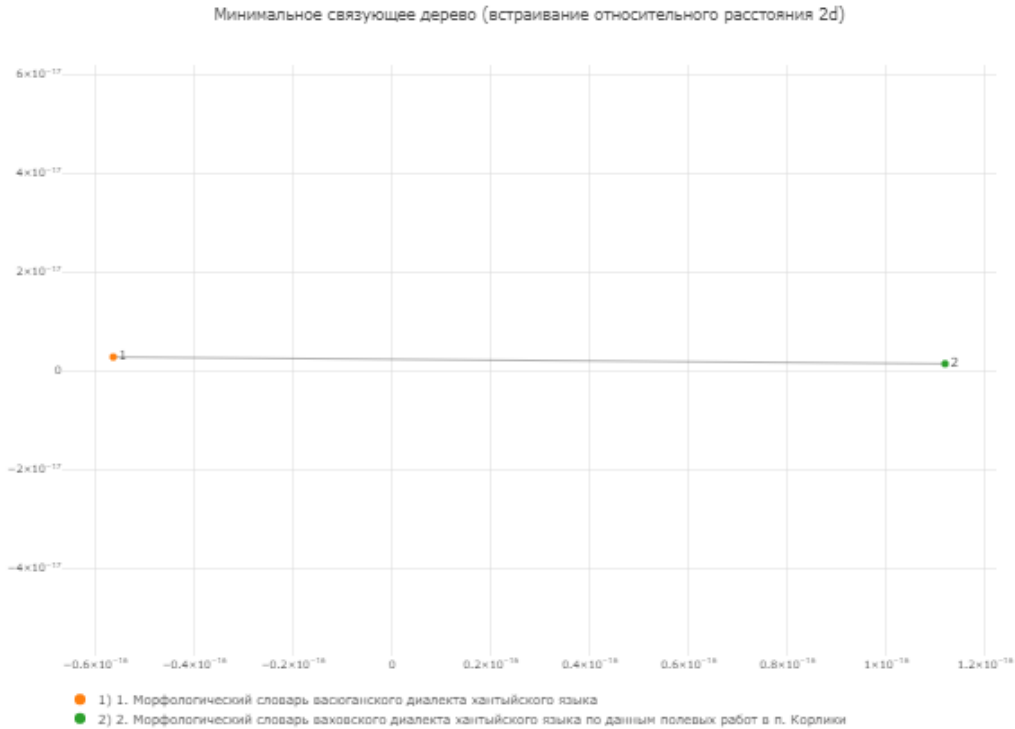


Fig. 3. Etymological tree of distance.

	1. Морфологический словарь васюганского диалекта хантыйского языка	2. Морфологический словарь ваховского диалекта хантыйского языка по данным полевых работ в п. Корлики
1	n/a	0.36 (98%)
2	0.36 (98%)	n/a

Fig. 4. Morphological calculation of the proximity of Vakh and Vasyugan Khanty in percentages.

Drawing on the analysis results of the two dictionaries, a common consolidated database of affixes is formed. The results of the calculation show that the Vakh and Vasyugan morphological dictionaries have 76 etymologically related affixes. All affixes that share etymologies are presented in Fig. 5a-f. The right column lists the affixes of the Vakh Khanty, and the left column – those of the Vasyugan Khanty.

1. Морфологический словарь васюганского диалекта хантыйского языка	2. Морфологический словарь ваховского диалекта хантыйского языка по данным полевых работ в п. Корлики
-a; -ä (LAT)	-ä; -a (LAT)
-at (PASS.3PL)	-ät; -at (PASS.3PL)
-i (ATTR)	-j (ATTR)
-i; -j (NST.PASS.3SG; PASS.3SG)	-j; -i (PASS.3SG; NST.PASS (только в бессуф. вр. PASS))
-i; -ы (IMP (со всеми ч. суб. и об. спр. кроме SG.S); IMP.SG<SG.S (только в SG об. спр., т.к. ч. суф. нулевой))	-j; -i (IMP.SG<SG.S (только в SG об. спр., т.к. ч. суф. нулевой); IMP (со всеми ч. суб. и об. спр. кроме SG.S))
-im; -im (SG<1SG.S)	-im; -im (NST.SG<1SG.S (только бессуф. вр.); SG<1SG.S)
-jm; -äm; -m; -am; -im; -am; -äm; -äm (NST.1SG.S (только в бессуф. вр.); 1SG; 1SG.S; POSS.1SG)	-äm; -äm; -äm; -im; -im; -jm; -äm; -m; -am (NST.1SG.S (только в бессуф. вр.); 1SG; 1SG; 1SG.S; 1SG.S; POSS.1SG)
-jw (SG<1PL.S)	-üy; -öy; -öy; -öy; -öy; -öy; -öy; -uü (POSS.1PL; 1PL; 1PL; 1PL.S; NST.1PL.S (только в бессуф. вр.); SG<1PL.S)
-öw; -aw (SG<1PL.S; 1PL; 1PL; 1PL.S; NST.1PL.S (только в бессуф. вр.); POSS.1PL)	
-jñ; -ñ; -añ; -äjñ (PROPR)	-əñ (PROPR)
-j; -n; -k; -y; -ñ (EP)	-j; -k; -ñ; -y (EP)
-kas; -yas (PST2.3SG.S; PST2)	-kas; -käš; -kas; -yäs; -yas (PST2 3SG S; PST2)

Fig 5a. Common affixes for Vakh and Vasyugan idioms

-kke; -əke; -ək; -iki (PRD)	-ək; -ək; -jki; -ki; -iki; -ikə; -əkj (PRD)
-käi; -yal; -käi; -yäl (PST3.3SG.S; PST3)	-yäl; -yal; -yal; -käi (PST3.3SG.S)
-kan; -yan; -ayen; -əken; -əyən; -iyən (NST.3SG.S)	-äyən; -äyən; -yən; -iyən (NST.3SG.S)
-l (NPST (кроме 3SG, 2SG, 3PL))	-l (NPST (кроме 3SG, 2SG, 3PL))
-lin; -lin (POSS.PL<2PL; POSS.PL<3DU; POSS.2DU.PL)	-lin; -lin (POSS.PL<3DU; POSS.2DU.PL; POSS.PL<2PL)
-l-ämən (POSS.PL<1DU)	-l-ämən (POSS.PL<1DU)
-l-öw (POSS.PL<1PL)	-l-öy; -öy (POSS.PL<1PL)
-li; -li; -lë; -el (MULT (acn. cyф.))	-li; -lië; -li (MULT (acn. cyф.))
-la (PL<SG.S (чис. cyф. IMP))	-lä; -la (PL<SG.S (чис. cyф. IMP))
-lij; -li (ATTR)	-äli; -äli; -ali; -ali (DIM)
-ali; -ali; -äli (DIM)	-li (ATTR)
-lte (CAUS (acn. cyф.))	-lte (CAUS (acn. cyф.))
-m (PFV.PTCP)	-m; -mä; -em; -äm; -më (PFV.PTCP)
-m; -emt; -mtë; -mt (SMF (acn. cyф.))	-m (SMF (acn. cyф.)) -mtë; -emt; -lte (MONT) -më; -m (SMF (acn. cyф.))
-min; -man (CVB)	-min; -min (CVB)
-mat (ORD)	-mät (ORD)
-na; -nä (COM)	-nä; -na; -nä (COM)
-nte; -ant; -nt (MULT.INTR (acn. cyф.))	-ënt; -nte (MULT.INTR (acn. cyф.))
-nä; -nə; -ən (LOC)	-nä; -n; -na; -nä; -nə; -ni; -në; -än; -ön; -ən; -ən (LOC)

Fig. 5b. Common affixes for Vakh and Vasyugan idioms.

The program for calculating the proximity between perspectives also distinguishes autonomous affixes and lists them in a separate group. In our case, there are only 6 of them, three affixes in each dictionary (see Fig. 6). The peculiarity of the Vasyugan idiom is three suffixes: the aspectual suffix with the meaning of instant intransitivity *-əyiŋ* (MONT.INTR), the word-formation suffix *-əw* (ATTR) and the reflexive suffix of pronouns *-nām*; *-nam* (REFL). The suffixes specific to the Vakh dialect alone are: the aspectual suffix of reflexivity *-ysə* (REFL), the case suffix *-ətəl*; *-ətəl*; *-təl*; *-əl* (DISTR (fall. suffix) and the reflexive suffix of pronouns *-ä*; *-a* (REFL).

-yäl-äm (POSS.DU<1SG)	-yäl-äm; -yäl-äm; -əyäl-äm; -yäl-äm; -käl-äm; -əkäl-äm (POSS.DU<1SG)
-a; -ä; -a (IMP.SG.S (только в SG суб. спр., т.к. ч. суф. нулевой))	-a; -ä; -ä; -ä (IMP.SG.S (только в SG суб. спр., т.к. ч. суф. нулевой))
-өлә (PL<3SG.S)	-l; -l; -l; -l; -l; -l; -l; -l (POSS.3SG; 3SG)
-la; -l; -äl; -äl; -al; -l; -l (3SG; POSS.3SG)	-t; -ä; -ät; -ät; -ä; -ä; -ä; -ä; -ä; -ä (PL)
-ət; -ə; -t (PL)	-l-ä; -өлә; -өлә; -l-ä; -өлә; -l-ä; -l-ä; -l-ä (POSS.PL<3SG; NST.PL<3SG.S (только в бессыф. вр.); PL<3SG.S)
-ла-а; -ла-а; -ла-а (POSS.PL<3PL; PL<3PL.S)	-l-ä; -äl-ä (NST.PL<3PL.S (только в бессыф. вр.); PL<3PL.S)
-ла-а; -ла-а; -ла-а (PL<1SG.S; POSS.PL<1SG)	-l-ä; -l-ä (POSS.PL<1SG; PL<1SG.S)
-ла-а (PL<2PL.S)	-l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä (PL<2DU.S; PL<PL.S (ч. суф. IMP); PL<3DU.S; PL<2PL.S; PL<DU.S (ч. суф. IMP))
-лих (ABS)	-l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä; -l-ä (ABS)
-ра; -р (REP)	-p-ä; -p-ä; -p-ä (REP)
-тах; -итах; -етах; -тых; -тыг (2PL.S)	-t-ä; -t-ä; -t-ä (PL.S (ч. суф. пов. накл.); 2PL.S)
-ты (PTCL)	-t-ä; -t-ä; -t-ä (PTCL)
-ыль; -л; -l (3PL; SG<3PL.S)	-l; -l; -l; -l; -l; -l; -l; -l (NST.SG<3PL.S (только в бессыф. вр.))

Fig. 5e. Common affixes for Vakh and Vasyugan idioms.

-ымынь; -ымынь; -əməñ; -məñ (POSS.1DU; 1DU.S)	-imän (SG<1DU.S)
-imän; -imän (SG<1DU.S)	-mən; -amən; -ämən; -əməñ; -məñ (1DU; POSS.1DU; NST.1DU.S (только в бессыф. вр.); 1DU.S)
-ətən; -tən; -ətən (SG<3DU.S; SG<DU.S (ч. суф. IMP); SG; 2DU.S; DU.S (ч. суф. IMP); POSS.2DU; POSS.3DU; NST.2DU.S (только в бессыф. вр.); POSS.3PL)	-imän; -imän (1SG<DU.S)
-əw (INDEF)	-tən; -ətən; -ətən (S<DU.S (ч. суф. IMP); DU.S (ч. суф. IMP); NST.2DU.S (только в бессыф. вр.); POSS.2DU; SG<3DU.S; POSS.2PL; POSS.3DU; 2DU.S)
	-əw (INDEF)

Fig. 5f. Common affixes for Vakh and Vasyugan idioms.

1. Морфологический словарь васюганского диалекта хантыйского языка	2. Морфологический словарь ваховского диалекта хантыйского языка по данным полевых работ в п. Корлики
-əyiŋ (MONT.INTR (асп. суф.))	
-əw (ATTR)	
-nām; -nam (REFL)	
	-ysə (REFL (асп. суф.))
	-ətəl; -ətəl; -təl; -əl (DISTR (пад. суф.))
	-ä; -a (REFL)

Fig. 6. Autonomous affixes of Vakh and Vasyugan idioms.

Thus, some minor differences in the grammar of Vakh and Vasyugan Khanty vernaculars have been experimentally identified. However, it is clear that such a low number of differing features of the two idioms cannot affect mutual intelligibility between ethnic communities living in separate groups. The analysis results speak in favour of the conclusion that Vakh and Vasyugan form a single dialectal continuum and represent one Vakh-Vasyugan dialect with its two varieties: Vakh and Vasyugan.

The differences revealed in the system of grammatical affixes, namely, the autonomous affixes, can not be regarded as final and decisive, since the Vakh and Vasyugan corpora are currently not balanced in volume. The Vakh corpus exceeds significantly that of Vasyugan, therefore, an error may surface arising from the fact that some aspectual affixes do not simply occur in these corpora or have not been not marked by the authors of these corpora. The distributive case, which is currently identified in the Vakh idiom alone, is rarely used due to its semantics, therefore it can be absent or unidentified in the Vasyugan texts uploaded to the LingvoDoc.

The discrepancy of the reflexive suffix of pronouns may arise from the lacking unification of the marking and designation of glosses. The autonomous morphological affixes should be further studied and analyzed. However, new findings can hardly affect the conclusion that Vakh and Vasyugan Khanty form a single dialect space, which has been confirmed by machine calculations.

3. Conclusion

The status of Vasyugan Khanty has not been clearly defined in scientific literature. To confirm its status, we conducted a mathematical analysis of the dictionaries of Vasyugan and Vakh Khanty, obtained from the corpus data, in the LingvoDoc online linguistic laboratory. For the analysis, we used one of the LingvoDoc tools “Degree of morphological proximity between dialects”. The results of the analysis show that the degree of morphological similarity between Vakh and Vasyugan is 98%, which indicates that Vakh and Vasyugan represent a single dialect space, and Vasyugan is a variety of the Vakh-Vasyugan dialect. Machine calculation of the morphological similarity of idioms of the Khanty language was carried out for the first time.

References

- [1]. V. N. Solovarov, G. L. Nakhracheva A. A. Shiyanova, “Dialects of Khanty”. Khanty-Mansiysk Izhevsk: Print-2 Publ., 2016, 349 p.
- [2]. M. Chepregi, “Surgut dialect of the Khanty language”. Khanty-Mansiysk: Pechatnyj mir g. Khanty-Mansiysk Publ., 2017, 275 p.
- [3]. E. Vértés, Einwandfrei Aufhebbare Widersprüche zweier Mitteilungen über die Vach-Mundart, *Linguistica Uralica*, 1965, vol. 1 (3), pp. 213–217.
- [4]. S. Sosa, Functions of morphosyntactic alternations, and information flow in Surgut Khanty Discourse. Helsinki, 2017. 254 p.
- [5]. P. Hajdú, “The Uralic languages and peoples”. Moscow: Progress, 1985, 425 p.
- [6]. L. Honti, Die ostjakische Sprache. The Uralic Languages. Description, History and Foreign Influences. *Handbuch der Orientalistik*, Ed. D. Sinor. Leiden; New York; København; Köln: Brill, 1988, pp. 172–196.
- [7]. K. F. Karjalainen, Grammatikalische Aufzeichnungen aus Ostjakischen Mundarten. Bearbeitet und herausgegeben von Edith Vertes, Helsinki: Suomalais-Ugrilainen Seura, 1964. 234 p.
- [8]. N. I. Tereshkin, Dictionary of the Eastern Khanty dialects. Leningrad: Nauka Publ., 1981, 544 p.
- [9]. A. Yu. Filchenko, A Grammar of Eastern Khanty. Doctor of Philology Thesis. Houston, Texas, 2007.
- [10]. V. Vorobeva, I. Novitskaya, Inflectional morphology of nouns in Eastern Khanty (Vakh, Vasyugan, Surgut, Salym). *Ural-Altaic Studies*, vol. 38, 2020, pp. 33-70.
- [11]. V.V. Vorobeva, I.V. Novitskaya, Functional and semantic analysis of the case markers in Vakh Khanty (a case study of the latest field data on LingvoDoc). *Proc. ISP RAS*, vol. 36, issue 2, 2024, pp. 199-210. DOI: 10.15514/ISPRAS-2023-35(2)-15.
- [12]. D. Crystal, A dictionary of linguistics and phonetics. 6th edition. Singapore, Blackwell Publishing Ltd, 2008. 529 p.
- [13]. Yu. V. Normanskaya, “Possibilities of “LingvoDoc” for identifying the degree of closeness of languages and dialects and describing graphic and phonetic features”. Information technologies in humanitarian research: Proc. Int. scientific-practical. conf. Krasnoyarsk, September 25-28, 2023, Ed. M. A. Lapteva.

- Krasnoyarsk: Sibirskij federalnyj universitet Publ., 2023, pp. 251–260. [Online]. Available: <https://doi.org/10.5281/zenodo.7385533>.
- [14]. I. P. Novak Distribution of anterior fricative consonants in the dialects of the Karelian language of Central Karelia (based on the application of the “cognate analysis” algorithm of the LingvoDoc linguistic platform). *Ural-Altai studies*. 2022. No. 02 (45). P. 79-105.
- [15]. https://lingvodoc.ispras.ru/language_databases (дата обращения 01.02.2025).
- [16]. Text corpus of Vakh Khanty. Author V.V. Vorobeva. URL: <https://lingvodoc.ispras.ru/dictionary/7349/279/perspective/7349/283/view/> (дата обращения 11.09.2024).
- [17]. Text corpus of Vakh Khanty. Author S. V. Kovylin. URL: <https://lingvodoc.ispras.ru/dictionary/3021/71/perspective/3021/75/view/> (дата обращения 05.09.2024).
- [18]. Texts corpus of Vasyugan Khanty. Author V.V. Vorobeva. URL: <https://lingvodoc.ispras.ru/dictionary/2488/8764/perspective/2488/8767/view/> (дата обращения 12.09.2024).
- [19]. Texts corpus of Vasyugan Khanty. Monument of St. Macarius (Nevsky). Author S.V. Kovylin. С. В. Ковылин.
- [20]. <https://www.lingvodoc.ispras.ru> (дата обращения 01.02.2025).
- [21]. Yu. V. Normanskaya, A. I. Avetisyan, I. B. Beloborodov, O. D. Borisenko, “The LingvoDOC software system and the possibilities it offers for documenting and analyzing Ob-Ugric languages”. / “Reports of the Russian Academy of Sciences. Mathematics, informatics, control processes”, 2022, no. 504 (1), pp. 60–82.

Информация об авторах / Information about authors

Виктория Владимировна ВОРОБЬЕВА – кандидат филологических наук, доцент ОИЯ ШОН Томского политехнического университета с 1999 года. В настоящее время также ведет исследования в лаборатории Лингвистических платформ ИСП им. В.П. Иванникова РАН. Сфера научных интересов: корпусная лингвистика, обско-угорские языки, документация исчезающих языков России.

Victoria Vladimirovna VOROBEOVA – Cand. Sci. (Philology), Associate Professor of Department of Foreign Languages, School of Social Sciences of National Research Tomsk Polytechnic University since 1999. At the present time she is a researcher at the Linguistic Platforms Laboratory of Information Systems Department as well. Research interests: comparative typological linguistics, Uralic family of languages, corpus linguistics, documentation and digital processing of endangered languages of Russia.

Ирина Владимировна НОВИЦКАЯ – доктор филологических наук, профессор кафедры английской филологии Томского государственного университета. Сфера научных интересов: морфология и семантика древнегерманских языков, сравнительные и типологические исследования морфологии языков, морфосинтаксические характеристики лексики хантыйского языка, прикладная лингвистика.

Irina Vladimirovna NOVITSKAYA – Dr. Sci. (Philology), professor at the Department of English Philology, National Research Tomsk State University. Research interests: morphology and semantics of the Old Germanic languages, typological studies, comparative studies, morpho-syntactical properties of Khanty lexical units, applied linguistics.