



Functional and Semantic Analysis of the Case Markers in Vakh Khanty (a Case Study of the Latest Field Data on LingvoDoc)

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

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

¹ Information Systems Department,
Ivannikov Institute for System Programming of the RAS
25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.

² National Research Tomsk Polytechnic University,
30, Lenina Avenue, Tomsk, 634050, Russia.

³ National Research Tomsk State University
36, Lenina Avenue, Tomsk, 634050, Russia.

Abstract. In this paper we argue that the category of nominal case in the Vakh Khanty dialect needs reviewing which is due to the availability of the recent field data obtained in 2019. This objective is considered to be topical because the results of the study will equip researches with all necessary data to develop a unified notation system for the case markers to be used in an annotated corpus of Vakh Khanty on the LingvoDoc platform. The declension system of the Eastern Khanty dialect under analysis is heterogeneous and relatively vast. The controversy around the synthetic means of expression of semantic relations concerns the terminology of case markers, their quantity, morphemic status and functional features. The purpose of this paper is to clarify the functional and semantic aspects of the Vakh Khanty case markers drawing on the recently obtained field material. This objective is achieved by means of the tools available on the LingvoDoc platform. The latest field data on the Vakh Khanty dialect was collected in the village of Korliki in 2019. More than 6,000 words of data are posted on the endangered languages documentation platform LingvoDoc. Part of the material used in the analysis is being processed and integrated on the LingvoDoc platform. The underlying opinion of this study draws on the seminal work by N.I. Tereshkin, as well as other prominent researchers of the Khanty language, and their description of the declension system of the Eastern dialect. The latest field data contains texts and questionnaires allowing researches to examine functioning of the morphological markers and compare results with the findings previously presented in the literature. Such approach entails observations leading to the systematization of the case category in this dialect at its current state of development. The examination of the range of meanings conveyed by the Vakh Khanty case markers at present has made it possible to distribute them into groups of semantic and syntactic case markers. As a result, the study confirmed that the nominal declension system encompasses such case markers as the abessive, ablative, allative, distributive, comitative, comparative, lative, locative, nominative, oblique, and translocative cases. The recent field data confirms the presence of the distributive case within the case paradigm and the status of the comparative element *nijit* as a postposition. Each case marker of the Vakh Khanty dialect is terminologically specified which is essential for developing a parser for this dialect on the LingvoDoc platform. The development of the parser in the future will speed up the process of processing and analyzing language material and lead to the elimination of errors.

Keywords: Khanty language; Vakh dialect; case category; case markers; field data; text corpora; LingvoDoc.

For citation: Vorobeva V.V., Novitskaya I.V. Functional and semantic analysis of the case markers in Vakh Khanty (a case study of the latest field data on LingvoDoc). *Trudy ISP RAN/Proc. ISP RAS*, vol. 36, issue 2, 2024. pp. 199-210. DOI: 10.15514/ISPRAS-2023-35(2)-15.

Acknowledgements. The authors are grateful to the anonymous reviewers for their careful reading of the work. The study was funded by Russian Science Foundation, project no. 20-18-00403 “Digital Description of Uralic Dialects Based on Big Data Analysis”.

Функционально-семантический анализ падежных маркеров ваховского хантыйского языка (на материале аннотированных полевых данных в системе ЛингвоДок)

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

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

¹ Институт системного программирования РАН,
109004, Россия, г. Москва, ул. А. Солженицына, д. 25.

² Национальный исследовательский Томский политехнический университет,
634050, Россия, Томск, пр. Ленина, д. 30.

³ Национальный исследовательский Томский государственный университет,
634050, Россия, Томск, пр. Ленина, д. 36.

Аннотация. В данной статье мы утверждаем, что описание категории именного склонения в ваховском хантыйском языке нуждается в пересмотре, в связи с наличием последних полевых данных, полученных в 2019 году. Эта задача считается актуальной, поскольку результаты исследования предоставят ученым необходимые данные для разработки единой системы обозначений падежных маркеров, которые будут использоваться при обработке текстового материала для размещения на платформе ЛингвоДок. Система склонения анализируемого диалекта восточных ханты неоднородна и относительно обширна. Дискуссия о синтетических средствах выражения семантических отношений в языковых единицах касается терминологии падежных маркеров, их количества, морфемного статуса и функциональных особенностей. Целью данной работы является уточнение функциональных и семантических аспектов падежных маркеров ваховских ханты на основе недавно полученного полевого материала. Эта цель достигается с помощью инструментов, доступных на платформе ЛингвоДок. Последние полевые данные по ваховскому хантыйскому диалекту были собраны в селе Корлики в 2019 году. Данные объемом более 6000 слов размещены на платформе документирования исчезающих языков ЛингвоДок. Часть материала, используемая при анализе, постепенно обрабатывается и интегрируется на платформу ЛингвоДок. В основе исследования лежат работы Н.И. Терёшкина (1961) и других исследователей (Я. Гуя, Л. Хонти и др.) хантыйского языка, которые ранее описывали систему склонения ваховского диалекта. Последние полевые данные содержат тексты и анкеты, позволяющие исследователям изучить функционирование морфологических маркеров и сравнить результаты с данными, ранее представленными в литературе. Такой подход предполагает наблюдение, ведущее к систематизации падежной категории в этом диалекте на современном этапе его развития. Изучение спектра значений, передаваемых падежными маркерами, в настоящее время позволило распределить их на группы семантических и синтаксических падежных маркеров. В результате исследования подтверждено, что система именного склонения включает в себя такие падежные маркеры, как аблативный, абессивный, аллативный, лативный, локативный, комитативный, дистрибутивный, обликативный и транслативный. Последние полевые данные подтверждают наличие дистрибутивного падежа в системе и статус сравнительного элемента *nijit* как послелoga. Каждый падежный маркер ваховско-хантыйского диалекта терминологически конкретизирован, что необходимо для разработки парсера этого диалекта. Развитие парсера в будущем ускорит процесс обработки и анализа языкового материала и приведет к устранению ошибок при разметке текстов.

Ключевые слова: хантыйский язык; ваховский диалект; категория падежа; падежные маркеры; полевые данные; корпуса текстов; ЛингвоДок.

Для цитирования: Воробьева В.В., Новицкая И.В. Функционально-семантический анализ падежных маркеров ваховского хантыйского языка (на материале аннотированных полевых данных в системе ЛингвоДок). Труды ИСП РАН, том 36, вып. 2, 2024 г., стр. 199–210 (на английском языке). DOI: 10.15514/ISPRAS-2024-36(2)-15.

Благодарности. Авторы статьи выражают благодарность анонимным рецензентам за тщательное и детальное прочтение работы. Исследование выполнено при поддержке РФФ, проект № 20-18-00403 «Цифровое описание диалектов уральских языков на основании анализа больших данных».

1. Introduction

In this paper we argue that the process of digital processing of the recently obtained field data on the Vakh Khanty dialect requires a review of its nominal declension system with an ensuing prospect of developing a valid parser for this idiom.

“Morphological case on nominals is a common device to express the syntactic and semantic relationships between clausal constituents” [1]. Some scientists distinguish synthetic (traditional) and analytical cases [2-3]. Regardless of the way of expression, the case is a meaningful category, and its expression is a secondary point. In this study we focus only on the synthetic form of expressing syntactic and semantic relations in sentences. The subsequent analysis of case markers in the Vakh-Khanty dialect is based on the concept of syntactic and semantic cases [4: 325, 5, 6: 122].

The category of case in the Khanty language has always aroused interest among researchers because the number of constituents that form the nominal declension system varies considerably among the group of dialects integrated by a geographical principle. In the Northern dialects of Khanty, for example, the case system includes three markers, but the system constituents may be different. For example, the Kazym dialect possesses the nominative, lative and locative case markers, while in the Priural dialect these markers are for the nominative, locative and translative cases. In the Southern dialects, which have now lost their vitality, researchers distinguish from five to six case markers. In the Eastern dialects, the nominal declension system may include from seven to eleven case markers. The diverging views on the number of case markers in Eastern Khanty reflect different linguists’ opinions about the status of some morphemes encoding case semantics, therefore, the same marker in grammatical essays can be termed as a case marker, and as a particle, and as a postposition [7-16].

Similar divergence of opinions is found with regard to the terms used to distinguish some of the identified constituents of the case category. Since most of the case markers are multifunctional, researchers face a challenging task of either selecting the most suitable term for a multifunctional case marker focusing on one of its predominant functions, or of using a dual term that reflects multiple functions of a single case marker. Thus, in his Vakh Khanty grammatical essay, N. I. Tereshkin argues that there are markers of such cases that he personally termed as “*napravitel’no-tselevoi*” (rus. lit. “*directional-target*”), “*tvoritel’no-ob’ektnyi*” (rus. lit. “*instrumental-objective*”), “*tvoritel’no-sovmestnyi*” (rus. lit. “*instrumental-comitative*”), “*otlozhitel’no-prodol’nyi*” (rus. lit. “*depositional-transitive*”) [13: 42–55].

It can be concluded that the constituents of the case category in the Vakh Khanty dialect, as they are presented in the scientific literature, are open to multiple interpretations resulting in ambiguous conclusions about representatives of individual cases. This fact complicates the process of digitizing the field data with the tools of the LingvoDoc platform [17].

The purpose of this paper is to clarify the functional and semantic aspects of the Vakh Khanty case markers drawing on the recently obtained field material. This objective is achieved by means of the tools available on the LingvoDoc platform.

While processing a text corpus of Vakh Khanty for posting on the LingvoDoc platform we faced such issues as aspectual, temporal markers and predicative markers, comparative constructions, words that have not yet been accepted by dictionaries. Certain issues we encountered concerned the Vakh Khanty case markers, namely, their terminology, number and functions. Digital processing of the data was grounded in the findings presented in the grammar by N. I. Tereshkin, a student of V. Steinitz. The researcher collected materials for his monograph during his scientific expeditions along the rivers Vakh, Vasyugan, Tromyegan and Salym that he made in the second half of the 20th century. His Khanty students at the Institute of the Peoples of the North in Leningrad [13: 5] also

contributed greatly with their speech samples. His monograph “Essays on dialects of the Khanty language” (1961) describes Vakh Khanty reliably and quite fully, and is, therefore, regarded as a starting point in the process of annotating the corpus. However, the issues faced while processing the field data remain unresolved. We also discovered that some grammatical phenomena were not described in full or were not mentioned at all in that grammar handbook, for example, the abessive case marker and some case functions, which will be discussed below. We also incorporated the works by foreign specialists of Vakh Khanty J. Gulya [9] and L. Honti [10] but in most cases the issues could not be solved and the task became even more complicated, since the scientists interpreted, labelled and described case functions differently.

All these have prompted us to write this article in which we aim to verify the constituents of the Vakh Khanty nominal case system. Standardization of the case marker labels is believed to be an important stage in developing a reliable and viable Vakh Khanty parser in the nearest future, and it will allow case morphemes to be glossed universally, in a standardized way. Besides, it will help errors to be excluded that, given the human factor, are inevitable in the manual processing of a large volume of data. It will also help achieve a completely different speed of text processing.

Research materials and methods

The study was conducted on the field data that was collected through the communication with speakers of the Vakh dialect by one of the authors in the village of Korliki (KhMAO-Yugra) in 2019. Language data was collected using traditional field research methods (survey, interview, experiment, observation, recording and description using various technical means and programs). Part of the collected material is digitized and presented on the LingvoDoc platform, namely: a fairy tale “A mouse’s song”, a text about deer, questionnaire “Case category” P. I, P. II, questionnaire “Tense, Mood, Aspect” P. I, II, III, IV” [AFM]. Some data that was also used in the analysis has not yet been digitally processed, but will soon appear on the same platform, these include: a fairy tale “Two birds - sister and brother” and a text about deer [AFM].

The analysis of the obtained language data on the Vakh dialect was carried out using descriptive-comparative and structural research methods, as well as corpus analysis methods (diversified search, compilation of contexts and concordances, counting, analysis of the text meta-marking).

2. The cases of the Vakh Khanty dialect

In this study we use the following terminology for case markers: the abessive, ablative, allative, distributive, comitative, comparative, lative, locative, nominative, oblique, and translocative cases. Two of the above case markers, namely, the nominative and locative, are involved in grammatical (syntactic) relations. The remaining markers, including the locative in one of its functions, specialize in the function of marking semantic relations. These are the latter that are the object of study in this article.

The case markers under analysis are divided into three subgroups in accordance with the conceptual domain their semantics belongs to. The first subgroup includes case markers that express spatial relations, namely, markers of the ablative, allative, lative and locative cases. The second subgroup includes markers of the abessive and comitative cases, which encode the joint or disjunctive nature of the designated action. The third subgroup includes distributive, objective and translocative cases markers, which are used to indicate adverbial or ‘indirect’ relations.

Let us now analyze the spectrum of meanings encoded by each marker.

The spatial case markers from the first group, namely the ablative, allative, lative, and locative markers, encode the meanings of direction of movement, static position, and position in time. Most of the case morphemes – the ablative, lative and locative markers – belonging to the analyzed group are multifunctional. The only case marker from this group – the allative marker – acts as a monofunctional one.

According to A. L. Malchukov's concept [18], spatial cases participate in the representation of basic semantic roles, namely, the semantic roles of Place, Path, Source (a starting point of movement), Goal (the end point of movement). The field material of Vakh Khanty dialect does not support a clear division of spatial cases according to these semantic roles. In this dialect of Khanty, each role can be expressed by one or two spatial cases. Thus, to represent the role of Place, the locative case is used, to implement the role of Path, the ablative case is used, which can also be used to represent the role of Source. The lative and allative cases are focused on representing the role of the Goal. Let us now dwell separately on each marker from the group of spatial cases.

2.1 Cases of spatial meaning

The ablative case marker *-oy/-öy* encodes the meaning of a starting point of some movement directed away from something or someone. It also marks movement or an action, and a concept or an event derived from an object. Let us illustrate these meanings with examples (1–6):

(1) *lüy-än män-t kōr-äm-öy pör-yas*
 3SG-LOC 1SG-ACC foot-POSS.1SG-ABL bite-PST2.SBJ.3SG
 'She bit me in the foot/leg.'

(2) *lüy-än män-t puylām-am-oy öymilä-s*
 3SG-LOC 1SG-ACC cheek-POSS.1SG-ABL kiss-PST1.SBJ.3SG
 'He kissed me on the cheek.'

(3) *lin paj-oy rīt wer-s-äyän*
 3DU aspen-ABL boat make-PST1-SBJ.3DU
 'They made a boat out of aspen.'

(4) *nūñ paḳi wer-yäs-än tōnəy-oy*
 2SG doll make-PST2-SBJ.2SG birch.bark-ABL
 'You made dolls out of birch bark?'

(5) *min wor ont-oy wiyəl-s-əmən*
 1DU woods inside-ABL come.out-PST1- SBJ.1DU
 'We came out of the woods.'

The ablative marker also encodes the prolative meaning of movement, expressing the meaning of the path, passage, or roadside "along, through" along which the action unfolds:

(6) *kiriw way-oy män-wäl*
 motorboat Vakh-ABL sail-NPST.SBJ.3SG
 'A motorboat sails along the Vakh river.'

(7) *lin lõk-öy pōḡl-a män-s-äyän*
 3DU road-ABL side-LAT walk-PST1-SBJ.3DU
 'They were walking along the side of the road.'

The allative case marker *-pa/-pä* is monofunctional and it usually indicates the final point of movement in the Vakh Khanty dialect. According to the field materials, the allative marker is regularly found in the speech samples of the Vakh Khanty, although in some cases it reveals its functional similarity to the marker of the lative case. This is illustrated by the following examples (8–9):

(8) *lüy lawka-j-pa jäl-äs päni n'än' wä-s*
 3SG store-EP-ALL go-PST1.SBJ.3SG and bread take-PST1.SBJ.3SG
 'He went to the store and bought bread.'

(9) *măčəy təḡi-l-pa mänä-s*
 farmland place-POSS.3SG-ALL go-PST1.SBJ.3SG
 'He went to the farmland.'

The marker of **the lative case -a/-ä** is multifunctional. One of its functions is to encode the final point of the movement trajectory, i.e. the meaning similar to that of the allative case. Let us illustrate these similar meanings with examples:

(10) *min wor ont-oy lar-a wiyəl-s-əmən*
1DU grove inside-ABL swamp-LAT come.out-PST1-SBJ.1DU
'We came out of the grove to the swamp.'

(11) *wälta təyi-l-a män-il-wəl*
live-INF place-POSS.3SG-LAT go-MULT-NPST.SBJ.3SG
'He is going to the farmland.'

Another function is to mark the lative meaning, which is to indicate the purpose to which the action is directed.

(12) *män-nä köy-nä-ti jöy-γäs-i möñkäm-ä*
1SG-LOC rock-COM-PRTC throw-PST2-PASS.3SG snake-LAT
'I threw a rock at the snake.'

(13) *lüy n'əñikiji-t-a-ti litot tu-yal*
3SG kid-PL-LAT-PRTC food take.away-PST3.SBJ.3SG
'She took food away to the kids.'

The lative case is often used to mark nouns that form semantically fused combinations with verbs, for example, to hunt, to matchmake, etc. Thus, one of the most frequently used phrases among speakers of Vakh Khanty 'to hunt' consists of the verb *kəñčta* 'to search' and the noun *wajəy* 'beast' in the lative case. The meaning of 'to matchmake' is conveyed by the verb *ärəytätä* 'to promise' and a noun in the lative case. Let us illustrate the use of these phrases with examples (14–16):

(14) *käčəy kotl wajy-a kəñč-ıl-wəl*
every day beast-LAT search-MULT-NPST.SBJ.3SG
'He hunts every day.'

(15) *änä, män-nä səwəs iki-j-ä ärəy-tə-s-uj-əm*
sister 1SG-LOC sevsiki old.man-LAT promise-TR-PST1-PASS-1SG
'Sister, I got you married to Sevsiki.'

The marker of the lative case -a/-ä can also be used to encode the recipient:

(16) *pəy-ali äni-l-ä t'uti jəy-iyən...*
boy-DIM sister-POSS.3SG-LAT this say-NST.SBJ.3SG
'The boy says this to his sister.'

As it follows from the description above, the lative case marker demonstrates a greater variety of functional applications. However, the most multifunctional case marker in the Vakh Khanty dialect appears to be **the locative case marker -nə/-nä** which is analyzed in detail in a separate article [15]. A new grammatical function of the locative marker has been identified which was not described previously in that work [15]. In this article, we give only few examples in which the locative case marker functions as a representative of the semantic role of Place and marks the location of an object in space:

(17) *köčəy satəw-ən-əki*
knife sheath-LOC-PRED
'The knife is in sheath.'

(18) *sar-γən jəñk-ən-äki-γən*
pike-DU water-LOC-PRED-3DU
'The pike are in the water.'

2.2 Cases of compatibility or deprivation meaning

The second subgroup of case markers includes morphemes of the abessive and comitative cases, which represent the meanings of compatibility or deprivation correspondingly. **The comitative case marker -na/-nā** expresses a comitative meaning, i.e. it describes multiple (two or more) participants of the situation fulfilling the same semantic role.

- (19) *pəy-ali-nə* *öy-äli-l-nā-ti* *kömän* *jaŋqa-l-əyən*
 boy-DIM-LOC girl-DIM-POSS.3SG-COM-PRTC outside play-NPST-SBJ.3DU
 ‘A boy and a girl are playing outside.’

- (20) *vasya* *marina-na-ti* *män-yäs*
 Vasya Marina-COM-PRTC go-PST2.SBJ.3SG
 ‘Vasya and Marina have left.’

The comitative case can convey an instrumental meaning by marking the subject which names the instrument and means of action:

- (21) *i-y-nə* *köt-äl-nā-ti* *joyim-yal-i* *wän ’əm-äl-öy*
 bear-LOC paw-POSS.3SG-COM-PRTC slap-PST3-PASS.3SG face-POSS.3SG-ABL
 ‘The bear slapped him in the face with his paw.’

- (22) *kolya* *rīt* *laijm-na-ti*, *suyəl-na-ti*, *kır-na-ti*
wer-yäs
 Kolya oblas axe-COM-PRTC scraper.axe-COM-PRTC side.axe-COM-PRTC
 make- PST2.SBJ.3SG
 ‘Kolya hollowed out the oblas with an axe, a scaper axe, a side axe.’

- (23) *tim-tom* *pajlaŋ-əl-na* *čoyət* *jerimil-l-ätä*
 this-that wing-POSS.3SG-COM snow streak-NPST-SBJ.3SG:OBJ.SG
 ‘[It] marks the snow with this or that wing.’

The comitative case also expresses a concomitant object:

- (24) *mara* *jö-yäs* *ul-na-ti*
 Mara come-PST2.SBJ.3SG berry-COM-PRTC
 ‘Mara came with berry.’

In the Vakh Khanty dialect, the comitative marker -na/-nā is also used to denote means of transportation, which can be considered the fourth function of the marker.

- (25) *iki* *ku-j-t* *wajəy* *kənç-il-tä* *weli-t-nā* *jäl-ilä-wäl-ät*
 old man-PL beast search-MULT-INF reindeer-PL-COM go-MULT-NPST-SBJ.3PL
 ‘Old men ride reindeer to hunt.’

- (26) *lūy rīt-na* *män-yäs*
 3SG boat-COM иди-ПСТ2.СБЖ.3СГ
 ‘He left by boat.’

- (27) *t’ukumənta*, *lūy* *män-yäs* *likər-nā-ti*
 turn.out 3SG go-PST2 sledge-COM-PRTC
 ‘It turns out that he left on a sledge.’

Field data show that the comitative marker attaches the particle -(ti)/-(ti) more often than other case morphemes, see examples (19–22, 24, 27).

Semantically, the comitative marker -na/-nā is opposed by the monofunctional **abessive case marker -ləy/-läy**.

- (28) *mä* *ämp-läy* *qoya-l-em*

1SG dog- ABS	walk-NPST-SBJ.1SG
'I am walking without a dog.'	
(29) <i>läy püčkän-läy</i>	<i>jäl-il-yäs-ät</i>
3PL gun- ABS	go-MULT-PST2-SBJ.3PL
'They went without weapons.'	

2.3 Cases of “oblique” meaning

The third subgroup of case morphemes includes markers of the oblique, translativ and distributive cases. These markers are grouped on the basis of the circumstantial (adverbial) or “indirect” semantics they convey, similar to the case meanings that are found only in languages with case paradigms including multiple members. The translativ and distributive cases are found in other languages of the Finno-Ugric group, for example, the translativ case is present in Finnish and Hungarian, but the distributive case is only found in Hungarian. The oblique case is a unique feature of the eastern dialects of the Khanty language.

The marker of **the oblique case** *-(t)ä/-(t)ä* conveys the meanings that are typically encoded by the instrumental case in Russian, however, it still differs in its ability to mark only the inanimate object of a transitive verb in the subject or object conjugation, active or passive. In the Russian literary translation, we use the accusative case (i.e., “Give me water”), although a literal translation implies the instrumental case “by water” (see examples 30-33):

(30) <i>män-t</i>	<i>äciy</i>	<i>jəŋk-ə</i>	<i>mäj-itəy</i>	
1SG-ACC	cold	water- OBL	give-IMP.SBJ.PL.	
'Give me some cold water.'				
(31) <i>lüy-ən</i>	<i>män-t</i>	<i>püčkän-ə</i>	<i>mäj-entil-wəl</i>	
3SG-LOC	1SG-ACC	rifle- OBL	give-MULT-NPST.SBJ.3SG	
'He gives me a rifle.'				
(32) <i>käčəŋ</i>	<i>ku</i>	<i>män-nə</i>	<i>kuł-ə</i>	<i>mə-s-i</i>
every	person	1SG-LOC	fish- OBL	give-PST1-PASS.3SG
'I gave everyone a fish.'				
(33) <i>käčəŋ</i>	<i>ku</i>	<i>män-nə</i>	<i>läŋki-tə</i>	<i>pit-s-i</i>
every	person	1SG-LOC	squirrel - OBL	help-PST1-PASS.3SG
'Everyone got a squirrel.'				

The marker of **the translativ case** *-y/-äy/-äy* is used to indicate a change in form or state. Typically, a noun with a translativ marker is used with the verbs *jäta* “to become” and *sälimtätä* “to transform”. This is illustrated by an example:

(34) <i>lüy jol-ta</i>	<i>ku-j-y</i>	<i>jä-yäs</i>	
3SG	shamanize-INF	man- TRL	become-PST2.SBJ.3SG
'He became a shaman.'			

The distributive case marker *-(ä)täl/-(ä)täl* is not identified by all linguists. As a case marker *-(ä)täl/-(ä)täl* is designated only in the Khanty reader by L. Honti [10]. The case marker encodes the meaning “everyone”; it is translated into Russian as “per everyone, per capita” and is not used with all words. Below are examples from the latest field materials that confirm the existence of this marker in the case paradigm of Vakh Khanty (see examples 35-36).

(35) <i>ejnäm</i>	<i>ej</i>	<i>n'än'</i>	<i>täyt-ali-täl</i>	<i>li-s-ät</i>
все	один	хлеб	кусок-DIM- DISTR	есть-PST1-SBJ.3PL
'Everyone ate a piece of bread.'				
(35) <i>käčəŋ</i>	<i>ku</i>	<i>kä</i>	<i>rit-ätäl</i>	<i>wä-yäs</i>
each	person	two	boat- DISTR	take-PST2.SBJ.3SG

‘Each took two boats.’

(36) *ejnäm* *ej* *kul-atal* *li-s-ät*
все один рыба-DISTR есть-PST1-SBJ.3PL

‘Everyone ate one fish.’

2.4 The Status of the element *nijit* and the summary of the nominal case markers of modern Vakh Khanty in a table

The comparative element *nijit*, which is singled out by J. Guya as an independent case marker [9], does not confirm its status as a case marker drawing on the results of the analysis of field data. It is used to denote the comparison of one object with another and functions as an autonomous auxiliary element - a postposition, which is confirmed by the violation of the vowel harmony of the element *nijit* with a preceding word. N. I. Tereshkin also defines the element *nijit* as a postposition, for example (37):

(37) *körliki* *larjyak* *nijit* *jäm-iki*
Korliki Larjyak compared.to good-PRED

‘Korliki is better than Larjyak.’

Table 1 shows markers of all cases of the nominal declension system, used to mark semantic relations in the present-day Vakh Khanty dialect.

Table 1. A set of markers of the semantic cases in Vakh Khanty based on the latest field data

Semantic groups of modern Vakh Khanty markers		
№	Case	Case marker
Group 1. Cases of spatial meaning		
1	Ablative (ABL)	oŷ/öŷ
2	Allative (ALL)	(a)pa/(ä)pä
3	Lative (LAT)	a/ä
4	Locative (LOC)	nə/nă
Group 2. Cases of compatibility or deprivation meaning		
5	Abessive (ABS)	ləŷ/lăŷ
6	Comitative (COM)	na/nă
Group 3. Cases “oblique” meaning		
7	Distributive (DISTR)	-(ə)təl/-(ə)täl
8	Oblique (OBL)	(t)ə/(t)ă
9	Translative (TRS)	ŷ/əŷ/ăŷ (< kə/kă)

Table 1 shows a summary of the unified data on case markers of Vakh Khanty, compiled on the basis of this study using materials from the latest expeditions of the current century. The table lists 9 case markers in unified terminological notations. We have proposed a one-component case terminology, where next to each case term there is also a variant of its designation for glossing in order to standardize glossing of texts in the Vakh dialect. Unification of notations is necessary for universal markup of texts, so that when using the tools of the LingvoDoc platform, an entire sample of combinations with case markers available in the corpus is selected.

3. Conclusion

Field data of the present-day Vakh Khanty dialect indicates that its nominal declension system includes ten case forms, nine of which specialize in encoding semantic relations. The analysis of the functional features of case markers was carried out drawing on a digitized corpus of linguistic data

of the Vakh Khanty dialect, available on the LingvoDoc platform. The corpus contains the latest language data on this dialect, collected in the village of Korliki, Nizhnevartovsk region, Khanty-Mansi Autonomous Okrug in 2019. The study results made it possible to confirm the stability of the system of nominal declension of the Vakh Khanty dialect, to clarify the functional uniqueness of each case marker and to identify the grounds for their preferential terminological designation with the following labels: 1) ablative, allative, lative, locative; 2) abessive, comitative; 3) distributive, oblique and translativ. The comparative element *niñit* did not confirm its status of the case marker. This study proposes a unified terminology of case markers for the development of parsers and markup of text corpora in Vakh Khanty.

In ensuing studies analyses of the analytical ways of expressing syntactic relations between objects will be conducted, namely, constructions with postpositions and postpositional nouns which are used to express spatial and temporal meanings along with synthetic cases.

4. List of abbreviations

ABL – ablative

ABS – abessive

ACC – accusative

ALL – allative

COM – comitative

DIM – diminutive

DISTR – distributive

DU – dual

EP – epenthetic

IMP – imperative

INF – infinitive

LAT – lative

LOC – locative

MULT – multiplicative

NPST – non-past

NST – suffixless

OBJ – objective

OBL – oblique

PASS – passive

PL – plural

POSS – possessive

PRED – predicative

PRTC – particle

PST1 past in -s

PST2 – past in -yas/-yäs

PST3 – past in -yal/-yäl

SBJ – subject

SG – singular

TRL – translativ

5. Field materials of the authors

AFM – Field materials of the authors. Polevye materialy avtora. Jekspedicija v p. Korliki HMAO – Jugry, leto 2019 g. [Author's field materials. Expedition to Khanty-Mansi Autonomous Okrug-Yugra. Summer 2019"] (speakers: T. F. Katkaleva, b. 1944, E. A. Kunina, b. 1939, Zh. A. Khokhlyankina, b. 1967).

References

- [1]. O.A. Iggesen, Number of cases. WALS Online. 2013. Eds. by Dryer Matthew S., Hasplmath Martin. Available at: <https://doi.org/10.5281/zenodo.7385533> (Accessed April, 2024).
- [2]. C. J. Fillmore, The case for case. Universals in linguistic theory. Eds. by E. Bach, R. T. Harms. New York, NY: Holt, Rinehart and Winston, 1968. Pp. 1–88.
- [3]. A. A. Zaliznyak, On the understanding of the term “case” in linguistic descriptions. Problemy grammaticheskogo modelirovaniya [Problems of grammatical modeling]. Ed. by A. A. Zaliznyak. Moscow: Nauka Publ., 1973. Pp. 53–87. (A.A. Зализняк. О понимании термина «падеж» в лингвистическом описании. Проблемы грамматического моделирования. М.: Наука, 1973. С. 53–87)
- [4]. I.A. Melchuk, Course of general morphology. Vol. 2. Moscow; Vienna: Languages of Russian culture, Vienna Slavic Almanac, Progress Publishing Group, 1998. 544 p. (И.А. Мельчук, Курс общей

- морфологии. Т. 2. Москва – Вена: Языки русской культуры. Венский славистический альманах. Из-во: Прогресс, 1998. 544 с.)
- [5]. V. A. Plungyan, *Studies in Grammar*, is. 2: Grammarization of spatial meanings in the languages of the world. M.: Russian dictionaries Publ., 2002. Pp. 57–99. (В. А. Плунгян, Исследования по теории грамматики, вып. 2: Грамматикализация пространственных наречий в языках мира. М.: Русские словари, 2002. С. 57–99).
- [6]. V.A. Plungyan, *General morphology. Introduction to the issue*. Moscow: Editorial URSS Publ., 2003. 384 p. (В. А. Плунгян. Общая морфология. Введение в проблематику. Москва: Эдиториал УРСС, 2003. 384 с.)
- [7]. M. Chepregi, *Surgut dialect of the Khanty language*. Khanty-Mansiysk, Printed World of Khanty-Mansiysk, 2017. 275 p. (М. Чепреги, Сургутский диалект хантыйского языка. Ханты-Мансийск: ООО «Печатный мир г. Ханты-Мансийск», 2017. 275 с.)
- [8]. A.Yu. Filchenko, *A Grammar of Eastern Khanty*. Houston, Texas, 2007. 615 p. Available at: <https://theswissbay.ch/pdf/Books/Linguistics/Mega%20linguistics%20pack/Uralic/Khanty%2C%20A%20Grammar%20of%20%28Filchenko%29.pdf> (Accessed April, 2024)
- [9]. J. Gulya, *Eastern Ostyak Chrestomathy*. Bloomington: The Hague, 1966. 207 p.
- [10]. L. Honti, *Chrestomathia Ostiastica*. Budapest: Tankonyvkiado, 1986. 432 p.
- [11]. K. F. Karjalainen, *Grammatikalische Aufzeichnungen aus Ostjakischen Mundarten*. Bearbeitet und herausgegeben von Edith Vertes. Helsinki: Suomalais-Ugrilainen Seura, 1964. 234 p.
- [12]. D. F. Mymrina, *Case category in the Khanty language (comparative aspect)*. Tomsk, 2006. 145 p. (Д. Ф. Мырина, Категория падежа в хантыйском языке (сопоставительный аспект). Автореферат диссертации на соискание ученой степени к.ф.н. Томск, 2006)
- [13]. N. I. Tereshkin, *Essays on dialects of the Khanty language (The Vakh dialect)*, L.: Nauka, 1961. (Н.И. Терёшкин, Очерки диалектов хантыйского языка (ваховский диалект). Л.: Наука, 1961.)
- [14]. V. Vorobeva, I. Novitskaya, *Inflectional morphology of nouns in Eastern Khanty (Vakh, Vasyugan, Surgut, Salym)*. *Ural-Altaic Studies*, vol. 38, 2020, pp. 33-70.
- [15]. V.V. Vorobeva, I.V. Novitskaya, Z.V. Fedorinova, *Syntactic cases of nouns in Vakh Khanty (based on the latest field data of LingvoDok / Kazan Science*, vol 9, 2023, pp. 157-159. (В.В. Воробьева, И.В. Новицкая, З.В. Федоринова, Синтаксические падежи имени существительного в ваховском хантыйском языке (на материале современных полевых данных платформы LingvoDok. Казанская наука, том 9, 2023, стр. 157-159).
- [16]. V.V., Vorobeva, I.V. Novitskaya, *Functional characteristics of the nominal case markers in Vakh Khanty (on the basis of the recent field language data of the LingvoDoc platform)*. *Trudy ISP RAN/Proc. ISP RAS*, vol. 35, issue 4, 2023. pp. 165-176 (В.В. Воробьева, И.В. Новицкая, Функциональные особенности падежных показателей в ваховском хантыйском языке (на материале базы современных полевых данных на платформе ЛингвоДок). Труды ИСП РАН, том 35, вып. 4, 2023 г., стр. 165–176).
- [17]. Vakh Khanty Corpus. Lingvodok Platform. Available at: <http://https://lingvodoc.ispras.ru/dictionary/3568/147/perspective/3568/151/view>. (Accessed 24 April, 2024). (Корпус ваховского диалекта хантыйского языка. Доступен на: <http://https://lingvodoc.ispras.ru/dictionary/3568/147/perspective/3568/151/view>. (Accessed 24 April, 2024)
- [18]. A. L. Malchukov, *Syntax of the Even language: structural, semantic, communicative aspects*. St. Petersburg, Nauka, 2008. 425 p. (Мальчуков А. Л. Синтаксис эвенского языка: структурные, семантические, коммуникативные аспекты. СПб, Наука, 2008. 425 с.)

Информация об авторах / Information about authors

Виктория Владимировна ВОРОБЬЕВА – кандидат филологических наук, доцент отделения иностранных языков Школы общественных наук Национального Исследовательского Томского политехнического университета с 1999 года. В настоящее время также работает старшим научным сотрудником в лаборатории Лингвистические платформы Института системного программирования РАН. Область научных интересов: сравнительно-типологическое языкознание, уральская языковая семья, корпусная лингвистика, документация и цифровая обработка исчезающих языков.

Victoria Vladimirovna VOROBÉVA – 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.

Ирина Владимировна НОВИЦКАЯ – доктор филологических наук, доцент, профессор кафедры английской филологии Национального исследовательского Томского государственного университета с 1997 года. Область научных интересов: морфология и семантика древнегерманских языков, типологическое языкознание, сравнительно-сопоставительные исследования, морфосинтаксис обско-угорских языков, прикладная лингвистика.

Irina Vladimirovna NOVITSKAYA – Dr. Sci. (Philology), professor at the Department of English Philology, at 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.