

The CALA 2019 Proceedings Paper 14 - 2

Language Documentation, Paper 2

**Lexicographic Documentation of an Endangered
Language: The Case of Ket**

Elizaveta Kotorova

University of Zielona Góra, Poland

National Research Tomsk Polytechnic University, Russia

Andrey Nefedov

Hamburg University, Germany

Lexicographic Documentation of an Endangered Language: The Case of Ket

Elizaveta Kotorova^a

^a*University of Zielona Góra, Poland; National Research Tomsk Polytechnic University, Russia*

Andrey Nefedov^b

^d*Hamburg University, Germany*

Abstract

This presentation aims at describing and analyzing the main problems faced in the course of making a dictionary of Ket, a unique and highly endangered language spoken in Northern Asia. Among the primary issues discussed are the following:

1. Target audience. There are three variants: (1) language community; (2) academic community; (3) both communities. The survey of existing Ket dictionaries and the present sociolinguistic situation in the Ket community has shown that scholars are by and large the main TA for the dictionary. This was also the determinant factor in dealing with various practical dictionary-related questions.

2. Basic vocabulary. The initial wordlist can be based on: (1) translation of the list of the most frequent words from a European language; (2) extraction of the wordlist from a corpus of texts; (3) thematic elicitation from native speakers. The main peculiarity in compiling a wordlist for the Ket dictionary is connected with the fact that it was initially created on the basis of a handwritten card file dictionary (compiled from a collection of field notes).

3. Dictionary entry. It includes two important components – a lemma and a commentary.

3.1. Lemma. Since the dictionary is primarily targeted at scholars, it uses notation based on IPA. Due to diverse orthographic notations used in the field notes as well as in other sources, dictionary representations of the Ket data required unification. As a result, Ket lemmata are provided in strict phonological transcription, while illustrative contexts are represented in a unified phonetic transcription reflecting dialectal differences.

3.2. Commentary. An obligatory and very important component of the commentary is a certain hierarchic arrangement of word meanings reflected in the corresponding meta-language translations. Linguists compiling dictionaries usually rely on the totality of contexts in which the given word can be found, and, if they are native speakers, upon their own intuition. Those who compile dictionaries of unwritten languages are generally not native speakers, and therefore contexts are of an utmost importance. In this case, each single meaning should be confirmed by an appropriate context. The corpus of illustrative examples for the dictionary is based both on published and unpublished sources. In many cases, the commentaries include encyclopedic information, as it helps to understand certain ethnospecific concepts.

Keywords: Ket, documentation, endangered languages, Lexicography

Introduction

This paper discusses key linguistic and sociolinguistic peculiarities that played a major role in the making of the Comprehensive Dictionary of Ket, a highly endangered language spoken in North Asia, the sole surviving member of the Yeniseian family.

The last remaining Ket speakers reside in the north of Russia's Krasnoyarsk province in remote settlements along the Yenisei river. The current sociolinguistic situation is characterized by the lack of monolingual speakers and the predominance of Russian in all spheres of communication. Speaking from the recent fieldwork experience, the present-day number of competent speakers in all of the Ket dialects combined does not exceed 20-30 people.¹ The average age of the majority of these speakers is over 65 years.

The drastic sociolinguistic situation made creating a comprehensive dictionary while there are still fluent speakers of Ket a task of utter importance. However, it also inevitably had its impact on certain aspects of the dictionary discussed below.

In what follows, we will first consider the Ket dictionaries that were published before the Comprehensive Dictionary of Ket (CDK for short) and then discuss the key distinctions of CDK and what factors were decisive in shaping the form of the dictionary.

Ket dictionaries: General overview

There exist several Ket dictionaries to date; however, all of these differ from the approach and mission of the Comprehensive Dictionary of Ket in many respects.

The first published Ket dictionary (not taking into account the century old lists of Ket words and various glossaries) is the Ket-Russian and Russian-Ket dictionary for school compiled by Heinrich Werner (Werner 1993). It is based on the Southern dialect of the language and employs the official Cyrillic-based Ket alphabet. Its main purpose was to facilitate the process of teaching Ket at school. The dictionary is rather small with about 2800 headwords in each part. The entries are very compact, mostly lacking any illustrative examples and extended contexts.

A similar dictionary based on the Central Ket dialect was compiled by the native Central Ket speaker Zoja Maksunova (Maksunova 2001). This dictionary contains only a Ket-Russian part with about 2500 lexical units; the entry is similar to that in the school dictionary of Werner.

Another important lexicographical work, the Comparative Dictionary of the Yeniseian languages (Werner 2002), represents a completely different type of dictionary. It is a solid three-volume work containing about 11500 entries belonging to all known Yeniseian languages. The main purpose of this dictionary is to compare words with similar meanings in these languages and, where possible, to reconstruct the corresponding Proto-Yeniseian forms. The meta-language of the dictionary is German with the Yeniseian data presented in a specific transcription, largely based on IPA. This fundamental work puts its main focus on the history of the Yeniseian family and is not intended to fully account for the semantics of Ket words at the synchronic level, since the author is more concerned with the Yeniseian etymology (cf. Kotorova and Nefedov 2004; Kotorova 2016).

¹ At the same time, according to the latest census, there are 1219 ethnic Kets.

Comprehensive Dictionary of Ket: Major distinctions

CDK differs from the aforementioned dictionaries in several respects.

First of all, it covers the largest amount of verified vocabulary from all Ket dialects (Southern, Central and Northern), which means that all the headwords in the dictionary were checked either by the native Ket speakers during fieldwork, or can be verified by contextual examples taken from both published and unpublished sources.² The total number of lexical entries in the dictionary is over 6000.

Second, the main objective of CDK is to reveal the semantics of words in Ket through illustrative contexts, explanations and comments. Therefore, it contains over 27000 illustrative contexts ranging from simple noun phrases to sentence strings, as well as supplementary encyclopedic descriptions. In addition, it also provides extensive grammatical information about the Ket words necessary to facilitate understanding the complexities of the language.

Third, CDK employs a unique approach to presenting the Ket material. For example, it uses a combination of phonological (headword) and phonetic (illustrative examples) representation of the Ket data, a formulaic representation of the polysynthetic Ket verb, etc. We will discuss some of the approaches in more detail in the sections below.

Forth, unlike the other Ket dictionaries which are bilingual, CDK includes parallel translations of Ket headwords into three languages: Russian, German and English. All the illustrative Ket examples in CDK, however, are provided with Russian translations only in order not to exceed a reasonable page limit for the printed version.

Finally, the dictionary consists of two separate volumes, one devoted to verbs only and the other one to nouns and rest of the parts of speech. The reason behind this decision is discussed in Section 4.2.

It should also be noted that the drastic sociolinguistic situation in the Ket community predetermined the choice of the target audience for the dictionary. While in many cases, the most obvious choice is making a dictionary that would serve both the interests of native speakers and linguists, in the case of CDK it was the linguist community that had to be chosen as the primary target audience (cf. Kotorova 2016 for more discussion). One of the most important decisions connected with this target audience was choosing the IPA-based transcription over the official Ket alphabet.

The statistical data show that only 2,8 % of the ethnic Kets are able to read and write the Cyrillic-based alphabet, and a somewhat larger percent, 10,5 %, can only read it (Krivonogov (2003: 86). Therefore, using this alphabet would render the dictionary much less useful in general, not to mention the limitations of the official Ket alphabet pointed out below.

Key peculiarities of Ket

In this section, we consider the key peculiarities of the Ket language that guided our decisions with respect to lexicographical representation of the Ket data in the dictionary. These peculiarities can be conventionally divided into phonetic and phonological, morphological and lexical-semantic ones.

²The only exception are words related to the sphere of native religious beliefs. This knowledge is almost completely lost among the modern Kets. Likewise, these words are usually not used in the Ket narratives. In such cases, the dictionary entry consists of only a headword and an extended ethnographic description; cf. also Section 4.3 below.

Phonetic and phonological peculiarities

The most prominent characteristic of Ket phonology is a system of four suprasegmental oppositions or tonemes in the domain of monosyllabic words: 1) high-even, 2) abrupt rising (laryngealized), 3) rising-falling, and 4) falling (cf. Nefedov and Vajda 2015: 28). The tonemes form numerous minimal pairs and even sets which differ in lexical or grammatical meaning.

Lexicographical representation of these tonemes differs from dictionary to dictionary. Both Verner (1993) and Maksunova (2001) are based on the official alphabet which captures the distinction between the second and the third tonemes with a special diacritic sign and a germinated vowel, respectively, but fails to distinguish between the first and fourth tonemes.³ In Werner (2002), on the other hand, all four tonemes are distinguished with a superscript number from 1 to 4 with additional suprasegmentals for the first three ones. In CDK, each of the four tonemes has its own distinct representation with a special tonal mark following Vajda (2004: 9-11). This way of representation of tonal distinctions is more or less conventional in the linguistic literature; therefore it would be more easily accessible for the target audience. Table 1 illustrates the differences between CDK and the other Ket dictionaries.

Table 1. Graphic presentation of Ket tonemes in the dictionaries.

CDK	Verner (1993)	Maksunova (2001)	Werner (2002)
qā ‘home’	қа	қа	¹ qa·
qaʔ ‘word’	қаʔ	қаʔ	² qaʔ
áàŋ ‘hot’	aaŋ	aaŋ	³ a:ŋ
àŋ ‘rope’	aŋ	aŋ	⁴ aŋ

A rather non-canonical lexicographic approach was employed with regard to presenting phonetic differences between Ket dialects. Ket distinguishes three major dialects: Southern, Central and Northern. They are further subdivided into subdialects named after the village each is spoken in. Among the most prominent phonetic differences between them are, for example, truncation of the final unstressed vowel in Southern Ket (e.g. SK sèl, CK sèle, NK sèli ‘reindeer’), rhotacism of intervocalic d > r in Southern and Northern Ket (e.g. CK tì:də, SK tìr, NK tì:ri ‘root’), spirantization of b > v in Southern and Northern Ket (e.g. CK do:ba, SK dova, NK do:va ‘a.k.o. fishtrap’), etc.⁴

As we mentioned in Section 2 above, two of the previously published dictionaries were based on one specific dialect – Verner (1993) on Southern Ket and Maksunova (2001) on Central Ket. In Werner (2002), the majority of Ket headwords are presented in their Southern Ket form. However, most of the entries provide information about the corresponding forms in the other dialects, as well. CDK takes a different approach influenced by the fact that the target audience is the linguistic community. In order to stay dialect-neutral, we decided to provide Ket headwords in their phonological form. At the same time, every entry shows dialectal variants, if they are considerably different from the phonological headword. Moreover, all corresponding illustrative contexts are provided in phonetic transcription to reflect dialectal differences. They are also marked with a village abbreviation to indicate the place each context was recorded in. This also provides a convenient way of introducing the reader to how phonology works in different Ket dialects. Below is an example of an entry from CDK.

èd1 m, edn; (sket. èrɪ, cket. èdə, nket. èri) соболь//Zobel//sable; sur. qɔʔk èd
 один соболь, sur. etn kurɪʔar связка (шкурок) соболей, sur. kiséŋ etn ənaŋ
 здесь соболей много, bak. hiɪl qaseŋ һана етн внизу (там) соболята
 (Kotorova and Nefedov 2015: 160)

Morphological peculiarities

As we mentioned above, the dictionary consists of two volumes, the first one containing all parts of speech except the verb and the second one being solely devoted to the Ket verb. The reason behind such unconventional division is that in CDK, the Ket verb headword is represented not with a lexical word, but with a special hyphenated stem formula, which is why it is not appropriate to list them together with the other parts of speech. The verb formula is based on the position class model developed by Edward Vajda. Table 2 illustrates the latest variant of the model from Nefedov and Vajda (2015: 35).

Table 2. Position classes in Modern Ket

P8	P7	P6	P5	P4	P3	P2	P1	P0	P-1
AGR or thematic valence reducing affix	1) left semantic head or 2) noun, adj, or adverb root	AGR	thematic consonant (most are seman- tically opaque)	<i>tense/</i> <i>mood</i> or AGR	AGR or thematic non- agreement affix	<i>past tense/</i> <i>imperative</i>	AGR or thematic valence reducing affix	1) right semantic head or 2) aspect/ voice auxiliary)	AGR (in verbs that use P8 for subject)

This position class model consists of ten separate positions. All the positions can be conventionally divided into three general types: lexical, tense/mood and agreement positions. The basic lexical stem is formed through a combination of morphemes in positions P7, P5 and P0. When present in a particular verb form, these positions remain unchanged throughout the whole paradigm, and therefore are responsible for the lexical meaning of the verb. Tense and mood distinctions are generally marked through a combination of morpheme shapes in positions P4 and P2. There are six productive tense/mood combinations in Modern Ket. Positions marked as 'AGR' are potential agreement positions.

The formulaic headword of the Ket verb in the dictionary consists of lexical morphemes marked by superscript numerals indicating position class. These morphemes that remain unchanged in all grammatical forms are the basis of each formula. Elements that sporadically appear or disappear across the stem's conjugated forms are placed in parentheses. Square brackets enclose tense-mood morphemes belonging to slots P4 or P2. An example of a verbal entry can be seen below.

ha(d)7-[s4]-[l2]-a0 vt1 iter кто-л режет что-л//jmd schneidet etw//smn cuts
 smth; kel. īsʲ āt thaʃa мясо я режу, kel. īsʲ āt haɪʲaɪa daqqajit рыбу я разрежу
 (на куски), пожарю

(Kotorova and Nefedov 2015: 500)

The formulaic representation of the Ket verb is a distinctive feature of CDK. The other Ket dictionaries, following the Indo-European lexicographic tradition, employ special non-finite forms referred to as ‘infinitives’ as a citation form for verbs. The reason for this is rather straightforward as these forms fulfill many of the functions typical of the Russian infinitive. However, if we consider all the factors including the functional range and the morphosyntactic properties inherent to these word forms, it becomes obvious that the term ‘action nominal’ would be more justified in this case (cf. Nefedov 2015: 27-30).

There are several reasons, why we did not use these non-finite forms as a citation form the Ket verb and decided to list them in the first volume with the non-verbal parts of speech. First of all, they are morphologically diverse and, in general, lack special marking (cf. Werner 1997: 175-180). Furthermore, they show a considerable degree of lexicalization, i.e. it is impossible in many cases to predict their form from the semantically corresponding finite verb and vice versa; consider the following example:

- (1) dbílabak
 d[i]8-b3-l2-bak0
 1sg8-3n3-pst2-drag0 t
 ‘I dragged it.’

The corresponding non-finite form for this verb is *bákdeŋ* ‘pulling’, not **bak* as one could expect (Werner 1997: 176). Some non-finites are in fully suppletive relation with the semantically corresponding finite verb, for example, *éjiŋ* ‘going’ and *bókatn* ‘I go’ [bo6-k5-a4-tn0 1sg6-th5-npst4-go0]. Finally, some finite verbs do not have a corresponding non-finite form at all, e.g. *dabátabet* ‘I understand’ [da8-ba6-t5-a4-b3-et0 ic8-1sg6-th5-npst4-3n3-understand0] – neither **et* nor anything else is the non-finite counterpart for this verb. In the latter case, Verner (1993) and Maksunova (2001) simply provide a fully finite verb in the first person singular as a citation form. Werner (2002) is closer to CDK in this respect, as it also employs a kind of formulaic verb stem consisting of lexical morphemes only (for example, for ‘understand’ it is *t...et*).

Lexical-semantic peculiarities

From the lexical-semantic point, there are the following lexicographic peculiarities in CDK. In the other Ket dictionaries, the verb headword is translated with either an infinitive or a finite verb form in the first person singular (depending on the headword form). In CDK, however, the translation of verbal headwords is done in the so-called “phrasal” format, e.g.: *ed(a)7-q5-a4-[l2]-da0 vt1 caus iter 1. кто-л посылает кого-л/что-л//jmd schickt jmdn/etw//smn sends smn/smith*. This translation format allows for a better understanding of the complex nature of the Ket verb and helps to capture the peculiarities connected with various restrictions coded in a verb form itself (for example, obligatory animacy or inanimacy of the subject/object, etc.). In cases, when the subject or object belongs to the animate class in Ket, but is inanimate in the target languages, its representation follows the source language, i.e. Ket, e.g.: *en7-t5-[l2]-da0 vt1 iter кто-л застёгивает кого-л (пуговицу)//jmd knöpft jmdn (den Knopf) zu//smn buttons*

smn (a button).

Another lexical-semantic peculiarity of CDK is its focus on presenting as much ethnocultural information in a dictionary entry as possible (cf. Kotorova and Nefedov 2016). In the lexicographic tradition of majority languages, ethnocultural information is usually presented in encyclopedic dictionaries which serve to describe ethnocultural realia, but not lexemes. Since Ket has no existing encyclopedic works describing ethnocultural realia, we decided to include this kind of information in addition to the translations in the dictionary in order to preserve the ethnic knowledge and give the reader a detailed understanding of such a lexeme. This kind of information is largely absent in the other Ket dictionaries, cf. Table 3 below.

Table 3. Kinds of hunting traps Ket

CDK	Verner 1993	Maksunova 2001	Werner 2002
dedaŋoks ‘čerkan (a type of jamming snare for fur-bearing animals)’	a trap for birds	-	a kind of trap
kī ‘past’ (a type of pressing trap for wolverines, hares, woodgrouses, polar foxes)’	kuljomka (a kind of trap)	kuljomka (a kind of trap)	a kind of hunting trap
kəntoks ‘kleptsa (a type of striking trap for foxes, wolverines, polar foxes)’	-	-	a kind of hunting trap

As we can see, the lexicographic description of lexemes with the meaning ‘trap’ in CDK reveals the semantics of these ethnocultural realia to a greater extent and highlights important differences between them.

To conclude, we would like to note that making a dictionary for an endangered language may require a creative and open-minded approach, as it often deals with tasks and problems going beyond traditional lexicography based on majority languages.

References

- Kotorova, E. (2016) “*Dictionary for a Minority Language: the Case of Ket*. In: *Lexicography and Linguistic Diversity*.” In Proceedings of the XVII EURALEX International Congress (Tbilisi 6 – 10 September, 2016), edited by Tinatin Margalitadze and George Meladze, 133-141. Tbilisi: Lexicographic Centre of Ivane Javakhishvili Tbilisi State University, 2016.
- Kotorova, E and Nefedov, A. (2004) “Ketskij proekt: sostojanie del, problemy, perspektivy.” In *Sravnitel’no-istoričeskie i tipologičeskie issledovanija jazyka i kul’tury: problemy i perspektivy. Sbornik naučnyx trudov Laboratorii jazykov narodov Sibiri*. Tom 2, 19-32. Tomsk: Centr učebno-metodičeskoj literatury Tomskogo gosudarstvennogo pedagogičeskogo universiteta.

- Kotorova, E. and Nefedov, A. (2015) *Comprehensive Dictionary of Ket* (with Russian, German and English Translations). Under the general supervision of E.G. Kotorova. Vol. 1 and 2. München: Lincom Europa.
- Kotorova, E. and Nefedov, A. (2016) Problema predstavleniia étnokul'turnyx realij v slovare minoritarnogo jazyka (na primere ketskoj leksiki). *Tomsk Journal of Linguistics and Anthropology*, 3, 24-32.
- Krivosnogov, V. and Kety, P. (2003) *desjat' let spustja (1991-2001)*. Krasnojarsk: Krasnojarskij gosudarstvennyj pedagogičeskij universitet.
- Maksunova, Z. (2001) *Kratkij ketsko-russkij slovar' (sredne-ketskij dialekt)*. Pod redaktsiej G.K. Vernera. Krasnojarsk.
- Nefedov, A. (2015) *Clause linkage in Ket*. Utrecht: LOT.
- Nefedov, A. and Edward Vajda. (2015) Grammatical sketch. In *Comprehensive Dictionary of Ket with Russian, German and English translations*, edited by Elizaveta Kotorova and Andrey Nefedov, 27-69. München: Lincom Europa.
- Vajda, E. (2000) *Ket prosodic phonology*. München: Lincom Europa.
- Vajda, E. (2004) *Ket*. München: Lincom Europa.
- Verner, G. (=Werner, Heinrich). (1993) *Slovar' ketsko-russkij i russko-ketskij (yuzhno-ketskij dialekt): Učeb. posobie dlya učashhikhsya nachal'nykh shkol*. SPb.: Prosveshhenie.
- Werner, H. (1997) *Die ketische Sprache*. Wiesbaden: Harrassowitz Verlag.
- Werner, H. (2002) Vergleichendes Wörterbuch der Jenissej-Sprachen. In *3 Bd.* Wiesbaden: Harrassowitz Verlag.