Annotation Guidelines for the Selkup Corpus (DFG)[1]

Josefina Budzisch

(29.11.2018)

1. Introduction

The corpus has been created within the DFG project *Syntactic description of the Central and Southern Selkup dialects: a corpus based analyses* (WA 3153/3-1). The primary goal of the project is to build a corpus and research syntactic structures on its base.

The corpus contains texts already published in written form; the aim is to make these texts digitally searchable according to the standards of a modern corpus.

The Selkup language is not entirely undescribed, especially phonological and morphological features are well described, lacking is a description of syntactic structures and their underlying causes. The focus lies therefore on filling these gaps in the existing research.

1.1. About the language

Selkup, also formerly known as Ostyak Samoyedic, is a language of the south Samoyedic branch of the Uralic language family. The Selkups inhabit Siberia, scattered between the rivers Ob and Yenissei. Two main groups are to be identified: the Selkups living in the northern parts of Siberia (settling close to the rivers Taz, Turukhan and Yenissei), the others living far more south in the vicinity of the river Ob.

According to the last Russian census in 2010, there are only 3,649 Selkups living in the Russian Federation. 1,023 people state, that they are active speakers of the Selkup language. The distribution of speakers is not uniform: results of more recent field research indicate that most of the speakers speak a variety of the Northern dialect; the Central and Southern dialects are close to extinction.

The distinction and classification of different dialects has proven difficult, there is no consensus between scientists. Nonetheless Selkup is at least to be divided into three bigger dialectal groups: Northern Selkup, Central Selkup and Southern Selkup, which are each to be subdivided into further subdialects, some researcher as e.g. Helimski (1998) considers Ket to be an own dialectal group with subdialects. Next to these groups, also mixed dialects appears in which characteristics of more than one dialect group can be found, mixed dialects in this corpus are a mix of Central/Southern group, namely Tym and Middle Ob dialects.

Table 1: Selkup dialects (following Glushkov 2013)

Northern Selkup	**Central Selkup**	**Southern Selkup**
Taz	Vakh	Middle Ob
Baikha	Tym	Upper Ob
Upper Tolka (Laryak)	Vasyugan	Ket
Yelogui	Narym	Chaya

[1] The annotation guidelines are based on Wagner-Nagy, Beáta – Sándor Szeverényi – Valentin Gusev 2018: User's Guide to Nganasan Spoken Language Corpus. *Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology* 1.

		Chulym (†)

1.2. Project members

Leader of the project: Prof. Dr. Beáta Wagner-Nagy, responsible for the glossing and annotating of Northern Selkup materials

Project members:

— Researchers:

Budzisch, Josefina, M.A. (October 2015 – September 2018), responsible for the glossing and annotating of Central Selkup materials

Harder, Anja, M.A. (October 2015 – September 2018), responsible for the glossing and annotating of Southern Selkup materials

— Student assistants:

Bui, Thao Van (March 2018 – September 2018), responsible for consistency checking of the Northern Selkup texts, translating texts into German and English

Jark, Florian (January 2017 – April 2018), responsible for translating the texts into German and English.

Krieg, Jacqueline (March 2018 – September 2018), responsible for consistency checking of the Central Selkup texts, translating texts into German and English

Otte, Felicitas (March 2018 – September 2018), responsible for consistency checking of the Southern Selkup texts, translating texts into German and English

1.3. Technical support

HZSK

Hedeland, Hanna – coordination

Ferger, Anne – data curation

1.4. The corpus

The corpus is based on written text, published in various sources beforehand. It contains text from all three dialectal groups. The corpus contains 144 glossed and annotated texts from 48 speakers, 9,156 utterances with 55,839 tokens can be found in the corpus.

Table 2: Corpus data

	speakers[2]	texts	utterances	tokens
Northern Selkup	9	26	1,410	7,815
Central Selkup	15	48	2,165	14,485
Southern Selkup	23	66	3,844	21,492
Mixed dialects	1	4	1,737	12,047
total	48	144	9,156	55,839

[2] 3 Northern Selkup, 4 Southern Selkup and 4 Mixed texts are from unknown speakers, which is here considered to be one speaker.

1.4.1. Basic information

For all texts the original Selkup text is given as well as a translation to English, furthermore most texts are also translated to Russian, German and some to Hungarian.

SIL Fieldworks Explorer (*FLEx*)[3] is used to gloss the texts morphological. Afterwards the texts are exported to *EXMARaLDA*, this is done by Dr. Alexandr Archipov and Beáta Wagner-Nagy. In *EXMARaLDA Partitur Editor*[4] annotations for syntactic functions (SyF), semantic roles (SeR) and information status (IST) are added as well as additional annotation for some texts. The data about the texts and the metadata about the speakers are managed with *EXMARaLDA Corpus Manager* (Coma)[5].

1.4.2. Citation

Budzisch, Josefina – Anja Harder – Beáta Wagner-Nagy 2019. *Selkup Language Corpus* (SLC). Archived in Hamburger Zentrum für Sprachkoropra.

All the authors have equally contributed to the creation of the corpus and are listed here in the alphabetical order.

1.4.3. Abbreviations of researcher

In the data about the texts in Coma, the main researchers adding to the corpus are marked by their abbreviations, given here in alphabetical order:

BJ: Budzisch, Josefina
HA: Harder, Anja
WNB: Wagner-Nagy, Beáta

1.4.4. Archiving

The transcriptions and metadata of the corpus are stored in EXMARaLDa format. The archiving and publication are taken care of by the *Hamburg Centre for Language Corpora* (HZSK).

1.4.5. Orthography in the corpus

The corpus is a compilation of published texts gathered by several researchers written in either cyrilic letters or Latin ones. In the corpus a unified, Latin based script is used in the transcription tier. Vowel length is marked with the IPA symbol <ː>, palatalization with <ʼ>. The *charis SIL font* is used throughout the corpus.

---

[3] http://software.sil.org/fieldworks/support/using-sendreceive/flex-bridge/

[4] http://exmaralda.org/en/partitur-editor-en/

[5] http://exmaralda.org/en/corpus-manager-en/

Table 3: Characters used in the corpus

Corpus	IPA	Cyrilic (original source)
a	a	а
ä	æ	ä
e	e	е
ɛ	ɛ	э
ə	ə	ӛ, ъ
i	i	и, u̇, і
ɨ	ɨ	ы
ı	ı	no cyrillic source
o	o	о
ɔ	ɔ	no cyrillic source
ö	ø	ӧ
u	u	у
ü	y	ӱ
č	t͡ʃ	ч
d	d	д
g	g	г
ɣ	ɣ	ҮҮ
h	h	х, ҳ
j	j	й
k	k	к
l	l	л, ӏ
m	m	м
n	n	н
ŋ	ŋ	ң, нг
p	p	п
q	q, G	қ, k, ҕ
r	r	р
s	s	с
š	ʃʲ	ш
t	t	т
w	v, β	в
z	z	з, ц
ʒ	ʒ	ж
ǯ	d͡ʒ	җ, дж

2. Metadata for the corpus

The metadata of the corpus is provided in *EXMARaLDA Corpus Manager* (*Coma*), here each text has an individual name and is linked to its speaker. The metadata for the texts contain basic metadata as the

place and date of recording as well as information about the researchers involved with the annotation of this text.

2.1. Naming conventions

The communications (texts) are all named the following way: the abbreviation of the speaker is given (first letter of respectively the first name, the patronymic and the last name), followed by the year of recording, a short title and the abbreviation of the genre.
For example:
Name: ChDN_1983_GirlAndIce_flk
Speaker code: ChDN
Year of recording: 1983
Short title: GirlAndIce
Genre: folklore

In the corpus, four genres can be found:
a) Folklore texts (flk)
b) Narrative texts (nar): stories about everyday life or biographies
c) Songs (song)
d) Translations (trans): translations from Russian to Selkup

2.2. Text metadata

Name: The name of the text, see 2.1.
Genre: The genre of the text (flk, nar, song or trans)
Recorded by: The researcher who recorded the text
Date of recording: The date of the recording (if known)
Dialect group: Information about the dialect group (Northern, Central, Southern)
Dialect: Information about the dialects (see Table 1 above)
Subdialect: Information about the subdialects
Transcribed by: The researcher who transcribed the text
Date of transcribing: The date of the transcribing, if known
Date of translation: The date of translation (for trans), if known
Speaker: Abbreviation of the speaker
Original speaker: Abbreviation of the speaker of the original (for trans)
Translation into Russian: The researcher who translated the text. [Here given is the original translation if available. Texts without Russian translation are mostly not translated into Russian.]
Translation into Selkup: The speaker who translated the text (for trans)
Translation into English: The researcher who translated the text
Translation into German: The researcher who translated the text
Translation into Hungarian: The researcher who translated the text. [Here given is the translation in the original source if available. Texts without Hungarian translation are not translated into Hungarian.]
Glossed by: The name of the researcher, who glossed the text

Annotation SeR: The name of the annotator for semantic roles

Annotation SyF: The name of the annotator for syntactic function

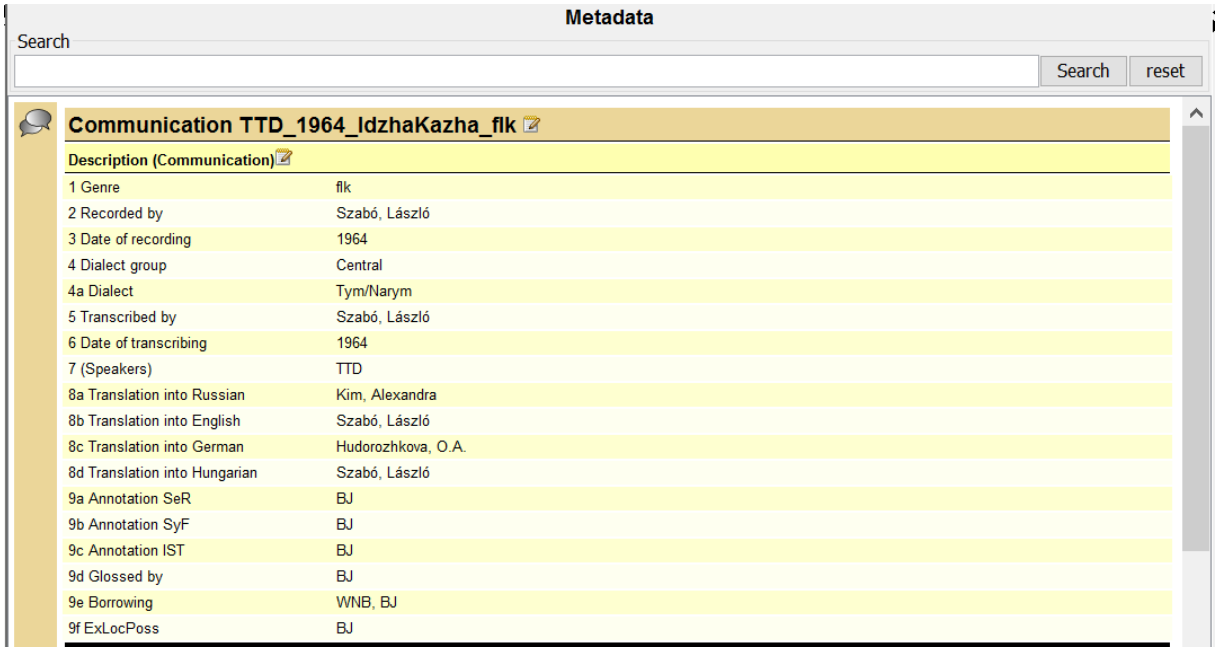
Annotation IST: The name of the annotator for information status

Annotation Borrowing: The name of the annotator of borrowed elements

Annotation ExLocPoss: The name of the annotator of existential/locative/possessive sentences

Annotation CVB: The name of the annotator of converbal constructions

Figure 1: Screenshot of text metadata



Additionally, there are given the following information:

Location:

City: the place where the text has been recorded (if known)

Country: the country where the texts was recorded (normally Russia)

LanguageCode: The language code of the text: sel - Selkup

Setting:

Archive: Information about the archive in which the text can be found, if it is known

Original text: Information about the source of the original text (for trans)

Published in: Information about previous publications. Here are all publications given in which the text was ever published.

Russian source: Information about the Russian source for texts based on Russian sources

Transcriptions: Fhe basic and segmented transcription are added here.

Files: e.g. copies of archive materials or publications. Files (pdfs) are named the following way:

a) **Publication**: Author_Year_ShortTitel_Genre_Pages

The year is referring to the year of publication, the short title is the same as for the text it is belonging to, from-to page numbers are indicated by <->, if the text is on separate pages, the numbers are divided by <_>.

Example: Kuzmina_1967_Mammoth_flk_320_328 (publication of the text KFN_1967_Mammoth_flk)

b) **Archive materials** : here are given materials from two archives

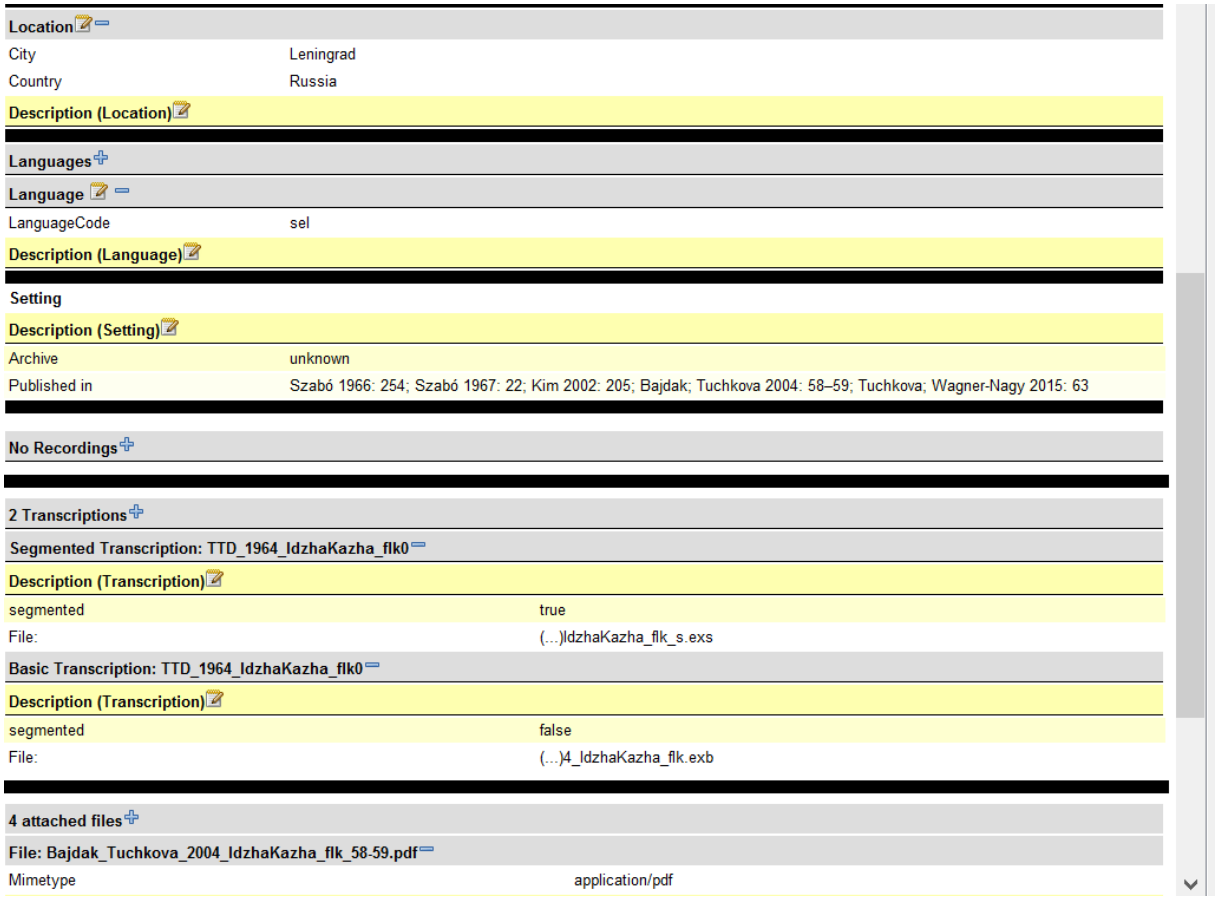
Kuzmina archive in Hamburg: Speaker_Year_Titel_Genre_Vol_Nt_Pages

Dulzon archive in Tomsk: Speaker_Year_Titel_Genre_Vol_Pages

Example (Kuzmina archive): KFN_1967_Mammoth_flk_Vol6_Nt4_75-76

(archive material of the text KFN_1967_Mammoth_flk)

Figure 2: Screenshot of text metadata - 2



2.3. Speaker metadata

Metadata related to the speakers include in all cases biographical information and the linguistic biography of the speaker. Further relevant data will also be included whenever it is available. The following information is available.

The following data is given (if known):

Description of speaker: The name of the speaker.

Given are: Family name, patronymic, given name

Education: Information about the education and occupation of the speaker.

Given are: Education, Higher education, Occupation (if it known)

Informant of: The researcher the speaker worked with.

Ethnicity: Background information about the ethnicity of the speaker and its relatives.

Given are: Ethnicity, Ethnicity of mother, Name of mother, Ethnicity of father, Name of father, Ethnicity of husband/wife, Name of husband/wife, Ethnicity of grandparents

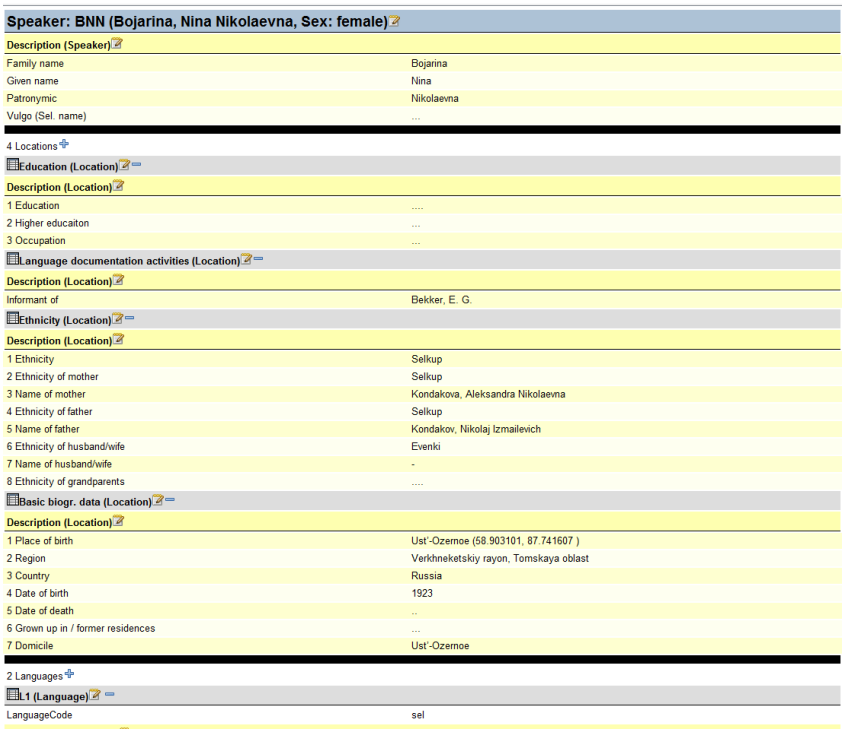
Basic biographical data: Information about the past and current places of residence and basic vital statistics; the domicile is always the current or last (in case of death) place of residence. These information are important to identify the dialect of the speaker.

Given are: Place of birth, Region, Country, Data of birth, Data of death, Grown up in /former residences, Domicile

Languages: The speaker's languages, all speakers speak Selkup (sel) and Russian (rus).

Given are: L1, L2

Figure 3: Screenshot of speaker metadata



Speaker: BNN (Bojarina, Nina Nikolaevna, Sex: female)	
Description (Speaker)	
Family name	Bojarina
Given name	Nina
Patronymic	Nikolaevna
Vulgo (Sel. name)	...
4 Locations	
Education (Location)	
Description (Location)	
1 Education	....
2 Higher educaiton	...
3 Occupation	...
Language documentation activities (Location)	
Description (Location)	
Informant of	Bekker, E. G.
Ethnicity (Location)	
Description (Location)	
1 Ethnicity	Selkup
2 Ethnicity of mother	Selkup
3 Name of mother	Kondakova, Aleksandra Nikolaevna
4 Ethnicity of father	Selkup
5 Name of father	Kondakov, Nikolaj Izmailevich
6 Ethnicity of husband/wife	Evenki
7 Name of husband/wife	-
8 Ethnicity of grandparents	....
Basic biogr. data (Location)	
Description (Location)	
1 Place of birth	Ust'-Ozernoe (58.903101, 87.741607)
2 Region	Verkhneketskiy rayon, Tomskaya oblast
3 Country	Russia
4 Date of birth	1923
5 Date of death	..
6 Grown up in / former residences	...
7 Domicile	Ust'-Ozernoe
2 Languages	
L1 (Language)	
LanguageCode	sel

3. Annotation

The morphological glossing (tier *ge* and *gr*) and the part of speech tagging for each morpheme (tier *mc*) are done in FLEx. The texts are then being converted to EXMARaLDA where the remaining annotations are done with the EXMARaLDA partitur editor.

3.1. Annotation tiers

Each transcription contains at least 12 tiers. The tiers are presented in the following table:

Table 3: Tiers in EXMARaLDA partitur editor

Name of tier	Description	Type	Category
ref	name of the communication	annotation	obligatory
tx	interlinearization	transcription	obligatory
mb	morpheme break	annotation	obligatory
mp	morphophonemes, underlying form	annotation	obligatory

ge	morphological glossing: English	annotation	obligatory
gr	morphological glossing: Russian	annotation	obligatory
mc	part of speech for each morpheme	annotation	obligatory
ps	part of speech for each word	annotation	obligatory
SyF	syntactic functions	annotation	obligatory
SeR	semantic roles	annotation	obligatory
CVB	converbs	annotation	optional
IST	information status	annotation	optional
BOR	borrowing	annotation	optional
ExLocPoss	existential, locative and possessive sentences	annotation	optional
fr	free translation: Russian	annotation	optional
fe	free translation: English	annotation	obligatory
fg	free translation: German	annotation	optional
fh	free translation: Hungarian	annotation	optional
fr-ed	edited free translation: Russian	annotation	optional
fe-ed	edited free translation: English	annotation	optional
fg-ed	edited free translation: German	annotation	optional
nt	notes	annotation	optional

ref - Reference

The tier *ref* gives information about the name of the text and the number of the sentence can also be found here. The tier is of type annotation and obligatory.

ts - Transcription

The tier *ts* contains the sentence as it was presented in the source. If there is sound, it is aligned with this tier. The tier is of type annotation, obligatory and always marked in green.

(1)

ref	ChDN_1983_HerosDaughter_flk.001
ts	Ugon ir wargɨmba madet puʒogɨt matur.

tx - Interlinearization

The tier *tx* is the basis for the morphological glossing, each cell contains one word. The tier is of type transcription, obligatory, linked to the speaker and marked in blue.

(2)

ref	ChDN_1983_HerosDaughter_flk.001					
ts	Ugon ir wargɨmba madet puʒogɨt matur.					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur

mb – Morpheme breaks

The tier *mb* shows a morpheme by morpheme break-up of the words, the morphemes are separated by hyphens; zero morphemes are left out in this tier. The tier is of type annotation and obligatory.

(3)

ref	ChDN_1983_HerosDaughter_flk.001					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur

mp - Morphophonemes

In the tier *mp* the underlying form of all morphs is presented. Selkup is a language with complex morphophonologial processes; hence words can have many allomorphs. Furthermore, Selkup is a non-standardised language with a vast dialectical continuum. Morphs may occur in several written forms. The tier is of type annotation and obligatory.

(4)

ref	ChDN_1983_HerosDaughter_flk.001					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	i:r	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur

gr, ge – Russian and English morpheme glossing

The tiers *gr* and *ge* are for the interlinear morpheme-by-morpheme glossing. The lexical meaning of the stem is given in either Russian or English, the glossing labels are the same for both languages, the Latin script is used here. The labelling follows international standards (mostly the Lepizig Glossing Rules); the additions made to this basic label set can be found in appendix 1.

A dot shows that the two (or more) components semantically belong together and is also used to separate stems in compounds, a dash separates alternative meanings, square brackets indicate non-overt morphemes. Combinations of person and number markings are combined in one gloss without a dot: e.g. 1PL for first person plural.

The unmarked category simple singular is only marked if the word is in nominative, then a complex gloss is used: [SG.NOM], apart from that singular is not marked in the corpus.

Selkup has two types of conjugation: a subjective and an objective one, in the glossing, this is marked by .S or .O following the person of the verb, in the plural the forms collapsed and are hence marked by S/O.

The tiers are of type annotation and obligatory.

(5)

ref	ChDN_1983_HerosDaughter_flk.001					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	iːr	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur
ge	earlier	long.ago	live-PST.REP.[3SG.S]	taiga-GEN	inside-LOC	hero.[SG.NOM]
gr	раньше	давно	жить-PST.REP.[3SG.S]	тайга-GEN	внутренность-LOC	герой.[SG.NOM]

mc – Morpheme class

The tier *mc* is used to indicate the morphological category of each morph – the part of speech of the lexical stem (see table 4) and the category of the suffixes (see table 5).

Table 4: Tags of lexical stems

tag	**description**
adj	adjective
adv	adverb
clit	clitic
conj	conjugation
dem	demonstrative
emph	emphatic pronouns
interj	interjection
interrog	interrogative pro-form
n	noun
num	numeral
ptcp	participle
ptcl	particle
pers	personal pronouns
pp	postposition
pro	pronoun
quant	quantifier
v	verb

Table 5: Tags for inflection

category	**tag**	**description**
nominal	num	number
	case	case
	poss	possessor
verbal	tense	tense
	mood	mood
	pn	personal ending

(6)

ref	ChDN_1983_HerosDaughter_flk.001					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	i:r	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur
ge	earlier	long.ago	live-PST.REP.[3SG.S]	taiga-GEN	inside-LOC	hero.[SG.NOM]
gr	раньше	давно	жить-PST.REP.[3SG.S]	тайга-GEN	внутренность-LOC	герой.[SG.NOM]
mc	adv	adv	v-v:mood-v:pn	n-n:case	n-n:case	n-n:case

ps – Part of speech

In the tier *ps* part of speech for each word form is tagged. The categorisation is syntax oriented.
Some classes are divided into subcategories: nouns are divided in common and proper nouns, particle, auxiliaries and verbs are divided into affirmative and negative categories with a special tag for the negative existential verb.
Cardinal numbers belong to the category QUANT while ordinal numerals are annotated as adjectives.
The pronominal class is split up: interrogative pronominals are tagged as QUE, adverbial pronominals as ADV, demonstrative pronouns as DEM and personal pronouns as PRONP. Personal pronouns in the function of a possessive pronoun are tagged with PRONPOS, pronouns being neither personal nor possessive are not further split and only marked as pronouns.

Table 6: Tags for part of speech

tag	**description**
N	common noun
NPR	proper noun
PRON	pronoun
PRONP	personal pronoun
PRONPOS	possessive pronoun
ADJ	adjective
ADV	adverb
V	affirmative verb
V.NEGEX	negative existential verb
CONJ	conjunction
DEM	demonstrative
INDF	indefinite
INTS	intensifier
INTERJ	interjection
NPI	negative polarity item
ONOM	onomatopoeia

PTCL	affirmative particle
PTCL.NEG	negative particle
PREP	preposition
PP	postposition
PREV	preverb
QUANT	quantifier
QUE	question word

(7)

ref	ChDN_1983_HerosDaughter_flk.001					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	i:r	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur
ge	earlier	long.ago	live-PST.REP.[3SG.S]	taiga-GEN	inside-LOC	hero.[SG.NOM]
gr	раньше	давно	жить-PST.REP.[3SG.S]	тайга-GEN	внутренность-LOC	герой.[SG.NOM]
mc	adv	adv	v-v:mood-v:pn	n-n:case	n-n:case	n-n:case
ps	ADV	ADV	V	N	N	N

fr, fe, fg, fh - Free Translations into Russian, English, German and Hungarian

The tiers *fr*, *fe, fg,* and *fh* are used for free translations into Russian, English, German and Hungarian. The English translation (*fe*) is obligatory for all texts, a Russian translation (*fr*) is provided for most texts (it is marked in red); a German translation (*fg*) as well as Hungarian (*fh*) is added if available.

(8)

ref	ChDN_1983_HerosDaughter_flk.001					
ts	Ugon ir wargɨmba madet puʒogɨt matur.					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	i:r	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur
ge	earlier	long.ago	live-PST.REP.[3SG.S]	taiga-GEN	inside-LOC	hero.[SG.NOM]
gr	раньше	давно	жить-PST.REP.[3SG.S]	тайга-GEN	внутренность-LOC	герой.[SG.NOM]
mc	adv	adv	v-v:mood-v:pn	n-n:case	n-n:case	n-n:case
ps	ADV	ADV	V	N	N	N
fr	Давным - давно жил в чаще леса богатырь.					
fe	Long ago, a hero lived in the forest.					
fg	Vor langer Zeit lebte ein Held im Wald.					

3.2 Annotation of Syntactic Function (SyF)

In the tier *SyF* the core syntactic functions subject, object and predicate are annotated. The tier is of type annotation and is obligatory. The annotation scheme corresponds with the scheme using for annotation of Nganasan corpus (Brykina et al. 2016). The form of annotation is <form.function>

Table 7: Tags for core syntactic functions

tag	description
S	subject
O	object
pred	predicate

3.2.1. Annotation of subject

The subject usually is in nominative, and either a common noun, a proper noun or a pronoun. But also adjectives can function as subject. If the subject is human or an anthropomorphised animal, it is marked human with <h>. As Selkup is a pro-drop language, the subject can be marked solely on the verb; it is therefore useful to annotate also covert subjects. In that case, dropped referent and predicate are annotated in the same cell.

Table 8: Tags for subjects

tag	description
np:S	nominal subject
pro:S	pronominal subject
np.h:S	nominal human subject
pro.h:S	pronominal human subject
0.1:S	dropped first person subject
0.2:S	dropped second person subject
0.3:S	dropped third person subject
0.1.h:S	dropped human first person subject
0.2.h:S	dropped human second person subject
0.3.h:S	dropped human third person subject

A dropped referent is shown in example (9)

(9)

ref	ChDN_1983_MistressOfFire_flk.089			
ts	Hel’d’ po:p pellag’e šogort.			
tx	Hel’d’	po:p	pellag’e	šogort.
mb	hel’d’	po:-p	pel-la-g’e	šogor-t
mp	hel’d’	po-m	pat-lä-k	šoɣor-nti

ge	seven	tree-ACC	put-OPT-1SG.S	stove-ILL
gr	семь	дерево-ACC	положить-OPT-1SG.S	печь-ILL
mc	num	n-n:case	v-v:mood-v:pn	n-n:case
ps	QUANT	N	V	N
SyF		np:O	0.1.h:S v:pred	
fr	Семь поленьев положу в печь.			
fe	I put seven logs in the stove.			

3.2.2. Annotation of object

Direct objects in Selkup are usually marked with accusative but can be found in e.g. nominative as well. Human objects (or humanlike animals) are marked as human with <h>.

Table 9: Tags for objects

tag	**description**
np:O	nominal object
pro:O	pronominal object
np.h:O	nominal human object
pro.h:O	pronominal human object
0.3:O	dropped third person object
0.3.h:O	dropped human third person object

Example (10) shows a direct object in accusative case.

(10)

ref	ChDN_1983_HerosDaughter_flk.008				
ts	Tab wargɨ hurum naj kwatkumbad.				
tx	Tab	wargɨ	hurum	naj	kwatkumbad
mb	tab	wargɨ	huru-m	naj	kwat-ku-mba-d
mp	tab	wargɨ	hurup-m	naj	kwat-ku-mbɨ-tɨ
ge	he.[NOM]	big	wild.animal-ACC	also	kill-ITER-PST.REP-3SG.O
gr	он.[NOM]	большой	зверь-ACC	тоже	убыть-ITER-PST.REP-3SG.O
mc	pers-n:case	adj	n-n:case	ptcl	v-v>v-v:mood-v:pn
ps	PRONP	ADJ	N	PTCL	V
SyF	pro.h:S		np:O		v:pred
fr	Она и на крупных зверей охотилась.				
fe	She also hunted big wild animals.				
fg	Sie jagte auch große wilde Tiere.				

3.2.3. Annotation of predicate

In the predicate position verbs, nouns, adjectives, participles and converbs can occur; nouns, adjectives and participles can be accompanied by copula but it is not necessarily the case. Converbs have to be accompanied either by copula or an auxiliary.

Table 10: Types of predicates

tag	description
v:pred	verbal predicate
n:pred	nominal predicate
adj:pred	attributive predicate
ptcl:pred	particle predicate
cvb:pred	converbal predicate

An example for a nominal predicate, accompanied by copula, is shown by the first part of the sentences in (11):

(11)

ref	ChDN_1983_ItjaStayedAlone_flk.002			
ts	Ad'ade eppɨmba menertɨl qup […]			
tx	ad'ade	eppɨmba	menertɨl	qup
mb	ad'a-de	e-ppɨ-mba	mene-r-tɨl	qup
mp	aǯ'a-tɨ	e:-mbɨ-mbɨ	mene-r-ntɨl'	qum
ge	father.[SG.NOM]-3SG	be-HAB-PST.REP.[3SG.S]	hunt-FRQ-PTCP.PRS	human.being.[SG.NOM]
gr	отец.[SG.NOM]-3SG	быть-HAB-PST.REP.[3SG.S]	охотиться-FRQ-PTCP.PRS	человек.[SG.NOM]
mc	n-n:case-n:poss	v-v>v-v:mood-v:pn	v-v>v-v>ptcp	n-n:case
ps	N	V	ADJ	N
SyF	np.h:S	cop		n:pred
fr	Отец был охотником, в лес ушел, из леса не вернулся.			
fe	The father was a hunter, went to the forest, did not return from the forest.			
fg	Der Vater war Jäger, ging in den Wald, kehrte aus dem Wald nicht zurück.			

3.2.4. Annotation of subordinate clauses

In the annotation of subordinate clauses five types are distinguished: adverbial, conditional, purpose, relative, temporal and complement. Table 14 shows the tagging set for these:

Table 11: Types of subordinate clauses

tag	description
s:adv	adverbial

s:cond	conditional
s:purp	purpose
s:rel	relative
s:temp	temporal
s:compl	complement

In example (12) a purpose clause is shown:

(12)

ref	ChDN_1983_GirlAndIce_flk.002			
ts	Podɨp aramum megu t'umba.			
tx	Podɨp	aramum	megu	t'umba.
mb	pod-ɨ-p	aramu-m	me-gu	t'u-mba
mp	p'ed'-ɨ-m	aramu-m	me-gu	tö:-mbɨ
ge	axe-EP-ACC	icehole-ACC	do-INF	come-PST.REP.[3SG.S]
gr	топор-EP-ACC	пробурь-ACC	делать-INF	прийти-PST.REP.[3SG.S]
mc	n-infl:ins-n:case	n-n:case	v-v:ninf	v-v:mood-v:pn
ps	N	N	V	V
SyF	np:O	s:purp		0.3.h:S v:pred
fr	Прорубь сделать пришел.			
fe	He brought an axe to make an ice hole.			

3.2.5 Converbal constructions (CVB)

To distinguish the different functions of converbal constructions, converbs and their accompanying finite verbs are annotated. Converbs are annotated according to their function and the finite verbs are separated into two groups: predicates of type 1 contain complex movements (a movement expressed by more than one verb) and phase verbs (e.g. begin), predicates of type 2 contain verbs with aspectual meaning and auxiliaries.

Table 12: Types of converbs

tag	**description**
adv	adverbial
s:coord	coordinated
s:sub	subordinated
cvb:pred	converbal predicate

Table 13: Finite Verb form

tag	description
v:pred1	complexe movements, phase verbs
v:pred2	aspectual meaning, auxiliaries

(13)

ref	PMP_1961_BodylessHead_flk.096			
ts	Ne:jqum pelgaliɨk warkl'e übəraŋ.			
tx	ne:jqum	pelgaliɨk	warkl'e	übəraŋ.
mb	ne:-j-qum	pel-galiɨk	wark-l'e	übə-r-a-ŋ
mp	ne:-l-qum	pelə-galiɨk	wargɨ-le	übə-r-ɨ-ŋ
ge	woman-ADJZ-person.[SG.NOM]	friend-CAR	live-CVB	begin-FRQ-EP-3SG.S
gr	женщина-ADJZ-человек.[SG.NOM]	друг-CAR	жить-CVB	начать-FRQ-EP-3SG.S
mc	n-n>adj-n-n:num-case	n-n>adv	v-v>adv	v-v>v-infl:ins-v:pn
ps	N	N	ADV	V
SyF	np.h:S			v:pred
CVB			cvb:pred	v:pred1
fr	Женщина одна стала жить.			
fe	The woman begins to live on her own.			
fg	Die Frau beginnt allein zu leben.			

3.3. Annotation of Semantic Roles (SeR)

The tier SeR is for the annotation of semantic roles. The tier is of type annotation and obligatory. The annotation scheme corresponds with the scheme using for annotation of Nganasan corpus (Brykina et al. 2016). The entry is build using the GRAID principle (Haig – Schnell 2017): <form.animacy:function> with some modifications. For now the following functions are implemented in the corpus:

Table 14: Tags for semantic roles - functions

abbreviation	description	comment
A	agent	Initiator of the action, in control of its action – it is causing and responsible for the happening.
B	beneficiary	Entity for whose benefit the action was performed or who is the beneficiary of the state, action or procedure.
Cau	cause	Entity causing the happening.
Com	comitative	Entity that convoys the participant of the action

E	experiencer	Entity that experiences or feels an action and is not responsible or in control of it - emotion, volition, cognition, perception (verbs like: *live, die, see, love, hate, understand, hear, taste, frighten, wish, want, think, remember, feel*)
G	goal	Location or entity towards something is moving
Ins	instrument	Entity by which the action is performed
L	location	Location in which something is situated
P	patient	Undergoer of the action, is changed by the action.
Path	path	Direction something is moving along
Poss	possessor	Entity who possesses something
R	recipient	Entity who receives something Addressee of a verb of speech (verbs like: *give, buy, bring, carry* and *say, be mad, shout at someone*)
So	source	Place of origin or original owner of something in a transfer
Th	theme	Entity which is moved by some action Entity whose location is specified (e.g. in existential and locative sentences) Entity about which a cognitive, communicative or emotional situation is about.
Time	time	Particular time Interval of time

3.2.1. Tagging of the referent

3.2.2. Form of referent

The form of the referent is marked in the corpus with the following tag set; Selkup is a pro-drop language hence it is advisable to mark also covert referents.

Table 15: Tags for semantic roles – form of referent

abbreviation	description
0	covert
adv	adverb
np	nominal phrase
pp	postposition
pro	pronoun

3.2.3. Properties of referent

In the corpus, all three persons are annotated. Furthermore, it is tagged if the referent is human or non-human: a human referent is marked with <h> while a non-human referent is not marked.

Anthropomorphise animals are also annotated as human, also groups in which at least one participant is human are marked as human.

Table 16: Tags for semantic roles – properties

abbreviation	description
1	first person
2	second person
3	third person
h	human referent

(14)

ref	ChDN_1983_HerosDaughter_flk.001					
ts	Ugon ir wargɨmba madet puʒogɨt matur.					
tx	Ugon	ir	wargɨmba	madet	puʒogɨt	matur
mb	ugon	ir	wargɨ-mba	made-t	puʒo-gɨt	matur
mp	ugon	i:r	wargɨ-mbɨ	maǯ'o-n	puǯo-qɨn	matur
ge	earlier	long.ago	live-PST.REP.[3SG.S]	taiga-GEN	inside-LOC	hero.[SG.NOM]
gr	раньше	давно	жить-PST.REP.[3SG.S]	тайга-GEN	внутренность-LOC	герой.[SG.NOM]
mc	adv	adv	v-v:mood-v:pn	n-n:case	n-n:case	n-n:case
ps	ADV	ADV	V	N	N	N
SeR	adv:Time	adv:Time		np:Poss	np:L	np.h:E
SyF			v:pred			np.h:S
fr	Давным - давно жил в чаще леса богатырь.					
fe	Long ago, a hero lived in the forest.					
fg	Vor langer Zeit lebte ein Held im Wald.					

3.3.6. Annotation of Information Status (IST)

The here used annotation of the information status is a combination of the guidelines taken from Götze et al. (2007) and some elements of the RefLex Scheme (Riester – Baumann 2017, first publications about that in 2014). The scheme was elaborated by Sándor Szeverényi[6].
The three core categories given, accessible and new are kept and further subdivided:
A new referent has not been mentioned in the discourse and is completely new to the hearer and cannot be determined via context.
A referent is given if mentioned in the discourse beforehand, it is marked as active if mentioned in the clause before; else it is marked as inactive.

6 The first version is described in Brykina et al. 2016. The here used version is published also in Brykina 2018.

The accessibility of a referent is further distinguished in four subcategories: situative (it is clear from the situation that the referent is needed and is therefore accessible), inferable (e.g. my hand, the door of a house), aggregational (two already mentioned referents emerge as one, e.g. my mother, my father – my parents), general (knowledge about the world, e.g. *sun*)

Table 17: Tags for Information Status

tag	description	category
giv	given (underspecified)	given
giv-active	given active	
giv-inactive	given inactive	
accs-sit	accessible situative	accessible
accs-inf	accessible inferable	
accs-agg	accessible aggregational	
accs-gen	accessible general	
new	new	new

Also in that annotation line, zero referents are marked by a leading 0.,e.g. 0.accs-inf, and referents appearing in a direct quotation are marked by a following -Q, e.g. accs-inf-Q. These two markers can also be combined as in e.g. 0.accs-inf-Q.
An example sentence with tagged information status can be seen here:

(15)

ref	TTD_1964_Squirrel_nar.004			
ts	Onek tabet qoːbɨmdɛ nʼiŋgleːbe.			
tx	Onek	tabet	qoːbɨmdɛ	nʼiŋgleːbe.
mb	onek	tabe-t	qoːbɨ-m-dɛ	nʼiŋg-leː-be
mp	onek	tapäk-n	kobɨ-m-tɨ	nʼiŋg-la-m
ge	myself.[NOM]	squirrel-GEN	skin-ACC-3SG	take.off-FUT-1SG.O
gr	я.сам.[NOM]	белок-GEN	шкура-ACC-3SG	снять-FUT-1SG.O
mc	emph	n-n:case	n-n:case-n:poss	v-v:tense-v:pn
ps	INTS	N	N	V
SyF	pro.h:S		np:O	v:pred
SeR	pro.h:A	np:Poss	np:P	
IST	giv-active	giv-inactive	accs-inf	
fe	I'll skin the squirrel myself.			

3.3.7. Annotation of Borrowing (BOR)

Borrowing is annotated in several tiers: BOR, BOR-Phon and BOR-Morph, the here used schema is taken from the Nganasan Corpus (Brykina 2018).

In the tier BOR the source language and the lexical type is annotated:

RUS: for Russian

DOL: for Dolgan, etc.

Different types of loanwords are annotaed according to Myers-Scotton (2002, 2006). It is distinguished between cultural borrowings and core borrowings. A further type is grammatical borrowing such as conjunctions (*i* 'and'). Additionally, borrowed discourse markers and modal words are annotaed. Table 16 below shows the annotation tags for the tier BOR.

Table 18: Tags for BOR

Annotation Tag	**Description**
cult	cultural borrowing
core	core borrowing
gram	grammatical borrowing
mod	modal word borrowed
disc	discourse marker borrowed

During the annotation, structural integration (phonetical / phonological and inflectional) of nouns and verbs is taken into consideration. This phenomenon is annotated in tier BOR-Phon.

Table 19: Tags for phonological adaptation strategies (Tier BOR-Phon)

Types of adaptation	**Tag**	**Comment**
deletion	inCdel	initial consonant deletion
	inVdel	initial vowel deletion (aphaeresis)
	medCsdel	medial consonant deletion
	medVdel	medial vowel deletion (syncope)
	finCdel	final consonant deletion
	finVdel	final vowel deletion (apocope)
insertion	inVins	initial vowel insertion
	medVins	medial vowel insertion
	finVins	final vowel insertion
substitution	Csub	consonant substitution
	Vsub	vowel substitution
lenition	lenition	weakening
fortition	fortition	strengthening

In case of verbal borrowings, tier BOR-Morph is used for further annotation, by applying Wohlgemuth's typology (2009). Wohlgemuth differentiates between the following categories:

a) direct insertion (no morphological adaptation),

b) indirect insertion (adaptation by affixation, etc.),

Table 20 shows the annotation tags for the tier BOR-Morph.

Table 20: Tags for morphological adaptation strategies (tier BOR-Morph)

Type	**Tag for strat.**	**Tag for inflextion**	**comment**
direct insertion	dir:	bare	direct insertion without any morphological adaptation
	dir:	infl	direct insertion with further inflection
indirect insertion	indir:	bare	insertion with morphological adaptation without further inflection
	indir:	infl	insertion with morphological adaptation with further inflection
paradigm insertion	parad:	bare	the verb is borrowed with verbal inflexion from the donor language, but is not further inflected
	parad:	infl	the verb is borrowed with verbal inflexion from the donor language and is not further inflected

Example (16) shows a Russian borrowing:

(16)

ref	ChDN_1983_BearCameIntoVillage_nar.004			
ts	Man akoškautɛ ponɛ manʒɛdegak.			
tx	Man	akoškautɛ	ponɛ	manʒɛdegak.
mb	man	akoška-utɛ	ponɛ	manʒɛ-de-ga-k
mp	man	akoška-un	po:ne	manti-nti-ŋi-k
ge	I.[NOM]	window-PROL	outward(s)	look-IPFV2-AOR-1SG.S
gr	я.[NOM]	окно-PROL	наружу	смотреть-IPFV2-AOR-1SG.S
mc	pers-n:case	n-n:case	adv	v-v > v-v:tense-v:pn
ps	PRONP	N	ADV	V
SyF	pro.h:S			v:pred
SeR	pro.h:E	np:Path		
BOR		RUS		
fr	Я из окна на улицу выглянула.			
fe	I looked out of the window.			

3.3.8. Annotation of existential, locative and possessive sentences (ExLocPoss)

In the tier *ExLocPoss* existential, locative and possessive sentences and the order of their components are annotated to make the sentences searchable through their word order. The scheme was elaborated by Josefina Budzisch.
At first the type of sentences (see table 21) is indicated after that the order of the components (see table 22) is marked; possessive suffixes on the theme are only marked in possessive sentences.

Table 21: Type of sentences

tag	description
Ex	existential sentence
Loc	locative sentence
Poss	possessive sentence

Table 22: Components of the sentences

tag	description
Th	theme
Loc	location
Cop	copula
(px)	possessive suffix

(17)

ref	MNS_1984_BrotherSister_flk.017	
ts	Hör ča:ŋgwa.	
tx	hör	ča:ŋgwa
mb	hör	ča:ŋg-wa
mp	hɛr	čaŋki-ŋɨ
ge	snow.[SG.NOM]	NEG.EX-AOR.[3SG.S]
gr	снег.[SG.NOM]	NEG.EX-AOR.[3SG.S]
mc	n-n:num-case	v-v:tense-v:pn
ps	N	V.NEGEX
SyF	np:S	v:pred
SeR	np:Th	
ExLocPoss	Ex: ThCop	
fe	There is no snow.	

4. Text sources

Bajdak et al. 2010: Bajdak, Alexandra – Nadezhda Fedotova – Natalya Maksimova 2010. Селькупские тексты. In: Filchenko, Andrey (ed.). *Annotated Folklore Prose Texts of Ob-Yenissey Language Area*. Tomsk: Veter, 133–184.

Bajdak; Maksimova 2002: Bajdak, Alexandra – Natalya Maksimova 2002. *Дидактизация оригинального текста: селькупский язык*. Tomsk.

Bajdak; Maksimova 2009: Bajdak, Alexandra – Natalya Maksimova 2009. Селькупский текст. In: The Department of Siberian Indigenous Languages (ed.). *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Tomsk: Veter, 45–90.

Bajdak; Maksimova 2012: Bajdak, Alexandra – Natalya Maksimova 2012. Селькупские тексты. In: The Department of Siberian Indigenous Languages (ed.). *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Tomsk: Veter, 72–100.

Bajdak; Maksimova 2013: Bajdak, Alexandra – Natalya Maksimova 2013. Селькупские тексты. In: The Department of Siberian Indigenous Languages (ed.). *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Tomsk: Veter, 153–201.

Bajdak; Maksimova 2015: Bajdak, Alexandra – Natalya Maksimova 2015. Селькупские тексты. In: The Department of Siberian Indigenous Languages (ed.). *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Tomsk: Veter, 108–149.

Bajdak; Tuchkova 2004: Bajdak, Alexandra – Natalya Tuchkova 2004. Эписоды "Эпоса об Итте" в чулылькупском диалектном ареале. In: *Коренные народы Сибири: проблемы историографии, истории, этнографии, лингвистики.* Tomsk, 51–64.

Bekker 1978: Bekker, Erika 1978. *Категория падежа в селькупском языке*. Tomsk: Издательство Томского университета.

Bekker 1980: Bekker, Erika 1980. Селькупские тексты. In: *Сказки народов Сибирского Севера*. Tomsk: Издательство Томского университета, 55–71.

Bykonia et al. 1996: Bykonia, Valentina – Alexandra Kim – Shimon Kuper – Natalya Maksimova – I. Ilyashenko 1996. *Сказки нарымских селькупов.* Tomsk: NTL.

Castrén 1855: Castrén, Matthias. *Nordische Reisen und Forschungen* 8. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Dulzon 1966a: Dulzon, Andreas 1966. *Кетские сказки*. Tomsk: Издательство Томского университета.

Dulzon 1966b: Dulzon, Andreas 1966. Селькупские сказки. In: *Языки и топонимия Сибири. Том 1*. Tomsk, 96–158.

Grigorovskij 1879: Grigorovskij, Nikolaj 1879: *Азбука сюссогой гулани*. Kazan: Типография Императорского Университета.

Grigorovskij 1883: Grigorovskij, Nikolaj 1883. Итя (Сказка обских самоедов). *Томские губернские ведомости* 24.

Janhunen 1975: Janhunen, Juha 1975. *Etä-sukukielet. Lapp, Volga-Finnic, Permian-Finnic, Ob-Ugrian, Samoyed.* Helsinki: Suomalaisen Kirjallisuuden Seura.

Katz 1975: Katz, Hartmut 1975. Materialien vom Tym. *Selcupica* 1. München.

Katz 1988: Katz, Hartmut 1988. Die Märchen in Grigorovskis Azbuka. Transkription, Übersetzung, Kommentar. *Selcupica* 4. München.

Kim 2002: Kim, Alexandra 2002. К истории изучения южноселькупского фольклора: тымские материалы Л. Сабо. In: *Образование в сибири: актуальные проблемы истории и современность*. Tomsk, 204–206.

Korobeynikova 2014: Korobeynikova, Irina 2014. Сказки и рассказы селькупки ирины. Tomsk: Veter.

Künnap 1992: Künnap, Ago 1992. Selkupin Iidjä-satu Tšajan murteela vuodelta 1913. *Fenno-ugristica* 18, 141–147.

Kuzmina 1967: Kuzmina, Angelina 1967. Диалектологические материалы по секупскому языку. In: *Исследования по языку и фольклору. Вып. 2.* Novosibirsk, 267–329.

Kuznecova et al. 1993: Kuznecova, A. – Olga Kazakevich – L. Ioffe – Eugen Helimski 1993: *Очерки по селькупски языку. Тазовский диалект.* Том 2. Moscow.

Lehtisalo 1960: Lehtisalo, Toivo 1960. Samojedische Sprachmaterialien. *Mémoires de la Société Finno-Ougrienne* 122. Helsinki.

Morev et al. 1981: Morev, Yuri – Nina Denning – Valentina Bykonia – G. Mikhenina – Alexandra Kim 1981. Самодийские тексты. Селькупские тексты. In: *Сказки народов Сибирского севера*. Tomsk: Издательство Томского университета, 122–143.

Skazki chainskih selkupov 2001: Сказки чаинских селькупов 2001. In: Земля чаинская: сборник научно-популярных очерков 100-летию села Подгорного. Tomsk, 115–126.

Szabó 1966: Szabó, Lászlo 1966. Szelkup szövegek szójegyzékkel (Tymi nyelvjárás). *Nyelvtudományi Közlemények* 68, 266–277.

Szabó 1967: Szabó, Lászlo 1967. Selkup texts with phonetic introduction and vocabulary. (*Uralic and Altaic Series* 75). Indiana: Bloomington.

Tuchkova 2002: Tuchkova, Natalya 2002. К вопросу о педагогических традициях селькупов. In: *Образование в сибири: актуальные проблемы истории и современность*. Tomsk, 195–204.

Tuchkova; Helimski 2010: Tuchkova, Natalya – Eugen Helimski 2010. *О материалах А. И. Кузьминой по селькупскому языку*. Hamburg: Institut für Finnougristik/Uralistik.

Tuchkova; Wagner-Nagy 2015: Tuchkova, Natalya – Beáta Wagner-Nagy 2015. *„sēľďe nūn qȫdi īťťe...“. «Семи богов мудростью обладающий Итте...» Тексты с героем Итя в селькупском фольклоре Часть 1. Итя-тексты.* Tomsk: Томский государственный педагогический университет.

References

Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor and Wagner-Nagy, Beáta. 2016. “Nganasan Spoken Language Corpus (NSLC).” Archived in Hamburger Zentrum für Sprachkorpora. Version 0.1. Publication date 2016-12-23. http://hdl.handle.net/11022/0000-0001-B36C-C.

Götze, Michael et al. 2007: Information structure, in Dipper, S., Götze, M. and S. Skopeteas (eds.): *Information Structure in Cross-Linguistic Corpora.* Interdisciplinary Studies on Information Structure 07 (2007): 147-187, Available online at http://edoc.hu-berlin.de/oa/reports/reQ5PntJcwYs/PDF/23TFAo8H6FW2.pdf [Accessed: 5.2.2013]

Haig, Geoffrey and Stefan Schnell 2014: *Annotations using GRAID (Grammatical relations and animacy in discourse)*, Introduction and guidelines for annotators, Version 7.0, Available online at https://www.uni-bamberg.de/fileadmin/aspra/Publications/GRAID7.0_manual.pdf [Accessed: 03.07.2016]

Helimski, Eugen 1998. Selkup. In: Abondolo, Daniel (ed.). *The Uralic Languages*. London – New York: Routledge, 548–579.

Glushkov, Sergej – Alexandra Bajdak – Natalya Maksimova 2013. Диалекты селькупского языка. In: Tuchkova, Natalya et al. (eds.). *Селькупы. Очерки традиционной культуры и селькупского языка*. Tomsk, 49–63.

Riester, Arndt – Stefan Baumann 2014: *RefLex Scheme – Annotation Guidelines.* http://www.ims.uni-stuttgart.de/institut/mitarbeiter/arndt/doc/RefLex-guidelines-01aug-2014.pdf

Russian Census 2010. *Всероссийская перепись населения* 2010. Том 4. Национальный состав и владение. http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm

Appendix

Tags for morpheme classes

1	first person	**DUR**	durative	**PROP**	proprietive
2	second person	**EMPH**	emphatic clitic	**PST**	past
3	third person	**EP**	epenthetic vowel	**PTCP**	participle
ABL	ablative	**EX**	existential verb	**REP**	reportative
ABST	abstract noun from verb	**FRQ**	frequentative	**RES**	resultative
ACC	accusative	**FUT**	future	**RFL**	reflexive
ACT	nomen actionis	**GEN**	gentive	**S**	subjective conjugation
ADJZ	adjektivizer	**HAB**	habituative	**SG**	singular
ADV	adverb	**ILL**	illative	**SING**	singulative
ALL	allative	**IMP**	imperative	**SUB**	subjunctive mood
AN	animate	**INCH**	inchoative	**TR**	transitive
AOR	aorist	**INDEF**	indefinitive	**TRL**	translative
ATTEN	attenuative	**INF**	infinitive	**US**	usative
CAP	captative	**INFER**	inferential	**VBLZ**	verbalizer
CAR	caritive	**INSTR**	instrumental		
CAUS	causative	**INT.PF**	intensive perfective		
COM	comitative	**IPFV**	imperfective suffix		
COND	conditional	**ITER**	iterative		
CONJ	conjunctive	**LOC**	locative		
COR	coordinative	**MULTS**	multisubjective		
CRC	connective reciproc	**NEG**	negative marker		
CVB	converb	**O**	objective conjugation		
DAT	dative	**OBL**	oblique case		
DES	desiderative	**OPT**	optative		
DETR	detransitive	**ORD**	ordinal numeral forming suffix		
DIM	diminutive	**PL**	plural		
DRV	derivational suffix	**PREV**	preverb		
DU	dual	**PROL**	prolative		