Beáta Wagner-Nagy – Sándor Szeverényi – Valentin Gusev

User's Guide to Nganasan Spoken Language Corpus

Working Papers in Corpus Linguistics
and Digital Technologies:
Analyses and Methodology
Vol. 1.

Beáta Wagner-Nagy Sándor Szeverényi Valentin Gusev

User's Guide to

Nganasan Spoken Language Corpus

Working Papers in Corpus Linguistics and Digital Technologies:

Analyses and Methodology

Vol. 1.

Szeged – Hamburg

2018

Working Papers in Corpus Linguistics and Digital Technologies: Analyses and methodology

Vol. 1

WPCL issues do not appear according to strict schedule.
© Copyrights of articles remain with the authors.
Vol. 1 (2018)

Editor-in-chief
Kristin Bührig (Universität Hamburg)

Series editors
Katalin Sipőcz (University of Szeged)
Sándor Szeverényi (University of Szeged)
Beáta Wagner-Nagy (Universität Hamburg)

Published by
University of Szeged, Department of Finno-Ugric Studies
Egyetem utca 2. 6722 Szeged
Universität Hamburg, Zentrum für Sprachkorpora
Max-Brauer-Allee 60 22765 Hamburg

Published 2018

ISBN 978-963-306-598-3 (pdf)

Contents

Tables and figures

1. Introduction

1.1 Objectives

The Nganasan Spoken Language Corpus (NSLC) was created as part of the project *Corpus based grammatical studies on Nganasan* at the Institute of Finno-Ugric/Uralic Studies of Universität Hamburg. The project was supported by the *Deutsche Forschungsgemeinschaft* under grant number WA3153/2-1 between 2014 and 2017. The primary goal of the project was to generate a digital, searchable corpus of spoken Nganasan. The language material to be integrated, glossed and annotated was collected by several researchers (see below in Section 1.4.2) and is available in audio format, most of it in video format as well.

On the one hand, the corpus contains materials collected by different researchers during earlier fieldworks, which were supported by several foundations (see Table 1 below). On the other hand, archive materials from Tomsk and St. Petersburg as well as published materials are used. The oldest text is from the beginning of the 20th century, which was collected by Prokofjev and published in 1933. In the second half of the 20th century, Natalya M. Tereshchenko has worked intensively on Nganasan. She collected several Nganasan texts, but they have remained unpublished up to this day. The materials are preserved in the Archive of the Institute for Linguistic Studies of Russian Academy of Sciences in St. Petersburg. We published the materials with the kind permission of the Institute. Bbetween 1968 and 1972, several scholars from the Duľson School from Tomsk carried out fieldwork among the Nganasans and collected grammatical data, word lists, and texts. Part of these texts were published in the series *Skazki Narodov Sibirskogo Severa* (SNSS: Tales of the Peoples of North Siberia, 1976, 1980, 1981) and in the series *Sbornik foľklornyh i bytovyh tekstov obsko-enisejskogo jazykovogo areala* (Annotated folklore and everyday texts from the Ob-Yenissei area, 2009, 2010, 2012, 2015), but a significant number of these materials still have yet to be analyzed and published. We published the texts with the kind permission of the Department of Siberian Indigenous Languages of Tomsk State Pedagogical University. The list of sources of the published texts is given in section 3 below/The sources of the published texts are listed in section 3 below.

Table 1 *List of supporters*

Year of fieldwork	**Supported by**
1992, 1994	Russian State University for Humanities
1994	University of Szeged
1996	Soros Foundation
2000	Russian Foundation for Humanities
2003-2005	Russian Academy of Sciences
2006-2011	National Science Foundation (USA)
2008	Hungarian Scientific Research Found (OTKA) Phonogrammarchive of Austrian Academy of Sciences FWF Der Wissenschaftsfond (Austria)
2016, 2017	DFG (German Research Grant)

1.2. The language

Nganasan is an agglutinative language displaying a series of inflectional features. It belongs to the Samoyedic branch of the Uralic language family, its closest relatives within the North Samoyedic group being Nenets and Enets. Today people speaking Nganasan solely live in some villages in the Taymyr Autonomous District, which is part of the Krasnoyarsk Krai of the Russian Federation. Nganasan is highly endangered: according to Russian census data of 2010[1] out of the total population of 807 people only about 125 speak Nganasan and there are no speakers under the age of 40, or, if there are, they can be considered semi-speakers[2] at most.

Commonly, two dialects of Nganasan are distinguished, Avam and Vadeyev, however these idioms do not differ significantly from each other (cf. Helimski 1998: 480–482).

Language codes are: ISO-639-3 code: **nio;** Glottolog code: **ngan1291**

1.3 Archiving

The corpus' transcription data as well as the metadata are stored in the EXMARaLDA format. The data curation, archiving, and publication are performed by the *Hamburg Centre for Language Corpora* (HZSK). The corpus is freely available under HZSK-ACA ("academic") license to registered HZSK users[3].

1.4 Citation

There are two versions of the corpus. The first version contains 55 glossed and annotated transcriptions from 15 different speakers. There are 4,331 utterances with 27,485 tokens in the corpus. This version is to be cited as follows:

> 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. Available online at http://hdl.handle.net/11022/0000-0001-B36C-C.

The second version of the corpus contains 176 glossed and partly annotated transcriptions from 33 different speakers. There are 21,723 utterances with 142,455 tokens (35,131 types) in the corpus. This version is to be cited as follows:

> Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor and Wagner-Nagy, Beáta. 2018. *Nganasan Spoken Language Corpus (NSLC)*. Archived in Hamburger Zentrum für Sprachkorpora. Version 0.2. Publication date 2018-06-12. Available online at http://hdl.handle.net/11022/0000-0007-C6F2-8

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

[1] http://www.gks.ru/free_doc/new_site/perepis2010/perepis_itogi1612.htm

[2] As defined in Grinevald – Bert 2011

[3] https://corpora.uni-hamburg.de/hzsk/de/korpusanfragen-lizenzen, last access: 07.02.2018.

1.5 Project members and involved researchers

The research was carried out at the Institute for Finno-Ugric/Uralic Studies (IFUU) of the Universität Hamburg (UHH). The project homepage can be visited at: https://www.slm.uni-hamburg.de/nganslc.html

The following researchers were involved in the compilation of the corpus:

Project members:

Prof. Beáta Wagner-Nagy	project leader	
Dr. Brykina, Maria	responsible for glossing	October 2014 – September 2015
Dr. Gusev, Valentin	responsible for glossing	October 2015 – August 2017
Dr. Szeverényi, Sándor	responsible for annotation	November 2014 – August 2017
Budzisch, Josefina	responsible for the alignment of transcriptions	January 2015 -– September 2015
Danilova, Victoria	responsible for the English translation	October 2015 – August 2017
Jawinsky, Gerrit	responsible for the alignment of transcriptions and for the English translation	October 2015 – August 2017
Jark, Florian	responsible for the English translation	January 2017 – August 2017

The abbreviations for contributors of the corpus are as follows:

BJ:	Budzisch, Josefina	LJL:	Lambert, Jean-Luc
BM:	Brykina, Maria	MT:	Mikola, Tibor
DM:	Daniel, Michael	SF:	Sobanski, Florian
DV:	Danilova, Victoria	SR:	Sutter, Regula
GV:	Gusev, Valentin	SzS:	Szeverényi, Sándor
HE:	Helimski, Eugene	VZS:	Várnai, Zsuzsa
JG:	Jawinski, Gerrit	WNB:	Wagner-Nagy, Beáta
JF:	Jark, Florian	ZR:	Zayzon, Réka

The technical infrastructure was provided by the *Hamburger Zentrum für Sprachkorpora* (HZSK). Hanna Hedeland coordinated the technical support, Heidemarie Sambale and Anne Ferger helped with the data curation.

2. The corpus

2.1 Basic information

The *Nganasan Spoken Language Corpus* is a multilingual parallel corpus, which contains the same text samples in at least three languages:

Original text: in Nganasan

Translation: mostly into Russian and English, sometimes also into German and Hungarian.

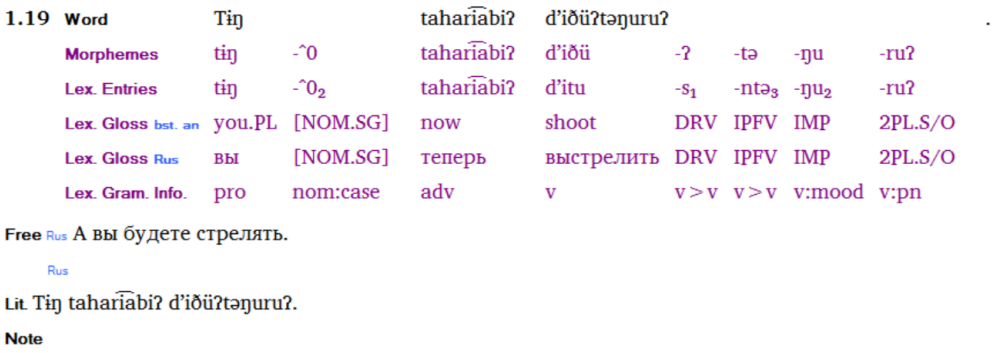
The language of the metadata is English.

The main annotation languages are English and Russian. Morpheme glosses in English and Russian are provided for lexical items; labels for grammatical morphemes are identical in the respective tiers and

are based on abbreviations of English terms, largely following the Leipzig Glossing Rules (see tiers **ge, gr**).

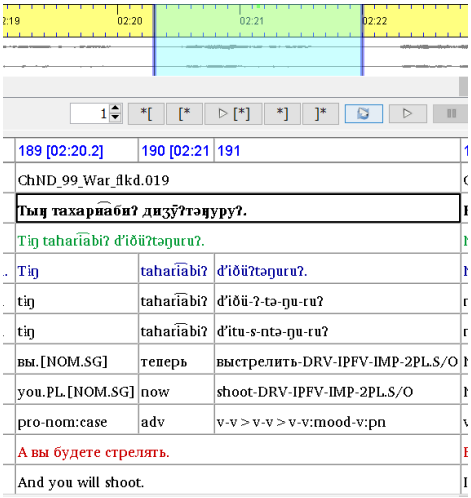
For the morphological glossing and the part of speech tagging we used the *Toolbox* software in our previous work and *SIL Fieldworks Language Explorer* (FLEx) for the text glossing in this project. The following screen shot shows glossing in FLEx.

Figure 1 *Text glossed in Flex*



The language materials processed with *Toolbox* and *FLEx* were imported into *EXMARaLDA*. The glossed transcripts were synchronised with the audio/video data with the help of *EXMARaLDA* *Partitur-Editor*[4]. The screen-shot below illustrates how this works. The text is the same as in the previous screen shot.

Figure 2 *Converted transcription in Partitur-Editor*



The texts are aligned sentence-by-sentence with the sound file (if available). The data of the corpus are managed by *EXMARaLDA Corpus Manager* (Coma).[5]

The correspondences between speakers and communications (texts) are provided in the corpus manager (Coma).

[4] http://www.exmaralda.org/partitureditor.html

[5] http://www.exmaralda.org/tool/corpus-manager-coma/

2.2 Corpus statistics

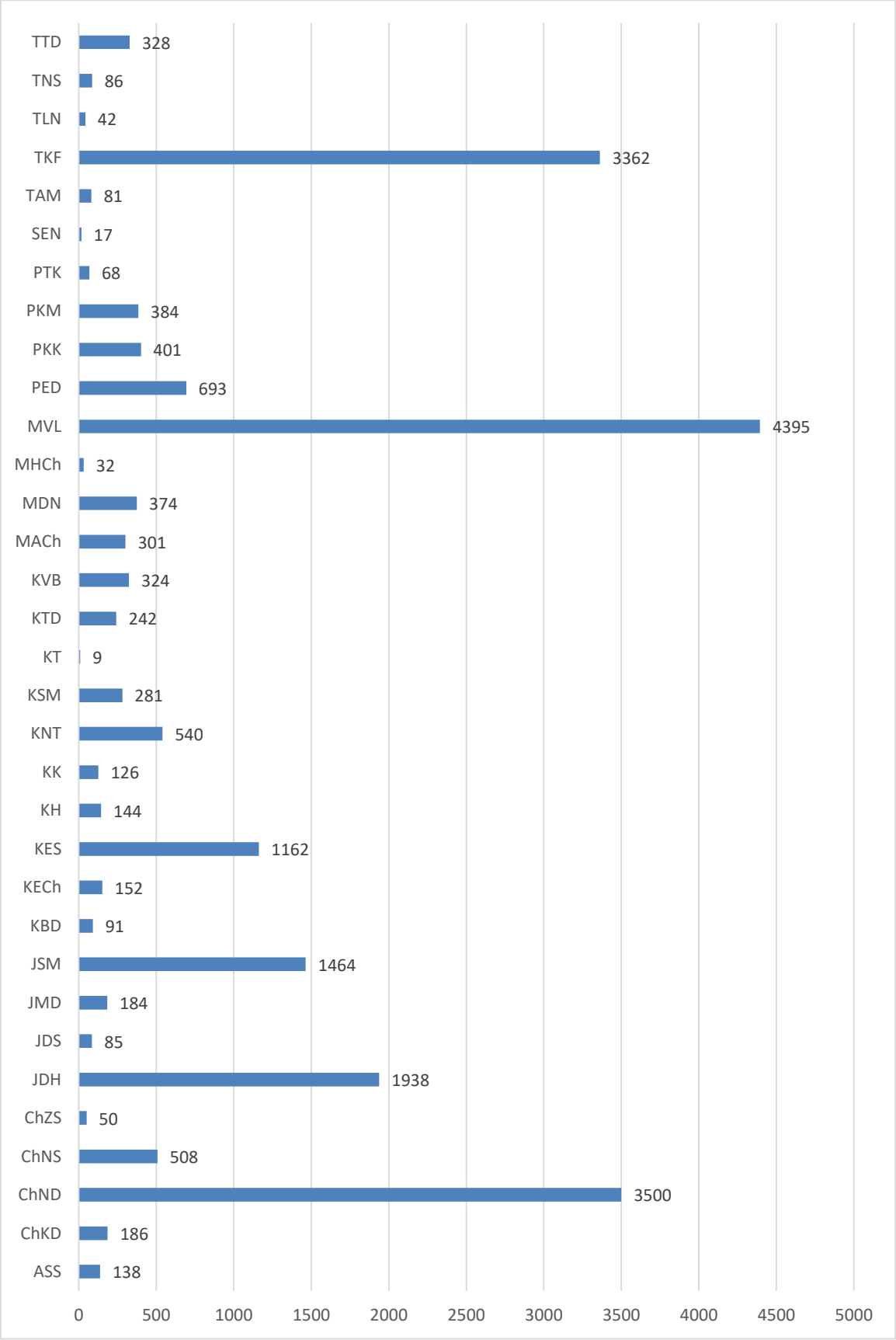
The corpus contains 176 transcriptions from 33 speakers. Additionally, metadata are provided for 34 speakers. There are 21,723 utterances with 142,455 tokens 35,131 types in the corpus. The oldest text is from 1933. The majority of the texts were recorded in the 2000s. The following table shows the distributions per year and speaker.

Table 2 *Distribution of the transcription*

Year of recording	Number of texts	Number of speakers
1933	1	1
1965	2	1
1968	3	1
1971	16	2
1972	1	1
1990	1	1
1992	1	1
1993	3	3
1994	3	2
1996	5	2
1997	12	4
1999	15	4
2000	8	1
2003	3	2
2004	10	4
2005	2	1
2006	21	8
2008	53	12
2009	4	2
2016	6	2
unknown	6	3

The most utterances are recorded from the speaker with the abbreviation MVL (4,395 utterances). The following diagram (Figure 8) shows the distribution of the utterances according to the speakers. The most texts are recorded from the speakers with the abbreviation ChND (26 communications), MVL (17) and TKF (16).

Figure 3. *Speakers and utterances*



2.3 Orthography in the corpus

Most of the transcriptions have a tier <st> (source transcription). This tier represents the text in Cyrillic writing system. Instead of the Cyrillic script, a Latin-based phonological transcription is used in the transcription and annotations tiers. Vowel length is marked by doubling the vowel letter: *n'aa* [ŋaː] 'Nganasan'. Palatalization is marked with the apostrophe symbol <'>. In the corpus the Charis font[6] is used. The following characters are used in the transcriptions:

Latin (phonological)	**IPA**	**Cyrillic**	**Meaning**
ə: *kəntə*	ə: *kəntə*	ə: *кəнтə*	'sledge'
a: *aba*	a: *aba*	а: *аба*	'sister'
o: *kou*	o: *kou*	о: *коу*	'ear'
e: *s'ejmɨ*	e: *sʲejmɨ*	е: *сеймы*	'eye'
i: *n'iliďi*	i: *ɲiliʲi*	и: *нилыди*	'to live'
i͡a: *ŋami͡aj*	i͡a: *ŋami͡aj*	иа: *ŋамиай*	'other'
ɨ: *ď'esɨ*	ɨ: *ʲesɨ*	ы: *десы*	'father'
u: *ŋua*	u: *ŋua*	у: *ŋуа*	'door'
u͡a: *kobtu͡a*	u͡a: *kobtu͡a*	уа: *кобтуа*	'girl, maid'
ü: *s'üar*; *anəl'ikü*	y: *sʲyar*, *anəlʲiky*	ю: *сюар* ӱ: *анəликӱ*	'friend' 'big'
b: *basa*	b: *basa*	б: *баса*	'iron'
g: *maagəl'it'ə*	g: *maːgəlʲicə*	г: *маагəличе*	'nothing'
d: *d'indüa*	d: *ʲindya*	д: *диндӱа*	'horse'
d': *d'esɨ*	ʲ: *ʲesɨ*	д: *десы*	'father'
ð: *l'iðiŋkə*	ð: *lʲiðiŋkə*	ʒ: *лизиӈкə*	'sable'
j: *kojkə*	j: *kojkə*	й: *койкə*	'idol'
k: *kou*	k: *kou*	к: *коу*	'ear'
l: *latəə*	l: *latəː*	л: *латəə*	'bone'
l': *l'ümü*	lʲ: *lʲymy*	л: *люмӱ*	'running'
m: *mənə*	m: *mənə*	м: *мəнə*	'I'
n: *nagür*	n: *nagyr*	н: *нагӱр*	'three'
n': *n'inɨ*	ɲ: *ɲinɨ*	н: нины	'older brother'
ŋ: *ŋarka*	ŋ: *ŋarka*	ӈ: *ӈарка*	'bear'
r: *sanirsa*	r: *sanirsa*	р: *санірса*	'to play'
s: *saü*	s: *say*	с: *саӱ*	'noise'
s': *s'iba*	sʲ: *sʲiba*	с: *сиба*	'servant'
h: *hu͡aa*	h: *hu͡aa*	х: *хуаа*	'tree'
t': *t'etua*	c: *cetua*	ч: *четуа*	'very'
ʔ: *l'üəʔsa*	ʔ: *lʲyəʔsa*	ʔ: *люоʔса*	'Russian'

[6] http://software.sil.org/charis/

2.4 Sound files

A great number of the transcriptions have sound recordings as well. The first recording is from 1965. The majority of the sound files are aligned with the text. The naming convention of the recordings is the same as the naming convention of the texts (see Section 3.1. below).

2.5 The use of the corpus

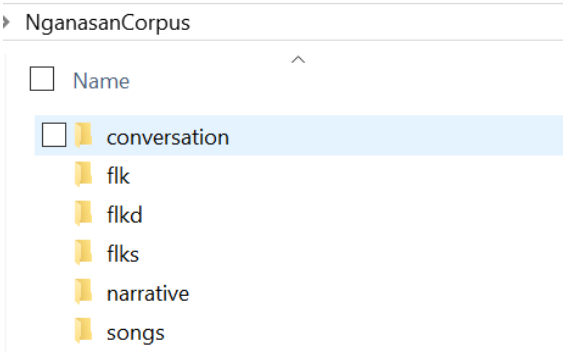
The corpus cannot be used online at the moment, but you can download the files and build the folder structure described in the following section.

2.5.1 Folder structure

The NSCL corpus is located in the folder *NganasanCorpus* that has the following subfolders.
Folders with the transcriptions and sound files are organized by genre:

- *conversation* (conversations)
- *flk* (folklore texts without specified folklore genre)
- *flkd* (folklore texts, dyürymy)
- *flks* (folklore texts, syteby)
- *narrative* (narrative texts)
- *songs* (texts of songs)

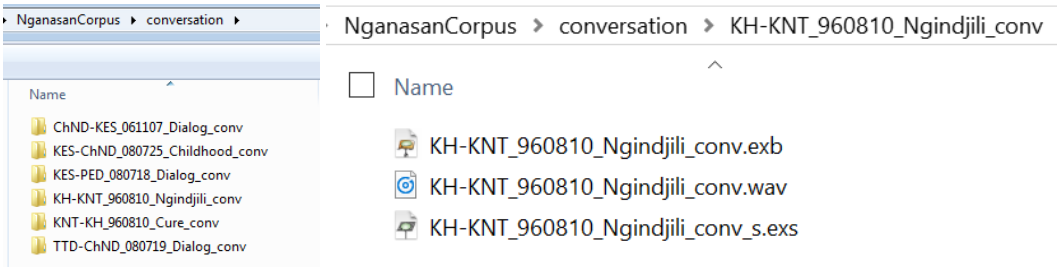
Figure 4 *Folder Structure*



Each of these genre folders contains one further subfolder per each communication, with a name identical to the communication name (See Figure 4). Each communication folder contains several files with the same filename identical to the communication name, and different extensions according to the file type (see Figure 5 below). The files are:

- the annotated transcript in EXMARaLDA format (*.exb)
- the segmented transcript (*.exb)
- optionally the glossed text as exported from FLEx, in FLEXTEXT format (*.flextext)
- optionally the scanned manuscript pages from the Dulson or Tereshchenko archive, in PDF (*.pdf) (only for texts from these archive)
- optionally scanned pages from publication
- if available sound file in WAV (*.wav) and in MP3 (only for texts with audio source)

Figure 5 *Subfolders and the content of the subfolder*



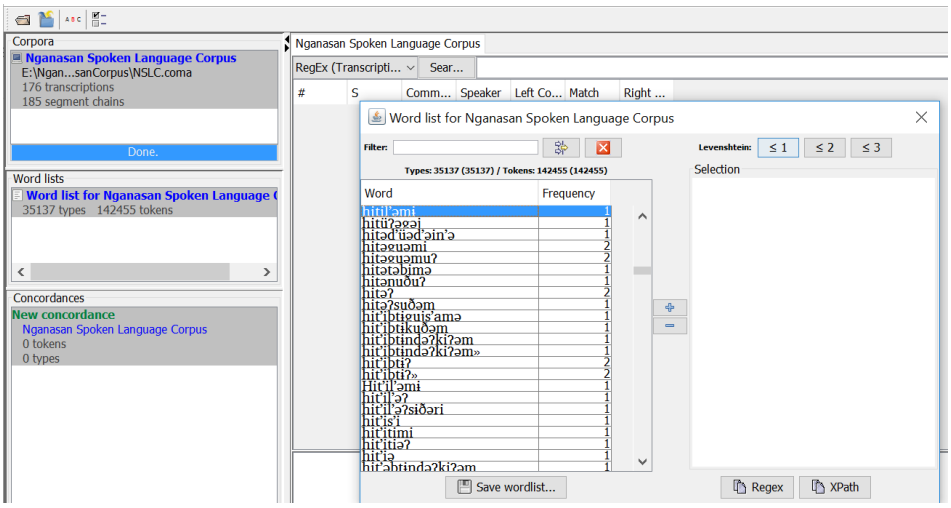
The NganasanCorpus folder contains some individual files:

- NSLC.coma (main metadata file, see Section 3 below)
- annotation-panel_nganasan.xml (annotation panel for use in EXMARaLDA Partitur Editor)
- NganCorpFormat.exf

2.5.2 Searching in the corpus

The corpus can be searched e.g. with EXAKT[7] (EXMARaLDA Analysis- and Concordance Tool). For searching, you must open the *NSLC.coma* file in EXAKT (File > Open corpus). With EXAKT you can generate a word list (see Figure 5) or you can create a concordance.

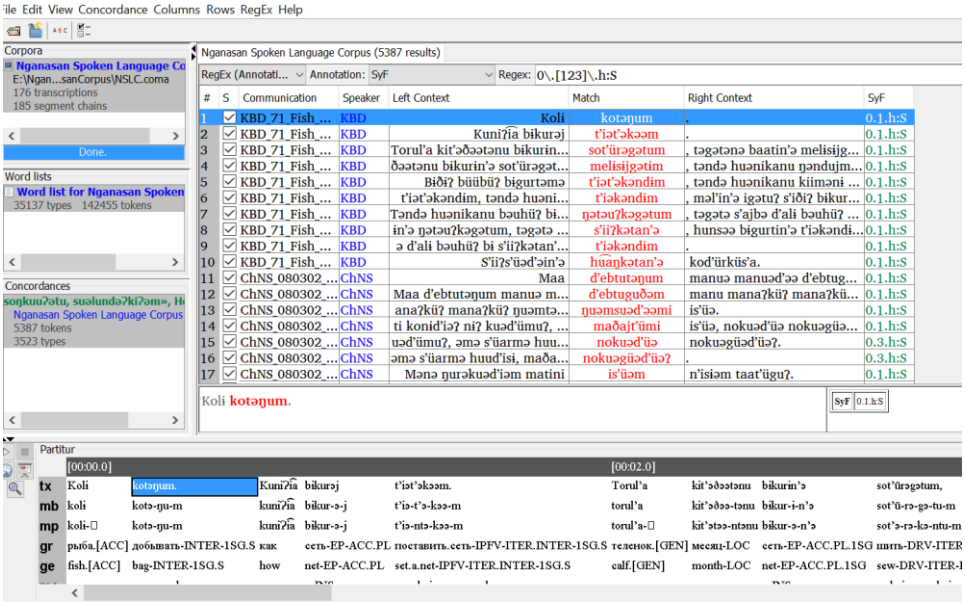
Figure 6 *Word list*



For creating a concordance, you can use all tiers. (For tiers see Section 4 below). The following example (Figure 6) shows the search for covert pronominal subject in the tier Sy(ntactic) F(unction). For a detailed description for using of EXAKT, see Schmidt (2017).

[7] http://exmaralda.org/en/exakt-en/

Figure 7 *Concordance*



3. Metadata for the corpus

The corpus is provided with metadata in the *EXMARaLDA Corpus Manager* (Coma) format and stored in the Coma XML format (file extension ".coma"). Réka Zayzon made the design for the corpus manager. The metadata include information on the Nganasan consultants/informants as well as on the recorded speech events (communication). One file contains the metadata for the whole corpus.

3.1 Naming conventions

3.1.1 Names of communications

All names of communications begin with the abbreviations of the speaker(s). If it is known, the exact date of recording is added: year, month, and day. If it is not known, only the year is marked or there are no date at all. This latter case is marked by NN. The name contains the short description of the text and provides the abbreviations of the genre.

Name: ChNS_080818_School_nar

Speaker: ChNS

Date of recording: 2008.08.18.

Title: School

Description: ChNS tells about her school years

Genre: narrative

The titles are built as follows: ChNS_20080818_School_nar *or* KBD_1971_Fish_nar

3.1.2 Names of speakers

The speaker codes are derived from the speaker's full names in the order "Family name — Given name — Patronymic". Most commonly, a code is thus composed of three initial capital letters, e.g. "ChND" stands for Chunanchar, Nina Demnimeevna. If the patronymic is not noted, only initials of the family name and of the first name are used, e.g. "MDa" for Mojbo, Dintade.

3.2 Communication metadata

Metadata on the communicative event include interaction type, location and time, and language used. The following pieces of information are given:

Name: The name given to the communication (ChNS_20080818_School_nar)

Description

- **Genre:** The genre of the communication
- **Recorded by**: Abbreviations or the full name of the person by whom the communication was recorded
- **Date of recording:** Here only the year of the recording is given.
- **Dialect:** Here we give information about the dialect of the speakers.
  Avam stands for the Avam dialect. Currently there are no communications from the Vadeyev dialect in the corpus.
- **Subdialect:** We distinguish here between the variant spoken in Ust'-Avam (marked as Avam) and the variant spoken in Volochanka (marked as Volochanka).
- **Transcribed by:** The name of the person(s) who provided the transcription. This person is mostly a linguist in cooperation with a Nganasan speaker. Here only the initials of the linguist are given. For abbreviations of the researcher, see section 1.5. above.
- **Date of transcribing:** The year (if it is known) of the transcription
- **Speaker(s):** The name(s) of the speaker(s).
- **Translation into Russian:** The name of the consultant with whose help the text was translated. The speaker metadata contains the metadata for this consultant as well. The language of this translation in not the standard Russian variant, but the language which was used by the Nganasan consultant.
- **Translation into Russian/edited:** if the communication has an edited, standard Russian translation, the name of the researcher, who translated the text, is given here. It must be noted that only very few communications are provided with standard Russian translation.
- **Translation into English:** the name(s) of the translator is/are given here. The majority of the communications is translated into English, but not all.
- **Translation into Hungarian:** the names of the translators are given here
- **Translation into German:** the names of the translators are given here

Location

City: The place of the recording. It must be mentioned here that it is not necessarily identical with the place of speakers' domicile. The geographic coordinates are also given.

Country: it is Russian in all cases

Languages

Language code: The language code of the communication.

nio - Nganasan

rus - Russian

Setting: In this section, we give some pieces of information about publications or archive materials as well as motives

Archive Volume: If the text is from the Tomsk or the Tereshchenko Archive, we give here the volume number. For texts the manuscript of which is not preserved in an archive, we give the notation *not in archive*.

Motive: for some folklore text we provide also the motive. Currently the following identifications are used: Berezina, Djajku, Kehy Luu, Hibula, Ojoloko, Reindeer/Lemming (they are all named after the hero of the text), War with Nenets, Dog Lake.

Published in: If the text was published, we indicate the source of the publication.

Variant: Some stories are recorded from two or more speakers or from the same speaker twice. In this case we refer to the variants. For example, ChND told the story with the hero Berezina twice: in 2008 [ChND_080729_Berizenaa_flks] and in 2006 [ChND-KES_061107_Berizenaa_flkd]

Recording

If the sound file(s) is/are available, it is/they are linked here to the corpus manager.

Transcriptions

The basic transcription (exb) and the segmented transcription (exs) are linked here to the corpus.

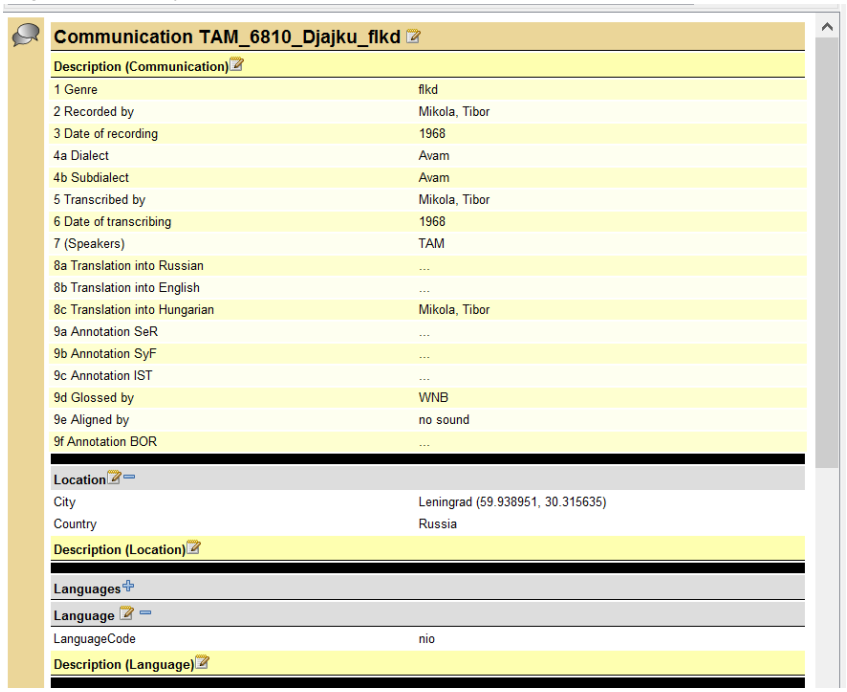
Attached file(s)

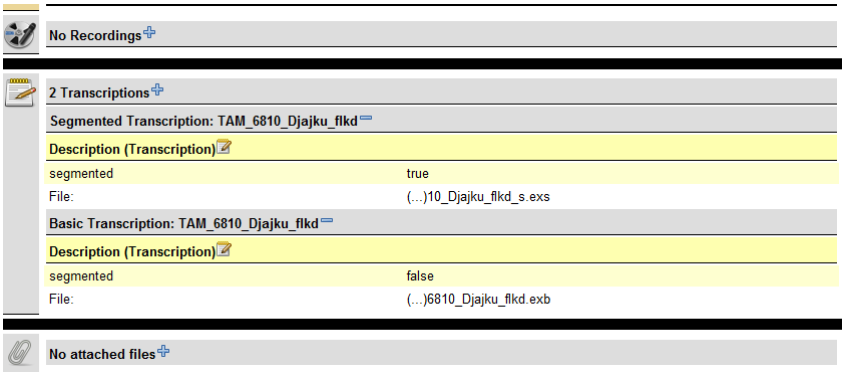
If additional files, e.g. copies of archive materials or copies of publication are available, they are in this section connected to the corpus.

"…" marks the field, which has not yet been translated or annotated

Missing pieces of information are marked with "unknown". The following screen shot shows the data for the communication TAM_6810_Djajku_flkd recorded by Tibor Mikola in 1968.

Figure 8 *Data for a communication*





3.3 Speaker metadata

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

Description of speaker

Family name

Patronymic

Surname

Vulgo (Nganasan name)

Education

Education

Higher education

Occupation

Informant of: In this section we give information with whom the consultant worked. See the abbreviation for the linguists in section 1.5. above.

Ethnicity: here information about the ethnicity of the given person and about the ethnicity of the family members is listed/given/provided

Ethnicity

Ethnicity of mother

Name of mother

Ethnicity of father

Name of father

Ethnicity of husband/wife

Name of husband/wife

Ethnicity of grandparents: this information is mostly unknown to the younger generation, thus this field is only seldom filled.

Basic biographical data

Place of birth

Region: It is always Taimyr Peninsula

Country: Russia

Data of birth

Data of death

Grown up in /former residences

Domicile: it is always the current domicile

Languages: here we give the language codes (*nio* notes Nganasan, *rus* Russian)

L1: The first language and the dialect are given here.

L2: it is mostly Russian, but in some cases, Nganasan is the second language

4. Annotation of the transcriptions

It is necessary for the data to be morphologically glossed and tagged for parts of speech (in the transcription the line mc) with Toolbox or FLEx, also further annotated, and processed with EXMARaLDA Partitur Editor. For this, it has been necessary to create a software tool for converting the data from Toolbox and Flex to EXMARaLDA, which keeps tokenization in accordance with EXMARaLDA format. This work was done by Alexandre Arkhipov.

4.1 Annotation tiers

Each communication contains at least 15 tiers. Some communications have more tiers. For the short description of the tiers see Table 1 below showing the tiers in EXMARaLDA.

Table 2 *Tiers in Nganasan Corpus*

TIERS	Comments	Type	Category
ref	Name of the communication	annotation	obligatory
so	Source text: origin	annotation	optional
st	Source texts: normally in Cyrillic transliteration	annotation	optional
ts	Transcription (what is heard)	annotation	obligatory
tx	Tier for interlinearization	transcription	obligatory
mb	Morpheme break	annotation	obligatory
mp	Morphophonemes, underlying forms	annotation	obligatory
gr	Morphological annotation: Russian gloss of each morpheme	annotation	obligatory
ge	Morphological annotation: English gloss of each morpheme	annotation	obligatory
mc	Part of speech of each morpheme	annotation	obligatory
ps	Part of speech of each word	annotation	obligatory
SeR	Annotation of semantic roles	annotation	obligatory
SyF	Annotation of syntactic function	annotation	obligatory
IST	Annotation of information status	annotation	optional
BOR	Annotation for borrowing	annotation	optional
BOR-Phon	Annotation for phonological adaptation	annotation	optional
BOR-Morph	Annotation for morphological adaptation	annotation	optional
CS	Annotation of code switching	annotation	optional
fr	Russian free translation	annotation	optional
fr_ed	edited Russian translation	annotation	optional
fe	English free translation	annotation	optional
fg	German free translation	annotation	optional
fh	Hungarian free translation	annotation	optional
nt	Notes on the text unit	annotation	optional

EXMARaLDA offers the possibility to insert a practically infinite number of tiers for annotation, which makes multiple-tier annotation easily doable.

The annotation scheme for syntactic functions and thematic roles applied to Nganasan has been developed on the basis of the GRAID: Grammatical Relations and Animacy in Discourse (cf. Haig & Schnell 2011, 2014), but it also differs from it. During the annotation, we take into account three factors: we annotate thematic roles and syntactic functions, and we provide information on their referents. For the annotation categories for syntactic function see section 4.3., for semantic roles see section 4.2.

4.1.1 Refrence (ref)

In the tier *ref*(erence) the name of the transcription and the number of the sentence are noted. It is an obligatory tier and it has a type of description.

4.1.2 Source origin (so)

This tier is optional. It is the tier for the original transcription according to the manuscript of the source as the chart below shows. This tier has a type description.

(1)

ref	PKK_71_OneTent_flkd.001	PKK_71_OneTent_flkd.002
so	**ӈуой ма.**	**м'ѐ́ðич'и тéйч'у.**
st	**Ӈуэиˀ маˀ.**	**Мыəзичи тəичу.**

4.1.3 Source texts (st)

This tier is optional if it is available, it is the tier for Cyrillic transliteration as the chart below shows. This tier has a type description.

(2)

ref	ChND_061023_School_nar.005
st	**Урубаакинюˀ нягэиˀ мелызысыəмыˀ, кӱзиахӱˀ серəніакəныˀ.**

4.1.4 Transcription (ts)

The tier *ts* contains the sentences that are heard. The tier *ts* is the tier that contains the original Nganasan text aligned with the audio/video files. This tier is always marked with green and has a type description.

(3)

ref	ChND_061023_School_nar.005
st	**Урубаакинюˀ нягэиˀ мелызысыəмыˀ, кӱзиахӱˀ серəніакəныˀ.**
ts	Urubaakin'üʔ n'aagəiʔ meliðɨsɨəmiʔ, küð͡iahüʔ s'erəniakəniʔ.

4.1.5 Text (tx): tier for interlinearization

The next tier is the tier *tx* (text) for interlinearization which provides the basis for glossing in Flex or Toolbox. This is the tier of type transcription. All other tiers containing additional analytic information about the transcription have a type a(nnotation). Every communication has one and only one tier of the type transcription for each speaker. This tier is always marked with blue. This tier is obligatory linked to the speaker.

(4)

ref	ChND_061023_School_nar.005				
st	**Урубаакинюˀ нягэиˀ мелызысыəмыˀ, кӱзиахӱˀ серəніакəныˀ.**				
ts	Urubaakin'üʔ n'aagəiʔ meliðɨsɨəmiʔ, küð͡iahüʔ s'erəniakəniʔ.				
tx	Urubaakin'üʔ	n'aagəiʔ	meliðɨsɨəmiʔ,	küð͡iahüʔ	s'erəniakəniʔ.

4.1.6 Morpheme breaks (mb)

In this tier the segmentable morphemes (incl. clitics) are separated by hyphens, as the following chart shows. The tier has a type a(nnotation). All words appear in separate cells. Zero morphemes are left out in the morpheme breaks tier.

(5)

ref	ChND_061023_School_nar.005			
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?
mb	urubaaki-n'ü?	n'aagəi-?	meliði-siə-mi?	küði͡a-hü?

4.1.7 Morphophonemes (mp)

In the morphophonemes tier the underlying forms of morphs (stems, suffixes, clitics, etc.) can be found. It is important because the Nganasan morphophonology is very complex. One morph can have up to 24 allomorphs. Zero morphemes, such as genitive or accusative suffix, are marked with -^{c} in the tier of morphophonemes, as the following sentence shows. This tier has a type annotation.

(6)

ChND_061023_School_nar.009			
tx	ŋanuə	ŋua	kad'ani.
mp	ŋanuə	ŋua-^{c}	kad'a-nu

The relation between the morpheme-breaks tier and morphophonemes tier is shown the following chart.

(7)

ref	ChND_061023_School_nar.005			
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?
mb	urubaaki-n'ü?	n'aagəi-?	meliði-siə-mi?	küði͡a-hü?
mp	urubaakə-n'ü?	n'aakəə-?	meliti-suə-mu?	kütia-hü?

4.1.8 Russian and English morpheme glosses (*gr* and *ge*)

These tiers serve for morphological analysis (interlinear morpheme-by-morpheme glossing). Here we identify all morphemes of all word forms. The lexical meaning of the stems is given in Russian and in English, written in Cyrillic and Latin alphabet. The labelling of grammatical morphemes follows the international standards, mostly according to the Leipzig Glossing Rules[8] with the addition of several items. For the list of abbreviations see Appendix 1 and Appendix 2. The glossing labels of grammatical morphemes are the same in the Russian and English tiers and are written in Latin alphabet. Semantic components of the same morpheme are separated by dot (x.y). Alternative meanings are separated by dash (X/Y). Non-overt morphemes are given in brackets, as [X]. Person and number combinations are marked as complex glosses without a dot, as 1SG, 1PL etc.

In complex grammatical glosses the following order is used:

Nominal Inflection

case-number: ACC.PL for accusative plural

case-number-person-number: ACC.PL.1PL for accusative plural possessive inflection for first person plural

[8] https://www.eva.mpg.de/lingua/pdf/LGR08.02.05.pdf

Verbal Inflection

tense/mood-person-number: PST-1PL.S/O for past, first person plural in subjective or objective conjugation.

Some unmarked categories (as e.g. nominative singular) are indicated it in the glosses in brackets, as [NOM.SG]. The following charts illustrate these tiers.

(8)

ref	ChND_061023_School_nar.005			
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?
mb	urubaaki-n'ü?	n'aagəi-ʔ	meliði-siə-mi?	küði͡a-hü?
mp	urubaakə-n'ü?	n'aakəə-ʔ	meliti-suə-mu?	küti͡a-hü?
gr	рубашка-ACC.PL.1PL	хороший-ADV	сделать-PST-1PL.S/O	встать-COND
ge	shirt-ACC.PL.1PL	good-ADV	make-PST-1PL.S/O	get.up-COND

(9)

ChND_061023_School_nar.009			
tx	ŋanuə	ŋua	kad'ani.
mb	ŋanuə	ŋua	kad'a-ni
mp	ŋanuə-^{c}	ŋua-^{c}	kad'a-nu
gr	настоящий.[GEN]	дверь.[GEN]	около-LOC.ADV
ge	real.[GEN]	door.[GEN]	near-LOC.ADV

4.1.9 Morpheme class (mc)

In the tier *mc*, the morphological category of all elements is given, thus the part of speech of the (lexical) stems (v, adj, adv, pp etc.) and the derivational and inflectional category of suffixes. The categorization of the derivational suffixes is not fully detailed. The following table show the tags of different inflectional categories and the tags for the lexical stems:

Table 3 *Tags of lexical stems*

Morpheme class	**Comment**
adj	adjective
adv	adverb
conj	conjunction
exl	exclamative
indef	indefinite
n	noun
num	numeral
pp	postposition
pr	pronoun
propr	proper name
ptcl	particle
v	verb

The following chart shows the relation between this tier and the tier *ge* (and *gr*).

(10)

ref	ChND_061023_School_nar.005			
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?
mb	urubaaki-n'ü?	n'aagəi-?	meliði-siə-mi?	küði͡a-hü?
ge	shirt-ACC.PL.1PL	good-ADV	make-PST-1PL.S/O	get.up-COND
mc	n-n.case-poss	adj-n.deriv.adv	v-v.tense-v.pn	v-v.nf

4.1.10 Part of speech (ps)

The tier *part of speech* (ps) specifies the grammatical categories of each word form. The categorization of part of speech is syntax-oriented. Some categories are divided into subclasses as noun into N 'common nouns' and NPR 'proper nouns' or auxiliary into AUX 'auxiliary verb' and AUX.NEG 'negative auxiliary'. We do not distinguish between intransitive, transitive, and ditransitive verbs, but we differentiate between common verbs, existential verbs and verbs with negative meaning like *ď'erusa* 'not know'.

Not only true particles are annotated as particles, but interjections, too. Particles with negative meaning like *ďaŋku* 'not' are annotated by using a special marker (PTCL.NEG).

Participles fall into the same category as adjectives; they are annotated as adjective, because from a syntactic point of view the participles behave as adjectives.

Numerals are treated as members of different categories:

– cardinal numerals are annotated as quantifiers (QUANT)

– ordinal numerals are annotated as adjectives (because their inflectional features are the same) (ADJ)

– adverbial numerals are annotated as adverbs (ADV)

The traditional pronominal categories are treated as members of different categories:

– interrogative pronouns are annotated as question word (QUE)

– demonstrative pronouns are annotated as demonstratives (DEM)

– pronominal adverbs are annotated as adverbs (ADV)

Personal pronouns in the function of possessive pronoun are annotated differently, they get/receive the label PRONPOS. The following chart shows the tier part of speech.

(11)

ref	ChND_061023_School_nar.005				
st	**Урубаакиню^{ʔ} няагəи^{ʔ} мелыʒысыəмы^{ʔ}, кӳʒиахӳ^{ʔ} серəниакəны^{ʔ}.**				
ts	Urubaakin'ü? n'aagəi? meliðisiəmi?, küði͡ahü? s'erəniakəni?.				
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?	s'erəni͡akəni?.
mb	urubaaki-n'ü?	n'aagəi-?	meliði-siə-mi?	küði͡a-hü?	s'erə-ni͡akə-ni?
mc	n-n.case-poss	adj-n.case	v-v.tense-v.pn	v-v.nf	v-v.nf-n.poss
ps	N	ADV	V	ADV	N

The following table shows the categories.

Table 4 *Tags for part of speech*

Tags	**Comments**
ADJ	adjective
ADV	adverb
AUX	auxiliary verb
AUX.NEG	negative auxiliary
CONJ	conjunction
COP	copula
DEM	demonstratives and determiners
INDF	indefinites
INTS	intensifiers
N	noun
NPI	negative polarity items
NPR	proper noun
PP	postposition
PRONP	personal pronoun
PRONPOS	possessive pronoun
PTCL	particle, interjection
PTCL.NEG	negation particle
QUE	question words
QUANT	quantifiers and numerals
V	verb
V.EX	existential verb
V.NEG	verb with negative semantic like *ď'eruď'a* 'not know'
V.QUE	question verb

4.1.11 Free translation into Russian and English (fr, fe)

In this tier, the free translations into Russian and into English are given. The tier *fr* is obligatory, while the tier *fe* is not obligatory, but it is mostly present. In very few texts, we give a German translation, too. The following charts illustrate the tiers.

(12)

ref	ChND_061023_School_nar.005				
st	**Урубаакиню^{ʔ} няагэи^{ʔ} мелыҙысыәмы^{ʔ}, кўҙиахў^{ʔ} серәниакәны^{ʔ}.**				
ts	Urubaakin'ü? n'aagəi? meliðɨsiəmɨ?, küði͡ahü? s'erəniakəni?.				
tx	Urubaakin'ü?	n'aagəi?	meliðɨsiəmɨ?,	küði͡ahü?	s'erəni͡akəni?.
mb	urubaaki-n'ü?	n'aagəi-?	meliði-siə-mɨ?	küði͡a-hü?	s'erə-ni͡akə-nɨ?
mc	n-n.case-poss	adj-n.case	v-v.tense-v.pn	v-v.nf	v-v.nf-n.poss
ps	N	ADV	V	ADV	N
fr	Одежду привели в порядок чтобы завтра одеть.				
fe	We put our clothes out in order to put them on tomorrow.				

(13)

ref	ChND_061023_School_nar.009		
tx	ŋanuə	ŋua	kad'aniɨ.
mb	ŋanuə	ŋua	kad'a-niɨ
gr	настоящий.[GEN]	дверь.[GEN]	около-LOC.ADV
ge	real.[GEN]	door.[GEN]	near-LOC.ADV
fr	возле двери		
fe	by the door		

4.1.12 Edited Russian translation (fr_ed)

There is a standard Russian translation provided for some communications. The following chart shows the differences between the two Russian translations.

(14)

ref	KNT-KH_960810_Cure_conv.006			
tx	Tə	lun'd'imtɨ	səbud'ü?ə	ərəkərəməniɨ.
mb	tə	lun'd'i-mtɨ	səbud'-ü-?ə	ərəkərə-məniɨ.
ge	well	spleen-ACC.SG.3SG	pull.out-EP-PF.[3SG.S]	beautiful-ADV
fr	он вытащил селезёнку красиво			
fr_ed	Селезенку вытащил красиво.			

4.1.13 Free translation for Hungarian and German

In some cases, there are translations available for Hungarian or for German. In these cases, we provide these translations.

4.2 Annotation of semantic roles (SeR)

The annotation of semantic (thematic) roles is given in tier labelled with *SeR*. This tier has a type annotation. The entry is built according to GRAID principle (Haig&Schnell 2014): <form.animacy:function> with some modifications.

To this day, no unified list of semantic roles exists despite the fact that argument structure and the assignment of semantic (thematic) roles are hot topics in the fields of semantics and syntax these days (cf. Dowty 1989, 1991, Grimshaw 1990, etc.). In our system of annotations, we have taken into account the thematic roles used in GRAID (Haig & Schnell 2014), but additionally we annotate some other semantic roles too, such as the Recipient (R), Benefactor (B) and Experiencer (E).

We differentiate between roles of Patient (P) and Theme (Th). Differentiating between a Recipient and a Goal is not unproblematic. One of the criteria for doing so is that if the verb expresses an actual or mental transfer, the argument at the other end is a Recipient. Naturally, the argument of verbs expressing a mental transfer is not a real recipient but only a recipient-like argument; this is not separately annotated. Several other semantic (thematic) roles such as Undergoer have not been included in the list at present, but can be included at a later stage. For annotating the semantic (thematic) roles, the following glosses are used:[9]

[9] We rely on Gawron (2007) for defining thematic roles.

Table 5 *Tags for semantic roles - functions*

Abbreviations	Comment
A	Agent: initiator (with volition) of the action, the participant is causing the action or it is responsible for something happening. It can be animate or inanimate.
B	Beneficent: entity for whose benefit the action was performed
Com	Comitative: animate entity that convoys a participant of the action (co-agent)
Cau	Cause: entity that causes an event
E	Experiencer: entity that experiences the action, it does not have control of an action or state -- emotion, volition, cognition, perception (verbs like: *see, love, hate, understand, hear, taste, frighten, wish, want, think, remember, feel*)
G	Goal: location or entity in the direction of which something moves
Ins	Instrument: medium by which the action or event is performed
L	Location: locative argument of verb, place in which something is situated (states location)
P	Patient: undergoer of the action arguments of verbs such as *die, sneeze, fall*
Path	Path
Poss	Possessor: indicate the possessor (who owns something)
R	Recipient: - animate recipient of transfer - addressee of verb of speech
So(urce)	- place of origin - original owner in a transfer
Th	Theme: - entity which is moved by some action (change of location or possession: object of *give*; subject of *walk*) - entity whose location is specified - an entity affected by the action (*Susan is in the house*.)
Time	a point or an interval of time

4.2.1 Form of referent

In the corpus, the form of the referent is annotated. Not all possible factors of such forms are provided, but noun phrase and pronominal referents are differentiated. In annotating the thematic role Locative, it also plays a role if the referent is adverbial, postpositional or nominal. The categories which are used in specifying the form of the referent are listed in Table 6 below. Given that Nganasan is a pro-drop language, it is useful to mark whether the referent in question is expressed overtly in the sentence or not. Accordingly, for the form of referential expressions the following glosses are used:

Table 6 *Tags for semantic roles – form of referent*

Abbreviations	Comment
pro	free pronoun
np	noun phrase
0	covert referent
adv	adverb
pp	postposition
v	verb

4.2.2 Properties of referent

Person is included in the inherent properties of the referent. We annotate all three persons. Semantically the referent can be a human or non-human. Human referents are annotated with the symbol <h>, while the non-human referents are bare. In Nganasan, the feature [±human] probably does not play a special role as far as thematic relations are concerned, however, we decided to include it in the annotation list anyway. The properties of the referent are linked to the form categories with the symbol <.>. We do not distinguish between a human and an anthropomorphized referent, the latter is annotated as human referent. Table 7 summarizes the properties of referents.

Table 7 *Tags for semantic roles – properties*

Abbreviations	Comment
1	first person
2	second person
3	third person
h	human referent

The following sentences show how the semantic role is annotated.

(15)

ref	ChND_061023_School_nar.005				
st	**Урубаакиню^{ʔ} няагэи^{ʔ} мелыӡысыэмы^{ʔ}, кӱӡиахӱ^{ʔ} серэниакэны^{ʔ}.**				
ts	Urubaakin'ü? n'aagəi? meliðisiəmi?, küði͡ahü? s'erəniakəni?.				
tx	Urubaakin'ü?	n'aagəi?	meliðisiəmi?,	küði͡ahü?	s'erəni͡akəni?.
ps	N	ADV	V	ADV	N
SeR	np:P		0.1.h:A	n:Time	
fr	Одежду привели в порядок чтобы завтра одеть.				
fe	We put our clothes out in order to put them on tomorrow.				

(16)

ref	ChND_061101_TwoTents_flkd.015	
tx	Maagüəðəmtə	ŋəði?əŋ?
SyF	pro:O	0.2.h:S v:pred
fe	'Did you find for you something?'	

4.3 Annotation of syntactic function (SyF)

In annotating grammatical relations, we focus only on the major syntactic functions as Subject and Object, as well as on the predicate, which can be nominal or verbal, making this distinction necessary to differentiate as well. The form of annotation is <form:function>.

Table 8 *Tags for core syntactic functions*

Main function	Tag
predicate	pred
subject	S
direct object	O

4.3.1 Annotation of the predicate

The predicate is the most important part of a Nganasan sentence. There are five kinds of predicates. A *simple verbal predicate* consists of one verb as predicate. A *complex verbal predicate* consists of more words, one of them is a verb in connegative form or in converb form, and the other one is an auxiliary. By annotating the Nganasan corpus we do not distinguish between simple and complex verbal predicates, both are marked as verbal predicates. A noun, an adjective and a particle can all appear in the predicate position. In present, they stand without any copula. In the non-present, they are complemented by a copula verb.

The first element of the abbreviation of a predicate refers to the type of predicate (nominal, particular or verbal), whereas the second one to the role played in the sentence – that is, in this case, that the given verb functions as a predicate. The verbal predicate is annotated as <v:pred>. As mentioned above, there are, however, predicates that go together with a copula which carries certain grammatical functions such as a modal or tense marker, as the sentence below shows. Here the actual predicate is annotated as a predicate, while the element bearing the tense marker receives the label copula, cf. sentence (17).

(17)

ref	KES_061020_MyLife_nar.011			
st	Бəхиа̂ исюə, нəŋхəмəны нилыдиəмы?.			
ts	Bəhia̯ is'uə, nəŋhəməni n'iliɨd'iəmi?.			
tx	Bəhia̯	is'üə,	nəŋhəməni	n'iliɨd'iəmi?.
mb	bəhia̯	i-s'üə	nəŋhə-məni	n'ili-d'iə-mi?
mp	bəhia̯	ij-suə	nəŋhə-mənu	n'ili-suə-mu?
gr	плохой.[3.SG]	быть-PST.[3SG.S]	плохой-ADV	жить-PST-1PL.S/O
ge	bad.[3.SG]	be-PST.[3SG.S]	bad-ADV	live-PST-1PL.S/O
mc	adj-v.pn	v-v.tense-v.pn	adj-adj.deriv.adv	v-v.tense-v.pn
ps	ADJ	V	ADV	V
SyF	adj:pred	cop		0.1.h:S v:pred
fr	Плохо было, плохо жили.			
fe	It was bad, we lived badly.			

As the sentence above demonstrates, an adjectival element can play the role of a predicate. However, as mentioned above, Nganasan has the characteristic that even nouns and particles can occur in this position (see Table 9 below). In addition to purely verbal predicates, auxiliaries are also differentiated. In sentences that contain a structure with an auxiliary, both elements of the predicate receive the annotation <v:pred>. The annotation scheme referring to the predicates is summarized in Table 9 below.

Table 9 *Tags for predicates*

Types of predicates	**Tag**
verbal predicate	v:pred
nominal predicate	n:pred
attributive predicate	adj:pred
particle predicate	ptcl:pred
pronominal predicate	pro:pred

4.3.2 Annotation of the subject

First, we shall clarify what is a subject in Nganasan. The most typical subject is a noun in nominative case or a pronoun in nominative, but the subject can be expressed by an adjective or demonstrative, too. Subjects expressed by demonstratives are annotated as pronominal subjects. If they refer to a human, they are marked as human.

If the subject is sentential like in the sentence *I was surprised that Kurumaku is hunting*, we annotated it as a subordinate clause. Equivalent/Similarily to semantic roles, human referents are annotated with the symbol <h>, while non-human referents are bare.

There are cases in which one element has to be assigned two syntactic roles (e.g. subject and predicate) during annotation. This can happen, for instance, when a pro-drop phenomenon occurs during which the pronominal subject is not expressed overtly. In this case, two syntactic functions have to be annotated and the given cell is annotated for both functions. Deleted referents are marked with the symbol <0>. The properties of the referent are linked to the form categories with the symbol <.>. The annotation scheme referring to the subject is summarized in Table 10 below.

Table 10 *Tags for subjects*

Tag	**Form of the referent**	**Inherent properties of the referent**	**Semantically specified individual form**
pro.h:S	full pronoun or demonstrative		human
0.1.h:S	deleted	first person	
0.2.h:S	deleted	second person	
0.3.h:S	deleted	third person	
np.h:S	noun phrase		
pro:S	full pronoun or demonstrative		non-human

Tag	Form of the referent	Inherent properties of the referent	Semantically specified individual form
0.3:S	deleted	third person	
np:S	noun phrase		

The sentences in examples (17) above and example (18) below illustrate the annotation. Both sentences demonstrate a case in which the subject is referred to only by the inflection on the verb. This is a frequent occurrence in Nganasan.

(18)

ref	KES_061020_MyLife_nar.020		
st	**Təti Летəмдендə чуу?əми.**		
ts	Təti L'etəmdə t'üü?əmi.		
tx	Təti	L'etəmdə	t'üü?əmi
mb	təti	L'etəm-də	t'üü-?ə-mi
mp	təti	L'etəmd'ə-ntə^{c}	t'üü-?ə-mi
gr	тот	Летовье-LAT.SG	достичь-PF-1DU.S/O
ge	that	Letovie-LAT.SG	reach-PF-1DU.S/O
mc	pr	propr-n.case	v-v.tense-v.pn
ps	DEM	NPR	V
SyR			0.1.h:S v:pred
SeR		np:G	
fr	Доехали до Летовья.		
fe	We reached Letovye.		

4.3.3 Annotation of the direct object

The direct object of a Nganasan clause can be represented by a bare noun, an adjective or a nominal phrase marked by accusative case. Pronouns in the direct object position and direct obejcts of imperative sentences are unmarked.

Deleted non-third person direct object does not appear in Nganasan, because the verbal inflection can only refer to third person object.

Table 11 *Tags for direct objects*

Tag	Form of the referent	Inherent properties of the referent	Semantically specified individual form
pro.h:O	full pronoun or demonstrative		human
0.3.h:O	deleted	third person	
np.h:O	noun phrase		
pro:O	full pronoun or demonstrative		non-human
np:O	noun phrase		
0.3:h	deleted	third person	

An example for a full pronoun human subject with an accusative-marked direct object is shown in sentence (19).

(19)

ref	KBD_71_Boat_nar.001			
st	**Куніʾи͡а мәнә ӈәнтум ӈусыӈым.**			
ts	Kuniʔi͡a mənə ŋəntum ŋusɨŋɨm.			
tx	Kuniʔi͡a	mənə	ŋəntum	ŋusɨŋɨm.
mb	kuniʔi͡a	mənə	ŋəntu-m	ŋusɨ-ŋɨ-m
ge	how	I	boat-ACC	sew-INTER-1SG.S
ps	QUE	PRONP	N	V
SeR		pro.h:A	np:P	
SyR		pro.h:S	np:O	v:pred
IST			new	
fr	Как я делаю ветку?			
fe	How to make a boat?			

4.3.4 Annotation of the subordinate clause

Most sentences in Nganasan are simple sentences from the point of view of their construction. We distinguish only five types: temporal, conditional, relative, adverbial, and purpose. Subordinate clauses with adverbial function referring to time are annotated as temporal subordination. Subordinate clauses with the function of subject or object arguments are annotated as relative clauses. This is the most uncommon construction. Table 10 summarizes the annotation for these clauses.

Table 12 *Tags for subordinate clauses*

Types of subordination	**Tag**
temporal	s:temp
conditional	s:cond
relative	s:rel
purpose	s:purp
adverbial	s:adv
complement	s:compl

The following sentences illustrate the annotation of a subordination.

(20)

ref	KECh_080214_Childhood_nar			
st	Натәмунудюәму^{ʔ} мыӈ ӈонәраану^{ʔ} ичуну^{ʔ}.			
ts	Natəmunud'üəmuʔ miŋ ŋo nəraanuʔ it'ünuʔ.			
tx	Natəmunud'üəmuʔ	miŋ	ŋonəraanuʔ	it'ünuʔ.
mb	natəmunu-d'üə-muʔ	miŋ	ŋonə-raa-nuʔ	i-t'ü-nuʔ
ge	think-PST-1PL.S/O	we	oneself-LIM-1PL	be-PRS-1PL.R
SyR	0.1.h:S v:pred	s:compl		
fr	Думали, что мы одни.			
fe	We thought that we are alone.			

4.4 Annotation of information status

Information structure can be conceived of in various ways and several layers of it can be differentiated. In annotating information status in our corpus, we follow the annotation guidelines presented in Götze et al. (2007), but we had to modify the used annotation scheme. Therefore, we adopted some elements of the RefLex Scheme (based on Riester and Baumann works, first of all 2014), but, at the same time, we kept the basis of the scheme in Götze et al.: new, given-active/inactive, accs (Götze et al. 2007). Here, we will apply only the Core Annotation Scheme including the annotation layers *Information Status* (with the corresponding tags 'given', 'accessible', and 'new'). Now we demonstrate the principles of annotating information status. In this case, the focus of the examination is what role the information plays in the discourse. In this annotation scheme, three notions are crucial: given, accessible, and new.

Given: an entity is given if it previously occurred in the discourse. This previous occurrence does not necessarily have to be in the immediately preceding sentence but can be a few sentences earlier and being activated again now.

In the extended annotation scheme, it is possible to differentiate between active *vs.* not active referents. A referent is active if it occurred in the previous sentence, while it is inactive if it did earlier than that.

The term *given-active* and *given-inactive* referentially given marks referents that are referentially given and co-referential with an antecedent in the previous discourse, in one of the following ways (as in Baumann – Riester 2014: 4):

a. Repetition of the same referent with the same content expression.
b. Repetition in a reduced, abbreviated or otherwise modified form
c. Pronominal reference
d. Repetition of the same referent with a different expression
e. Rhetorical devices expressing co-reference, e.g. metonymy, synecdoche

According to the system by Götze et al., we tagged a subcategory of referential givenness, the so-called situative. It has two types: given-active-sit (antecedents are in 2 clauses) and given-inactive-sit (antecedents further away than 2 clauses)

Accessible: a referent is accessible if it has not been mentioned before but can be identified, for instance, from the context of the situation, general knowledge, or the course the discourse takes subsequently. According to Götze's system (2007: 157–160), it is possible to annotate exactly what is known. We do not go into such details and use core annotation instead.

New: an element is new in a sentence if it conveys new information in the sentence.
Table 11 below summarizes the main abbreviations used for annotating Nganasan information status.

Table 13 *Basic tags for information status*

Information status	Annotation	Comment
Given	giv	unspecified
	giv-active	given active: referred within the last or in current sentence
	giv-active-sit	given active situative
	giv-inactive	given inactive
	giv-inactive-sit	given inactive situative
Accessible	accs (underspecified)	
New	new	

The following sentence illustrates the annotation system of information status in the corpus.
(21)

ASS_161023_Djajku2_flkd.037								
Тэниа тахаря тэті Дяйкӯрэ маагэльчета ӈӯнарба?а ися тэниа бии?иа͡								
Təni?i͡a taharii͡aa təti D'ajkürə maagəl't'ətə ŋünarba?a is'a təni?i͡a bii?i͡a.								
Təni?i͡a	tahari͡aa	təti	D'ajkürə	maagəl't'ətə	ŋünarba?a	is'a	təni?i͡a	bii?i͡a.
təni?i͡a	tahari͡aa	təti	D'ajkü-rə	maa-gəl't'ə-tə	ŋüna-rba?a	i-s'a	təni?i͡a	bii-?i͡a
təni?i͡a	tahari͡aa	təti	D'ajku-rə	maa-kəl'it'ə-tə	ŋüni͡a-rba?a	ij-sa	təni?i͡a	biu-?ə
так	теперь	тот.[NOM.SG]	Дяйку-NOM.SG.2SG	что-EMPH-ABLADV	богатый-AUG	быть-INF	так	уйти-PF.[3SG.S]
so	now	that.[NOM.SG]	Dyaiku-NOM.SG.2SG	what-EMPH-ABLADV	rich-AUG	be-INF	so	go.away-PF.[3SG.S]
ADV	ADV	DEM	NPR	ADV	ADJ	V	ADV	V
		giv-inactive			accs			
так Дяйкӯ богатым уехал.								
Dyaiku as a rich person, left.								
or: Dyaiku left as a rich person.								

As it is stated in RefLex Guidelines, "elements which occur in direct speech are not co-referential with elements that have occurred before the direct speech section. Thus, direct speech is treated as separate, embedded, discourse" (Baumann & Riester 2014: 15). In the case of the Nganasan Corpus, we have to annotate the elements in direct speech as well, because our Corpus contains mostly narratives including direct speech constructions. We intend to use distinctive markers for referents occurring in quotation. It is important because there is no indirect speech strategy in the Nganasan language, the speakers use direct speech constructions. In direct speech constructions, the information status of a referent can change due to the change of perspective.

In the tier of IST, we mark the utterance predicates that introduce an utterance and a change of perspective. It is required because the corpus contains many quotations. A quotative verb usually precedes the utterance, but it can stand after the utterance or within the utterance. A quotative verb as utterance predicate always marks a change of perspective. In Nganasan, the subject of a sentence can

be expressed by personal endings, the pronoun is not obligatory. Principally, the first appearance of an entity is always expressed by lexical means; we keep tags for these untypical cases. This tag occurs in the cell containing the inflected verb. If the utterance predicate occurs in a quotation, the marker gets -Q segment. Table 12 shows the markers for referents in a quotation.

Table 14 *Markers for referents in a quotation*

Information status	Annotation	Quoted	Quoted and zero
given	giv (underspecified)	giv-Q	giv-Q_0
	giv-active	giv-active-Q	giv-active-Q_0
	giv-inactive	giv-inactive-Q	giv-inactive-Q_0
accessible	accs (underspecified)	accs-Q	accs-Q_0
new	new		new-Q_0

The following sentence illustrates the extended annotation system of information status in the corpus.
(22)

KNT_960809_WildAnimals_flkd.116							
Тахаряа манакӳ^1 бинтісиʲа туйхӳʔ мунугэҧ: »Әмтырә тәнә ху͡ара^{1}ам.«							
Tahari͡aa manaküʔ bintisʲiʔa tujhüʔ munugəŋ: »əmtɨrə tənə hu͡araʔam.«							
Tahari͡aa	manaküʔ	bintiʔsʲiʔa	tujhüʔ	munugəŋ:	əmtɨrə	tənə	hu͡araʔam.
tahari͡aa	manaküʔ	bintiʔsʲi-ʔa	tuj-hüʔ	munu-gəŋ	əmtɨ-rə	tənə	hu͡ara-ʔa-m
теперь	только.что	росомаха-AUG.[GEN.SG]	прийти-COND	говорить-IMP.FUT.2SG.S	этот-NOM.SG.2SG	ты	взяться-PF-1SG.S
now	just.now	wolverine-AUG.[GEN.SG]	come-COND	say-IMP.FUT.2SG.S	this-NOM.SG.2SG	you.SG	take.up-PF-1SG.S
adv	adv	n-n.augm-n.case	v-v.nf	v-v.mood-v.pn	pr-n.case-poss	pr	v-v.tense-v.pn
ADV	ADV	N	N	V	DEM	PRONP	V
s:temp				0.2.h:S v:pred			0.1.h:S v:pred
				0.2.h:A			
		giv-active-Q		quot-S-Q	giv-active-Q	giv-active-Q	giv-active-Q_0
И так только росомаха придёт скажи: "Я тебе сейчас дам!"							
And right when the large wolverine comes, you'll say: "Now I give it to you!"							

4.5 Annotation of Borrowing (BOR)

Borrowing in Nganasan is not a well-studied phenomenon. The reason for this is certainly the missing of an annotated corpus. Borrowing is annotated in several tiers: BOR, BOR-Phon and BOR-Morph. The schema was prepared in cooperation with Alexandre Arkhipov.

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

RUS: for Russian

DOL: for Dolgan, etc.

We annotate different types of loanwords, according to Myers-Scotton (2002, 2006). We distinguish between cultural borrowings (in Ngan, *kola* ‘school’) and core borrowings (in Ngan. *dumairsʲa* ‘think’).

Up to now, we have not encountered any loan translations or loan shifts in Nganasan. A further type is grammatical borrowing such as conjunctions (*i* 'and'). Additionally, we annotate borrowed discourse markers and modal words. Table 15 below shows the annotation tags for the tier BOR.

Table 15 *Annotation tags for the tier BOR*

Source Langues	**Annotation Tag**	**Comment**
RUS	:cult	cultural borrowing from Russian
	:core	core borrowing from Russian
	:gram	grammatical borrowing from Russian
	:mod	modal word borrowed from Russian
	:disc	discourse marker borrowed from Russian

During the annotation, we take into consideration the structural integration (phonetical / phonological and inflectional) of nouns and verbs. There are transcriptions in the corpus in which the speakers use for example the word 'school' borrowed from Russian in two different forms: *kola-mu?* 'our school' and *škola-mu?* 'our school' (the latter is an example of insertion of lexical material according to Muysken 2000: 3). The first case involves phonetic adaptation (deletion of the initial consonant). Some words, for example the word for 'table' (Russ. *stol*) appears also in different forms in the corpus: *istolə, əstolə*. In all cases, the speakers insert a vowel to change a not-permitted consonant in onset to vowel. However, for the word-initial clusters the speakers used two different strategies: vowel prothesis and non-adaptation. This phenomenon is annotated in tier BOR-Phon.

Table 16 *Annotation tags for phonological adaptation strategies*

Tier	**Types of adaptation**	**Tag**	**Comment**
BOR-Phon	**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

Tier	Types of adaptation	Tag	Comment
	lenition	lenition	weakening
	fortition	fortition	strengthening

In case of verbal borrowings, we use further annotation in the tier BOR-Morph, 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.),
In Nganasan, it seems that indirect insertion is the most frequent strategy in Nganasan: Russ *duma-* > *duma-**ir-ii*** 's/he thinks'. A further parameter, the inflection in the matrix language, was introduced. Table 17 shows the annotation tags for the tier BOR-Morph.

Table 17 *Annotation tags for morphological adaptation strategies*

Type	**Tag for strat.**	**Tag for inflection**	**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 further inflected

The following example shows the annotation of borrowing.

(23)

ASS_161023_Djajku2_flkd.032				
Хемоу, хасибаку мыӈ ни? котэ?.				
Hemou, hu͡as'ibaku miŋ n'i? kotə?.				
Hemou,	hu͡as'ibaku	miŋ	n'i?	kotə?.
hemou	hu͡as'iba-ku	miŋ	n'i-?	kotə-?
hiəmə	hu͡as'ibə-?ku	miŋ	n'i-?	kotə-?
EXCL	спасибо-DIM	мы.[ACC.SG]	NEG-IMP.2SG.S	убивать-CNG
EXCL	thanks-DIM	we.[ACC.SG]	NEG-IMP.2SG.S	kill-CNG
PTCL	PTCL	PRONP	AUX.NEG	V
	RUS:cult			
	inCdel Csub Vsub			
	dir:infl			
Ой спасибо, не убивай нас.				
Oh, please, do not kill us!				

4.6 Annotation of Code Switching (CS)

The purpose of annotation of code switching is to mark the foreign elements in Nganasan. It is well known that a strict differentiation between borrowing and code switching is hardly possible. Words that are phonologically or morphologically adapted to Nganasan, we annotate as borrowing. The status of non-inflected forms is determined by their meaning and by their phonology:

- Lexemes expressing new concepts (e.g. cinema, brigade) are considered Russian borrowing.
- Lexemes expressing new concepts, but the borrowed form has been built into Nganasan morphology and phonology (e.g. *kolə* 'school', *školamiʔ* 'our school') are categorized as borrowing.
- Functional lexemes expressing function not expressed in Nganasan with lexical means (e.g. *il'i* 'or') (See Section 4.5. above).

Annotation for code switching is used for

- marking clauses in Russian (sentence external code switching: ext)
- marking single lexemes or compounds inflected Russian (sentence internal)
- marking lexemes expressing concepts that are also expressed in Nganasan (sentence internal)

Table 18 shows the annotation tags for code switching.

Table 18 *Annotation tags for code switching*

Type	Annotation tag	Comment
sentence external	ext	language change on the clause or utterance borders without interference
sentence internal	int: ins	a fragment is inserted into the matrix language (usually an NP or adjunct)
	int: alt	sentence internal alternation between structures, the fragment in the embedded language does not form a syntactic unit

Sentence (24) shows a sentence-external/inter-sentential code switching, while sentence (25) presents a sentence internal/intra-sentential insertion.

(24)

ref	KECh_080214_Childhood_nar.049				
tx	Nu radostʼ bɨla nu i	konɨŋkətɨmɨʔ	ŋonəiʔ	maðunuʔ	dʼa.
CS	ext				
fe	It was a pleasure, because we returned? home.				

(25)

ref	MVL_080225_Ojoloko_flkd.028					
tx	Təti	tənə	mununtuŋ	təʔ	korotʼenʼkij	iŋəə
ge	that.[NOM.SG]	you.SG.[NOM.SG]	say-PRS-2SG.S	you.know	short	be-IMP.[3SG.S]
CS					int:ins	
fe	You said it had to be short…					

Published texts

Gusev, Valentin 2017: Нганасанский миф «Две птицы». *Родной язык*, 1(6): 169-198.

Filchenko, Andrey (ed.) 2009: *Сборник аннотированных фольклорных и бытовых текстов обско-енисейского языкового ареала*. Tomsk: TGPU

Filchenko, Andrey (ed.) 2010: *Сборник аннотированных фольклорных и бытовых текстов обско-енисейского языкового ареала*. Tomsk: Veter

Filchenko, Andrey (ed.) 2012: *Сборник аннотированных фольклорных и бытовых текстов обско-енисейского языкового ареала* 2. Tomsk: TGPU

Filchenko, Andrey 2015: *Сборник аннотированных фольклорных и бытовых текстов обско-енисейского языкового ареала* 4. Tomsk: Agraf Press

Mikola, Tibor 1970: Adalékok a nganaszan nyelv ismeretéhez. *Nyelvtudományi Közlemények* 72: 60–93.

Prokof'jev, G. N. 1937: Nganasanskij (tavgijskij) dialekt. [The Nganasan (Tavgi) dialect]. In: *Jazyki i pismennost' narodov Severa. Chasť 1.* 53–77. Moscow–Leningrad: Uchpedgiz.SNNS 1980: *Skazki narodov sibirskogo Severa.*. Tomsk: Izd. Tomsk. Univ.

SNNS 1981: *Skazki narodov sibirskogo Severa.*. Tomsk: Izd. Tomsk. Univ.

Wagner-Nagy, Beáta (ed.) 2002: *Chrestomathia Nganasanica*. Szeged-Budapest: SztE Finnugor Tanszék – MTA Nyelvtudományi Intézet

References

Dowty, David 1989: On the Semantic Content of the Notion “Thematic Role”. In *Properties, Types, and Meanings*. Vol. II. Ed. by G. Chierchia, B. H. Partee, and R. Turner. Dordrecht: Kluwer. 69-130.

Dowty, David 1991: Thematic Proto-Roles and Argument Selection. *Language* 67. 547-619.

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]

Grinevald, Colette and Michel Bert 2011: Speakers communities, in Austin, K. Peter and Julia Sallabank (eds): The Cambridge Handbook of endangered Languages, Cambridge: CUP, 45–65.

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]

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

Schmidt, Thomas 2017: EXMARaLDA EXAKT Manual. Version 1.3. Available under http://www.exmaralda.org/pdf/EXAKT_Manual.pdf [accessed 2017-02-25]

Appendix 1: *Tags for morpheme classes*

Abbreviation	Comment
1	first person
2	second person
3	third person
ABES	abessive
ABL	ablative
ABSTR	abstract noun
ACC	accusative
ACTN	action noun
ADJREL	adjectivizer for relative adjective
ADJZ	adjectivizer
ADJZ.LOC	locative adjective
ADMON	admonitive
ADV	adverbial ending
AGN	agent noun
ALL	allative
ANDV	andative
ATT	attenuative
AUD	auditive
AUG	augmentative
CAP	captative
CAR	caritive
CAUS	causative
CNG	connegative
COM	comitative suffix
COND	conditional
COND.FUT	future conditional
CONTR	contrastive
CRC	connective-reciprocal suffix
DEB	debitive
DEF	definite
DIM	diminutive
DISTR	distributive
DRV	derivational suffix
DST	destinative
DUB	dubitive
DUR	durative
EMPH	emphatic
EP	epenthetic vowel

Abbreviation	Comment
EXCL	exclamative
FEM	feminine
FRQ	frequentative
GEN	genitive
HAB	habituative
IMM	immediate
IMP	imperative
INCH	inchoative
INDF	indefinite marker
INF	infinitive
INFER	inferential
INSTR	instrumental noun
INT	intentional
INTER	interrogative
INZ	inceptive?
IPFV	imperfective
IRR	irreal
ITER	iterative
LAT	lative
LATADV	adverbial lative
LATPRON	pronominal lative
LIM	limitative suffix
LOC	locative
LOCN	locative noun
LOCPRON	pronominal locative
MOM	momentaneous suffix
NAR	narrative
NMLZ	nominalizer
NMLZIPF	imperfective nominalizer
NMLZPF	nominalizer perfect
NOM	nominative
OPT	optative
ORD	ordinal numeral suffix
PASS	passive
PF	perfective
PR.CL	pronominal clitic
PROL	prolative
PRS	present
PST	past
PSTPF	past perfect

Abbreviation	Comment
PTCP	participle
PUNCT	punc(ta)tive
QUAL	qualitative
REC	reciprocal suffix
RES	resultative
SEL	selective
SIM	simulative
SOC	sociative
SPEC	speculative mood
SRLAT	superlative
SUP	supine
TEMP	temporal
TR	transitive
TRANSF	transformative
TRL	translative
VBLZ	verbalizer
VN	verbal noun
VOC	vocative
VOL	volitive

Appendix 2: *Morphemes in Nganasan in alphabetical order*

Marker	**Abbreviation**	**Function**
-alu	DISTR	distributive
-biə	PR.CL	pronominal clitic
-bsa^{n}	ACTN	action noun
-bsa^{n}	LOCN	locative noun
-bsa^{n}	INSTR	instrumental noun
-bsutə	DEB	debitive
-bta	INDF	indefinite marker
-btə	SRLAT	superlative
-btu	INSTR	instrumental
-btu	VBLZ	verbalizer
-btu	ATT	attenuative
-btu	CAUS	causative
-bturu	CAUS	causative
-btüt	DRV	derivational suffix
-bu	DRV	derivational suffix
-büə	NMLZ	nominalizer
-^{c}	[LAT.ADV]	covert adverbial lative
-^{c}	[ACC.SG]	covert accusative singular
-^{c}	[GEN.SG]	covert genitive singular
-d’a	ALL	allative
-d’aʔa	PEJOR	pejorative
-d’ə(ə)	NPST	derivational suffix:
-d’ə	ADJREL	adjectivizer
-d’ətə	ADJZ	adjectivizer
-d’üm	SEL	selective
-ə	EP	epenthetic vowel
-ə	AUG	augmentative
-ə	ADJREL	adjectivizer
-ə	VBLZ	verbalizer
-ə	DRV	derivational suffix
-əu	EXCL	exclamative
-haa	OPT	optative
-haatəə	IRREAL	irreal
-haŋhu	NAR	narrative
-hatu	INFER	inferential
-hu	INTER.PST	interrogative past
-hu͡a	NAR.INTER	narrative interrogative
-huə	ADJZ	adjectivizer

Marker	Abbreviation	Function
-hüʔ	COND	conditional
-hüʔə	NMLZPF	nominalizer perfect
-hüʔnü	COND.FUT	future conditional
-iŋ	FEM	feminine
-ir	VBLZ	verbalizer
-j	ACC.PL	accusative plural
-j	VOC.SG	vocative singular
-jt'ü	DRV	derivational suffix
-jan'd'ə	PTCP.IMM	immediate particle
-jka	DRV	derivations suffix
-ka	DRV	derivational suffix
-kaj	CAR	caritive
-kal'i	CAR	caritive
-kə	DRV	derivational suffix
-kə	ITER	iterative
-kə	IMP.FUT	imperative future
-kəə	ADJZ	adjectivizer
-kəə	ITER.INTER	iterative interrogative
-kəə	ADMON	admonitive
-kəin'i	1DU.ODU	first person dual subjective conjugation
-kəit'i	2/3DU.ODU	second/third person dual for dual object conjugation
-kəj	NOM.DU	number marker
-kəj	3DU.S	third person dual subjective conjugation
-kəl	TEMP	temporal
-kəl'it'ə	EMPH	emphatic
-kəᶜtə	ABL.SG	ablative singular
-kiᶜ	ACC.DU	accusative dual
-kiᶜ	GEN.DU	genitive
-kiᶜtə	ABL.PL	ablative plural
-ku	DRV	derivational suffix
-ku	IMP	imperative
-kuə	QUAL	qualitative
-kuə	IMP.FUT	imperative future
-küə	EMPH	emphatic
-kuj	DUR	durative
-kümu	EMPH	emphatic
-kutə	PTCP.ITER	iterative participle
-l	DRV	derivational suffix
-l	DRV	derivational suffix
-lə	INCH	inchoative

Marker	Abbreviation	Function
-ləmu	NMLZ.APPR	nominalizer
-lɨ	DUB	dubitive
-l'iť'ə	EMPH	emphatic
-l'iᶜku	QUAL	qualitative
-l'ü	INDF	indefinite
-m	TRL	translative
-m	ACC.SG	accusative singular
-m	1SG.S	first person verbal ending
-mə	VBLZ	verbalizer
-mə	EXCL	exclamative
-mə	NOM.SG.1SG	possessive suffix first person singular nominative
-mə	ACC.SG.1SG	possessive suffix first person singular accusative
-mə	1SG.O	first person singular verbal ending for objective conjugation
-məbtu	DRV	derivational suffix
-məə	PTCP.PASS	passive participle
-məənu	PROL.ADV	adverbial prolative
-miə	1DU.S/O.EXCL	first person dual verbal ending for subjective and objective conjugation with exclamative
-mənu	PROL.SG	prolative singular
-mətumaʔa	ABES	abessive
-mətumaʔa	PTCP.ABES	participle abessive
-mi	NOM.SG.1DU	possessive suffix first person dual nominative
-mi	ACC.SG.1DU	possessive suffix first person dual accusative
-mi	1DU.S/O	first person dual verbal ending for subjective and objective conjugation
-mi͡aku	DIM	diminutive
-mtə	ACC.SG.2SG	possessive suffix second person singular accusative
-mti	ACC.SG.2DU	possessive suffix second person dual accusative
-mti	ACC.SG.3DU	possessive suffix third person dual accusative
-mtu	TRANSF	transformative
-mtu	ACC.SG.3SG	possessive suffix second person singular accusative
-mtuə	ORD	ordinal numeral suffix
-mtuŋ	ACC.SG.3PL	possessive suffix third person plural accusative
-mtuʔ	ACC.SG.2PL	possessive suffix second person plural accusative
-mun	DRV	derivational suffix
-muⁿ	VN.IPF	imperfective verbal noun
-munəʔ	AUD	auditive
-munuj	AUD	auditive
-muŋhaᶜ	HAB	habituative
-müs	MOM	momentan suffix

Marker	Abbreviation	Function
-muʔ	NOM.SG.1PL	possessive suffix first person plural nominative
-muʔ	1PL.S/O	first person plural verbal ending for subjective and objective conjugation
-nakə	SUP	supine
-nantu	VOL	volitive
-naʔa	AUG	augmentative
-nduŋ	GEN.SG.3PL	possessive suffix third person plural genitive, singular possessed
-nə	GEN.SG.1SG	possessive suffix first person singular genitive, singular possessed
-nə	GEN.PL.1SG	possessive suffix first person singular genitive, plural possessed
-nə	1SG.R	first person singular verbal ending for reflexive conjugation
-n'ə	NOM.PL.1SG	possessive suffix first person singular nominative, plural possessed
-n'ə	ACC.PL.1SG	possessive suffix first person singular accusative, plural possessed
-n'ə	IMP.2SG.OP	second person singular verbal ending for objective conjugation with plural object in imperative
-n'ə	1SG.OP	first person singular verbal ending for objective conjugation with plural object
-ni	GEN.SG.1DU	possessive suffix first person dual genitive, singular possessed
-ni	GEN.PL.1DU	possessive suffix first person dual genitive, plural possessed
-ni	1DU.R	first person dual verbal ending for reflexive conjugation
-ni	LOC.PRON	pronominal locative
-n'i	NOM.PL.1DU	possessive suffix first person dual nominative, plural possessed
-n'i	ACC.PL.1DU	possessive suffix first person dual accusative, plural possessed
-n'i	1DU.OP	first person dual verbal ending for objective conjugation with more than one object
-nta	PRS.R	present marker for reflexive conjugation
-ntə	GEN.SG.2SG	possessive suffix second person singular nominative, singular possessed
-ntə	ADJZ.LOC	locative adjective
-ntə	IPFV	imperfective
-ntənu	LOC.SG	locative singular
-ntəʔ	3PL.R	third person plural verbal ending for reflexive conjugation
-ntəʔkə	INZ	inceptive
-ntəᶜ	LAT.SG	lative singular
-ntiʔ	LAT.PL	lative plural
-ntiᶜ	GEN.SG.2DU	possessive suffix second person dual genitive, singular possessed
-ntiᶜ	GEN.SG.3DU	possessive suffix third person dual genitive, singular possessed
-ntiᶜ	2DU.R	second person dual verbal ending for reflexive conjugation
-ntiᶜ	3DU.R	third person dual verbal ending for reflexive conjugation
-ntiᶜnu	LOC.PL	plural locative
-ntu	GEN.SG.3SG	possessive suffix third person plural genitive, singular possessed

Marker	Abbreviation	Function
-ntu	NMLZIPF	imperfective nominalizer
-ntu	PRS	present
-ntuə	PTCP.PRS	present participle
-ntuərəku	SPEC	speculative moos [present participle + simulative]
ntuərəkəu	SPEC.EXCL	speculative moos [present participle + simulative] with exclamative
-ntuəuʔ	2PL.R.EXCL	second person plural verbal ending for reflexive conjugation with exclamative
-ntuʔ	GEN.SG.2PL	possessive suffix second person plural genitive, singular possessed
-ntuʔ	2PL.R	second person plural verbal ending for reflexive conjugation
-nu	LOC.ADV	adverbial locative
-n'üə	VBLZ	verbalizer
-nuʔ	GEN.SG.1PL	possessive suffix first person plural genitive, singular possessed
-n'üʔ	NOM.PL.1PL	possessive suffix first person plural nominative, plural possessed
-n'üʔ	ACC.PL.1PL	possessive suffix first person plural accusative, plural possessed
-n'üʔ	1PL.OP	third person plural verbal ending for objective conjugation with more than one object
-n'üʔ	1PL.R	third person plural verbal ending for reflexive conjugation
-ŋ	2SG.S	second person verbal ending for subjective conjugation
-ŋ	2SG.R	second person verbal ending for reflexive conjugation
-ŋalə	EMPH	emphatic
-ŋəə	IMP	imperative
-ŋəu	INTER.EXCL	interrogative exclamative
-ŋku	DIM	diminutive
-ŋu	IMP	imperative
-ŋu	INTER	interrogative
-ŋuľü	INDF	indefinite marker
-ŋutu	DEF	definite
-r	FRQ	frequentative
-ra	DRV	derivational suffix
-raa	LIM	limitative suffix
-raʔa	ADJZ	adjectivizer
-rbaʔa	AUG	augmentative
-rə	VBLZ	verbalizer
-rə	NOM.SG.2SG	possessive suffix second person singular nominative, singular possessed
-rə	2SG.O	second person verbal ending for objective conjugation
-rəə	DRV	derivational suffix
-rəku	SIM	simulative
-rəmu	LOCN	locative noun
-ri	NOM.SG.2DU	possessive suffix second person dual nominative, singular possessed

Marker	Abbreviation	Function
-ri	2DU.S/O	second person dual verbal ending for subjective and objective conjugation
-ru	CAUS	causative
-ru	PASS	passive
-ruə	DRV	derivational suffix
-ruʔ	NOM.SG.2PL	possessive suffix second person plural nominative, singular possessed
-ruʔ	2PL.S/O	second person plural verbal ending for subjective and objective conjugation
-s	VBLZ	verbalizer
-s	ANDV	andative
-sa	INF	infinitive
-sə	CRC	connective-reciprocal suffix
-səbtə	SOC	sociative suffix
-səmu	LOCN	locative noun
-sərə	CONTR	contrastive
-sɨdʼə	REC	reciprocal suffix
-sʼiə	EMPH	emphatic marker
-sɨj	DRV	derivational suffix
-suə	PST	past
-suədʼəə	PTCP.PST	past participle
-suədʼəə	PSTPF	past perfect
-ta	DRV	derivational suffix
-tə	ABL.ADV	adverbial ablative
-tə	DRV	derivational suffix
-tə	AUG	augmentative
-tə	GEN.PL.2SG	possessive suffix second person singular genitive, singular possessed
-tə	DST	destinative
-tə	IMP.2SG.O	second person verbal ending for objective conjugation in imperative
-təm	1SG.S	first person verbal ending for subjective conjugation
-tətə(təə)dʼəə	DST.IRREAL	destinative irrealis
-tə^{c}	3SG.R	third person dual verbal ending for reflexive conjugation
-təə	DRV	derivational suffix
-ti	3DU.O	third person dual verbal ending for objective conjugation
-t͡iaj	DRV	derivational suffix
-tiŋ	IMP.2SG.R	imperative second person singular reflexive conjugation
-tu	CAP	captative
-tu	TR	transitive
-tur	DRV	derivational suffix
-tur	FRQ	frequentative

Marker	Abbreviation	Function
-tüʔ	ABSTR	abstract noun
-t’ə	EMPH	emphatic marker
-u	REC	reciprocal suffix
-ʔ	NOM.PL	nominative plural
-ʔ	3PL.S	third person plural verbal ending for subjective conjugation
-ʔ	ADV	adverbium suffix
-ʔ	CNG	connegative
-ʔ	IMP.2SG.S	second person singular verbal ending for subjective conjugation in imperative
-ʔa	LATPRON	pronominal lative
-ʔbalə	ADJZ	adjectivizer
-ʔhan	INT	intentional
-ʔkə	RES	resultative
ʔmis	PUNCT	punc(ta)tive
-ʔmuə	VN.PF	perfective verbal noun
-ʔmun	VN.IPF	imperfective verbal noun
-ʔnar	FRQ	frequentative
-ʔsa	SUP	supine
-ʔsan	NMLZ	nominalizer
-ʔs’i	AGN	agent noun
-^{c}	[VBLZ]	covert verbalizer