INEL Nganasan corpus

User documentation

Maria Brykina, Valentin Gusev, Sándor Szeverényi, Beáta Wagner-Nagy, 01.05.2025

1. Introduction

1.1. Objective of the corpus

The present corpus of the Nganasan language has been developed as part of the long-term research INEL project ("Grammatical Descriptions, Corpora and Language Technology for Indigenous Northern Eurasian Languages"). It is largely based on the Nganasan Spoken Language Corpus (NSLC), which has been adapted to the INEL standards and supplemented with new texts. The INEL Nganasan corpus makes possible typologically oriented corpus-based research on Nganasan and expands the documentation of the lesser described indigenous languages of Northern Eurasia.

1.2. Nganasan language

1.2.1. Description

Nganasan belongs to the Samoyedic branch of the Uralic language family, its closest relatives within the North Samoyedic group being Enets and Nenets. Today people speaking Nganasan live in several villages in the Taimyr Autonomous District, which is part of the Krasnoyarsk Krai of the Russian Federation. Nganasan is highly endangered: according to Russian census data of 2010 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 at most.

The typological profile of Nganasan is rather typical for the Samoyedic languages. It has mainly agglutinative morphology and head-final syntax.

1.2.2. Language Codes

ISO-639-3 code: **nio**

Glottolog code: **ngan1291**

1.2.3. Dialectal subdivisions

Commonly, two dialects of Nganasan are distinguished, Avam and Vadeyev, however these idioms do not differ significantly from each other (cf. Helimski 1998: 481–482).

1.3. Archiving

The INEL Nganasan corpus consists of two parts: a) texts provided with source media files (whenever available) and annotated transcripts in *EXMARaLDA*[1] transcript format (searchable part of the corpus) and b) texts without *EXMARaLDA* transcripts represented either by audio recordings (optionally – with preliminary transcriptions) or

[1] http://exmaralda.org/en/, last access: 28.04.2025.

scanned pages of the manuscripts or publications (archival part of the corpus). Not all available transcriptions are attached to non-glossed texts, because of their preliminary quality.

Texts from both parts of the corpus are provided with metadata descriptions in *EXMARaLDA* Coma format (see 2.8 for details).

For the searchable part of the corpus a copy of *EXMARaLDA* transcripts in ELAN[2] EAF format is also provided as an alternative for ELAN users. A copy of transcripts in ISO/TEI format is provided for use in compatible tools, in particular for the Tsakorpus online search platform.

The corpus is archived and published by the Research Data Repository of the University of Hamburg[3] under open-access conditions with Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0).[4]

The corpus is available for download in four packages of different size (see 2.5.2 for details):

- The "video" package includes sound files in WAV format and video files in MPEG or MP4 format.
- The "standard" package includes sound files in WAV format and no video files.
- The "mp3" package includes sound files in MP3 format and no video files.
- The "lite" package does not include any sound, video or PDF files.

Besides the downloadable packages, the corpus is accessible online through Tsakorpus,[5] an open-source search platform for linguistic corpora (see 3.4.2 for details). The current version of the corpus can be accessed at https://inel.corpora.uni-hamburg.de/NganasanCorpus/search.

1.4. Citation

The corpus is to be cited as follows:

Brykina, Maria; Gusev, Valentin; Szeverényi, Sándor; Wagner-Nagy, Beáta. INEL Nganasan Corpus. Version 1.0. Publication date 2025-05-02. https://hdl.handle.net/11022/0000-0007-FE63-C. Archived at Universität Hamburg. In: *The INEL corpora of indigenous Northern Eurasian languages*. https://hdl.handle.net/11022/0000-0007-F45A-1

Note that the authorship of the corpus refers to linguistic analysis of included texts. Many other people contributed to the corpus. First of all, needless to say, this refers to Nganasan speakers who acted as storytellers. Besides, this refers to those who did recording, transcribing, translating, editing and technical processing of the data included into the corpus. Everyone's input is acknowledged throughout the corresponding sections of this document and in the metadata.[6]

Additional reference to published (e.g. Mikola 1970) or manuscript archive source (e.g. archive of the Department of Siberian Indigenous Languages of Tomsk State Pedagogical University) is welcome when citing sentences coming from there.

1.5. Project members

Project summary information

The INEL Nganasan corpus has been created within the long-term INEL project ("Grammatical Descriptions, Corpora and Language Technology for Indigenous Northern Eurasian Languages"), 2016–2033. For an overview of the project, see (Arkhipov & Däbritz 2018). The project homepage can be visited at: https://inel.corpora.uni-hamburg.de/.

The Nganasan subproject spanned three and a half years from January 2022 to May 2025. The research was carried out at the Institute for Finno-Ugric/Uralic Studies (IFUU) of the Hamburg University (UHH).

[2] https://tla.mpi.nl/tools/tla-tools/elan/, last access: 28.04.2025.
[3] https://www.fdr.uni-hamburg.de/communities/inel, last access: 28.04.2025.
[4] https://creativecommons.org/licenses/by-nc-sa/4.0/, last access: 28.04.2025.
[5] https://github.com/timarkh/tsakorpus, last accessed: 28.04.2025.
[6] Brykina, as one of the main developers of the corpus in terms of INEL project, has the main responsibility for remaining errors, inconsistencies and further shortcomings of the corpus.

Project leader

Beáta Wagner-Nagy

Editors

Beáta Wagner-Nagy, Alexandre Arkhipov

Main corpus authors

Maria Brykina

Valentin Gusev

Sándor Szeverényi, University of Szeged (member of the NSLC Project team)

Beáta Wagner-Nagy

Contributions of Nganasan speakers and of researchers not mentioned here are acknowledged in more detail in section 1.6 and in the metadata to the corpus (see 2.8).

Technical developers

Elena Lazarenko

Aleksandr Riaposov

Timm Lehmberg (technical coordinator; until January 2023)

Student assistants

Alena Kulikova (until September 2022)

Daniil Lokshin (July 2023 – September 2024)

Symbat Saldyr (August 2023 – April 2025)

Christina Seibel (January 2025 – April 2025)

The team of the NSLC

Of course, the whole team of the Nganasan Spoken Language Corpus, which was integrated into the INEL Nganasan corpus, must be considered as contributors to the INEL Nganasan corpus as well:

Prof. Beáta Wagner-Nagy (project leader; glossing, annotation)

Dr. Brykina, Maria (glossing)

Dr. Gusev, Valentin (glossing)

Dr. Szeverényi, Sándor (annotation)

Budzisch, Josefina (alignment of transcriptions)

Danilova, Victoria (translation into English)

Jawinsky, Gerrit (alignment of transcriptions; translation into English)

Jark, Florian (translation into English)

1.6. Acknowledgements

Funding

The INEL Nganasan corpus has been produced in 2022–2025 in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities. The project was applied for by Beáta Wagner-Nagy, Michael Rießler, Hanna Hedeland, and Timm Lehmberg.

The Nganasan Spoken Language Corpus (NSLC), which was integrated into the INEL Nganasan corpus, 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 corpus contains materials collected by different researchers during earlier fieldwork sessions, which were supported by several foundations (see Table 1 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)

Providing legacy data

Significant amount of data included into the corpus became available due to courtesy of institutions and private individuals who shared their archives.

The Department of Siberian Indigenous Languages of Tomsk State Pedagogical University (Russian: Кафедра языков народов Сибири Томского государственного педагогического университета; abbreviated as Tomsk in the metadata) generously provided access to their archive. Personally, we are grateful to *Andrey Filchenko* and *Elena Kryukova*.

The Institute for Linguistic Studies RAS (Russian: *Институт лингвистических исследований РАН*) made it possible to access the manuscript archives of *Natalia Mitrofanovna Tereshchenko*.

The Dudinka branch of GTRK "Norilsk" (Russian: *Дудинское отделение ГТРК «Норильск»*; also known informally as "Taimyr radio"; abbreviated as "STRC Taimyr, Dudnika" in the metadata) generously provided access to the Nganasan part of its extensive audio archive (most recordings were done in 1988 – 2005 by *Larisa Yandipteevna Turdagina*).

We are very grateful to the Taimyr House of National Arts (Russian: *Таймырский дом народного творчества*; also known in English as Taimyr House of Folk Culture) and to the City Centre of National Arts (Russian: *Городской центр народного творчества*) who helped and supported us during our field trips to the Nganasans.

We would also like to thank *Oksana Eduardovna Dobzhanskaya* and *Jean-Luc Lambert* who kindly shared with us audio recordings from their private archives.

Storytelling[7]

We want to express our deepest gratitude to Nganasan speakers who generously shared their unique knowledge of Nganasan and without whom this corpus could not exist. Some of them told their stories, some wrote them down. All Nganasan speakers who contributed to the corpus through storytelling (except for those who unfortunately remain unidentified) are listed in Appendix A1.

Text recording, analysis, processing the data

First of all, we would like to thank all speakers of the Nganasan language who spent days and sometimes months working with us or with our colleagues **transcribing and translating** previously recorded stories and answering questions about Nganasan words and sentences:

[7] We use the term storytelling broadly referring to producing a text of any genre for recording.

Aleksandr Chaleevich Momde
Aleksej Molkovich Turdagin
Aleksej Nikolaevich Chunanchar
Anna Alekseevna Momde
Dar`ya Chaixoreevna Momde
Dilimyaku Sochupteevna Yarockaya
Ekaterina Chubovna Kokore
Ekaterina Nikolayevna Sovalova
Ekaterina Subobteevna Kosterkina
Evdokiya Demnimeevna Porbina
Galina Fenakovna Porotova
Lodun Nadeevna Turdagina
Mariya Dyuntovna Yarotskaya
Mariya Madyureevna Porbina
Nadezhda Kajbyureevna Turdagina
Nadezhda Tubyakovna Kosterkina
Nikolai Sankevich Chunanchar
Nina Dentumeevna Chunanchar
Serafima Mudimeevna Kupchik
Svetlana Moibovna Kudryakova
Svetlana Sy`gakovna Aksyonova
Sy`ku Madyureevna Yarotskaya
Tamara Sochubteevna Turdagina
Tat`yana Demnimeevna Turkina
Tat`yana Ton`dyuleevna Kuzenko
Vasilij Fenakovich Porbin
Vera Lajmareevna Momde
Yuliya Madyureevna Goricheva
Zinaida Solobteevna Chebodaeva
Zoya Chuboevna Porbina

Our very special thanks to all those who contributed to **text recording and analysis**.

Nina Dentumeevna Chunanchar accomplished a huge amount of work spanning 20 years, not only helping us to transcribe and translate the rarest words and most difficult texts, but also transcribing and translating many texts on her own, including those of the *sitaby* genre.

Nadezhda Tubyakovna Kosterkina, a Nganasan speaker and linguist, has recorded, transcribed and translated an entire range of texts.

Svetlana Sy`gakovna Aksyonova recorded 10 of her own conversations with Nina Dentumeevna Chunanchar in 2019 (only three of them have been analyzed so far).

Larisa Yandipteevna Turdagina, N.A. Popov and *A. Aksyonova* had been recording interviews with the Nganasan people from the late 1960s to the early 2000s at the Dudinka branch of GTRK “Norilsk” (Taimyr radio).

Eugene Helimski, the organizer of several linguistic field trips in Taimyr, had been recording, transcribing, translating into Russian and glossing Nganasan texts, including some long *sitaby*, since 1982.

Tatjana Zhdanova, a participant of those field trips, recorded several texts in 1997, and also transcribed, translated into Russian and glossed previously recorded ones.

A lot of texts were recorded in the 1990s by the French ethnographer *Jean-Luc Lambert*. Starting from the late 1980s, musicologist *Oksana Eduardovna Dobzhanskaya* also made many recordings during her fieldwork, two of which are glossed and 12 are included in the archival part of the corpus.

Our colleagues *Florian Sobanski, Zsúzsa Varnai, Michael Daniel, Peter Voliak, Marina Dmitrievna Lyublinskaya, Josefina Budzisch* and *Eugénie Stapert* also recorded texts during their fieldwork and shared their recordings with us.

Published materials also became part of the corpus:

In the beginning of the 20th century *Georgij Nikolaevich Prokofiev* worked on Nganasan. In 1933 he published one text, which is the oldest in our collection.

In the second half of the 20th century, *Natalia Mitrofanovna Tereshchenko* worked intensively on Nganasan. She collected several Nganasan texts, which are now available online through the efforts of the Institute for Linguistic Studies of Russian Academy of Sciences in St. Petersburg.

In 1968, the Hungarian linguist *Tibor Mikola* recorded (and later published) three texts by Aleksej Molkovich Turdagin.

Between 1968 and 1972, several scholars from Tomsk, students of Andrej Dul´zon, including *Ekaterina Petrovna Boldt, I.E. Machkinis, E.P. Nojfeld* and *A.K. Stolyarova*, carried out fieldwork among the Nganasans and collected grammatical data, word lists, and texts. Only some of these texts were later published.

Several texts in the corpus were **translated into English** by *Réka Zayzon* and *Regula Sutter*. One of the published texts was translated into English by *Juha Janhunen* and one text by *Andrey Filchenko*.

We would also like to thank *Alexandre Arkhipov* for technical support and valuable advice on all stages of producing the corpus.

2. The corpus

2.1. The language(s) of the corpus

2.1.1. Content

The language of content in the corpus is Nganasan, but audio recordings contain instances of code-switching into Russian and Govorka (part of them are left untranscribed in the corpus).

There is always only one main transcription tier (per speaker), using the common INEL transcription style (see 3.2.4).

2.1.2. Annotations

The main annotation language in the corpus is English.

The content is translated into English and Russian (tiers **fe/lte** (see 3.3.17 and 3.3.19) and **fr** respectively). Single texts are translated into German and Hungarian (tiers **fg** and **fh** respectively).

For texts from the published sources and/or manuscript archives, original translation into Russian is given as provided in the sources (see tier **ltr**, 3.3.18). For texts transcribed during fieldwork, translation into Russian written down during oral transcription sessions is also given in tier **ltr**.

Morpheme glosses in English and Russian are provided for lexical stems; labels for grammatical morphemes are identical in the respective tiers and are based on abbreviations of English terms, largely following Leipzig Glossing Rules (see tiers **ge, gr**, 3.3.4).

For more information on annotations in the corpus see 3.3.

2.1.3. Metadata

The main language of the metadata is English. Russian spellings of personal names and place names are also provided. On transliteration of names, see 2.5.6.

2.2. Sources

Apart from archive materials (see 1.6, Providing legacy data), published texts (see 2.5.5, Publications) and recordings made by researchers acknowledged in 1.6, numerous recordings were made by the authors of the corpus (Maria Brykina, Valentin Gusev, Sándor Szeverényi and Beáta Wagner-Nagy) in the fieldtrips to Ust`-Avam, Volochanka, Ugarnaya, Potapovo and Dudinka in the period from 1994 to 2016.

2.3. Content

The corpus contains texts of various genres, which are broadly classified as folklore (*sitaby*,[8] *dyurymy*[9] and unspecified folklore), narrative (monologues that are neither folklore texts nor songs), conversation and song. The only text directly translated from Russian is classified as translation apart from the other genres. For an overview of story types and subtypes (motives) that occur in the corpus most frequently see Appendix A2.

2.4. Corpus size

Table 2 contains the summary of number of texts that have EXMARaLDA transcripts, number of sentences, duration of recordings and number of tokens in total and specified by genre. Texts were recorded from 94 speakers.

[8] *Sitaby* (a genre of Nganasan folklore) is an epic tale usually having multiple sung fragments.

[9] *Dyurymy* is a a specific genre of Ngansan folklore. Unlike sitaby, tales of this genre do not contain sung fragments.

Table 2. Size of the searchable (glossed) part of the corpus

Genre		**Texts**	**Sentences**	**Recordings**	**Tokens**
folklore texts	*sitaby*	23	8,969	12:05:50	55,319
	dyurymy	100	11,270	15:30:08	73,428
	unspecified	11	6,462	10:07:49	45,154
narratives		78	5,651	9:20:40	36,065
conversations		14	2,194	1:59:26	10,306
songs		9	249	0:49:24	985
translations		1	77	no audio	490
Total		**236**	**34,872**	**49:53:17**	**221,747**

The archival part of the corpus contains 98 hours of audio material (210 texts; optionally – with preliminary transcriptions) and 30 manuscripts.

2.5. Naming Conventions

2.5.1. Name of the corpus

The name of the corpus is INEL Nganasan Corpus.

2.5.2. Folder structure and filenames

The entire corpus is contained in the folder "nganasan" which has the following files and subfolders.

Folders with text transcripts, organized by genre:

- "conv" (conversations)
- "flk" (folklore texts, unspecified)
- "flkd" (folklore texts, *dyurymy*)
- "flks" (folklore texts, *sitaby*)
- "nar" (narrative texts)
- "song" (songs)
- "transl" (a text translated into Nganasan from Russian)

Each of these genre folders contains one further subfolder per text ("communication"), named identically to the text name (see 2.5.3). Each text folder contains one or several files with different extensions according to the file type (see 2.6 for details on file formats):

- annotated transcript in EXMARaLDA EXB and EXS formats (*.exb, *_s.exs)
- annotated transcript converted into ELAN format (*.eaf)
- annotated transcript converted into ISO/TEI format (*_tei.xml)
- sound file with the recording in WAV format (*.wav) ["standard" and "video" packages] or MP3 format (*.mp3) ["mp3" package]
- video file with the simultaneous video recording in MPEG (*.mp4 or *.mpeg) format ["video" package]
- original version coming from published or archive sources in PDF (*.pdf) format (scanned images or text depending on the available source; transcriptions originally written down as MS Word files and handwritten transcriptions are also kept here) [all packages except "lite"]

Annotated transcripts and original audio and video files have the file names identical to the text name (see 2.5.3), except for "_s" and "_tei" suffixes.

Supplementary folders:

- "documentation" (contains the present document)
- "corpus-utilities" (contains annotation panel files that can be opened in EXMARaLDA Partitur Editor):
    - "annotation-panel-inel.xml": annotation values (along with short descriptions) used in tiers SeR, SyF, IST, BOR, BOR-Phon, BOR-Morph, CS, ExLocPoss (in this and other currently developed INEL corpora; thus includes values not encountered in the present corpus; see 3.3.8–3.3.12 and 3.3.14)
    - "gloss-panel-nganasan.xml": annotation values used in the part-of-speech tier (**ps**, see 3.3.6) and glossing labels for grammatical meanings used in tiers **ge**, **gr** (see 3.3.4), along with short descriptions

Individual files:

- "nganasan.coma" (main metadata file; see 2.8)
- "coma_overview.html" (a browser-readable overview of the main metadata file)

2.5.3. Names of texts (communications)

The names of the texts which are used as their IDs throughout the corpus are composed of the following components: main speaker code (see 2.5.4); further speaker codes (optional); date of recording; short title; genre abbreviation. These components are joined by underscore ("_").

The exact date is mentioned in the communication code if known, in the format YYMMDD. If the day or both the day and the month are unknown, they are omitted (thus YYMM or YY). If the year of recording is only approximate or altogether unknown, a placeholder character "X" is used to fill the missing digits (e.g. “XX“).

The genre abbreviations used are listed in 2.5.5.

In what follows an example of a text name can be seen:

Text name: ChND_061023_UnrequitedLove_song

Speaker code: ChND (i.e. Chunanchar, Nina Dentumeevna)

Date of recording: 2006.10.23

Short title: UnrequitedLove (i.e. “Unrequited Love”)

Genre: song

2.5.4. Speaker codes

The speaker codes are derived from the speaker’s full names in the order “Family name — First name — Patronymic” in their INEL Latin transliteration. Most commonly, a code is thus composed of three initial capital letters, e.g. “PED” stands for Porbina, Evdokiya Demnimeevna (Порбина, Евдокия Демнимеевна). Still, several letter combinations like Ch are used for the corresponding transliterated letters (Ч), e.g. “ChND” for Chunanchar, Nina Dentumeevna (Чунанчар, Нина Дентумеевна). If a code is already assigned to a different speaker, additional letters are used from one or more of the name parts, e.g. “KSM” for Kupchik, Serafima Mudimeevna (Купчик, Серафима Мудимеевна) and “KuSM” for Kudryakova, Svetlana Mojbovna (Кудрякова, Светлана Мойбовна).

For some speakers, the first capital letter does not correspond to the family name, e.g. “JSM” for Yarockaya, Sy`ku Madyureevna (Яроцкая, Сыку Мадюреевна), but this allows for consistency with the NSLC.

Appendix 1 contains the full list of Nganasan speakers who contributed to the corpus along with their codes.

For three texts, the speakers are unidentified; the placeholder “NN” stands for the speaker’s name in such a case.

2.5.5. Abbreviations used in metadata

Contributors

The following codes are used for members of the INEL Project and for the team of the Nganasan Spoken Language Corpus:

BJ – Budzisch, Josefina (student assistant)
BrM – Brykina, Maria
BuK – Bui, Kitty (student assistant)
DCh – Däbritz, Chris Lasse
DV – Danilova, Victoria (student assistant)
GVY – Gusev, Valentin
JF – Jark, Florian
JG – Jawinsky, Gerrit
KAY – Kulikova, Alena (student assistant)
LD – Lokshin, Daniil (student assistant)
LE – Lazarenko, Elena
OtB – Ottenberg, Bianca (student assistant)
SaS – Saldyr, Symbat (student assistant)
SCh – Seibel, Christina (student assistant)
SzS – Szeverényi, Sándor
WNB – Wagner-Nagy, Beáta

There are also several Nganasan speakers who were involved into activities other than composing texts, storytelling and oral transcription (their codes are the same as their speakers' codes):

ASS – Aksyonova, Svetlana Sy`gakovna
ChND – Chunanchar, Nina Dentumeevna
KNT – Kosterkina, Nadezhda Tubyakovna
TLJ – Turdagina, Larisa Yandipteevna

Archives

The following abbreviations for archival data are used in the corpus:

Archive of the STRC Taimyr, Dudinka – archive of the Dudinka branch of GTRK "Norilsk" (*Дудинское отделение ГТРК «Норильск»*; also known informally as "Taimyr radio" or "Dudinka radio")

Helimski archive – the archive of Eugene Helimski

Tereshchenko archive – the archive of the Institute for Linguistic Studies RAS (Institut lingvisticheskix issledovanij RAN / Институт лингвистических исследований РАН)

Tomsk – the archive of the Department of Siberian Indigenous Languages of Tomsk State Pedagogical University (Russian: *Кафедра языков народов Сибири Томского государственного педагогического университета*)

Publications

The following codes are used for publications containing texts included into the corpus (see References for details):

Mikola1970 – Mikola 1970
Nagy 1996 – Nagy 1996
Wagner-Nagy2002 – Wagner-Nagy 2002
Больдт1981 – Boldt 1981
Вагнер-Надь 2020 – Wagner-Nagy & Brykina 2020
Вагнер-Надь2017 – Wagner-Nagy & Gusev 2017
Гусев2010 – Gusev 2010
Гусев2012 – Gusev 2012
Гусев2017 – Gusev 2017
Мачкинис1980 – Machkinis 1980
Прокофьев1937 – Prokofiev 1937
Столярова1981 – Stolyarova 1981

If a text has been previously published online the link to the corresponding online resource is provided:

http://www.iling-ran.ru/gusev/Nganasan/texts/index.php – A corpus of Nganasan folklore texts (Russian: *Корпус нганасанских фольклорных текстов*; создан при поддержке целевой программы Президиума РАН «Этнокультурное взаимодействие в Евразии» в 2003-2005 гг.).

https://www.philol.msu.ru/~languedoc/rus/index.php – LangueDOC center for digital language archives[10]

http://www.univie.ac.at/negation/sprachen/nganasana.html – project "Typology of Negation in Ob-Ugric and Samoyedic Languages"[11]

Nganasan Spoken Language Corpus(http://hdl.handle.net/11022/0000-0007-C6F2-8) – Brykina et al. 2018

Genres

The following abbreviations are used for text genres:

flk – folklore texts, unspecified
flkd – folklore texts, *dyurymy*
flks – folklore texts, *sitaby*
nar – narrative texts
song – texts of songs
conv – conversations
transl – texts translated into Nganasan from Russian

2.5.6. Transliteration of Cyrillic names

In the metadata fields referring to personal names (see 2.8), a romanized spelling is used alongside their Cyrillic spelling according to the Russian orthography. Transliterations follow the GOST 7.79–2000 System B transliteration standard (GOST 2001). Elsewhere, e.g. in text titles, English glosses and free translation, romanization is performed at the authors' discretion.

2.6. Technical formats

2.6.1. Transcripts

The transcripts in the corpus are provided in several formats, all of them in XML. The main working format is EXMARaLDA EXB, while the other formats are derived from EXB to provide search functionalities and alternative ways of access to the data.

EXMARaLDA EXB and EXS

The annotated transcripts are delivered in the formats of the EXMARaLDA software suite. The main transcript file which can be used for browsing the transcript with the EXMARaLDA Partitur Editor is the "basic transcription" format (EXB). From the basic transcription, a supplementary "segmented transcription" (EXS) is automatically generated which is necessary to make searches across the corpus with the EXMARaLDA EXAKT corpus search tool and to provide word and sentence counts. (Note that the segmented transcription files are **not** to be opened with the Partitur Editor.) The respective file extensions are ".exb" and ".exs".

Please refer to EXMARaLDA documentation for introduction to the use of this software: https://exmaralda.org/en/quickstart-documents/.

Time alignment (synchronization)

The transcripts in the corpus are time-aligned with the available sound recordings. Please be aware that the time alignment is only valid at sentence level (**ts** tier; see 3.1). Technically, time values are also present at word level (**tx** tier), however they should be disregarded as arbitrary. Time values are also technically present in transcripts without any available sound; these are completely arbitrary and should likewise be disregarded.

Video files are not time-aligned with the sound recordings and are not linked in the transcripts.

[10] supported by NSF grant #0553546 «Five languages of Eurasia» under the Documenting Endangered Languages Program, by RFBR grants № 05-06-80351 «Minority languages and cultures: On the verge of extinction» and № 08-06-00345 «Multimedia corpora for endangered languages»

[11] funded by the Austrian Science Fund FWF and supported by the Division of Finno-Ugric Studies at the DECLS, University of Vienna

ELAN EAF

Additionally, the annotated transcripts are converted into ELAN format (".eaf"), which makes the downloaded corpus also browsable and searchable locally using ELAN.

ELAN transcripts differ from the original EXB transcripts in tier structure due to inherent differences between the two data models. In EXB transcripts, the main transcription tier is the tier **tx** (with subdivision into words), and all other tiers are dependent on **tx** (see 3.2). In ELAN transcripts, the main transcription tier is the tier **ts** (sentence-level), and all other tiers are dependent on **ts**. Furthermore, annotations on each dependent tier are time-aligned independently of the other tiers, therefore in case of modification of time-alignment and/or merging or splitting annotations the initial alignment between tiers could be broken.

Please be aware that the ELAN versions of the transcripts are provided for compatibility only and are not specifically tested or curated.

ISO/TEI XML

ISO/TEI is an ISO standard (ISO 24624:2016 "Language resource management — Transcription of spoken language"[12]) for representation of spoken data, and at the same time a TEI[13] compliant XML format. It is used, among other, as a source format for the Tsakorpus platform which provides online search over INEL corpora.

2.6.2. Metadata

The corpus metadata are created in the EXMARaLDA Coma (Corpus Manager) and stored in the Coma XML format (file extension ".coma"). One file holds the metadata for the whole corpus. The fields used are explained in 2.8.

2.6.3. Media

For texts with audio sources, sound files are provided in Linear PCM WAV format (file extension ".wav"). MP3 versions of all sound files are also provided as a light-weight option (44.1kHz, 192kbps).

For texts with available video recordings, video files are provided in MPEG and MPEG4 format (file extensions ".mpeg" and ".mp4").

2.6.4. Other data

PDF format ("*.pdf") corresponds to original Word transcriptions of the recordings and to scans of original publications or manuscripts.

2.7. Workflow of the source files

While the main part of analysis of all texts was done in the *SIL Fieldworks Language Explorer* (FLEx),[14] and most of the transcripts were exported into the EXB format via the FLEx's Flextext XML format, the previous workflow was different depending on the source of the text.

- Texts from published sources and manuscript archives were typed and rewritten in the INEL standard transcription. Time alignment was performed in EXMARaLDA Partitur-Editor after export from FLEx.
- Transcriptions made during the field sessions in ELAN were time-aligned there and then exported to FLEx.
- Transcriptions made during field sessions in MS Word were either first imported into ELAN, where they were checked and time-aligned; or first imported into FLEx and later time-aligned already in EXMARaLDA Partitur-Editor.
- Texts, that had earlier been glossed in Toolbox, were converted directly into EXMARaLDA EXB format (without being reglossed in FLEx). The time alignment was performed in EXMARaLDA Partitur-Editor.

For most of the texts, the morphological analysis (interlinear glossing) was done in FLEx. At this stage, all the morpheme-level tiers (**mb, mp, ge, gr, mc**) and the part-of-speech tier (**ps**) are created. As soon as glossing is

[12] https://www.iso.org/standard/37338.html, last access: 28.04.2025
[13] https://tei-c.org/, last access: 28.04.2025
[14] https://software.sil.org/fieldworks/, last access: 28.04.2025.

complete, the text is exported from FLEx as FLEXTEXT XML and converted to EXMARaLDA EXB format. During this conversion, the **ref** tier is created which combines communication code and sentence numbering (see 3.3.1).

All further annotation (and editing) is then done in the EXMARaLDA Partitur-Editor.

2.8. Metadata for the corpus

The metadata for the corpus are stored in *EXMARaLDA Coma* format. It is an XML-based format with separate interlinked descriptions for texts ("communications" in EXMARaLDA; analogous to IMDI "sessions") and speakers. The main fields contained in the descriptions are listed in the following sections. This includes for example the location and date of recording, but also information on which part of the processing and analysis was done by whom. Metadata about speakers contains some (unfortunately, quite sparse) biographical data.

Personal names are given in their romanized spelling alongside Cyrillic spelling according to Russian orthography.

2.8.1. General corpus metadata

The general metadata about the whole corpus include the corpus name ("INEL Nganasan Corpus") and some basic metadata fields complying with the standards of DC (Dublin Core) and OLAC (Open Language Archive Community).

2.8.2. Text ("communication") metadata

Name: The code which is assigned to the text (see 2.5.3)

Description (Communication):

- **0a Title:** Short title (in English)
- **0b Title (RU):** Short title (in Russian)[15]
- **1 Genre:** Abbreviation of the genre of the text (conv=conversation, flk = folklore (unspecified folklore texts), flkd = *dyurymy* (folklore texts without sung fragments), flks = *sitaby* (epic folklore texts, usually containing sung fragments), nar = narrative, song = song, transl = translation).
- **2a Recorded by**: Abbreviation of the person by whom the text was recorded (see 2.5.5)
- **2b Year of recording:** Here the year of recording is given.
- **3a Dialect:** Nganasan dialect used by the speaker(s) is given here (Avam or Vadeyev).
- **3b Subdialect:** The subdialect of the Avam dialect used by the speaker(s) is given here (Ust`-Avam or Volochanka).
- **4 Speakers:** Code(s) of the speaker(s)
- **5a Transcribed by:** Code(s) of the person(s) who did the transcription. During field sessions the transcription (and translation into Russian) was often made by a researcher in cooperation with a Nganasan speaker. Here only the code or name of the researcher are given. The code of the Nganasan speaker is given in the field "7a Translation into Russian". The notation "no" in this field denotes the non-transcribed recordings.
- **5b Year of transcribing:** Year(s) of transcribing
- **6a Typed by:** Code of the person who did the typing from the manuscript or printed text (only if this was a separate stage distinguishable from other processing stages)
- **6b Time-aligned by:** Code(s) of the person(s) who time-aligned the transcription
- **7a Translation into Russian:** Code(s) of the person(s) who did the first available translation into Russian
- **7b Translation into Russian (edited):** Code(s) of the person(s) who edited the Russian translation. It must be noted that not all communications are provided with standard Russian translation.
- **7c Translation into English:** Code(s) of the person(s) who did the translation into English or edited it
- **7d Translation into German:** Code(s) of the person(s) who did the translation into German

[15] this field is filled in only for the searchable part of the corpus, i.e. for communications containing annotated transcript in EXMARaLDA

- **7e Translation into Hungarian:** Code(s) of the person(s) who did the translation into Hungarian
- **8 Glossed by:** Code(s) of the person(s) who did the morphological glossing
- **9a-h Annotation SeR / SyF / IST / BOR / CS / IS / ExLocPoss / cltyp_vsem:** Codes of the person(s) who annotated the respective tiers
- **C1a-b Story type / Story subtype:** The type/subtype of the text is given here to provide a first impression of its content (see Appendix A2 for details).
- **C2 Variants**: The name of the text ("communication") that is a variant of the current story. The symbol "=" at the beginning means that the same recording has been transcribed once more. If the same story was told by another speaker, its name is given in brackets. If the same story was told earlier (or later) by the same speaker, no brackets are used.
- **T1 Transcription comment:** Comments on the quality of transcription. The notation "not precisely" is used for texts that were glossed without checking the first transcription made during the fieldwork. Other notations, written in free form, indicate, for example, that not all slips (false starts, unintelligible fragments) were transcribed or that the second speaker was not transcribed.
- **T2 Glosses comment:** the notation "to check" is used for the texts, which still contain some inconsistency in the glosses in the current version of the corpus.

Location: The following fields specify the location where the text was recorded.

- **Country:** All the texts were recorded in Russia (until 1991, as a part of USSR)
- **Region:** The great majority of the texts originate from Taimyr (currently, Taimyr region is part of the Krasnoyarsk Krai). Exceptions are texts recorded in Moscow, Leningrad and Novosibirsk.
- **Settlement:** The place of the recording
- **Settlement (RU):** The place of the recording in Russian
- **Settlement (LatLng):** Geographic coordinates (latitude, longitude) of the settlement

Languages:

- **Language code:** The ISO-code of the language of the text. It is always "nio" – Nganasan.

Setting: In this section some information about legacy sources and existing publications is given as well as information about the presence/absence of the audio and video recordings.

- **1 Archive:** The abbreviation of an archive (see 2.5.5) is given here supplemented by the volume number and pages where appropriate. For texts the manuscript of which is not preserved in an archive, we give the notation "not in archive".
- **2a Published in:** If the text has been published, the abbreviation of the publication or the link to the online resource is provided here (see 2.5.5).
- **2b Published as:** If the text has been published online, the name of the corresponding file is given here.
- **3a Has audio:** Marked "yes" if a sound recording is available
- **3b Has video:** Marked "yes" if a video recording is available

Recording: If an audio file is available, it is linked to the text description.

Transcriptions: The basic transcription (.exb) and the segmented transcription (_s.exs) are linked here to the text description; the latter is needed for searching the corpus.

Attached file(s): All other relevant files are linked to the text description here if available: PDFs with original publication(s) / scanned manuscripts / transcription made during fieldwork sessions; video recordings.

2.8.3. Speaker metadata

Metadata of speaker(s) include primarily biographical information of the speaker (unfortunately, rather sparse). Name fields exist both in Russian (RU) and English (translated or transliterated) version. The following fields are defined:

Sigle: Speaker code as defined in 2.5.4
Pseudo: Name shown in Coma's main view (using family name, first name and patronymic)
Sex: male or female
Description:

- **1a-b Family name (EN, RU):** For married women who have/had their husband's married family names, their maiden names are given in parentheses if known
- **2a-b Given name (EN, RU)**
- **3a-b Patronymic (EN, RU)**
- **4a-b Nganasan name (LAT, CYR):** Unofficial Nganasan names are given here.

Basic biographic data: Here basic biographical data of the speaker are provided.

- **1a-b Place of birth (EN, RU)**
- **1c Place of birth (LatLng)**: Geographic coordinates (latitude, longitude) of place of birth
- **2 Region**: (always "Taimyr Peninsula")
- **3 Country**: (always "Russia")
- **4 Date of birth**
- **5 Date of death**
- **6a-b Former residences (EN, RU):** If former residences prior to the work with the linguist are known and differ significantly from place of birth, they are mentioned here
- **7a-b Current residence (EN, RU)**: The current (i.e. at the time of the recording) place of residence of the speaker if known
- **7c Current residence (LatLng)**: Geographic coordinates (latitude, longitude) of the current residence
- **8a-b Other information (EN, RU)**

Education: Here information – if available – is given on the speaker's education and occupation/profession.

- **1a-b Education (EN, RU):** Here information on basic education (i.e. school) of the speaker is given if known.
- **2a-b Higher education (EN, RU):** If the speaker had a higher education, it is sometimes mentioned here.
- **3a-b Occupation (EN, RU):** Here the profession and/or occupation of the speaker is sometimes mentioned if known.

Family: Here information about the ethnicity of the speaker and their family members is given if known. For women, their maiden names are given in parentheses if known. If a speaker's relative is also a speaker with metadata on their own, their code is also given in parentheses.

- **1 Ethnicity**
- **2a-b-c Ethnicity of mother / Name of mother / Name of mother (RU)**
- **3a-b-c Ethnicity of father / Name of father / Name of father (RU)**
- **4a-b-c Ethnicity of husband/wife / Name of husband/wife / Name of husband/wife (RU)**
- **5a-b Ethnicity of grandparents / Names of grandparents**
- **6a-b Family / Family (RU):** other information about family

Languages: Here the language repertoire of the speaker is given; Russian is not always mentioned.
L1/L2/L3 (First language / Second language / Third language)

- **Language code:** Here the ISO code is given.
- **1 First language:** The name of the language
- **2a-b Dialect / Subdialect**

3. Transcription and annotation

The general principles of transcription, annotation and translation in many respects go back to the Nganasan Spoken Language Corpus (Brykina et al. 2018), documented in the respective user guidelines (Wagner-Nagy et al. 2018). They largely follow those of the INEL project as described in Arkhipov (2020).

3.1. Tier layout

Table 3. Overview of the annotation tiers used in EXMARaLDA-transcripts.

Tier label	Tier full name	Description	Unit	Optionality
ref	Reference	Text ID + sentence number Text ID + speaker code + sentence number (for texts with multiple speakers)	sentence	obligatory
so	Source text (Cyrillic)	Original transcription according to the manuscript of the source	sentence	optional
st	Source transcription (Cyrillic)	Original transcription of the source in Cyrillic alphabet	sentence	optional
stl	Source transcription (Latin)	Original transcription of the source if written in Latin alphabet, or original Cyrillic transcription (so or st) converted into Latin	sentence	optional
ts	Text (sentence)	Main transcription adapted from stl tier for glossing and annotation.	sentence	obligatory
tx	Text (word)	Main transcription segmented by word	word	obligatory
mb	Morpheme breaks	Morpheme breakdown of words (hyphen-separated morphemes for each word)	morph	obligatory
mp	Morphemes (morphophonological)	Morphophonemes, underlying forms (see notes in 3.3.3)	morph	obligatory
ge	Gloss (English)	Morphological glosses (with lexical glosses in English)	morph	obligatory
gr	Gloss (Russian)	Morphological glosses (with lexical glosses in Russian)	morph	obligatory
mc	Morphological category	Morphological category/part of speech for each morpheme	morph	obligatory
ps	Part of speech	Part of speech for each word	word	obligatory (see notes in 3.3.6)
geo	Geographical coordinates	Coordinates of named places and geographical objects	word	optional
SyF	Syntactic function	Syntactic functions of predicates and arguments, as well as for subordinate clauses	word / group of words	optional
SeR	Semantic role	Semantic (thematic) roles	word / group of words	optional
BOR	Borrowing	Borrowings (source language and borrowing type)	word	optional
BOR-Phon	Borrowing phonology	Phonological adaptations in borrowings	word	optional

Tier label	Tier full name	Description	Unit	Optionality
BOR-Morph	Borrowing morphology	Morphological adaptations in borrowings	word	optional
CS	Code switching	Code switching and calques (source language and type)	word / group of words	optional
IST	Information status	Information status for major NPs (given/new/accessible)	word	optional
Foc	Focus	Focus-background-structure	word / group of words	optional
Top	Topic	Topic-comment-structure	word / group of words	optional
ExLocPoss	Existential, locative and possessive structures	Existential, locative and possessive structures	word / group of words	optional
cltyp	Clause type	Clause type	word / group of words	optional
vsem	Verb semantics	Verb meanings characterized in terms of frames	word	optional
fe	Free translation (English)	Free English translation	sentence	optional
fr[16]	Free translation (Russian)	Free Russian translation	sentence	obligatory
fg	Free translation (German)	Free translation (German)	sentence	optional
fh	Free translation (Hungarian)	Free translation (Hungarian)	sentence	optional
ltr[17]	Original translation (Russian)	Original Russian translation, as provided in the source (including draft translation written down during a session of oral transcription and translation)	sentence	optional
lte	Automatic translation (English)	Automatic translation into English produced with the DeepL software[18] from **fr**-tier	sentence	optional
nt	Notes	Notes (in English or in Russian)	sentence	optional

Figure 1 gives an example of how a sentence looks like in the corpus (empty tiers are omitted):

[16] corresponds to the tier **fr_ed** in the NSLC
[17] corresponds to the tier **fr** in the NSLC
[18] https://www.deepl.com

Figure 1. A sample transcript fragment showing the tier layout

ref	ChND_080719_Chunanchar_flkd.019				
st	**тə-тə нумəйкаʔа кумуңтəңу бəйкаʔаиʔ дяа.**				
ts	Тə-тə, numajkaʔa kumuntəŋu bəjkaʔaiʔ d'a.				
tx	Тə-тə,	numajkaʔa	kumuntəŋu	bəjkaʔaiʔ	d'a.
mb	тə-тə	numajkaʔa	kumun-тə-ŋu	bəjka-ʔai-ʔ	d'a
mp	тə-тə	numajkaʔa	kumun-ntə-ŋu	bəjku-ʔa-ʔ	d'a
gr	ну-ну	юноша.[NOM.SG]	что.сказать-FUT2-INTER.[3SG.S]	старик-AUG-GEN.PL	ALL
ge	well-well	young.man.[NOM.SG]	say.what-FUT2-INTER.[3SG.S]	old.man-AUG-GEN.PL	ALL
mc	intj-intj	n.[n:case]	v-v>v-v:mood.[v:pn]	n-n>n-n:case	pp
ps	intj	n	v	n	pp
SyF		np.h:S	v:pred		
SeR		np.h:A		np.h:R	
IST		giv-active		giv-active	
fr	*Ну, юноша что скажет старикам.*				
ltr	*Но, юноша что скажет старикам.*				
fe	*What can the young man answer to old men.*				

3.2. Transcription tiers

3.2.1. Main transcription tiers (tx and ts)

The tiers **tx** and **ts** are the main transcription tiers. They use the INEL transcription (see 3.2.4). The major difference between them is that **ts** presents transcriptions of entire sentences, while in **tx** transcriptions are divided into words. Technically speaking, in EXMARaLDA format it is only the **tx** tier which has the type "transcription", all other tiers being of the type "annotation". It is thus the **tx** tier which serves as the basis for segmentation (in "segmented transcription" format, EXS), which is relevant for search using the EXAKT tool and for all sentence and word counts.

The treatment of some special cases and phenomena such as uncertainties, unintelligible fragments and false starts largely follow general INEL conventions described in (Arkhipov 2020: 12–17). Still, due to the fact that the Nganasan corpus imported a large number of texts transcribed and glossed in NSLC, these conventions are not always followed consistently. The main differences are as follows.

- Sometimes the content of the tier **ts** differs from that presented in **tx**. It may contain:
    - false starts and markers of unintelligible fragments, which are absent in **tx**,
    - words, transcribed differently than in **tx**,
    - the sequence of characters "((F))" used to mark false starts,
    - the sequence of characters "((XXX))" used for unintelligible fragments or false starts.
- **tx** contains two abbreviations for frequently used words: "tah" stands for "tahar͡iaa"; "tb" stands for "tahar͡iabə".
- Ellipsis character surrounded with double round brackets, ((…)) sequences, is sometimes used in the Nganasan corpus not only for unintelligible fragments, but also for false starts and untranscribed code-switching.
- Mismatches between original texts and translation are sometimes not marked in any form. Information missing in the source sentence but important for interpretation can be added in parentheses.

3.2.2. Source texts tier (so)

The source text tier (**so**) contains the original Cyrillic version of the text in question, if available. This is relevant for the texts that were typed from the manuscripts of the archive of the Department of Siberian Indigenous Languages of Tomsk State Pedagogical University.

3.2.3. Source transcription tiers (st and stl)

The source transcription tier (**st**) contains the original Cyrillic notation of the text, made during field sessions. The **stl** tier contains this phonetic transcription converted into Latin alphabet. An example for these two tiers in this configuration may be found in 0. For the texts coming from a source using Latin transcription, the **stl** tier contains the original transcription.

3.2.4. Transcription and orthographical conventions in the corpus

For transliteration of Cyrillic names in the metadata, see 2.5.6.

INEL transcription system

A unified phonological transcription is used throughout the corpus. Vowel length is marked by doubling the vowel letter: *n'aa* [ŋaː] 'Nganasan'. Palatalization is marked with the apostrophe symbol <'>. The following characters are used in the transcriptions:

Table 4. Main characters used in the transcriptions and their IPA correspondences

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: *ɲilɨjɨ*	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* ÿ: *анəликÿ*	'friend' 'big'
b: *basa*	b: *basa*	б: *баса*	'iron'
g: *maagəl'it'ə*	g: *maːgəlʲicə*	г: *маагəличе*	'nothing'
d: *d'indü͡a*	d: *jind͡ya*	д: *диндÿа*	'horse'
d': *d'esɨ*	j: *jesɨ*	д: *десы*	'father'
ð: *l'iðiŋkə*	ð: *lʲiðiŋkə*	ʒ: *лизиŋкə*	'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: *люмÿ*	'running'

Latin (phonological)	IPA	Cyrillic	Meaning
m: *mənə*	m: *mənə*	м: *мәнә*	‘I’
n: *nagür*	n: *nagyr*	н: *нагүр*	‘three’
n’: *n’inɨ*	ɲ: *ɲinɨ*	н: нины	‘older brother’
ŋ: *ŋarka*	ŋ: *ŋarka*	ң: *ңарка*	‘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: *h͡uaa*	h: *h͡uaa*	х: *хуаа*	‘tree’
t’: *t’etua*	c: *cetua*	ч: *четуа*	‘very’
ʔ: *l’üəʔsa*	ʔ: *lʲyəʔsa*	ʔ: *люоʔса*	‘Russian’

Additionally, the following characters are used in the transcriptions of the words borrowed from Russian: *c, c’, š, š’, č, č’, p, f, v, z, ž, ž’*.

Code-switching

Transcriptions of code-switching in Russian and Govorka do not follow any particular conventions. Sometimes they are given in Cyrillic in Russian orthography, sometimes they are transcribed phonetically, and sometimes they are transcribed in Russian orthography converted into Latin alphabet.

Capitalization and punctuation

Following the INEL project standards, sentence-initial words and proper names are capitalized, and all sentences have traditional punctuation marks at the end (full stop, question mark, exclamation mark, ellipsis) in all texts. Other punctuation marks are used in the corpus at the authors’ discretion. For songs and sung fragments of *sitaby*, two slashes “//” are used to mark the end of a sung phrase that doesn’t correspond to the end of the sentence.

3.3. Annotation tiers

3.3.1. Reference (ref)

The reference tier (**ref**) for each sentence contains the text name and the number of the sentence, separated by a full stop. The sentences are numbered throughout the entire text. The sentence numbers are zero-padded up to 3 digits. This part of the **ref** tier should be used for citation of a specific sentence coming from the corpus.

In brackets, the numbering according to the FLEx scheme is given (*paragraph_number*.*sentence_number*). The FLEx number is only kept for internal error tracking and has neither linguistic value nor relevance for citation. Thus e.g. the sentences in the example below should be cited as “TKF_97_Tejbulaa_flkd.TKF.002” and “TKF_97_Tejbulaa_flkd.TKF.003” respectively:

ref	TKF_97_Tejbulaa_flkd.TKF.002 (003)	TKF_97_Tejbulaa_flkd.TKF.003 (004)

In texts recorded from multiple speakers, the speaker code is additionally provided between the text code and the sentence number, separated by dots. The numbering is consecutive within each speaker separately, starting from 001. (The Flex numbering is consecutive across speakers.)

ref-ASS	ASS_ChND_190725_Batu_conv .ASS.001 (001)		ASS_ChND_190725_Batu_conv. ASS.002 (003)
ref-BAP		ASS_ChND_190725_Batu_conv.ChND.001 (002)	

3.3.2. Morpheme breaks (mb)

The morpheme breaks tier (**mb**) breaks words into segmentable morphs. Each word, according to the tier **tx**, appears in a separate cell. The morphs are represented in their surface form and are separated from each other by hyphens. Zero morphs are not represented in this tier. For an example see Figure 1.

3.3.3. Morphemes (morphophonological) (mp)

The underlying morphemes tier (**mp**) shows an approximation to the morphophonological representation of the morphs (morphophonemes), both stems and affixes, which appear in the **mb** tier.

For affixes, their morphophonological representation is known from previous research, except for several rare occurring derivations. The list of all morphophonological representations of affixes is provided in Table 28 (Appendix A3). Due to technical reasons, zero morphs, such as genitive or accusative suffix, could not be represented in this tier, even in case when they should have been represented by a zero-consonant -^{c} (also referred to as empty slot, cf. Wagner-Nagy 2018:67-71).

For lexical stems, the situation is more complicated. Frequently occurring lexical stems also get a morphophonological representation in the tier **mp**, whereas more rare ones remain in phonological or sometimes even phonetic transcription. Thus, with respect to the lexical stems, the tier **mp** provides material for further (morphophonological and etymological) research.

All morphemes within a word are separated by hyphens. Zero morphs are not represented in this tier. For an example see Figure 1.

3.3.4. Gloss (ge, gr)

The gloss tiers (**ge**, **gr**) contain the English and Russian glossing of the morphemes in **mb** and **mp**. Stems receive their respective lexical glosses in the two languages, while affixes are glossed identically in capital Latin letters and mostly according to the Leipzig Glossing Rules.[19] For the list of abbreviations used see Appendix A3.

Glosses for all morphemes within a word are separated by hyphens. Non-overt morphemes are given in square brackets preceded by a dot (e.g. ".[GEN.SG]", ".[3SG.S]").

If a gloss contains two or more semantic components, these are separated by a dot. For more convenient reading the dot is omitted in combinations of person and number (e.g. "IMP.2SG").

Alternative meanings are separated by a slash (S/SG.O for the subjective conjugation/objective conjugation with singular object). If the gloss of a morpheme is uncertain, the gloss is preceded by one percent sign. Morphemes with an unknown meaning are glossed with two percent signs (%%).

Interjections, exclamations and ideophones are sometimes glossed as INTERJ, EXCLAM and IDEOPHON respectively, and sometimes they are translated with Russian/English interjections, exclamations and ideophones.

Epenthetic syllables in songs that serve to harmonise the rhythmic structure are sometimes glossed (as SONG if they are separate words or as EP.SONG if they occur within a word), sometimes left unglossed or not transcribed.

Derivations are reflected variably throughout the corpus. In some derived forms derivational markers are glossed separately, while other similar forms are not subdivided and glossed as a whole.

For an example of gloss tiers (**ge**, **gr**) see Figure 1.

3.3.5. Morphological category (mc)

The **mc** tier indicates the morphological category of both lexical stems (i.e. the part of speech) and affixes (i.e. the inflectional category or the derivational process). Table 5 and Table 6 show the tags used for lexical stems, inflectional categories and other affixes. Tags for inflectional categories are marked as *x:cat* or *x:(cat)*, where *x* is the corresponding lexical stem tag, *cat* is a tag for the category filling an obligatory slot, *(cat)* is a tag filling an optional slot (e.g. *n:case* – case of nominals, *n:(num)* – number of nominals). Derivational processes are marked as *x>y*, *x* and *y* being the tags for lexical stems (e.g. *n>v* for verbalizers deriving verbs from nominals). The morphological category

[19] https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf, last access: 28.04.2025.

of zero morphs is indicated within square brackets (e.g. *.[n:case]* for a zero case suffix). Elements with unknown meaning are marked with two percent signs ("%%").

Table 5. Tags for lexical stems

Tag	**Description**
adj	adjective
adv	adverb
aux	auxiliary
conj	conjunction
demadj	demonstrative adjective
demadv	demonstrative adverb
dempro	demonstrative pronoun
intens	intensifier
intj	interjection
n	noun
npr	proper noun
num	numeral
pers	personal pronoun
pp	postposition
ptcl	particle
ptcl.neg	negative particle
qadj	interrogative adjective
qadv	interrogative adverb
qpro	interrogative pronoun
quant	quantifier
reln	relational noun
v	verb
%%	unknown

Table 6. Tags for affixes

Inflection	
infl:num	number marker
infl:poss	possessive marker
Inflection of nominals	
n:case	nominal case marker
n:case.poss	cumulative nominal marker of case and possession
n:dst	nominal destinative marker
Inflection of verbs	
v:mood	verbal mood marker
v:mood.pn	cumulative verbal marker of mood and person-number

v:tense	verbal tense marker
v:pn	verbal marker of person-number
Inflection of adverbials	
adv:case	adverbial case marker
Other	
ins	insertion (epenthetic elements)
cl	clitic

For an example see Figure 1.

3.3.6. Part of speech (ps)

The part of speech tier (**ps**) contains information about the grammatical category of each word form. Hence, the outcome of derivational processes is marked here. Table 7 shows the list of possible parts of speech. The tags used here are slightly different from those used in the morphological category tier **mc** (see Table 5). Auxiliary verb (**mc**: *aux*) is tagged as verb (*v*) in the **ps** tier; negative particles (**mc**: *ptcl.neg*) are tagged as particle (*ptcl*). Demonstrative and interrogative adjectives, adverbs and pronouns (**mc**: *demadj, demadv, dempro, qadj, qadv, qpro*) are tagged as adjectives (*adj*), adverbs (*adv*) and pronouns (*pro*) respectively. The tag *cvb* is used to mark converbs derived from verbs.

The **ps** tier is meant to be an obligatory one. However, some texts imported from NSLC didn't have this tier, so we filled in most of the missing values, but for 608 words the **ps** tier still remains empty.

Table 7. Tags for part of speech

Tag	**Description**
adj	adjective
adv	adverb
conj	conjunction
cvb	converb
intens	intensifier
intj	interjection
n	noun
npr	proper noun
num	numeral
pers	personal pronoun
pp	postposition
pro	pronoun (demonstrative, interrogative, negative)
ptcl	particle
quant	quantifier
reln	relational noun
v	verb
%%	unknown

For an example see Figure 1.

3.3.7. Geographical coordinates (geo)

For some placenames encountered in the corpus, geographical coordinates in "latitude, longitude" format are provided in the **geo** tier, e.g. "71.114071, 92.820798" for Ust`-Avam.

3.3.8. Syntactic function (SyF)

The Syntactic function tier is annotated in texts coming from the Nganasan Spoken Language Corpus according to (Wagner-Nagy et al. 2018: 24ff.).

In the **SyF** tier, basic syntactic functions (i.e. subject, direct object, predicate) are tagged. Copulae in complex predicates are also tagged. As well as semantic roles (see below), syntactic functions are only tagged in main clauses, with an exception for complement clauses in the form of direct speech. But subordinate clauses themselves are being tagged, the cells belonging to the subordinate clause are merged.

The set of tags for SyF tier is provided in the following table.

Table 8. Tags for syntactic functions

Tag	Description
Main arguments	
S	subject
O	direct object
Predicate	
v:pred	verbal predicate
n:pred	nominal predicate
adj:pred	attributive/adjectival predicate
adv:pred	adverbial predicate
pro:pred	pronominal predicate
ptcl:pred	particle predicate
cop	overt copular verb
Subordinate clauses	
s:compl	complement clause
s:rel	relative clause
s:temp	temporal clause
s:cond	conditional clause
s:adv	adverbial clause
s:purp	purpose clause

There are two other points that concern annotating both semantic roles and syntactic functions. First, we place the annotations (in corresponding tiers) on the head of the noun phrase, on the noun in the prepositional phrase and on the whole clause if it is a subordinate clause; for covert referents, the annotation is placed on the predicate. Second, we annotate the properties of the referent (if relevant) in both annotation layers. These properties are thus duplicated, for the more comfortable perception of tagging and also to make the search a bit easier. These properties are annotated before the main tag and are separated with a colon (<:>).

For each referent, we mark whether it is covert (<0>) or not (no special tag). In case the referent is covert (for example, it is a pro-drop subject), we indicate three possible values of the grammatical category "person": first person (<1>), second person (<2>) and third person (<3>). If the referent is overt, we annotate its form: whether it is a personal or demonstrative pronoun (<pro>), a noun phrase (<np>), postpositional phrase (<pp>) or an adverbial phrase (<adv>). For both overt and covert referents we annotate, whether they are human (<h>) or non-human (no tag). Tags for different properties of a referent are separated with a dot (<.>). For an example see Figure 1.

Table 9. Tags for referent expressions

Tag	Description
0.1	zero/covert first-person referent
0.2	zero/covert second-person referent

Tag	Description
0.3	zero/covert third-person referent
adv	adverbial referent
np	nominal referent (noun phrase)
pp	postpositional phrase
pro	pronominal referent
.h	human referent
v	verb

3.3.9. Semantic role (SeR)

The Semantic role tier (**SeR**) is annotated in texts coming from the Nganasan Spoken Language Corpus. The annotation is based on GRAID principles (Haig & Schnell 2014) with some further developments by Beáta Wagner-Nagy and Sándor Szeverényi (Wagner-Nagy et al. 2018: 21ff.).

The full set of tags for semantic roles is listed in the following table.

Table 10. Tags for semantic roles

Tag	Description
A	Agent: volitional initiator of the action, the participant is causing the action or is responsible for something happening.
B	Beneficiary: entity for whose benefit the action is performed.
Com	Comitative: an entity which is involved in the situation in the same way as some other participant (e.g. co-agent, co-patient, etc.).
Cau	Cause: entity (mostly non-human) that causes an event without volition.
E	Experiencer: living entity that undergoes a sensory, cognitive, or emotional experience (first argument of the verbs of emotion, volition, cognition, perception).
G	Goal: location or entity towards which the movement happens.
Ins	Instrument: entity used to perform an action, which does not control the situation.
L	Location: location or entity where an event takes place or where something is located (i.e. stative location).
P	Patient: entity, which undergoes physical changes, is created or destroyed.
Path	Path: entity or location along or through which the movement takes place.
Poss	Possessor: entity which owns something.
R	Recipient: animate recipient of transfer or addressee of verb of speech
So	Source: location or entity from which the movement originates.
St[20]	Stimulus: entity perceived by experiencer.
Th	Theme: entity whose location is specified, which is moved or affected by some action without physical change (change of location or possession: object of give; subject of walk); the content of mental verbs and verba dicendi (think, say etc.).
Time	Time: time point or time interval.

This list does not pretend to cover all possible semantic functions, since we do not aim to tag every noun phrase in the text, we tag the most frequent ones. For example, we do not tag semantic roles for depictives or translatives.

As well as for syntactic functions, we do not annotate semantic roles inside subordinate clauses. Referent annotation follows the same rules as in the SyF tier (see 3.3.8). For an example see Figure 1.

3.3.10. Borrowings (BOR, BOR-Phon, BOR-Morph)

The Borrowing tier (**BOR**) contains the annotation of borrowed lexical items. Both the donor language of the item in question and the type of borrowing are annotated. The tags are made up as follows: <LANGUAGE:type>. At the beginning of the NSLC project, the **BOR** tier was annotated manually, later the annotations were implemented in the

[20] The tag "St" does not occur in the corpus.

FLEx lexicon and automatically exported to EXMARaLDA. Borrowings from Russian (RUS) are marked in the corpus most consistently, the rest (DOLG – Dolgan, EV – Evenki, TURK - Turkic) are represented by single occurrences.

For the type of borrowing the following tags are used (cf. also Arkhipov 2020):

Table 11. Tags for annotating borrowings

Tag	**Description**
:cult	cultural lexicon (also used for borrowed names)
:core	core lexicon
:gram	grammatical devices (e.g. conjunctions)
:disc	discourse markers
:mod	modal words

The tier **BOR-Phon** contains the annotation of phonological processes in borrowing. It is annotated in texts coming from the Nganasan Spoken Language Corpus. The tag set is the following:

Table 12. Tags for phonological processes in borrowings

Tag	**Description**
Deletions	
inCdel	initial consonant deletion
inVdel	initial vowel deletion (aphaeresis)
medCdel	medial consonant deletion
medVdel	medial vowel deletion (syncope)
finCdel	final consonant deletion
finVdel	final vowel deletion (apocope)
Insertions	
inCins	initial consonant insertion
inVins	initial vowel insertion
medCins	medial consonant insertion
medVins	medial vowel insertion
finVins	final vowel insertion
Substitutions	
Csub	consonant substitution
Vsub	vowel substitution
Other	
lenition	lenition (weakening)
fortition	fortition (strengthening)

The tier **BOR-Morph** contains the annotation of morphological processes in borrowing. It is annotated in texts coming from the Nganasan Spoken Language Corpus. The tags are made up as follows: <Strategy:Inflection>. The tag set is the following:

Table 13. Tags for annotating morphological processes in borrowings

Tag	**Description**
Adaptation strategies	
dir	direct insertion (i.e. insertion without morphological adaptation)
indir	indirect insertion (i.e. insertion with morphological adaptation)
parad	paradigm insertion (i.e. an inflected paradigm item is borrowed)
Further inflection (in the matrix language)	
:bare	no inflection
:infl	further inflection

The following example illustrates the annotation of borrowings from Russian:

Figure 2. Annotation of borrowings

ref	ChND_061105_Nenets_nar.009					
tx	N’üənti	maaŋaləðəj	n’akələsɨəgəj	ibahu,	kəŋhiətəʔ,	peť’ən’iʔ.
mb	n’üə-nti	maa-ŋalə-ðə-j	n’akələ-sɨə-gəj	i-bahu	kəŋhiətə-ʔ	peť’ən’i-ʔ
mp	n’üə-nti^{c}	maa-ŋalə-tə-j	n’akələ-suə-kəj	ij-haŋhu	kəŋhiətə-ʔ	peť’ən’i-ʔ
ge	child-GEN.SG.3DU	what-EVEN-DST1-ACC.PL	take-PST-3DU.S	be-NAR.[3SG.S]	bonbon-NOM.PL	cookie-NOM.PL
BOR					RUS:cult	RUS:cult
BOR-Phon					Csub Vsub	
BOR-Morph					dir:infl	dir:infl
fe	*They bought, it is said, candies, cookies for the children.*					

3.3.11.Code-switching (CS)

The Code-switching tier (**CS**) contains the annotation of code-switching. Whereas borrowings concern single words, code-switching (mostly) concerns sequences of two or more words. Both language of the code-switch and type of the code-switch are annotated, namely according to the scheme <LANGUAGE:type>. The languages of code-switching are Russian and Govorka; both are currently tagged as “RUS”. The tag set for the type of code-switch is the following:

Table 14. Tags for annotating code-switching

Tag	**Description**
Sentence-external code-switching	
:ext	languages change at sentence (clause, utterance) borders
Sentence-internal code-switching	
:int.ins	languages change at phrase borders (e.g. an NP or a PP is inserted)
:int.alt	the point of change is somewhere at an arbitrary point in the sentence
:int	distinguishing between subtypes is problematic

In the current version of the corpus, some instances of code-switching are not transcribed and not annotated.

The following example illustrates the annotation of code-switching with Russian:

Figure 3. Annotation of code-switching with Russian

ref	JMD_080219_MyLife_nar.078						
tx	küüʔə,	ŋamɪ͡aj	təti	taharɪ͡aa	vɨros	mal’t’ik	bəðuʔu͡a.
mb	küü-ʔə	ŋamɪ͡aj	təti	taharɪ͡aa			bəðu-ʔu͡a
mp	kuə-ʔə	ŋamɪ͡aj	təti	taharɪ͡aa			bətu͡a^{c}-ʔə
ge	die-PF.[3SG.S]	other.[NOM.SG]	that.[NOM.SG]	now			grow-PF.[3SG.S]
CS					RUS:int.alt		
lte	*[The girl] died, and the other one, the boy, grew up.*						

3.3.12.Information status (IST)

The Information status tier (**IST**) is annotated in texts coming from the Nganasan Spoken Language Corpus. The annotation is based on the annotation guidelines for information structure and information status in Götze et al. (2007), some minor changes were nevertheless done. The principles of annotation and the annotation scheme itself were developed by Wagner-Nagy & Szeverényi (Wagner-Nagy et al. 2018: 28ff.). According to Götze et al. (2007: 150)

the information status (a.k.a. activation, cognitive status, givenness) of a discourse referent reflects its retrievability within the discourse in question. A referent can be either given, accessible or new which can be determined by using the parameters [±discourse-old] and [±hearer-old]:

Table 15. Parameters for determining information status

	+discourse-old	**–discourse-old**
+hearer-old	given	accessible
–hearer-old	—	new

In detail that means that given referents are necessarily and by default aforementioned in the discourse while accessible and new referents are not. Accessible referents can be somehow (see below) inferred by the "hearer" of the discourse. Hence, new referents are neither aforementioned nor inferable for the hearer. The basic tags for annotating information status are *giv*, *accs* and *new*, the extended tag set can be seen from the following table:

Table 16. Basic tags for annotating information status

Tag	**Description**
Given referents	
giv-active	given and active referent (i.e. mentioned in the current or previous sentence)
giv-active-sit	given and active referent, situational
giv-inactive	given and inactive referent (i.e. mentioned before the previous sentence)
giv-inactive-sit	given and inactive referent, situational
Accessible referents	
accs	accessible referent (unspecified)
accs-sit	referent, accessible through the situation (e.g. having breakfast: "Give me the butter, please.")
New referents	
new	new referent

Another problem which was dealt with is the issue of direct speech. 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" (Riester & Baumann 2014: 15). In the corpus, distinctive markers for referents occurring in quotation are used. It is important because there is no indirect speech strategy in the Nganasan language, the speakers use direct speech constructions. In order to be aware of possible changes of perspective, the tag <-Q> was proposed by Wagner-Nagy et al. (2018: 29). Furthermore, so-called utterance predicates are tagged by the tag *quot* and it is distinguished between speech and thought (*quot-sp* vs. *quot-th*) (ibid.). The following table provides an overview of all tags used in the tier **IST** in the corpus. For an example see Figure 1.

Table 17. Tags for information status

Tag	**Description**
New referent	
new	new referent
new-Q	new referent in direct quotation
0.new	covert new referent
0.new-sit	covert new referent, situational
0.new-Q	covert new referent in direct quotation
Given referent	
giv-active	given and active referent (i.e. mentioned in the current or previous sentence)
giv-active-Q	given and active referent in direct quotation
giv-active-sit	given and active referent, situational
giv-active-sit-Q	given and active referent, situational, in direct quotation
giv-inactive	given and inactive referent (i.e. mentioned in the current or previous sentence)

Tag	Description
giv-inactive-Q	given and inactive referent in direct quotation
giv-inactive-sit	given and inactive referent, situational
giv-inactive-sit-Q	given and inactive referent, situational, in direct quotation
0.giv-active	covert given and active referent (i.e. mentioned in the current or previous sentence)
0.giv-active-Q	covert given and active referent in direct quotation
0.giv-active-sit-Q	covert given and active referent, situational, in direct quotation
0.giv-active-sit	covert given and active referent, situational
0.giv-inactive	covert given and inactive referent (i.e. mentioned in the current or previous sentence)
0.giv-inactive-Q	covert given and inactive referent in direct quotation
0.giv-inactive-sit	covert given and inactive referent, situational
Accessible referent	
accs	accessible referent (unspecified)
accs-Q	accessible referent, direct quotation
accs-sit-Q	referent, accessible through the situation, in direct quotation
0.accs	accessible covert referent
0.accs-Q	accessible covert referent, in direct quotation
Utterance predicate	
quot-sp	speech predicate (introducing direct speech)
0.quot-sp	speech predicate with zero subject
quot-sp-Q	speech predicate, in direct quotation
0.quot-sp-Q	speech predicate, in direct quotation with zero subject

3.3.13.Information structure (Foc, Top)

The Topic-comment tier (**Top**) and Focus-background tier (**Foc**) contain the annotation of information structure. This tier was annotated by Chris Lasse Däbritz, who is also the author of this section.

The tag set and the annotation principles were developed from the Leipzig Model of Information Structure (LM; cf. Junghanns & Zybatow 2009, Däbritz 2021: Ch. 5). The LM operates in the theoretical framework of the Minimalist Program (cf. Chomsky 1995) and was developed to describe the information structure of Slavic languages. However, it is flexible enough to adapt to other languages and language families. The main idea of the LM is that "information structuring is a pragmatically – through the situation of the communication, the context – determined ordering principle through which elements of the sentence get a certain communicative stress." (Junghanns & Zybatow 2009: 687). Within the LM, there are two information structural levels: On the one hand, the topic-comment-structure and, on the other hand, the focus-background-structure (Junghanns & Zybatow 2009: 688). That means that topic and focus are not complementary in the sentence, both being the salient component on their respective level. *Topic* is understood as an aboutness topic in an Aristotelic sense, i.e. the part of the sentence what the predication is about. In contrast, focus is understood as the part of the sentence important for the speaker (ibid.).

Topics are divided into external topics and internal topics, the former standing outside the syntactic structure of the clause (e.g. *That man – he stole my car*.) and the latter standing inside the syntactic structure of the clause (e.g. *My mother worked for the social services*.). Internal topics can be concrete (i.e. having a clearly identifiable referent) or abstract (i.e. situational, so-called frame-setting topics). A special case of a topic is a contrastive topic (e.g. *My mother worked for the social services, but my father worked at TV*.). The tag set developed for topics is the following:

Table 18. Tags for annotating topics

Tag	Description
External topics	
top.ext	external topic
Internal topics	
top.int.concr	concrete internal topic
top.int.concr.contr	concrete contrastive internal topic

Tag	Description
top.int.abstr	abstract internal topic

As topical referents can be deleted, covert topics are tagged with <0.> at the predicate of the clause.

Focus is divided into natural focus (a.k.a. informational focus) and special focus. Within natural foci, it is distinguished between wide, intermediate and narrow focus: A wide focus contains the whole clause, an intermediate focus contains the VP of the clause, and a narrow focus contains a single constituent smaller than the VP. Special foci are contrastive foci (e.g. "Since when do you live in Berlin?" – "I live in Dresden now.") and verum foci (e.g. "Did you buy butter?" – "Yes, I did."). The tag set developed for foci is the following:

Table 19. Tags for annotating focus

Tag	Description
Natural focus	
foc.wid	wide natural focus
foc.int	intermediate natural focus
foc.nar	narrow natural focus
Special focus	
foc.contr	contrastive focus
foc.ver	verum focus

Figure 4 illustrates the annotation of information structure (see also Figure 5):

Figure 4. Annotation of information structure

ref	KES_061020_MyLife_nar.011			
tx	Bəhɪa͡	is'üə,	nəŋhəmənɨ	n'ilɨd'iəmɨʔ.
mb	bəhɪa͡	i-s'üə	nəŋhə-mənɨ	n'ilɨ-d'iə-mɨʔ
mp	bəhɪa͡	ij-suə	nəŋhə-mənu	n'ilɨ-suə-muʔ
ge	bad.[3SG.S]	be-PST.[3SG.S]	bad-ADVZ2	live-PST-1PL.S/SG.O
Top				0.top.int.concr
Foc	foc.nar		foc.nar	
fe	*It was bad, we lived badly.*			

3.3.14. Existential, locative and possessive predication (ExLocPoss)

The **ExLocPoss** tier provides the annotation of existential, locative, and possessive predication. This annotation was integrated in the corpus by Chris Lasse Däbritz specifically for his dedicated project on existential, locative, and possessive predication in languages of the Ob-Yenisei area. Chris Lasse Däbritz is also the author of this section.

Existential and locative predications express the temporary presence/absence of a referent X (figure) at a place Y (ground). In locative predication, the figure serves as the starting point for the perspectivization of the state of affairs, whereas it is the ground in existential predication. As a corollary, the figure is prototypically definite and topical in locative predication, whereas it is prototypically indefinite in existential predication, belonging to the focus domain. Possessive predications express that one referent Y (possessee) belongs to another referent X (possessor); prototypically, this relationship is again temporary, and the possessor has control over the possessee. In the case of inalienable possession (mostly, kinship and body terms), the latter does not hold. The core structures are thus the translational equivalents of:

- "At Y, there is (no) X" (existential)
- "X is (not) at Y" (locative)

- "X has (no) Y" (possessive)

Koch's (2012) generic existentials and Haspelmath's (2025) hyparctic clauses (e.g. English *there are (many) unhappy people; there is a/no God*) are subsumed under existential clauses in the annotation. Inverted possessive clauses, i.e. appertentive clauses in Haspelmath's (2025) terms, (e.g. English *the book is John's*) are subsumed under possessive clauses in the annotation.

The annotation scheme includes the three functional domains existential (Ex), locative (Loc) and possessive (Poss), the coding strategy (see below), as well as the polarity (Aff or Neg) of the clause. The annotation tags have the format <Domain:Strategy.Polarity>. Table 20 lists the tags used for annotating existential, locative and possessive predication.

Table 20. Tags for existential, locative, and possessive predication

Tag	**Description**
Functional domain	
Ex	existential predication
Loc	locative predication
Poss	possessive predication
Coding strategy and polarity	
:Zero.Aff	no lexical linking element, does not exclude pn-marking of figure/possessor at ground/possessee; affirmative
:Zero.Neg	no lexical linking element, does not exclude pn-marking of figure/possessor at ground/possessee; negative
:Cop.Aff	copula as linking element (*e-* and *bu͡ol-); affirmative*
:Cop.Neg	copula as linking element (*e-* and *bu͡ol-); negative*
:Ex.Aff	affirmative existential item (ba:*r)* as linking element
:Ex.Neg	negative existential item (*hu͡ok)* as linking element
:PosV.Aff	posture verb (e.g. *tur-* 'stand') as linking element; affirmative

The following example illustrates the annotation of locative clause:

Figure 5. Annotation of locative clause

ref-ChND	ChND_041213_Reminiscence_nar.ChND.068 (072)			
tx-ChND	Mənə	abaʔanə	kad'anɨ	in'antugətum.
mb-ChND	mənə	aba-ʔa-nə	kad'a-nɨ	i-n'antu-gə-tu-m
mp-ChND	mənə	aba-ʔa-nə	kad'a-nu	ij-nantu-kə-ntu-m
ge-ChND	PRO1SG.[NOM.SG]	mummy-AUG-GEN.SG.1SG	near-LOC.ADV	be-VOL-ITER-PRS-1SG.S
Top-ChND	top.int.concr			
Foc-ChND		foc.int		
ExLocPoss-ChND	Loc:Cop.Aff			
fe-ChND	*I wanted to stay by my mother.*			

3.3.15.Clause type (cltyp)

The annotation of clause type is given in the tier labelled **cltyp**. This tier was annotated in 16 texts by Nikita Muravyev, who is also the author of this section.

Clause type is a bundle of descriptive morphosyntactic, semantic, and pragmatic parameters used in typologically oriented grammatical descriptions, see (Aikhenvald 2014 225–264). The annotation tagset includes the following parameters: transitivity (transitive, intransitive), finiteness (finite, infinitive, conditional, temporal), speech act (declarative, interrogative, imperative, exclamative), evidentiality (neutral, evidential), modality (neutral, modal), negation (positive, negative). The entry is built in the following fashion:

<transitivity:finiteness:speech.act:evidentiality:modality:negation>

For annotating the clause type, the following glosses are used:

Table 21. Tags for clause types

Tag	**Description**
Transitivity	
trns	transitive clause
intr	intransitive clause
Speech act	
decl	declarative clause
q	interrogative clause (“question”)
excl	exclamative clause
imp	imperative clause
Finiteness	
cond	conditional clause
temp	temporal clause
inf	infinitive clause
Fvidentiality, negation and modality	
evid	evidential clause
mod	modal clause
neg	negated clause

Neutral values, such as main syntactic status or direct modality, are omitted in annotation for brevity with the exception of transitivity and speech act. Transitivity lacks a neutral value and speech act has the neutral value “declarative” which importantly implies the neutral value for finiteness as well. Thus, a transitive finite declarative neutral evidential modal positive clause will be tagged as trns:decl:mod. Intransitive conditional modally and evidentially neutral negative clause will then receive a tag intr:cond:neg.

Embedded clauses, i.e. those that fit inside the boundaries of a main clause, are annotated in parentheses directly following the main clause tag. Thus, a main intransitive declarative clause containing an embedded transitive infinitive clause will be tagged as intr:decl(trns:inf).

The following example illustrates the annotation of clause type:

Figure 6. Annotation of clause type

ref	ChNS_080214_Hibula_flkd.062		
tx	Tə,	d'eŋkun'i	ŋəðəʔ.
mb	tə	d'eŋku-n'i	ŋəðə-ʔ
mp	tə	d'eŋkuj-n'i	ŋətə-ʔ
ge	well	trap-ACC.PL.1DU	examine-IMP.2SG.S
fe	*"Check our traps.*		
cltyp	trns:imp		
vsem			scrutiny:examining

3.3.16.Verb semantics (vsem)

The annotation of verb semantics is given in the tier labelled **vsem**. This tier was annotated in 12 texts by Nikita Muravyev, who is also the author of this section.

Verb meanings are best characterized in terms of frames, or generic situations, including the event description and a set of roles of the event participants. A typical verbal frame features the event itself expressed with the verb as well as a set of core and peripheral participants of the event usually expressed with noun phrases. The annotation of frames relies on the FrameNet database (Fillmore et al. 2003; Fillmore & Baker 2010) with some additions and corrections. Since the only purpose of this tagset is to specify the event type, it does not feature any details about the participants, see separate annotation tagsets **SeR** and **SyF** for semantic roles and syntactic functions respectively.

Subframes specify the exact manner of event described by the frame or certain participant types. Thus, watching and listening are both active perception events, but the former employs the visual whereas the latter the auditory channel. Reading is similar to watching in that it is also visual but is restricted to textual stimuli. Subframes also profile a certain subset of participants. For example, Commerce frame has four core participants: Buyer, Seller, Money and Goods. Selling focuses on the Seller and the Goods, whereas Buying highlights the Buyer and the Goods, and Costing highlights the Goods and the Money. Some frames have versions involving an external causer, e.g. eating ~ feeding. Causative versions of frames are not simply included in those frames but are annotated separately. Such versions feature "Cause" as part of the label and the cs_ prefix in the tag. Thus, feeding is not simply Ingestion but Cause ingestion, hence tagged as cs_ingestion.

For the list of tags for annotating the clause type . The entry is built in the following way: <frame:subframe>. A subframe most often uniquely belongs to the superordinate frames, though multiple inheritance is sometimes also possible, e.g. transfer:losing and perception:losing. Original FrameNet labels are given in parentheses. The list is non-exhaustive and can be expanded if necessary. It should also be noted that the current version of the list contains only transitive frames, while intransitive ones may be added in future versions upon necessity.

3.3.17.Free translation (fe, fr, fg, fh)

The free translation tiers **fe** and **fr** give free translation of the utterance in question into English and Russian respectively. The translations are free, i.e. they usually do not reflect morphological and syntactical properties of the Ngansan original. The translations follow the common guidelines presented in Arkhipov (2020: Ch. 3), except for the texts coming from the NSLC.

One text was translated into German (tier **fg**), three texts have translation into Hungarian (tier **fh**).

3.3.18.Original Russian translation (ltr)

The tiers **ltr** contains the Russian original translation of texts included into the corpus as provided in the source (including draft translation written down during a session of oral transcription).

3.3.19. Automatic English translation (lte)

The tier **lte** contains the automatic translation into English produced with the DeepL software, in case the corresponding **fe** tier is empty.

3.3.20. Notes (nt)

The Notes tier (**nt**) contains notes related to the sentence. Sometimes the notes begin with the indication of who made the note (abbreviation as listed in 2.5.5) in square brackets. If this abbreviation is absent, it is unknown who wrote the notes.

3.4. Searching the corpus

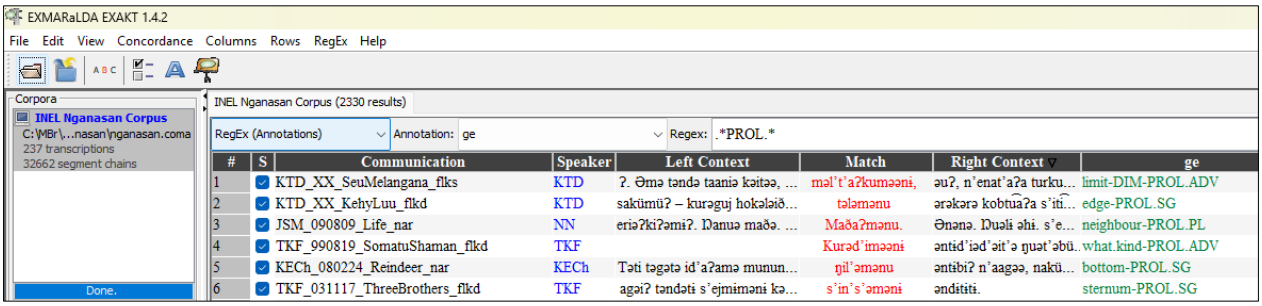
3.4.1. Search with EXMARaLDA EXAKT

The EXMARaLDA software suite includes EXAKT, an analysis and concordance tool.

In order to perform a search on the downloaded corpus files locally, the main metadata file (**nganasan.coma**) should be opened with "File > Open Corpus" command. (Creating a word list is optional.)

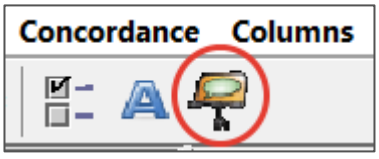
- One of the tiers should be selected in the main concordance window: either one of the annotation tiers (recommended; use "RegEx (Annotations)"; select any of tiers except **tx** under "Annotation") or the transcription tier (**tx**; use "RegEx (Transcription)").
- A search expression (interpreted as a regular expression[21]) should be specified in the **Regex** field. The matching results will be displayed in a column corresponding to the selected tier, e.g. "ge". Please refer to sections 3.2–3.3 and Appendix A3 for annotations used in the corpus.
- Note that only the part matching the search expression will be displayed in the column. E.g. when searching for a prolative case marker with "PROL" in tier **ge**, only "PROL" will be shown in the "ge" column. in order to have the complete word gloss displayed in the "ge" column, enter ".*PROL.*" as search expression.

Figure 7. EXAKT search window



- The "Match" column represents the content of the **tx** tier (word or sentence) corresponding to the annotation found in the specified tier. Double-click the entry in the "Match" column to display a portion of the entire transcript containing the example found (all tiers) in the lower part of the screen. After that, a click on the "Open Partitur" button will open the entire transcript in EXMARaLDA Partitur Editor.

  Figure 8. EXAKT: "Open Partitur" button



Please refer to EXMARaLDA manuals[22] for further details on using EXAKT and Partitur Editor.

[21] https://www.regular-expressions.info/, last accessed 28.04.2025.

[22] https://exmaralda.org/en/quickstart-documents/, last accessed: 28.04.2025.

3.4.2. Online search in Tsakorpus

Online search in the corpus is provided via Tsakorpus, an open-source search platform for linguistic corpora. The current version of the corpus can be accessed at https://inel.corpora.uni-hamburg.de/NganasanCorpus/search. The interface of online search is available in English and in Russian.

Tsakorpus offers the following possibilities:

- Search in multiple annotation tiers
- Search for substring, simple patterns (using *) or regular expressions
- Multi-word search (with or without distance restrictions)
- Negative queries (sentences which do NOT have a word with specified parameters)
- Search for sentences, words (wordforms), lemmas
- Search in a subcorpus
- Exporting search results as CSV/XLSX

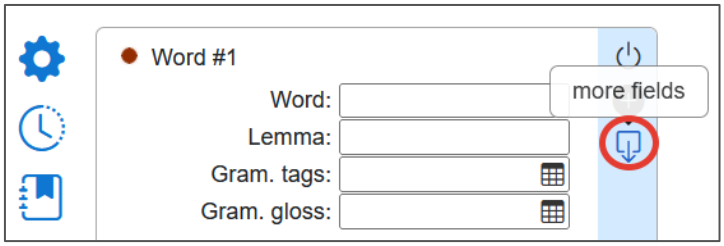
To run a search in the main transcription tier (**tx**) or in the word- and morph-level annotation tiers, "Language/tier" field should be set to "Nganasan" and the search expression(s) entered in one or more corresponding fields.

Table 22. Tsakorpus search fields and EXMARaLDA tiers: main transcription and word-/morph-level annotation

Tsakorpus search field	**Corresponding tier in EXMARaLDA**
Word	**tx**
Lemma	**mp** (stem)
Gram. tags	**ps**; grammar tags generated from grammatical glosses (**ge, gr**)
Gram. gloss	grammatical (i.e. affix) glosses (**ge, gr**)
Lexical gloss (en)*	lexical (i.e. stem) glosses (**ge**)
Lexical gloss (ru)*	lexical (i.e. stem) glosses (**gr**)
Morph. slot*	**mc**
Semantic role*	**SeR**
Syntactic function*	**SyF**
Inform. status*	**IST**
Borrowing*	**BOR**
Bor. phonetics*	**BOR-Phon**
Bor. morphology*	**BOR-Morph**
Code-switching*	**CS**
Geogr. coordinates*	**geo**

*To display search fields marked with *, click on "More fields" button next to "Word" and "Lemma" fields.

Figure 9. Tsakorpus interface: Show more fields



Please refer to sections 3.2–3.3 and Appendix A3 for annotations used in the corpus.

Lexical and grammatical glosses in Tsakorpus

Each word in Tsakorpus is internally split into stems (lexical items) and affixes (grammatical morphs).

The stem can be found by searching for its underlying (**mp**) form (e.g. “mat”) in the **Lemma** field, or by searching for its lexical gloss (e.g. “tent” / “чум”) in **Lex. gloss (en)** or **Lex. gloss (ru)** fields.

The affixes can be found by searching for the complete gloss (e.g. “LOC.SG”) in the **Gram. gloss** field, or with corresponding grammar tags (e.g. “loc,sg”) in the **Gram. tags** field (see next section for details on grammar tags).

To find only a particular allomorph, its form can be specified in curly braces following the gloss in the **Gram. gloss** field: “LOC.SG{tənu}”.

In case there exist more than one underlying form of an affix in **mp** tier, a particular underlying form can be specified in curly braces with underscore in the **Gram. gloss** field: “1SG.S{_təm}”.

Table 23. Stems and affixes in Tsakorpus

EXMARaLDA tier	**Word**	**Stem**	**Search field**	**Affix**	**Search field**
tx	matənu				
mb	ma-tənu	ma		tənu	Gram. gloss: LOC.SG{tənu}
mp	mat-ntənu	mat	Lemma: mat	ntənu	Gram. gloss: LOC.SG{_ ntənu }
ge	tent-LOC.SG	tent	Lex. gloss (en): tent	LOC.SG	Gram. gloss: LOC.SG Gram. tags: loc,sg
gr	чум-LOC.SG	чум	Lex. gloss (ru): чум	LOC.SG	Gram. gloss: LOC.SG Gram. tags: loc,sg
ps	n		Gram. tags: n		

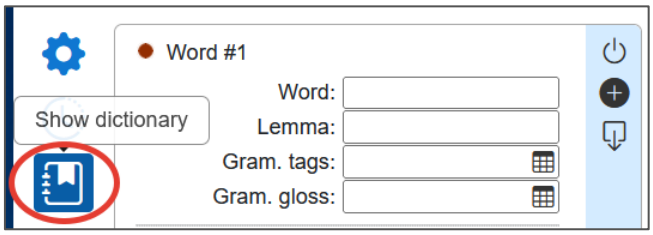
Note that some stems may have a glossing label similar to a grammatical gloss, e.g. NEG for “negative verb”. In such cases, this gloss should be entered into **Lex. gloss (en)** or **Lex. gloss (ru)** field, since it will not be treated as a grammatical gloss by Tsakorpus. It will however get a grammar tag (see next section) and can be found by searching for this tag. Such glosses are marked as “lexical” in Comment columns in table Table 27 (Appendix A3).

Table 24. Stems and affixes in Tsakorpus: stems with grammatical glosses

EXMARaLDA tier	**Word**	**Stem**	**Search field**	**Affix**	**Search field**
tx	n’iŋɨm				
mb	n’i-ŋɨ-m	n’i		ŋɨ m	Gram. gloss: INTER{ŋɨ} Gram. gloss: 1SG.S{m}
mp	n’i-ŋu-m	n’i	Lemma: n’i	ŋu m	Gram. gloss: INTER{_ŋu} Gram. gloss: 1SG.S{_m}
ge	NEG-INTER-1SG.S	NEG	Lex. gloss (en): NEG Gram. tags: neg	INTER 1SG.S	Gram. gloss: INTER Gram. tags: inter Gram. gloss: 1SG.S Gram. tags: pn1,pnsg,s
gr	NEG-INTER-1SG.S	NEG	Lex. gloss (ru): NEG Gram. tags: neg	INTER 1SG.S	Gram. gloss: INTER Gram. tags: inter Gram. gloss: 1SG.S Gram. tags: pn1,pnsg,s
ps	v		Gram. tags: v		

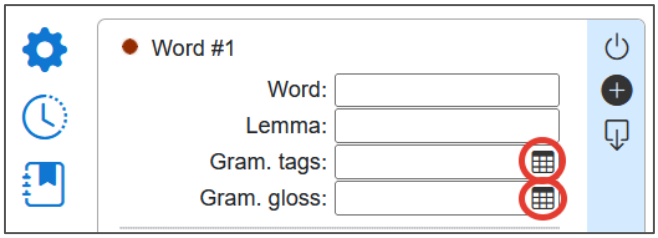
A list of lemmas (i.e. underlying forms of stems as represented in **mp** tier) along with their translations (lexical glosses) can be displayed with “Show dictionary” button.

Figure 10. Tsakorpus interface: Show dictionary



For most word- and morph-level annotation tiers, such as grammar tags, grammatical glosses, borrowings, one can either type in the search expression directly or choose from the list of available values. To open the list of values, click on the icon in the search field.

Figure 11. Tsakorpus interface: Show list of values



Grammatical glosses and grammar tags in Tsakorpus

In addition to grammatical glosses as present in tiers **ge**, **gr**, Tsakorpus provides another search possibility called "grammar tags." Grammar tags are generated by rules based on part of speech and glosses. For a complete list of glosses and grammar tags please refer to Table 27 (Appendix A3).

- Tags are assigned to an entire word and not to a particular morpheme in a word.
- By default, grammar tags are identical to a lower-case version of the corresponding gloss or part of speech label, e.g. (part of speech) "v" => "v", (gloss) "DU" => "du". Exceptions are mostly due to avoiding overlapping.
- Parts of speech can only be found with grammar tags since they do not have a corresponding gloss.
- Stems with glossing labels similar to a grammatical gloss, e.g. "NEG" for "negative verb" (see previous section), will also be assigned grammar tags. Such glosses are marked as "lexical" in Comments columns in Table 27 (Appendix A3). They can be found with either **Gram. tags** or **Lex. gloss (en)** / **Lex. gloss (ru)** fields, but not with **Gram. gloss** field.
- A group of related glosses can get more than one tag each to allow different ways of searching. E.g. of the two hortative markers, "CAR1" will get tags "car,car1" and "CAR2" will get tags "car,car2". Therefore, each of them separately can be found with their distinctive tags (resp. "car1" and "car2"), while searching for "car" will find both of them. Please refer to Table 27 (Appendix A3) for complete lists of tags.
- When a gloss consists of multiple components, such as "ABL.SG" or "3SG.MD", each of them is usually translated into a tag, e.g. "ABL.SG" => "abl" (ablative), "sg" (singular); "3SG.MD" => "pn3" (3 person), "pnsg" (singular number), "md" (middle conjugation). A search for tag "pn3" will return all words with any of glosses "3SG", "3DU", "3PL", "3SG.MD", etc.; a search for tag "abl" will return all words with any of glosses "ABL.SG", "ABL.PL1", "ABL.PL2", etc.
- When searching with glosses, the entire gloss should be entered as a search expression. E.g. a search for "PROL" will not find "PROL.SG". Use grammar tags if you need to search for a component of a complex gloss.
- Zero morphs have no overt segment in **mb**, **mp** tiers, and their glosses are shown in square brackets preceded by a dot in **ge**, **gr** tiers. In Tsakorpus, they can only be found with corresponding grammar tags.

E.g. a search for a gloss “GEN.SG” will not find a zero morph in “tent.[GEN.SG]”. Such wordforms can only be found with corresponding grammar tags, e.g. “gen,sg” in this case.

- When specifying more than one tag in a search expression, they can be combined with logical operators: AND (“,”), OR (“|”) and NOT (“~”), e.g. “v,inch,md” or “n,(abl|loc)”. When selecting tags from the list of values, multiple tags which are listed as belonging to the same Tsakorpus category (see Table 27 in Appendix A3) will be by default joined by OR (“|”), e.g. “(abl|loc)”. Multiple tags which are listed as belonging to different Tsakorpus categories will be by default joined by AND (“,”), e.g. “v,inch,md”.

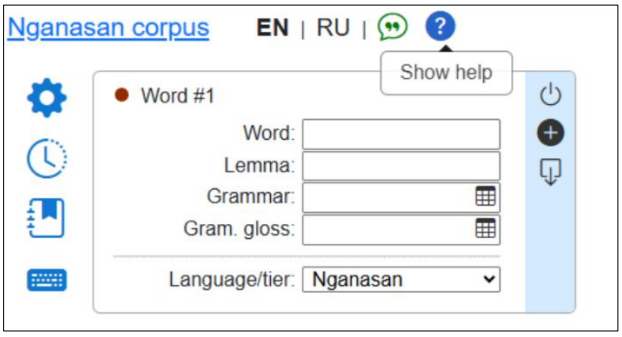
To search in one of the sentence-level annotation tiers, the search expression should be entered into “Word” field, and “Language/tier” field set to one of the following:

Table 25. Tsakorpus search fields and EXMARaLDA tiers: sentence-level annotation

Language/tier label	**Corresponding tier in EXMARaLDA**
Source text (Cyr.)	**so**
Nganasan transcription (Cyr.)	**st**
Nganasan transcription (Lat.)	**stl**
Russian translation	**fr**
English translation	**fe**
Orig. Russian translation	**ltr**
Automatic translation (English)	**lte**
Notes	**nt**
Sentence ID	**ref**

For further details please refer to Tsakorpus online help.

Figure 12. Tsakorpus interface: Show help



References

Aikhenvald A. Y. 2014. *The art of grammar: A practical guide*. Oxford University Press.

Arkhipov, Alexandre V. & Däbritz, Chris Lasse. 2018. Hamburg corpora for indigenous Northern Eurasian languages. *Tomsk Journal of Linguistics and Anthropology* 3(21). 9–18. https://doi.org/10.23951/2307-6119-2018-3-9-18

Arkhipov, Alexandre. 2020. *INEL corpora general transcription and annotation principles*. Szeged – Hamburg: University of Szeged & University of Hamburg. (Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 5.) https://doi.org/10.14232/wpcl.2020.5

Boldt, Ekaterina 1981. Nganasanskie teksty. In: *Skazki narodov Sibirskogo Severa* 4 [Nganasan texts: The stories of the peoples of the Siberian north, 4]. 159–177. Tomsk: Izdateľstvo Tomskogo universiteta.

Brykina, Maria & Gusev, Valentin & Szeverényi, Sándor & Wagner-Nagy, Beáta. 2018. *Nganasan Spoken Language Corpus (NSLC)*. Archived in Hamburger Zentrum für Sprachkorpora. Version 0.2. Publication date 12.06.2018. http://hdl.handle.net/11022/0000-0007-C6F2-8

Däbritz, Chris Lasse 2021. *Topik, Fokus und Informationsstatus: Modellierung am Material nordwestsibirischer Sprachen.* (Language, Context and Cognition 17). Berlin & Boston: De Gruyter.

Fillmore, Ch. J. & C. F. Baker. 2010. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 313–339. Oxford: Oxford University Press.

Fillmore, Ch. J., Ch. R. Johnson & M. R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16(3). 235–250.

GOST 2001 — *GOST 7.79–2000. Sistema stadartov po informacii, bibliotečnomu i izdatel'skomu delu. Pravila transliteracii kirillovskogo pis'ma latinskim alfavitom* [GOST 7.79–2000.The system of standards in information, librarianship and publishing. Rules of transliteration of Cyrillic letters into the Latin alphabet]. Minsk: Mežgosudarstvennyj sovet po standartizacii, metrologii i sertifikacii. https://protect.gost.ru/document.aspx?control=7&id=130715, https://ifap.ru/library/gost/7792000.pdf

Götze, Michael & Weskott, Thomas & Endriss, Cornelia & Fiedler, Ines & Hinterwimmer, Stefan & Petrova, Svetlana & Schwarz, Anne & Skopeteas, Stavros & Stoel, Ruben. 2007. Information structure. In Dipper, Stephanie, et al. (eds.), *Information structure in cross-linguistic corpora*, 147–187. Potsdam: Universitätsverlag Potsdam. (Interdisciplinary Studies on Information Structure 07) https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/2036/file/Kapitel6_07.pdf.

Gusev, Valentin 2010. Nganasan texts [Нганасанские тексты]. In: Filchenko, Andrey (ed.): *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Volume 1. Tomsk: Vajar.

Gusev, Valentin 2012. Nganasan texts [Нганасанские тексты]. In: Filchenko, Andrey (ed.): *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Volume 2. Tomsk: Vajar.

Gusev, Valentin 2017: Nganasanskij mif „Dve pticy" [Нганасанский миф «Две птицы»]. *Rodnoj jazyk [Родной язык]*, 1(6): 169-198.

Haig, Geoffrey & Schnell, Stefan. 2014. *Annotations using GRAID (Grammatical Relations and Animacy In Discourse): Introduction and guidelines for annotators*. Version 7.0. Available online at https://multicast.aspra.uni-bamberg.de/data/pubs/graid/Haig+Schnell2014_GRAID-manual_v7.0.pdf.

Haspelmath, Martin 2025. Nonverbal clause constructions. To appear in *Language and Linguistics Compass*. Lingbuzz/006673. https://ling.auf.net/lingbuzz/006673 [Accessed: 28.04.2025].

Helimski, Eugen 1998. Nganasan. In Abondolo, Daniel (ed.): *The Uralic Languages*. London: Routledge, 480–515.

Junghanns, Uwe & Zybatow, Gerhild. 2009. Grammatik und Informationsstruktur. In: Gutschmidt, Karl et al. (eds.). *Die slavischen Sprachen.* Handbücher zur Sprach- und Kommunikationswissenschaft. Bd. 32, 2. Berlin: De Gruyter, 684–707.

Koch, Peter 2012. Location, existence, and possession: A constructional-typological exploration. *Linguistics* 50(3), 533–603.

Machkinis I. E. 1980. Nganasanskie teksty. In: *Skazki narodov Sibirskogo Severa* 3 [Nganasan texts: The stories of the peoples of the Siberian north, 3]. 3-54. Tomsk: Izdateľstvo Tomskogo universiteta.

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

Nagy, Beáta B. 1996. A Sámán-tó legendája. In: *Néprajz és Nyelvtudomány*. Vol. 37. 219-222.

Prokofiev, G. N. 1937. Nganasanskij (tavgijskij) dialekt. [The Nganasan (Tavgi) dialect]. In: *Jazyki i pismennost' narodov Severa.* Chasť 1. 53–77. Moscow–Leningrad: Uchpedgiz.

Riester, Arndt & Baumann, Stefan 2014: *RefLex Scheme – Annotation Guidelines*. http://dx.doi.org/10.18419/opus-9011

Stolyarova A. K. 1981. Nganasanskie teksty. In: *Skazki narodov Sibirskogo Severa* 4 [Nganasan texts: The stories of the peoples of the Siberian north, 4]. 159–177. Tomsk: Izdateľstvo Tomskogo universiteta.

Wagner-Nagy, Beáta & Brykina, Maria 2020. Nganasan texts [Нганасанские тексты]. In: Filchenko, Andrey (ed.): *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Volume 6. Tomsk: Agraf-Press, Vajar, p. 152–232.

Wagner-Nagy, Beáta & Szeverényi, Sándor & Gusev, Valentin. 2018. *User's guide to Nganasan Spoken Language Corpus*. Szeged – Hamburg: University of Szeged & University of Hamburg. (Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 1.) https://doi.org/10.14232/wpcl.2018.1.

Wagner-Nagy, Beáta & Gusev, Valentin 2017. Nganasan texts [Нганасанские тексты]. In: Filchenko, Andrey (ed.): *Annotated Folk and Daily Prose Texts in the Languages of Ob-Yenissei Linguistic Area*. Volume 5. Tomsk: Vajar, 249–216.

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

Wagner-Nagy, Beáta 2018. *A grammar of Nganasan*. Leiden: Brill.

Appendix A1. Nganasan speakers

Appendix A1 contains the full lists of speakers who contributed to the corpus as storytellers and/or who provided oral transcription of texts with their codes, except for those who unfortunately remain unidentified. As mentioned in the section 1.6, we gratefully acknowledge everyone's contribution and realize that this corpus could never be possible, unless as a result of a joint effort of all the people listed here.

Table 26. Nganasan speakers

Name	**Code**	**Dialect**
Aksyonova, Svetlana Sy`gakovna	ASS	Avam
Chebodaeva, Zinaida Solobteevna	ChZS	Avam
Chunanachar, Aleksej Nikolaevich	ChAN	Avam
Chunanchar, Kejkumte Dyalevich	ChKD	Avam
Chunanchar, Nambu Chutovich	ChNCh	unknown
Chunanchar, Nikolaj Sankevich	ChNS	Avam
Chunanchar, Nina Dentumeevna	ChND	Avam
Goricheva, Yuliya Madyureevna	GJM	Avam
Kokore, Berime Neamovna	KBN	Avam
Kokore, Ekaterina Chubovna	KECh	Avam
Kokore, Karabej Xidibtievich	KKH	unknown
Kokore, Korore	KK	Avam
Kokore, Nyantu Dyubeevich	KND	Vadeyev
Kokore, Toryukyusi	KoT	unknown
Kosterkin, Boris Dyuxodoevich	KBD	Avam
Kosterkin, Demnime Dyuxodoevich	KDD	Avam
Kosterkin, Dyul`simyaku Demnimeevich	KDDe	Avam
Kosterkin, Igor` Demnimeevich	KID	Avam
Kosterkin, Leonid Tubyakovich	KLT	Avam
Kosterkin, Torunte	KT	Avam
Kosterkin, Tubyaku Dyuxodoevich	KTD	Avam
Kosterkina, Ekaterina Subobteevna	KES	Avam
Kosterkina, Nadezhda Tubyakovna	KNT	Avam
Kosterkina, Sandymyaku Chajxoreevna	KSCh	unknown
Kosterkina, Valentina Bintaleevna	KVB	Avam
Kosterkina, Xolyamte Dyatovna	KH	Avam
Kudryakova, Svetlana Mojbovna	KuSM	Avam
Kupchik, Dar`ya	KD	Avam
Kupchik, Serafima Mudimeevna	KSM	Avam
Kuzenko, Tat`yana Ton`dyuleevna	KTT	Avam
Logvinova, Nina Demnimeevna	LND	Avam
Lyubov` Salirovna		Avam
Mirny`x, Denchude Nuteevich	MDN	Avam
Momde, Aleksandr Chaleevich	MACh	Avam

Name	Code	Dialect
Momde, Dintade Bul`chuevich	MDB	Vadeyev
Momde, Ekaterina Ton`dyuleevna	MET	Avam
Momde, Sajbore Togubteevich	MST	Vadeyev
Momde, Vera Lajmareevna	MVL	Avam
Momde, Xosyu Chaleevich	MHCh	Avam
Okko, Mukalyuo Xidibteevich	OMX	Vadeyev
Porbin, Katamu Kondakovich	PKK	Avam
Porbin, Kyudapte Mory`chesovich	PKM	Avam
Porbin, Salira My`dovich	PSM	Avam
Porbin, Tolomu Komsapteevich	PTK	Avam
Porbin, Use	PU	Avam
Porbin, Vasilij Fenakovich	PVF	Avam
Porbina, Evdokiya Demnimeevna	PED	Avam
Porbina, Mariya Madyureevna	PMM	Avam
Porbina, Mariya Yagulaevna	PMJ	Avam
Porbina, Nadezhda Kajboreevna	PNK	Avam
Porbina, Nelyutasi Fominichna	PNF	Avam
Porbina, Oktyabrina Yaguleevna	POJ	Avam
Porbina, Turgule Xilipteevna	PTH	Avam
Porbina, Xella Lajmareevna	PHL	Avam
Porbina, Zoya Chuboevna	PZCh	Avam
Porotova, Galina Fenakovna	PGF	Avam
Porotova, Tat`yana Chajxoreevna	PTCh	Vadeyev
Sidel`nikova, Evgeniya Chebyakovna	SE	Avam
Sovalova, Ekaterina Nikolaevna	SEN	Avam
Truchin Mola Tulageevich	TMT	Avam
Tuglakov, Kurumaku Fyodorovich	TKF	Avam
Turdagin, Aleksej Molkovich	TAM	Avam
Turdagin, Dentume	TD	Avam
Turdagin, Lomkaj Xursabteevich	TLH	Avam
Turdagin, Molitasi Singimeevich	TMS	Avam
Turdagin, Motyumyaku Sochupteevich	TMoS	Avam
Turdagin, Nikolaj Y`saevich	TNY	Avam
Turdagin, Nirie Nikolaevich	TNN	Avam
Turdagin, Nobyaku Sogumovich	TNS	Avam
Turdagin, Numumu Xursabteevich	TNH	Avam
Turdagin, Xonyaku Lamkaevich	THL	Avam
Turdagin, Y`saj	TY	Vadeyev
Turdagina, Ladalyu Samnareevna	TLS	Avam
Turdagina, Larisa Yandipteevna	TLJ	Vadeyev

Name	Code	Dialect
Turdagina, Lodun Nadeevna	TLN	Avam
Turdagina, Lyubov` Muruyevna	TLM	Avam
Turdagina, Mariya Pulkovna	TMP	Avam
Turdagina, Nadezhda Kajbyureevna	TNK	Avam
Turdagina, Natal`ya Kundupteevna	TNKu	Avam
Turdagina, Nina Ton`dyuleevna	TNT	unknown
Turdagina, Tamara Sochubteevna	TTS	Avam
Turdagina, Zoya Dentumeevna	TZD	Avam
Turkina, Tat`yana Demnimeevna	TTD	Avam
Xorbi, Anatolij Xuchurevich	HAH	Vadeyev
Yarockaya, Dilimyaku Sochupteevna	JDS	Avam
Yarockaya, E. Ch.	JECh	unknown
Yarockaya, Marina Hosyuvna	JMH	unknown
Yarockaya, Mariya Dyuntovna	JMD	Avam
Yarockaya, Sy`ku Madyureevna	JSM	Avam
Yarockij, Apche By`ngapteevich	JAB	Avam
Yarockij, Dyalamte Xorsabteevich	JDH	Avam
Yarockij, Numare	JN	Avam
Zharkova, Larisa Xotureevna	ZhLH	Vadeyev
unknown	NN	unknown

Appendix A2. Story types

The list below provides an overview of story types and subtypes (motives) that occur most frequently or reflect best the content of the texts. Story subtypes are separated from story types by colon.

Folklore texts (flks, flkd, flk):

- Sitaby (epic tales)
- Sitaby fragment
- Legend: War with Nenets, Dog lake, Chukchis
- Myth: Creating Earth, Tuu Benke brings summer, Getting fire
- Story about shaman: Becoming shaman, Diving shaman, Shaman form the Stone Clan, Two shaman brothers, Woman shaman, …
- Short funny tale: Short story about Djajku, Short story about Ibula
- Tale about animals: Comparisons (of wild and domestic animals), Fox, Reindeer and lemming, Three big animals and three small animals
- Tale about barusi (devil)
- Tale about sigie (ogre): Creating mosquitos, Turning into an ogre, Killing an ogre and his children, Going to Jamal
- Short tale: Bad wife, Bringing warm weather, Eating a child, Playing hide-and-seek with the eyes, Stealing wife, …
- Long tale: A story about tundra chiefs and Russian tsar, Brother and sister in squirrel clothings, Dolgans killing a woman eating people, Handicapped young man, Kidnapping Nganasan's wife, Nenets sacrifices, Seeking for a wife, Two brothers and a hole, Two brothers and their children, Two brothers Momde, …
- Fairytale with elements from Russian folklore: Two horses, Nyomu Kamlegu Ny, …

Narratives (nar):

- Biographical: Family, Marrying off, Modern life, Name, Own life, Shamans, Wandering with reindeer, …
- Procedural (a storyteller explains, how he usually does things, so-called "how-to" texts): Boat, Bread, Coat, Dried fish, Fishing, Knife, Marrying off, Mittens, Net, Parka, Reindeer, Sledge, Tent, Traps for polar foxes, …
- Real story: Accident, Illness, Lightning, Meeting Russians, Shaman, Wandering, …
- Puzzles
- Describing pictures

Conversations (conv)

- Biographical
- Procedural
- Conversation (not specified)

Songs (song)

- Child song
- Keinyrsja song (allegorical song)
- Personal song

Appendix A3. Morpheme glossing labels (tiers ge, gr) and Tsakorpus grammar tags

Table 27 lists the glossing labels used in tiers **ge**, **gr** and corresponding grammar tags for use in Tsakorpus online search. It is sorted by category.

Table 28 presents a list of affixes and clitics in their morphophonological form (tier **mp**), sorted alphabetically by gloss. The so-called zero-consonant -^{c} (also referred to as empty slot, cf. Wagner-Nagy 2018:67-71) is listed in this table as a separate morpheme, even though it couldn't be represented in the **mp** tier for technical reasons, since it is always realized by zero morphs. The same holds true for two other morphemes: *-s* (ANDV, VBLZs, DRVs), and *-r* (FRQ), which are shown in the **mp** tier if they have overt allomorphs and absent there in case a zero allomorph is used. Morphemes that may need further research in respect to their morphophonological form and function are shown in grey.

Table 27. List of morpheme glossing labels by category

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
Person and number				
1SG	1 person singular	pn1,pnsg	persnum-person,persnum-number	only in combinations
1SG.S	1 person singular, subjective conjugation	pn1,pnsg,s	persnum-person,persnum-number,v-conj	
1SG.S.EXCL	1 person singular, subjective conjugation, exclamative	pn1,pnsg,s,excl	persnum-person,persnum-number,v-conj,cl	
1SG.MD	1 person singular, middle conjugation	pn1,pnsg,md	persnum-person,persnum-number,v-conj	
1SG.SG.O	1 person singular, objective conjugation, singular object	pn1,pnsg,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
1SG.NSG.O	1 person singular, objective conjugation, non-singular object	pn1,pnsg,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
1DU	1 person dual	pn1,pndu	persnum-person,persnum-number	only in combinations
1DU.S/SG.O	1 person dual, subjective conjugation / objective conjugation, singular object	pn1,pndu,s,o,objsg	persnum-person,persnum-number,v-conj,v-conj,persnum-number	
1DU.S/SG.O.EXCL	1 person dual, subjective conjugation / objective conjugation, singular object, exclamative	pn1,pndu,s,o,objsg,excl	persnum-person,persnum-number,v-conj,v-conj,persnum-number,cl	
1DU.MD	1 person dual, middle conjugation	pn1,pndu,md	persnum-person,persnum-number,v-conj	
1DU.NSG.O	1 person dual, objective conjugation, non-singular object	pn1,pndu,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
1PL	1 person plural	pn1,pnpl	persnum-person,persnum-number	only in combinations
1PL.S/SG.O	1 person plural, subjective conjugation / objective conjugation, singular object	pn1,pnpl,s,o,objsg	persnum-person,persnum-number,v-conj,v-conj,persnum-number	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
1PL.S/SG.O.EXCL	1 person plural, subjective conjugation / objective conjugation, singular object, exclamative	pn1,pnpl,s,o,objsg,excl	persnum-person,persnum-number,v-conj,v-conj,persnum-number,cl	
1PL.MD	1 person plural, middle conjugation	pn1,pnpl,md	persnum-person,persnum-number,v-conj	
1PL.MD.EXCL	1 person plural, middle conjugation, exclamative	pn1,pnpl,md,excl	persnum-person,persnum-number,v-conj,cl	
1PL.NSG.O	1 person plural, objective conjugation, non-singular object	pn1,pnpl,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
1PL.NSG.O.EXCL	1 person plural, objective conjugation, non-singular object, exclamative	pn1,pnpl,o,objnsg,excl	persnum-person,persnum-number,v-conj, persnum-number,cl	
2SG	2 person singular	pn2,pnsg	persnum-person,persnum-number	only in combinations
2SG.S	2 person singular, subjective conjugation	pn2,pnsg,s	persnum-person,persnum-number,v-conj	
2SG.MD	2 person singular, middle conjugation	pn2,pnsg,md	persnum-person,persnum-number,v-conj	
2SG.SG.O	2 person singular, objective conjugation, singular object	pn2,pnsg,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
2SG.NSG.O	2 person singular, objective conjugation, non-singular object	pn2,pnsg,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
2DU	2 person dual	pn2,pndu	persnum-person,persnum-number	only in combinations
2DU.S/SG.O	2 person dual, subjective conjugation / objective conjugation, singular object	pn2,pndu,s,o,objsg	persnum-person,persnum-number,v-conj,v-conj,persnum-number	
2DU.MD	2 person dual, middle conjugation	pn2,pndu,md	persnum-person,persnum-number,v-conj	
2DU.MD.EXCL	2 person dual, middle conjugation, exclamative	pn2,pndu,md,excl	persnum-person,persnum-number,v-conj,cl	
2DU.NSG.O	2 person dual, objective conjugation, non-singular object	pn2,pndu,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
2PL	2 person plural	pn2,pnpl	persnum-person,persnum-number	only in combinations
2PL.MD	2 person plural, middle conjugation	pn2,pnpl,md	persnum-person,persnum-number,v-conj	
2PL.MD.EXCL	2 person plural, middle conjugation, exclamative	pn2,pnpl,md,excl	persnum-person,persnum-number,v-conj,cl	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
2PL.S/SG.O	2 person plural, subjective conjugation / objective conjugation, singular object	pn2,pnpl,s,o,objsg	persnum-person,persnum-number,v-conj,v-conj,persnum-number	
2PL.S/SG.O.EXCL	2 person plural, subjective conjugation / objective conjugation, singular object, exclamative	pn2,pnpl,s,o,objsg,excl	persnum-person,persnum-number,v-conj,v-conj,persnum-number,cl	
2PL.NSG.O	2 person plural, objective conjugation, non-singular object	pn2,pnpl,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
3SG	3 person singular	pn3,pnsg	persnum-person,persnum-number	only in combinations
3SG.S	3 person singular, subjective conjugation	pn3,pnsg,s	persnum-person,persnum-number,v-conj	
3SG.MD	3 person singular, middle conjugation	pn3,pnsg,md	persnum-person,persnum-number,v-conj	
3SG.SG.O	3 person singular, objective conjugation, singular object	pn3,pnsg,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
3SG.NSG.O	3 person singular, objective conjugation, non-singular object	pn3,pnsg,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
3DU	3 person dual	pn3,pndu	persnum-person,persnum-number	only in combinations
3DU.S	3 person dual, subjective conjugation	pn3,pndu,s	persnum-person,persnum-number,v-conj	
3DU.S.EXCL.INT	3 person dual, subjective conjugation prosodical exclamative	pn3,pndu,s excl,excl.int	persnum-person,persnum-number,v-conj,cl,cl	
3DU.MD	3 person dual, middle conjugation	pn3,pndu,md	persnum-person,persnum-number,v-conj	
3DU.SG.O	3 person dual, objective conjugation, singular object	pn3,pndu,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
3DU.NSG.O	3 person dual, objective conjugation, non-singular object	pn3,pndu,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
3PL	3 person plural	pn3,pnpl	persnum-person,persnum-number	only in combinations
3PL.S	3 person plural, subjective conjugation	pn3,pnpl,s	persnum-person,persnum-number,v-conj	
3PL.MD	3 person plural, middle conjugation	pn3,pnpl,md	persnum-person,persnum-number,v-conj	
3PL.MD.EXCL	3 person plural, middle conjugation, exclamative	pn3,pnpl,md,excl	persnum-person,persnum-number,v-conj,cl	
3PL.SG.O	3 person plural, objective conjugation, singular object	pn3,pnpl,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
3PL.NSG.O	3 person plural, objective conjugation, non-singular object	pn3,pnpl,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
3PL.NSG.O.EXCL	3 person plural, objective conjugation, non-singular object, exclamative	pn3,pnpl,o,objnsg,excl	persnum-person,persnum-number,v-conj, persnum-number,cl	
Personal pronouns				
PRO1	personal pronoun, 1 person	pn1	persnum-person	lexical
PRO1SG	personal pronoun, 1 person singular	pn1,pnsg	persnum-person,persnum-number	lexical
PRO1DU	personal pronoun, 1 person dual	pn1,pndu	persnum-person,persnum-number	lexical
PRO1PL	personal pronoun, 1 person plural	pn1,pnpl	persnum-person,persnum-number	lexical
PRO2	personal pronoun, 2 person	pn2	persnum-person	lexical
PRO2SG	personal pronoun, 2 person singular	pn2,pnsg	persnum-person,persnum-number	lexical
PRO2DU	personal pronoun, 2 person dual	pn2,pndu	persnum-person,persnum-number	lexical
PRO2PL	personal pronoun, 2 person plural	pn2,pnpl	persnum-person,persnum-number	lexical
PRO3	personal pronoun, 3 person	pn3	persnum-person	lexical
PRO3SG	personal pronoun, 3 person singular	pn3,pnsg	persnum-person,persnum-number	lexical
PRO3DU	personal pronoun, 3 person dual	pn3,pndu	persnum-person,persnum-number	lexical
PRO3PL	personal pronoun, 3 person plural	pn3,pnpl	persnum-person,persnum-number	lexical
Nominal categories				
Number				
SG	singular number	sg	n-num	only in combinations
DU	dual number	du	n-num	
PL	plural number	pl	n-num	only in combinations
Possession				
	possessive	poss	n-poss	only as a tag
OBL	oblique	obl,poss	n:case,n-poss	only in combinations
OBL.1SG	oblique, 1 person, singular	obl,poss,pn1,pnsg	n:case,n-poss,persnum-person,persnum-number	
OBL.1DU	oblique, 1 person dual	obl,poss,pn1,pndu	n:case,n-poss,persnum-person,persnum-number	
OBL.1PL	oblique, 1 person plural	obl,poss,pn1,pnpl	n:case,n-poss,persnum-person,persnum-number	
OBL.1PL.EXCL	oblique, 1 person plural, exclamative	obl,poss,pn1,pnpl,excl	n:case,n-poss,persnum-person,persnum-number,cl	
OBL.2SG	oblique, 2 person singular	obl,poss,pn2,pnsg	n:case,n-poss,persnum-person,persnum-number	
OBL.2DU	oblique, 2 person dual	obl,poss,pn2,pndu	n:case,n-poss,persnum-person,persnum-number	
OBL.2PL	oblique, 2 person plural	obl,poss,pn2,pnpl	n:case,n-poss,persnum-person,persnum-number	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
OBL.2PL.EXCL	oblique, 2 person plural, exclamative	obl,poss,pn2,pnpl,excl	n:case,n-poss,persnum-person,persnum-number,cl	
OBL.3SG	oblique, 3 person singular	obl,poss,pn3,pnsg	n:case,n-poss,persnum-person,persnum-number	
OBL.3DU	oblique, 3 person dual	obl,poss,pn3,pndu	n:case,n-poss,persnum-person,persnum-number	
OBL.3PL	oblique, 3 person plural	obl,poss,pn3,pnpl	n:case,n-poss,persnum-person,persnum-number	
Case				
NOM	nominative case	nom	n:case	only in combinations
NOM.SG	nominative case, singular number	nom,sg	n:case,n-num	
NOM.SG.1SG	nominative case, singular number, 1 person, singular	nom,sg,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.1DU	nominative case, singular number, 1 person dual	nom,sg,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.1PL	nominative case, singular number, 1 person plural	nom,sg,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.1PL.EXCL	nominative case, singular number, 1 person plural, exclamative	nom,sg,poss,pn1,pnpl,excl	n:case,n-num,n-poss,persnum-person,persnum-number,cl	
NOM.SG.2SG	nominative case, singular number, 2 person singular	nom,sg,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.2DU	nominative case, singular number, 2 person dual	nom,sg,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.2PL	nominative case, singular number, 2 person plural	nom,sg,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.2PL.EXCL	nominative case, singular number, 2 person plural, exclamative	nom,sg,poss,pn2,pnpl,excl	n:case,n-num,n-poss,persnum-person,persnum-number,cl	
NOM.SG.3SG	nominative case, singular number, 3 person singular	nom,sg,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.3DU	nominative case, singular number, 3 person dual	nom,sg,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.3PL	nominative case, singular number, 3 person plural	nom,sg,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.SG.3PL.EXCL	nominative case, singular number, 3 person plural, exclamative	nom,sg,poss,pn3,pnpl,excl	n:case,n-num,n-poss,persnum-person,persnum-number,cl	
NOM.DU	nominative case, dual number	nom,du	n:case,n-num	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
NOM.PL	nominative case, plural number	nom,pl	n:case,n-num	
NOM.PL.1SG	nominative case, plural number, 1 person, singular	nom,pl,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.1DU	nominative case, plural number, 1 person dual	nom,pl,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.1PL	nominative case, plural number, 1 person plural	nom,pl,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.2SG	nominative case, plural number, 2 person singular	nom,pl,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.2DU	nominative case, plural number, 2 person dual	nom,pl,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.2PL	nominative case, plural number, 2 person plural	nom,pl,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.2PL.EXCL	nominative case, plural number, 2 person plural, exclamative	nom,pl,poss,pn2,pnpl,excl	n:case,n-num,n-poss,persnum-person,persnum-number,cl	
NOM.PL.3SG	nominative case, plural number, 3 person singular	nom,pl,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.3DU	nominative case, plural number, 3 person dual	nom,pl,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.3PL	nominative case, plural number, 3 person plural	nom,pl,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
NOM.PL.3PL.EXCL	nominative case, plural number, 3 person plural, exclamative	nom,pl,poss,pn3,pnpl,excl	n:case,n-num,n-poss,persnum-person,persnum-number,cl	
GEN	genitive case	gen	n:case	only in combinations
GEN.SG	genitive case, singular number	gen,sg	n:case,n-num	
GEN.SG.1SG	genitive case, singular number, 1 person, singular	gen,sg,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.1DU	genitive case, singular number, 1 person dual	gen,sg,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.1PL	genitive case, singular number, 1 person plural	gen,sg,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.2SG	genitive case, singular number, 2 person singular	gen,sg,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.2DU	genitive case, singular number, 2 person dual	gen,sg,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
GEN.SG.2PL	genitive case, singular number, 2 person plural	gen,sg,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.3SG	genitive case, singular number, 3 person singular	gen,sg,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.3DU	genitive case, singular number, 3 person dual	gen,sg,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.SG.3PL	genitive case, singular number, 3 person plural	gen,sg,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.DU	genitive case, dual number	gen,du	n:case,n-num	
GEN.PL	genitive case, plural number	gen,pl	n:case,n-num	
GEN.PL.1SG	genitive case, plural number, 1 person, singular	gen,pl,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.1DU	genitive case, plural number, 1 person dual	gen,pl,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.1PL	genitive case, plural number, 1 person plural	gen,pl,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.2SG	genitive case, plural number, 2 person singular	gen,pl,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.2DU	genitive case, plural number, 2 person dual	gen,pl,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.2PL	genitive case, plural number, 2 person plural	gen,pl,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.3SG	genitive case, plural number, 3 person singular	gen,pl,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.3DU	genitive case, plural number, 3 person dual	gen,pl,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
GEN.PL.3PL	genitive case, plural number, 3 person plural	gen,pl,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC	accusative case	acc	n:case	only in combinations
ACC.SG	accusative case, singular number	acc,sg	n:case,n-num	
ACC.SG.1SG	accusative case, singular number, 1 person, singular	acc,sg,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.1DU	accusative case, singular number, 1 person dual	acc,sg,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.1PL	accusative case, singular number, 1 person plural	acc,sg,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.2SG	accusative case, singular number, 2 person singular	acc,sg,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
ACC.SG.2DU	accusative case, singular number, 2 person dual	acc,sg,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.2PL	accusative case, singular number, 2 person plural	acc,sg,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.3SG	accusative case, singular number, 3 person singular	acc,sg,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.3DU	accusative case, singular number, 3 person dual	acc,sg,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.SG.3PL	accusative case, singular number, 3 person plural	acc,sg,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.DU	accusative case, dual number	acc,du	n:case,n-num	
ACC.PL	accusative case, plural number	acc,pl	n:case,n-num	
ACC.PL.1SG	accusative case, plural number, 1 person, singular	acc,pl,poss,pn1,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.1DU	accusative case, plural number, 1 person dual	acc,pl,poss,pn1,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.1PL	accusative case, plural number, 1 person plural	acc,pl,poss,pn1,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.2SG	accusative case, plural number, 2 person singular	acc,pl,poss,pn2,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.2DU	accusative case, plural number, 2 person dual	acc,pl,poss,pn2,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.2PL	accusative case, plural number, 2 person plural	acc,pl,poss,pn2,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.3SG	accusative case, plural number, 3 person singular	acc,pl,poss,pn3,pnsg	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.3DU	accusative case, plural number, 3 person dual	acc,pl,poss,pn3,pndu	n:case,n-num,n-poss,persnum-person,persnum-number	
ACC.PL.3PL	accusative case, plural number, 3 person plural	acc,pl,poss,pn3,pnpl	n:case,n-num,n-poss,persnum-person,persnum-number	
LOC	locative case	loc	n:case	only in combinations
LOC.SG	locative case, singular number	loc,sg	n:case,n-num	
LOC.PL	locative case, plural number	loc,pl	n:case,n-num	
LAT	lative case	lat	n:case	only in combinations
LAT.SG	lative case, singular number	lat,sg	n:case,n-num	
LAT.PL	lative case, plural number	lat,pl	n:case,n-num	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
LAT.PL.EXCL	lative case, plural, exclamative	lat,pnpl,excl	n:case,persnum-number,cl	
ABL	ablative case	abl	n:case	only in combinations
ABL.SG	ablative case, singular number	abl,sg	n:case,n-num	
ABL.PL	ablative case, plural number	abl,pl	n:case,n-num	
ABL.PL2	ablative case, plural	abl.pl2	n:case	
PROL	prolative case	prol	n:case	only in combinations
PROL.SG	prolative case, singular number	prol,sg	n:case,n-num	
PROL.PL	prolative case, plural number	prol,pl	n:case,n-num	
COM	comitative case	com	n:case	only in combinations
COM.SG	comitative case, singular number	com,sg	n:case,n-num	
COM.PL	comitative case, plural number	com,pl	n:case,n-num	
VOC	vocative case	voc	n:case	only in combinations
VOC.SG	vocative case, singular number	voc,sg	n:case,n-num	
Case of adverbials				
LOC.ADV	locative case of adverbials	loc,advcase	n:case,n:misc	
PROL.ADV	prolative case of adverbials	prol,advcase	n:case,n:misc	
	lative case of adverbials	lat,latadv,advcase	n:case,n:misc,n:misc	only as a tag
LAT.ADV1	lative case of adverbials 1	lat,latadv,latadv1,advcase	n:case,n:misc,n:misc,n:misc	
LAT.ADV2	lative case of adverbials 2	lat,latadv,latadv2,advcase	n:case,n:misc,n:misc,n:misc	
ABL.ADV	ablative case of adverbials	abl,advcase,abladv	n:case,n:misc,n:misc	
Nominalizations				
	auditive (any)	aud	deriv-n	only as a tag
AUD1	auditive 1	aud,aud1	deriv-n,deriv-n	
AUD2	auditive 2	aud,aud2	deriv-n,deriv-n	
AUD2.EXCL	auditive 2, exclamative	aud,aud2,excl	deriv-n,deriv-n,cl	
%NMLZ	nominalization (possible)	%nmlz	deriv-n	
ABSTR	abstract nominal	abstr	deriv-n	
ACT1	action nominal 1	act1	deriv-n	
ACT2	action nominal 2	act2	deriv-n	
ACT3	action nominal 3	act3	deriv-n	
ACT4	action nominal 4	act4	deriv-n	
AG	agent nominal	ag	deriv-n	
APPR	approximative nominal	appr	deriv-n	
NMLZ.INSTR	instrumental nominal	nmlz.instr	deriv-n	
NMLZ	nominalization	nmlz	deriv-n	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
NMLZ.LOC	locative nominal	nmlz.loc	deriv-n	
NMLZt	nominalization in -t	nmlzt	deriv-n	
	temporal nominal (any)	tmp	deriv-n	only as a tag
TMP1	temporal nominal 1	tmp,tmp1	deriv-n,deriv-n	
TMP2	temporal nominal 2	tmp,tmp2	deriv-n,deriv-n	
NMLZ.INDF	indefinite nominal	nmlz.indf	deriv-n	
Destinative				
	destinative	dst	n-misc	only as a tag
DST1	destinative 1	dst,dst1	n-misc,n-misc	
DST1.PL	destinative 1, plural number	dst,dst1,pl	n-misc,n-misc,n-num	
DST2	destinative 2	dst,dst2	n-misc,n-misc	
DST.IRR3	irreal destinative	dst,dstirr3	n-misc,v:mood	
Other nominal categories				
ANT	nominal anteriority	ant	n-misc	
AUG	augmentative	aug	n-misc	
DIM	diminutive	dim	n-misc	
DYA	dyadic	dya	n-misc	
SRLAT	superlative	srlat	n-misc	
SEL	selective	sel	n-misc	
PEJOR	pejorative	pejor	n-misc	
FEM	feminitive	fem	n-misc	
N.ABSTR	abstract nominal	n.abstr	n-misc	
N.CAR	caritive nominal	n.car	n-misc	
N.LOC	locative nominal	n.loc	n-misc	
RELNZ	relational noun derivation	relnz	n-misc	
Verbal categories				
Conjugation				
S	subjective conjugation	s	v-conj	only in combinations
O	objective conjugation	o	v-conj	only in combinations
MD	middle conjugation	md	v-conj	only in combinations
SG.O	objective conjugation, singular object	o,objsg	v-conj,persnum-number	only in combinations
NSG.O	objective conjugation, non-singular object	o,objnsg	v-conj,persnum-number	only in combinations
S/SG.O	subjective conjugation / objective conjugation, singular object	s,o,objsg	v-conj,v-conj,persnum-number	only in combinations
Tense				
PF	perfect tense	pf	v:tense	
PF.EXCL	perfect tense, exclamative	pf,excl	v:tense,cl	
PRS	present tense	prs	v:tense	
PRS.EXCL	present tense, exclamative	prs,excl	v:tense,cl	
	future tense (any)	fut	v:tense	only as a tag
FUT.PF	future perfect tense	fut,fut.pf	v:tense,v:tense	
FUT1	future tense	fut,fut1	v:tense,v:tense	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
PST	past tense	pst	v:tense	
PST.PF	past perfect tense	pst.pf	v:tense	
Mood				
INTER	interrogative	inter	v:mood	
INTER.PST	interrogative, past tense	inter,pst	v:mood,v:tense	
INTER.ITER	interrogative, iterative	inter.iter	v:mood	
NAR	narrative	nar	v:mood	
NAR.INTER	narrative, interrogative	nar,inter	v:mood,v:mood	
NAR.EXCL	narrative, exclamative	nar,excl	v:mood,cl	
INFER	inferential	infer	v:mood	
IRR	irrealis	irr	v:mood	
ADMON	admonitive	admon	v:mood	
	dubitative (any)	dub	v:mood	only as a tag
DUB1	dubitative 1	dub,dub1	v:mood,v:mood	
DUB1.EXCL	dubitative 1, exclamative	dub,dub1,excl	v:mood,v:mood,cl	
DUB2	dubitative 2	dub,dub1	v:mood,v:mood	
ABES	abessive (verbal or nominal)	abes	v:mood	
OPT	optative	opt	v:mood	
DEB	debitive	deb	v:mood	
SPEC	speculative	spec	v:mood	
SPEC.PRS	speculative, present tense	spec,prs	v:mood,v:tense	
SPEC.FUT	speculative, future tense (any)	spec,fut	v:mood,v:tense	
HAB.QUAL	qualitative habitual	hab.qual	v:mood	
IMP	imperative	imp	v:mood	
IMP.2SG.S	imperative, 2 person singular, subjective conjugation	imp,pn2,pnsg,s	v:mood,persnum-person,persnum-number,v-conj	
IMP.2SG.MD	imperative, 2 person singular, middle conjugation	imp,pn2,pnsg,md	v:mood,persnum-person,persnum-number,v-conj	
IMP.2SG.MD.EXCL	imperative, 2 person singular, middle conjugation, exclamative	imp,pn2,pnsg,md,excl	v:mood,persnum-person,persnum-number,v-conj,cl	
IMP.2SG.SG.O	imperative, 2 person singular, objective conjugation, singular object	imp,pn2,pnsg,o,objsg	persnum-person,persnum-number,v-conj, persnum-number	
IMP.2SG.NSG.O	imperative, 2 person singular, objective conjugation, non-singular object	imp,pn2,pnsg,o,objnsg	persnum-person,persnum-number,v-conj, persnum-number	
IMP.FUT	imperative, future tense	imp,fut	v:mood,v:tense	
IMP.FUT.2SG.S	imperative, future tense, 2 person singular, subjective conjugation	imp,fut,pn2,pnsg,s	v:mood,v:tense,persnum-person,persnum-number,v-conj	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
IMP.FUT.2SG.MD	imperative, future tense, 2 person singular, middle conjugation	imp,fut,pn2,pnsg,md	v:mood,v:tense,persnum-person,persnum-number,v-conj	
Verbalizations				
	verbalizer (any)	vblz_any	deriv-v	only as a tag
TRL	translative	vblz_any,trl	deriv-v,deriv-v	
VBLZ	verbalizer	vblz_any,vblz	deriv-v,deriv-v	
VBLZs	verbalizer in -s	vblz_any,vblzs	deriv-v,deriv-v	
%VBLZ	verbalizer (possible)	vblz_any,%vblz	deriv-v,deriv-v	
CAP	captative	vblz_any,cap	deriv-v,deriv-v	
VBLZ.POSS	possessive verbalizer	vblz_any,vblz.poss	deriv-v,deriv-v	
PRIV	privative	vblz_any,priv	deriv-v,deriv-v	
Other verbal categories				
DRVs	verbal derivation in -s	drvs	v-misc	
STAT	stative	stat	v-misc	
ANDV	andative	andv	v-misc	
IPFV	imperfective	ipfv	v-misc	
FUT2	future (interrogative)	fut,fut2	v:tense,v-misc	
RES	resultative	res	v-misc	
ITER	iterative	iter	v-misc	
INCH	inchoative	inch	v-misc	
FRQ	frequentative	frq	v-misc	
DUR	durative	dur	v-misc	
DISTR	distributive	distr	v-misc	
HAB	habitual	hab	v-misc	
MOM	momentaneous	mom	v-misc	
SMLF	semelfactive	smlf	v-misc	
ATT	attenuative	att	v-misc	
TEMP	temporative	temp	v-misc	
TR	transitive	tr	v-misc	
PASS	passive	pass	v-misc	
CAUS	causative	caus	v-misc	
REC	reciprocal	rec	v-misc	
VOL	volitional	vol	v-misc	
INT	intentional	int	v-misc	
PROS	prospective	pros	v-misc	
Participles				
	participle (any)	ptcp	v-nfin	only as a tag
PTCP.PRS	present participle	ptcp,ptcp.prs	v-nfin,v-nfin	
PTCP.PST	past participle	ptcp,ptcp.pst	v-nfin,v-nfin	
PTCP.PASS	passive participle	ptcp,ptcp.pass	v-nfin,v-nfin	
PTCP.DEB	debitive participle	ptcp,ptcp.deb	v-nfin,v-nfin	
PTCP.ABES	abessive participle	ptcp,ptcp.abes	v-nfin,v-nfin	
PTCP.ITER	iterative participle	ptcp,ptcp.iter	v-nfin,v-nfin	
PTCP.IMM	immediate participle	ptcp,ptcp.imm	v-nfin,v-nfin	
PTCP.IRR	irreal participle	ptcp,ptcp.irr	v-nfin,v-nfin	
Converbs				
	conditional converb (any)	cond	v-nfin	only as a tag
COND1	conditional converb 1	cond,cond1	v-nfin,v-nfin	
COND2	conditional converb 2	cond,cond2	v-nfin,v-nfin	
	supine (any)	sup	v-nfin	only as a tag
SUP1	supine 1	sup,sup1	v-nfin,v-nfin	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
SUP2	supine 2	sup,sup2	v-nfin,v-nfin	
INF	infinitive	inf	v-nfin	
CNG	connegative	cng	v-nfin	
INF.IMM	immediate infinitive	inf.imm	v-nfin	
Adverbial categories				
	caritive (any)	car	deriv-misc	only as a tag
CAR1	caritive 1	car,car1	deriv-misc,deriv-misc	
CAR2	caritive 2	car,car2	deriv-misc,deriv-misc	
SOC	sociative	soc	deriv-misc	
	adverbalizer (any)	advz	deriv-misc	only as a tag
ADVZ1	adverbalizer 1	advz,advz1	deriv-misc,deriv-misc	
ADVZ2	adverbalizer 2	advz,advz2	deriv-misc,deriv-misc	
LOCZ	locativizer	locz	deriv-misc	
Adjectival categories				
ADJZ	adjectivizer	adjz	deriv-misc	
%ADJZ	adjectivizer (possible)	%adjz	deriv-misc	
QUAL1	qualitative adjective 1	qual1	deriv-misc	
QUAL2	qualitative adjective 2	qual2	deriv-misc	
QUAL3	qualitative adjective 3	qual3	deriv-misc	
ORD	ordinal numeral	ord	deriv-misc	
ADJZ.LOC	locative adjectivizer	adjz.loc	deriv-misc	
ADJZ.LOC.PL	locative adjectivizer, plural number	adjz.loc,pl	deriv-misc,n-num	
SMELL	having smell and/or taste of	smell	deriv-misc	
Clitics				
	emphatic marker (any)	emph	cl	only as a tag
EMPH1	emphatic marker 1	emph,emph1	cl,cl	
EMPH2	emphatic marker 2	emph,emph2	cl,cl	
EMPH3	emphatic marker 3	emph,emph3	cl,cl	
EMPH4	emphatic marker 4	emph,emph4	cl,cl	
EMPH5	emphatic marker 5	emph,emph5	cl,cl	
EMPH6	emphatic marker 6	emph,emph6	cl,cl	
EMPH7	emphatic marker 7	emph,emph7	cl,cl	
EMPH8	emphatic marker 8	emph,emph8	cl,cl	
EMPH.PR	pronominal emphatic marker	emph,emph.pr	cl,cl	
EXCL	exclamative	excl	cl	
EXCL.INT	prosodical exclamative	excl,excl.int	cl,cl	
EXCL.INTER	exclamative, interrgoative	excl,excl.inter	cl,cl	
EVEN	"even"	even	cl	
LIM	limitative	lim	cl	
CTR	"unexpectedly"	ctr	cl	
SIM	similative	sim	cl	
TOP	topical marker	top	cl	
UNCERT	marker of uncertainty	uncert	cl	
	indefinite marker (any)	indf	cl	only as a tag
INDF1	indefinite marker 1	indf,indf1	cl,cl	
INDF2	indefinite marker 2	indf,indf2	cl,cl	
%INDF	indefinite marker (possible)	indf,%indf	cl,cl	
Miscellaneous				
EP	epenthetic vowel	ep	misc	

Gloss	Description	Tsakorpus grammar tags	Tsakorpus category	Comment
EP.SONG	epenthetic syllable in songs	ep.song	misc	
DRV	unspecified derivation	drv	misc	
%%	unknown morph	%%	misc	
Lexical				
ALL	allative	all	pp	lexical
NEG	negative verb	neg	aux	lexical
PLC	placeholder	plc	n/v	lexical
HES	hesitation	hes	intj	lexical
INTERJ	interjection	interj	intj	lexical
SONG	epenthetic syllable in songs	song	intj	lexical
EXCLAM	exclamation	Eclat	intj	lexical
IDEOPHON	ideophone	ideophon	intj	lexical

Table 28. List of affixes and clitics by glosses

Gloss	Morpheme (mp)	Morphological category (mc)	Description
1DU.MD	*ni*	v:pn	1 person dual, middle conjugation
1DU.NSG.O	*n'i*	v:pn	1 person dual, objective conjugation, non-singular object
1DU.S/SG.O	*mi*	v:pn	1 person dual, subjective conjugation / objective conjugation, singular object
1DU.S/SG.O.EXCL	*mei*	v:pn	1 person dual, subjective conjugation / objective conjugation, singular object, exclamative
1PL.MD	*nuʔ*	v:pn	1 person plural, middle conjugation
1PL.MD.EXCL	*nuəuʔ*	v:pn	1 person plural, middle conjugation, exclamative
1PL.NSG.O	*n'üʔ*	v:pn	1 person plural, objective conjugation, non-singular object
1PL.NSG.O.EXCL	*n'üəuʔ*	v:pn	1 person plural, objective conjugation, non-singular object, exclamative
1PL.S/SG.O	*muʔ*	v:pn	1 person plural, subjective conjugation / objective conjugation, singular object
1PL.S/SG.O.EXCL	*muəuʔ*	v:pn	1 person plural, subjective conjugation / objective conjugation, singular object, exclamative
1SG.MD	*nə*	v:pn	1 person, singular, middle conjugation
1SG.NSG.O	*n'ə*	v:pn	1 person singular, objective conjugation, non-singular object
1SG.S	*m*	v:pn	1 person, singular, subjective conjugation
1SG.S	*təm*	v:pn	1 person, singular, subjective conjugation
1SG.S.EXCL	*təum*	v:pn	1 person, singular, subjective conjugation, exclamative
1SG.SG.O	*mə*	v:pn	1 person singular, objective conjugation, singular object
2DU.MD	*nti*^{c}	v:pn	2 person dual, middle conjugation
2DU.MD.EXCL	*ntiəu*^{c}	v:pn	2 person dual, middle conjugation, exclamative
2DU.NSG.O	*t'i*	v:pn	2 person dual, objective conjugation, non-singular object
2DU.S/SG.O	*ri*	v:pn	2 person dual, subjective conjugation / objective conjugation, singular object
2PL.MD	*ntuʔ*	v:pn	2 person plural, middle conjugation
2PL.MD.EXCL	*ntuəuʔ*	v:pn	2 person plural, middle conjugation, exclamative
2PL.NSG.O	*t'üʔ*	v:pn	2 person plural, objective conjugation, non-singular object
2PL.S/SG.O	*ruʔ*	v:pn	2 person plural, subjective conjugation / objective conjugation, singular object
2PL.S/SG.O.EXCL	*ruəuʔ*	v:pn	2 person plural, subjective conjugation / objective conjugation, singular object, exclamative

Gloss	Morpheme (mp)	Morphological category (mc)	Description
2SG.MD	*ŋ*	v:pn	2 person singular, middle conjugation
2SG.NSG.O	*t'ə*	v:pn	2 person singular, objective conjugation, non-singular object
2SG.S	*ŋ*	v:pn	2 person singular, subjective conjugation
2SG.SG.O	*rə*	v:pn	2 person singular, objective conjugation, singular object
3DU.MD	*ntiᶜ*	v:pn	3 person dual, middle conjugation
3DU.NSG.O	*t'i*	v:pn	3 person dual, objective conjugation, non-singular object
3DU.S	*kəj*	v:pn	3 person dual, subjective conjugation
3DU.S.EXCL.INT	*kəəj*	v:pn	3 person dual, subjective conjugation prosodical exclamative
3DU.SG.O	*ti*	v:pn	3 person dual, objective conjugation, singular object
3PL.MD	*ntəʔ*	v:pn	3 person plural, middle conjugation
3PL.MD.EXCL	*ntəəuʔ*	v:pn	3 person plural, middle conjugation, exclamative
3PL.NSG.O	*t'üŋ*	v:pn	3 person plural, objective conjugation, non-singular object
3PL.NSG.O.EXCL	*t'üəuŋ*	v:pn	3 person plural, objective conjugation, non-singular object, exclamative
3PL.S	*ʔ*	v:pn	3 person plural, subjective conjugation
3PL.SG.O	*tuŋ*	v:pn	3 person plural, objective conjugation, singular object
3SG.MD	*təᶜ*	v:pn	3 person singular, middle conjugation
3SG.MD	*ʔ*	v:pn	3 person singular, middle conjugation
3SG.NSG.O	*t'ü*	v:pn	3 person singular, objective conjugation, non-singular object
[3SG.S]	*Ø*	v:pn	3 person singular, subjective conjugation
3SG.SG.O	*tu*	v:pn	3 person singular, objective conjugation, singular object
ABES	*btumu*	n>n	abessive (verbal or nominal)
ABES	*mətumaʔa*	v:mood	abessive (verbal or nominal)
ABL.ADV	*tə*	adv:case	ablative case of adverbials
ABL.PL	*kiᶜtə*	n:case	ablative case, plural number
ABL.PL2	*giʔ*	n:case	ablative case, plural
ABL.SG	*kəᶜtə*	n:case	ablative case, singular number
ABSTR	*bs'in*	v>n	abstract nominal
ABSTR	*tüt*	v>n	abstract nominal
ABSTR	*ði*	v>n	abstract nominal
ACC.DU	*kiᶜ*	n:case	accusative case, dual number
ACC.PL	*j*	n:case	accusative case, plural number
ACC.PL.1DU	*n'i*	n:case.poss	accusative case, plural number, 1 person dual
ACC.PL.1PL	*n'üʔ*	n:case.poss	accusative case, plural number, 1 person plural
ACC.PL.1SG	*n'ə*	n:case.poss	accusative case, plural number, 1 person, singular
ACC.PL.2DU	*t'i*	n:case.poss	accusative case, plural number, 2 person dual
ACC.PL.2PL	*t'üʔ*	n:case.poss	accusative case, plural number, 2 person plural
ACC.PL.2SG	*t'ə*	n:case.poss	accusative case, plural number, 2 person singular
ACC.PL.3DU	*t'i*	n:case.poss	accusative case, plural number, 3 person dual
ACC.PL.3PL	*t'üŋ*	n:case.poss	accusative case, plural number, 3 person plural
ACC.PL.3SG	*t'ü*	n:case.poss	accusative case, plural number, 3 person singular

Gloss	Morpheme (mp)	Morphological category (mc)	Description
[ACC.SG]	ᶜ	n:case	accusative case, singular number
ACC.SG	*m*	n:case	accusative case, singular number
ACC.SG.1DU	*mi*	n:case.poss	accusative case, singular number, 1 person dual
ACC.SG.1PL	*muʔ*	n:case.poss	accusative case, singular number, 1 person plural
ACC.SG.1SG	*mə*	n:case.poss	accusative case, singular number, 1 person, singular
ACC.SG.2DU	*mti*	n:case.poss	accusative case, singular number, 2 person dual
ACC.SG.2PL	*mtuʔ*	n:case.poss	accusative case, singular number, 2 person plural
ACC.SG.2SG	*mtə*	n:case.poss	accusative case, singular number, 2 person singular
ACC.SG.3DU	*mti*	n:case.poss	accusative case, singular number, 3 person dual
ACC.SG.3PL	*mtuŋ*	n:case.poss	accusative case, singular number, 3 person plural
ACC.SG.3SG	*mtu*	n:case.poss	accusative case, singular number, 3 person singular
ACT1	*mun*	v>n	action nominal 1
ACT2	*bsan*	v>n	action nominal 2
ACT3	*ʔsan*	v>n	action nominal 3
ACT4	*ʔmuə*	v>n	action nominal 4
ADJZ	*d'ə*	n>adj	adjectivizer
ADJZ	*d'əə*	n>adj	adjectivizer
ADJZ	*huə*	v>adj	adjectivizer
ADJZ	*jkuə*	v>adj, adj>adj	adjectivizer
ADJZ	*kuə*	n>adj	adjectivizer
ADJZ	*kəə*	n>adj, v>adj, adj>adj	adjectivizer
ADJZ	*lɑʔɑ*	n>adj	adjectivizer
ADJZ	*lkəə*	n>adj	adjectivizer
ADJZ	*ltəə*	v>adj	adjectivizer
ADJZ	*l'aŋkə*	n>adj	adjectivizer
ADJZ	*muə*	v>adj	adjectivizer
ADJZ	*rəə*	v>adj	adjectivizer
ADJZ	*sətə*	n>adj	adjectivizer
ADJZ	*s'i*	n>adj	adjectivizer
ADJZ	*s'iə*	v>adj	adjectivizer
ADJZ	*tə*	n>adj	adjectivizer
ADJZ	*tətə*	n>adj	adjectivizer
ADJZ	*təə*	n>adj	adjectivizer
ADJZ	*ðə*	n>adj	adjectivizer
ADJZ	*ʔkəə*	n>adj	adjectivizer
ADJZ	*ə*	n>adj	adjectivizer
ADJZ	*əke*	n>adj	adjectivizer
ADJZ	*ᶜkuə*	v>adj	adjectivizer
ADJZ.LOC	*ntəə*	n>adj	locative adjectivizer

Gloss	Morpheme (mp)	Morphological category (mc)	Description
ADJZ.LOC	*təə*	n>adj	locative adjectivizer
ADJZ.LOC.PL	*ntiə*	n>adj	locative adjectivizer, plural number
ADMON	*kəə*	v:mood	admonitive
ADVZ1	*ʔ*	adj>adv	adverbalizer 1
ADVZ2	*mənu*	adj>adv	adverbalizer 2
AG	*ʔs'i*	v>n	agent nominal
[ANDV]	*s*	v>v	andative
ANDV	*s*	v>v	andative
ANT	*d'əə*	n>n	nominal anteriority
APPR	*ləmu*	v>n	approximative nominal
ATT	*bsəə*	v>v	attenuative
ATT	*bta*	v>v	attenuative
AUD1	*munuj*	v>n	auditive 1
AUD2	*munəʔ*	v>n	auditive 2
AUD2.EXCL	*munəəuʔ*	v>n	auditive 2, exclamative
AUG	*kaʔa*	n>n	augmentative
AUG	*naʔa*	n>n	augmentative
AUG	*raʔa*	n>n	augmentative
AUG	*rbaʔa*	n>n	augmentative
AUG	*rbə*	n>n	augmentative
AUG	*tə*	n>n	augmentative
AUG	*ʔa*	n>n	augmentative
AUG	*ʔə*	n>n	augmentative
AUG	*ʔəə*	n>n	augmentative
CAP	*tu*	n>v	captative
CAR1	*kal'i*	n>adv	caritive 1
CAR2	*kaj*	n>adv	caritive 2
CAUS	*btu*	v>v	causative
CAUS	*ru*	v>v	causative
CAUS	*rubtu*	v>v	causative
CNG	*ʔ*	v>cvb	connegative
COM.PL	*ʔna*	n:case	comitative case, plural number
COM.SG	*na*	n:case	comitative case, singular number
COND1	*hüʔ*	v>cvb	conditional converb 1
COND2	*hüʔnü*	v>cvb	conditional converb 2
CTR	*sərə*	cl	"unexpectedly"
DEB	*bsutə*	v:mood	debitive
DIM	*aʔku*	n>n	diminutive
DIM	*r̄ɑŋku*	n>n	diminutive
DIM	*ʔku*	n>n	diminutive

Gloss	Morpheme (mp)	Morphological category (mc)	Description
DISTR	*alu*	v>v	distributive
DRV	*bsəə*	v>v	unspecified derivation
DRV	*btu*	v>v	unspecified derivation
DRV	*bu*	v>v	unspecified derivation
DRV	*bə*	v>v	unspecified derivation
DRV	*d'ü*	v>v	unspecified derivation
DRV	*d'ə*	v>v, n>n	unspecified derivation
DRV	*h͡ɪɑlə*	%%	unspecified derivation
DRV	*hu*	v>adj	unspecified derivation
DRV	*həð͡ɪɑ*	n>n	unspecified derivation
DRV	*ir*	v>v	unspecified derivation
DRV	*j*	v>v, %%	unspecified derivation
DRV	*jkɑ*	n>n	unspecified derivation
DRV	*jk͡ɪɑ*	n>n, adj>adj	unspecified derivation
DRV	*jku*	n>n	unspecified derivation
DRV	*js'i*	adj>adj, n>n	unspecified derivation
DRV	*jtu*	v>v	unspecified derivation
DRV	*jt'ü*	v>v, n>v	unspecified derivation
DRV	*ku*	v>n	unspecified derivation
DRV	*kutə*	v>v	unspecified derivation
DRV	*kə*	n>n	unspecified derivation
DRV	*kəə*	v>v	unspecified derivation
DRV	*l*	v>v	unspecified derivation
DRV	*ltu*	v>v	unspecified derivation
DRV	*l'aŋkə*	n>adj	unspecified derivation
DRV	*l'ir*	n>n	unspecified derivation
DRV	*l'ü*	v>v	unspecified derivation
DRV	*m*	v>v	unspecified derivation
DRV	*mi*	n>n	unspecified derivation
DRV	*mtar*	n>n	unspecified derivation
DRV	*mu*	v>v	unspecified derivation
DRV	*mə*	v>v	unspecified derivation
DRV	*məbtu*	v>v	unspecified derivation
DRV	*nam*	v>v	unspecified derivation
DRV	*ntu*	v>v	unspecified derivation
DRV	*nə*	v>v	unspecified derivation
DRV	*nɨ*	v>v	unspecified derivation
DRV	*nɨr*	v>v	unspecified derivation
DRV	*n'd'ə*	n>n	unspecified derivation
DRV	*n'üə*	v>v	unspecified derivation

Gloss	Morpheme (mp)	Morphological category (mc)	Description
DRV	*rɑ*	n>n	unspecified derivation
DRV	*ruə*	v>v	unspecified derivation
DRV	*rübtü*	v>v	unspecified derivation
DRV	*rə*	v>v	unspecified derivation
DRV	*rəə*	n>n	unspecified derivation
DRV	*s*	v>v	unspecified derivation
DRV	*su*	v>v, n>v	unspecified derivation
DRV	*sə*	n>n	unspecified derivation
DRV	*sətə*	n>adj	unspecified derivation
DRV	*sɨj*	v>v	unspecified derivation
DRV	*s'i*	v>v	unspecified derivation
DRV	*tɑ*	v>v	unspecified derivation
DRV	*ti*	v>v	unspecified derivation
DRV	*t͡ɪɑ*	v>v	unspecified derivation
DRV	*t͡ɪɑj*	n>n, v>v	unspecified derivation
DRV	*tu*	v>v, n>v	unspecified derivation
DRV	*tukə*	n>n	unspecified derivation
DRV	*turə*	n>n	unspecified derivation
DRV	*tü*	v>v	unspecified derivation
DRV	*tüʔ*	n>n	unspecified derivation
DRV	*tə*	v>v, n>n, n>adj	unspecified derivation
DRV	*təə*	n>n, v>n	unspecified derivation
DRV	*u*	v>v	unspecified derivation
DRV	*ðə*	v>v, n>adj, n>v, n>n	unspecified derivation
DRV	*ŋ*	v>v	unspecified derivation
DRV	*ʔkür*	v>v	unspecified derivation
DRV	*ʔmə*	v>v	unspecified derivation
DRV	*ʔs'i*	n>n	unspecified derivation
DRV	*ʔə*	v>v	unspecified derivation
DRV	*ə*	v>v, v>n	unspecified derivation
[DRVs]	*s*	v>v	verbal derivation in -s
DRVs	*s*	v>v	verbal derivation in -s
DST.IRR3	*tətəd'əə*	n:dst	irreal destinative
DST1	*tə*	n:dst	destinative 1
DST1.PL	*ti*	n:dst	destinative 1, plural number
DST2	*tətə*	n:dst	destinative 2
DU	*kəj*	infl:num	dual number
DUB1	*lɨ*	v:mood	dubitative 1
DUB1.EXCL	*l'əi*	v:mood	dubitative 1, exclamative

Gloss	Morpheme (mp)	Morphological category (mc)	Description
DUB2	*hʊ͡aləlɨ*	v:mood	dubitative 2
DUR	*kuj*	v>v	durative
DYA	*sə*	n>n	dyadic
EMPH.PR	*bɪ͡a*	cl	pronominal emphatic marker
EMPH1	*ᶜt'ə*	cl	emphatic marker 1
EMPH2	*kəl'it'ə*	cl	emphatic marker 2
EMPH3	*küə*	cl	emphatic marker 3
EMPH4	*kümu*	cl	emphatic marker 4
EMPH5	*s'iə*	cl	emphatic marker 5
EMPH6	*kəl'i*	cl	emphatic marker 6
EMPH7	*kələ*	cl	emphatic marker 7
EMPH8	*l'it'ə*	cl	emphatic marker 8
EP	*ə*	ins	epenthetic vowel
EP.SONG	*CV*	ins	epenthetic syllable in songs
EVEN	*ŋalə*	cl	"even"
EXCL	*mə*	cl	exclamative
EXCL	*məəu*	cl	exclamative
EXCL	*əu*	cl	exclamative
EXCL.INT	*u*	cl	prosodical exclamative
EXCL.INTER	*ŋ*	cl	exclamative, interrgoative
FEM	*iŋ*	n>n	feminitive
[FRQ]	*r*	v>v	frequentative
FRQ	*r*	v>v	frequentative
FRQ	*ʔnar*	v>v	frequentative
FRQ	*ʔtur*	v>v	frequentative
FUT.PF	*ʔsutəd'əə*	v:tense	future perfect tense
FUT1	*ʔsutə*	v:tense	future tense
FUT2	*ntə*	v>v	future (interrogative)
GEN.DU	*kiᶜ*	n:case	genitive case, dual number
GEN.PL	*ʔ*	n:case	genitive case, plural number
GEN.PL.1DU	*ni*	n:case.poss	genitive case, plural number, 1 person dual
GEN.PL.1PL	*nuʔ*	n:case.poss	genitive case, plural number, 1 person plural
GEN.PL.1SG	*nə*	n:case.poss	genitive case, plural number, 1 person, singular
GEN.PL.2DU	*ti*	n:case.poss	genitive case, plural number, 2 person dual
GEN.PL.2PL	*tuʔ*	n:case.poss	genitive case, plural number, 2 person plural
GEN.PL.2SG	*tə*	n:case.poss	genitive case, plural number, 2 person singular
GEN.PL.3DU	*ti*	n:case.poss	genitive case, plural number, 3 person dual
GEN.PL.3PL	*tuŋ*	n:case.poss	genitive case, plural number, 3 person plural
GEN.PL.3SG	*tu*	n:case.poss	genitive case, plural number, 3 person singular
[GEN.SG]	*ᶜ*	n:case	genitive case, singular number

Gloss	Morpheme (mp)	Morphological category (mc)	Description
GEN.SG	*ŋ*	n:case	genitive case, singular number
GEN.SG.1DU	*ni*	n:case.poss	genitive case, singular number, 1 person dual
GEN.SG.1PL	*nuʔ*	n:case.poss	genitive case, singular number, 1 person plural
GEN.SG.1SG	*nə*	n:case.poss	genitive case, singular number, 1 person, singular
GEN.SG.2DU	*ntiᶜ*	n:case.poss	genitive case, singular number, 2 person dual
GEN.SG.2PL	*ntuʔ*	n:case.poss	genitive case, singular number, 2 person plural
GEN.SG.2SG	*ntə*	n:case.poss	genitive case, singular number, 2 person singular
GEN.SG.3DU	*ntiᶜ*	n:case.poss	genitive case, singular number, 3 person dual
GEN.SG.3PL	*ntuŋ*	n:case.poss	genitive case, singular number, 3 person plural
GEN.SG.3SG	*ntu*	n:case.poss	genitive case, singular number, 3 person singular
HAB	*muŋhaᶜ*	v>v	habitual
HAB.QUAL	*ntuʔa*	v:mood	qualitative habitual
IMP	*ku*	v:mood	imperative
IMP	*ŋu*	v:mood	imperative
IMP	*ŋəə*	v:mood	imperative
IMP.2SG.MD	*tiŋ*	v:mood.pn	imperative, 2 person singular, middle conjugation
IMP.2SG.MD.EXCL	*tiəuŋ*	v:mood.pn	imperative, 2 person singular, middle conjugation, exclamative
IMP.2SG.NSG.O	*n'ə*	v:mood.pn	imperative, 2 person singular, objective conjugation, non-singular object
IMP.2SG.S	*ʔ*	v:mood.pn	imperative, 2 person singular, subjective conjugation
IMP.2SG.SG.O	*tə*	v:mood.pn	imperative, 2 person singular, objective conjugation, singular object
IMP.FUT	*kuə*	v:mood	imperative, future tense (any)
IMP.FUT	*kə*	v:mood	imperative, future tense (any)
IMP.FUT.2SG.MD	*kiᶜ*	v:mood.pn	imperative, future tense, 2 person singular, middle conjugation
IMP.FUT.2SG.S	*kəᶜ*	v:mood.pn	imperative, future tense, 2 person singular, subjective conjugation
INCH	*lə*	v>v	inchoative
INDF1	*bta*	cl	indefinite marker 1
INDF2	*l'ü*	cl	indefinite marker 2
INF	*sa*	v>cvb	infinitive
INF.IMM	*kajs'a*	v>cvb	immediate infinitive
INFER	*hatu*	v:mood	inferential
INT	*ʔhan*	v>v	intentional
INTER	*ŋu*	v:mood	interrogative
INTER.ITER	*kəə*	v:mood	interrogative, iterative
INTER.PST	*hu*	v:mood	interrogative, past tense
IPFV	*ntə*	v>v	imperfective
IRR	*haatəə*	v:mood	irrealis
ITER	*kə*	v>v	iterative
[LAT.ADV1]	ᶜ	adv:case	lative case of adverbials 1
LAT.ADV2	*ʔa*	adv:case	lative case of adverbials 2
LAT.PL	*ntiʔ*	n:case	lative case, plural number

Gloss	Morpheme (mp)	Morphological category (mc)	Description
LAT.PL.EXCL	*ntiəuʔ*	n:case	lative case, plural, exclamative
LAT.SG	*ntəᶜ*	n:case	lative case, singular number
LIM	*raa*	cl	limitative
LOC.ADV	*nu*	adv:case	locative case of adverbials
LOC.PL	*ntiᶜnu*	n:case	locative case, plural number
LOC.SG	*ntənu*	n:case	locative case, singular number
LOCZ	*ni*	reln>adv	locativizer
MOM	*kutə*	v>v	momentaneous
MOM	*l'müs*	v>v	momentaneous
MOM	*mə*	v>v	momentaneous
MOM	*ʔmüs*	v>v	momentaneous
N.ABSTR	*tüt*	n>n	abstract nominal
N.CAR	*kalətüt*	n>n	caritive nominal
N.LOC	*s'əmu*	n>n	locative nominal
NAR	*haŋhu*	v:mood	narrative
NAR.EXCL	*haŋhuəu*	v:mood	narrative, exclamative
NAR.INTER	*ha*	v:mood	narrative, interrogative
[NMLZ]	*ᶜ*	v>n	nominalization
NMLZ	*bsan*	v>n	nominalization
NMLZ	*btüt*	v>n	nominalization
NMLZ	*btüʔə*	v>n	nominalization
NMLZ	*ku*	v>n	nominalization
NMLZ	*kə*	v>n	nominalization
NMLZ	*lu*	v>n	nominalization
NMLZ	*mu*	v>n	nominalization
NMLZ	*n*	v>n	nominalization
NMLZ	*r*	v>n	nominalization
NMLZ	*sɨ*	v>n	nominalization
NMLZ	*t*	v>n	nominalization
NMLZ	*ta*	v>n	nominalization
NMLZ	*tu*	v>n	nominalization
NMLZ	*tü*	adj>n	nominalization
NMLZ	*tə*	v>n	nominalization
NMLZ	*təə*	v>n	nominalization
NMLZ	*ðə*	v>n	nominalization
NMLZ	*ʔkə*	v>n	nominalization
NMLZ	*ʔsan*	v>n	nominalization
NMLZ	*ʔə*	v>n	nominalization
NMLZ	*ə*	v>n	nominalization
NMLZ	*əə*	v>n	nominalization

Gloss	Morpheme (mp)	Morphological category (mc)	Description
NMLZ.INDF	*nalə*	v>n	indefinite nominal
NMLZ.INSTR	*bsan*	v>n	instrumental nominal
NMLZ.INSTR	*bs'in*	v>n	instrumental nominal
NMLZ.INSTR	*ʔsan*	v>n	instrumental nominal
NMLZ.LOC	*bsan*	v>n	locative nominal
NMLZ.LOC	*rəmu*	v>n	locative nominal
NMLZ.LOC	*s'əmu*	v>n	locative nominal
NMLZ.LOC	*ʔsan*	v>n	locative nominal
NMLZ.LOC	*ʔsəmu*	v>n	locative nominal
NOM.DU	*kəj*	n:case	nominative case, dual number
NOM.PL	*ʔ*	n:case	nominative case, plural number
NOM.PL.1DU	*n'i*	n:case.poss	nominative case, plural number, 1 person dual
NOM.PL.1PL	*n'üʔ*	n:case.poss	nominative case, plural number, 1 person plural
NOM.PL.1SG	*n'ə*	n:case.poss	nominative case, plural number, 1 person, singular
NOM.PL.2DU	*t'i*	n:case.poss	nominative case, plural number, 2 person dual
NOM.PL.2PL	*t'üʔ*	n:case.poss	nominative case, plural number, 2 person plural
NOM.PL.2PL.EXCL	*t'iəuʔ*	n:case.poss	nominative case, plural number, 2 person plural, exclamative
NOM.PL.2SG	*t'ə*	n:case.poss	nominative case, plural number, 2 person singular
NOM.PL.3DU	*t'i*	n:case.poss	nominative case, plural number, 3 person dual
NOM.PL.3PL	*t'üŋ*	n:case.poss	nominative case, plural number, 3 person plural
NOM.PL.3PL.EXCL	*t'üəuŋ*	n:case.poss	nominative case, plural number, 3 person plural, exclamative
NOM.PL.3SG	*t'ü*	n:case.poss	nominative case, plural number, 3 person singular
[NOM.SG]	*Ø*	n:case	nominative case, singular number
NOM.SG.1DU	*mi*	n:case.poss	nominative case, singular number, 1 person dual
NOM.SG.1PL	*muʔ*	n:case.poss	nominative case, singular number, 1 person plural
NOM.SG.1PL.EXCL	*muəuʔ*	n:case.poss	nominative case, singular number, 1 person plural, exclamative
NOM.SG.1SG	*mə*	n:case.poss	nominative case, singular number, 1 person, singular
NOM.SG.2DU	*ri*	n:case.poss	nominative case, singular number, 2 person dual
NOM.SG.2PL	*ruʔ*	n:case.poss	nominative case, singular number, 2 person plural
NOM.SG.2PL.EXCL	*ruəuʔ*	n:case.poss	nominative case, singular number, 2 person plural, exclamative
NOM.SG.2SG	*rə*	n:case.poss	nominative case, singular number, 2 person singular
NOM.SG.3DU	*ti*	n:case.poss	nominative case, singular number, 3 person dual
NOM.SG.3PL	*tuŋ*	n:case.poss	nominative case, singular number, 3 person plural
NOM.SG.3PL.EXCL	*tuəuŋ*	n:case.poss	nominative case, singular number, 3 person plural, exclamative
NOM.SG.3SG	*tu*	n:case.poss	nominative case, singular number, 3 person singular
OBL.1DU	*ni*	infl:poss	oblique, 1 person dual
OBL.1PL	*nuʔ*	infl:poss	oblique, 1 person plural
OBL.1PL.EXCL	*nəuʔ*	infl:poss	oblique, 1 person plural, exclamative
OBL.1SG	*nə*	infl:poss	oblique, 1 person, singular
OBL.2DU	*nti^{c}*	infl:poss	oblique, 2 person dual

Gloss	Morpheme (mp)	Morphological category (mc)	Description
OBL.2PL	*ntuʔ*	infl:poss	oblique, 2 person plural
OBL.2PL.EXCL	*ntuəuʔ*	infl:poss	oblique, 2 person plural, exclamative
OBL.2SG	*ntə*	infl:poss	oblique, 2 person singular
OBL.3DU	*ntiᶜ*	infl:poss	oblique, 3 person dual
OBL.3PL	*ntuŋ*	infl:poss	oblique, 3 person plural
OBL.3SG	*ntu*	infl:poss	oblique, 3 person singular
OPT	*haa*	v:mood	optative
ORD	*mtuə*	num>adj	ordinal numeral
PASS	*ru*	v>v	passive
PEJOR	*d'aʔa*	n>n	pejorative
PF	*ʔə*	v:tense	perfect tense
PF.EXCL	*ʔəu*	v:tense	perfect tense, exclamative
PRIV	*ʔsaʔ*	n>v	privative
PROL.ADV	*məənu*	adv:case	prolative case of adverbials
PROL.PL	*ʔmənu*	n:case	prolative case, plural number
PROL.SG	*mənu*	n:case	prolative case, singular number
PROS	*d'akuj*	v>v	prospective
PRS	*ntu*	v:tense	present tense
PRS.EXCL	*ntuəu*	v:tense	present tense, exclamative
PST	*suə*	v:tense	past tense
PST.PF	*suəd'əə*	v:tense	past perfect tense
PTCP.ABES	*mətumaʔa*	v>adj	abessive participle
PTCP.DEB	*ʔsutə*	v>adj	debitive participle
PTCP.IMM	*jən'd'ə*	v>adj	immediate participle
PTCP.IRR	*ʔsutətəə*	v>adj	irreal participle
PTCP.ITER	*kutə*	v>adj	iterative participle
PTCP.PASS	*məə*	v>adj	passive participle
PTCP.PRS	*ntuə*	v>adj	present participle
PTCP.PST	*suəd'əə*	v>adj	past participle
QUAL1	*l'iᶜku*	adj>adj, n>adj	qualitative adjective 1
QUAL2	*ᶜkuə*	adj>adj, n>adj	qualitative adjective 2
QUAL3	*ʔbalə*	n>adj	qualitative adjective 3
REC	*u*	v>v	reciprocal
RELNZ	*kə*	adv>reln	relational noun derivation
RES	*ʔkə*	v>v	resultative
SEL	*d'um*	n>n	selective
SIM	*rəku*	cl	similative
SMELL	*n'əəkə*	n>adj	having smell and/or taste of
SMLF	*ʔaltu*	v>v	semelfactive
SOC	*səbtə*	n>adv	sociative

Gloss	Morpheme (mp)	Morphological category (mc)	Description
SPEC.FUT	*ʔsutərəku*	v:mood	speculative, future tense (any)
SPEC.PRS	*ntuərəku*	v:mood	speculative, present tense
SRLAT	*btə*	n>n, adj>adj	superlative
[STAT]	ᶜ	v>v	stative
STAT	*n*	v>v	stative
SUP1	*nɑkə*	v>cvb, v>n	supine 1
SUP2	*ʔsɑ*	v>cvb, v>n	supine 2
TEMP	*kəl*	v>v	temporative
TMP1	*ntu*	v>n	temporal nominal 1
TMP2	*hüʔə*	v>n	temporal nominal 2
TOP	*ŋutu*	cl	topical marker
TR	*tu*	v>v	transitive
TRL	*m*	v>v, n>v	translative
UNCERT	*btərəku*	cl	marker of uncertainty
[VBLZ]	ᶜ	n>v	verbalizer
VBLZ	*btu*	n>v	verbalizer
VBLZ	*bu*	n>v	verbalizer
VBLZ	*d'ə*	n>v	verbalizer
VBLZ	*ir*	n>v	verbalizer
VBLZ	*j*	n>v	verbalizer
VBLZ	*jt'ü*	n>v	verbalizer
VBLZ	*kə*	n>v	verbalizer
VBLZ	*kəə*	n>v	verbalizer
VBLZ	*l*	n>v	verbalizer
VBLZ	*ltə*	n>v	verbalizer
VBLZ	*lə*	n>v	verbalizer
VBLZ	*l'ir*	n>v	verbalizer
VBLZ	*min*	n>v	verbalizer
VBLZ	*miʔ*	n>v	verbalizer
VBLZ	*mə*	n>v	verbalizer
VBLZ	*mɨn*	n>v	verbalizer
VBLZ	*n*	n>v	verbalizer
VBLZ	*nɑm*	adj>v, n>v	verbalizer
VBLZ	*ntu*	n>v	verbalizer
VBLZ	*n'üə*	n>v	verbalizer
VBLZ	*r*	n>v	verbalizer
VBLZ	*ri*	n>v	verbalizer
VBLZ	*ruə*	n>v	verbalizer
VBLZ	*rə*	n>v	verbalizer
VBLZ	*rɨr*	n>v	verbalizer

Gloss	Morpheme (mp)	Morphological category (mc)	Description
VBLZ	*s*	n>v	verbalizer
VBLZ	*su*	n>v	verbalizer
VBLZ	*tu*	n>v	verbalizer
VBLZ	*tuə*	n>v	verbalizer
VBLZ	*tə*	n>v	verbalizer
VBLZ	*u*	n>v	verbalizer
VBLZ	*ur*	n>v	verbalizer
VBLZ	*ðə*	n>v	verbalizer
VBLZ	*ʔ*	n>v	verbalizer
VBLZ	*ə*	n>v	verbalizer
VBLZ	*ɨ*	n>v	verbalizer
VBLZ.POSS	*ʔtə*	n>v	possessive verbalizer
[VBLZs]	*s*	v>v	verbalizer in -s
VBLZs	*s*	n>v	verbalizer in -s
[VOC.SG]	ᶜ	n:case	vocative case, singular number
VOC.SG	*j*	n:case	vocative case, singular number
VOL	*nantu*	v>v	volitional
%ADJZ	*ðɨmə*	n>adj	adjectivizer (possible)
%INDF	*rəbtə*	cl	indefinite marker (possible)
%NMLZ	*bsɨmɨ*	v>n	nominalization (possible)
%NMLZ	*tɨmɨ*	v>n	nominalization (possible)
%VBLZ	*rɨm*	n>v	verbalizer (possible)

Appendix A4. Tags for verb meanings

*Table 29. Tags for verb meanings used in **vsem** tier*

Tag	Description
act_perception	Active perception frame (Perception_active). FrameNet description: This frame contains perception words whose perceivers intentionally direct their attention to some entity or phenomenon in order to have a perceptual experience. Subframes: ignoring, listening, reading, watching.
activity	Activity frame. FrameNet description: This is an abstract frame for durative activities, in which the Agent enters an ongoing state of the Activity, remains in this state for some Duration of Time, and leaves this state either by finishing or by stopping. The Agent's Activity should be intentional. Subframes: doing, finishing, preparing, starting, stopping, waiting.
assistance	Assistance frame. FrameNet description: A Helper benefits a Benefited_party by enabling the culmination of a Goal that the Benefited_party has. A Focal_entity that is involved in reaching the Goal may stand in for it. Subframes: helping, rescuing.
attachment	Attachment frame (Attaching). FrameNet description: The Attachment frame covers two situations: a scene in which somebody causes one thing to be physically connected to something else; or a scene in which somebody causes two things to be connected to each other. In the first, the frame includes an Agent who attaches an Item to a Goal by manipulating a Connector, creating an asymmetric relationship between the Item and the Goal. In the second, the Agent attaches two Items to each other, where each serves as a Goal for the other, creating a symmetric relationship between the two Items. In both cases, the Connector remains to bind the two entities (either Item and Goal, or two Items), without creating a new entity. Subframes: attaching, harnessing, hooking, joining, loosening, matching, separating, sticking, tying, unharnessing, unhooking, untying.
awareness	Awareness frame. FrameNet description: A Cognizer has a piece of Content in their model of the world. The Content is not necessarily present due to immediate perception, but usually, rather, due to deduction from perceivables. In some cases, the deduction of the Content is implicitly based on confidence in sources of information (believe), in some cases based on logic (think), and in other cases the source of the deduction is deprofiled (know). Subframes: discovering, finding, forgetting, forgetting, knowing, learning, losing, recognizing, remembering, teaching.
closure	Closure frame. FrameNet description: An Agent manipulates a Fastener to open or close a Containing_object (e.g. coat, jar). Sometimes an Enclosed_region or a Container_portal may be expressed. Since the Manipulator is syntactically omissible, many verbs in this frame incorporate the Fastener. Subframes: closing, covering, opening.
clothing	Clothing frame. FrameNet description: This frame refers to clothing and its characteristics, including anything that people conventionally wear. The Wearer wears a Garment that has a certain Style, is made of a certain Material, and can be otherwise characterized by a Descriptor. In some cases, the Body_location (where the garment is worn) and Subregion of the clothing may also be specified. Subframes: dressing, wearing.
cogitation	Cogitation frame. FrameNet description: A person, the Cognizer, thinks about a Topic over a period of time. What is thought about may be a course of action that the person might take, or something more general. Subframes: believing, choosing, counting, expecting, imagining, judging, thinking.
commerce	Commerce frame (Commercial_transaction). FrameNet description: basic commercial transactions involving a Buyer and a Seller who exchange Money and Good Subframes: buying, paying, selling.
communication	Communication frame. FrameNet description: A Communicator conveys a Message to an Addressee; the Topic and Medium of the communication also may be expressed.

Tag	Description
	Subframes: allowing, blaming, complaining, deceiving, promising, questioning, requesting, scolding, slandering.
competition	Competition frame. FrameNet description: This frame is concerned with the idea that people (Participant_1, Participant_2, or Participants) participate in an organized, rule-governed activity (the Competition) in order to achieve some advantageous outcome (often the Prize). Subframes: winning.
conduct	Conduct frame. FrameNet description: An Agent acts in a certain Manner either generally or under some particular Circumstances. Subframes: caring, treating.
contingency	Contingency frame. FrameNet description: The answer to one open question, the Outcome, is dictated (partially or completely) by the answer to another open question, the Determinant, i.e. if the answer to the Determinant is known, the answer to the Outcome can be predicted. Subframes: obeying, relying, supporting, trusting.
creation	Creation frame (Creating). FrameNet description: A Cause leads to the formation of a Created_entity. Subframes: boiling, building, cooking, braiding, furrowing, giving_birth, making, planing, setting_fire, sewing, writing.
cs_awareness	Cause awareness frame. FrameNet description: none. Description: An Agent causes a Cognizer to receive a piece of Content for their model of the world. Subframes: notifying, raising, teaching, letting_think.
cs_dir_motion	Cause directed motion frame. FrameNet description (general for cause motion): An Agent causes a Theme to move from a Source, along a Path, to a Goal. Subframes: blowing, breathing, carrying, drawing, driving, dropping, falling, gathering, holding_out, leading, lifting, moving, pointing, pouring, pulling, pushing, riding, sending, shoving, sinking, sliding, strewing, throwing.
cs_emotion	Cause emotion frame (Cause_to_experience). FrameNet description: An Agent intentionally seeks to bring about an internal mental or emotional state in the Experiencer. This frame in closely related to Experiencer_obj. Subframes: bustling, calming, charming, disturbing, entertaining, exhausting, satisfying, torturing, upsetting.
cs_fragment	Cause fragment frame (Cause_to_fragment). FrameNet description: An Agent suddenly and often violently separates the Whole_patient into two or more smaller Pieces, resulting in the Whole_patient no longer existing as such. Subframes: breaking, carving, chopping, cutting, dismantling, grinding, hacking, pulverizing, raking, separation, shattering, shredding, tearing.
cs_ingestion	Cause ingestion frame. FrameNet description for Ingestion: An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. Subframes: feeding.
cs_light_emission	Cause light emission frame. FrameNet description: none. Description: in this frame, a Source emits Light. Subframes: illumination.
cs_moisture_change	Cause moisture change frame (Cause_to_be_dry and Cause_to_be_wet). FrameNet description: An Agent causes a Dryee (either a surface or an entire entity, inside and out) to become dry or a Patient to be wet with a Liquid. Subframes: drying, soaking.
cs_motion	Cause motion frame. FrameNet description (general for cause motion): An Agent causes a Theme to move from a Source, along a Path, to a Goal. Comment: this frame in contrast to Cause directed motion gathers motion types that do not involve a distinct path to a goal while focusing instead on the manner of motion.

Tag	Description
	Subframes: moving, raking, reeling, rotating, scattering, shaking, shaking, slipping, spilling, stirring, swinging, turning, unrolling, whirling, winging.
cs_perception	Cause perception frame. FrameNet description: An Agent, Actor, Entity, or Medium causes a Phenomenon to be perceived by a Perceiver. Subframes: distracting, hiding, pointing, showing.
cs_phase_change	Cause phase change frame. FrameNet description for Change_of_phase: In this frame a Patient undergoes a change of phase. Subframes: boiling, freezing, melting.
cs_recovery	Cause recovery frame. FrameNet description for Cure: This frame deals with a Healer treating and curing an Affliction (the injuries, disease, or pain) of the Patient, sometimes also mentioning the use of a particular Treatment or Medication. Subframes: healing, repairing, reviving.
cs_shape_change	Cause shape change frame. FrameNet description: none. Description: in this frame, an Agent changes the shape of an Item. Subframes: bending, crooking, deformation, sharpening, straightening.
cs_size_change	Cause size change frame (Cause_expansion). FrameNet description: An Agent or non-human Cause causes an Item to change its physical size. Subframes: growing, shortening, stretching.
cs_sound_emission	Cause sound emission frame. FrameNet description: none. Description: in this frame, a Source emits a Sound. Subframes: crackling, crying, sounding.
cs_temp_change	Cause temperature change frame. FrameNet description: In this frame, an Agent changes the temperature of an Item. Subframes: heating, warming.
destruction	Destruction frame (Destroying). FrameNet description: A Destroyer (a conscious entity) or Cause (an event, or an entity involved in such an event) affects the Patient negatively so that the Patient no longer exists. Subframes: destroying, killing.
employment	Employment frame. FrameNet description: An Employer employs an Employee whose Position entails that the Employee perform certain Tasks in exchange for Compensation. Subframes: hiring.
excretion	Excretion frame (Excreting). FrameNet description: An Excreter excretes Excreta. Subframes: spitting.
exp_emotion	Experience emotion frame (Experiencer_focused_emotion). FrameNet description: The words in this frame describe an Experiencer's emotions with respect to some Content. Subframes: being_nasty, envying, fearing, disliking, loving, being_sorry.
filling	Filling frame. FrameNet description: These are words relating to filling containers and covering areas with some thing, things or substance, the Theme. Subframes: charging, emptying, filling, loading, occupying.
hostile_encounter	Hostile encounter frame. FrameNet description: This frame consists of words that describe a hostile encounter between opposing forces (Side_1 and Side_2, collectively conceptualizable as Sides) over a disputed Issue and/or in order to reach a specific Purpose. Subframes: attacking, fighting.
impacting	Impacting frame (Impact). FrameNet description: While in motion, an Impactor makes sudden, forcible contact with the Impactee, or two Impactors both move, mutually making forcible contact.

Tag	Description
	Subframes: beating, biting, bleeding, boiling, boiling, burning, contaminating, digging, dirtying, gnawing, hitting, kicking, piercing, pricking, processing, scratching, shooting, spoiling, spuring, strangling, troubling, washing, weakening, whipping, wounding.
ingestion	Ingestion frame. FrameNet description: An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. Subframes: drinking, eating, inhaling, sniffing.
inh_motion	Inhibited motion frame (Inhibit_movement). FrameNet description: An Agent restricts the movement of a Theme to within the vicinity of the Holding_location, despite the Theme's desire, plan, or tendency towards motion. Subframes: hindering, releasing, stopping.
leisure	Leisure frame. FrameNet description: none. Description: in this frame An Agent interacts with a Theme for entertainment purposes. Subframes: playing.
manipulation	Manipulation frame. FrameNet description: The words in this frame describe the manipulation of an Entity by an Agent. Generally, this implies that the Entity is not deeply or permanently physically affected, nor is it overall moved from one place to another. Subframes: caressing, clapping, grabbing, holding, kissing, leveling, licking, manipulating, pressing, pushing, rubbing, stretching, touching.
modal	Modal frame. FrameNet description (for Desiring, Needing and Capability): An Experiencer desires that an Event occur (Desiring). The speaker believes that some state of affairs or entity (the Requirement) must be present in order to cause some other dependent state of affairs to occur (the Dependent) (Needing). An Entity meets the pre-conditions for participating in an Event (Capability). Subframes: ability, managing, needing, wanting.
motion	Motion frame. FrameNet description: Some entity (Theme) starts out in one place (Source) and ends up in some other place (Goal), having covered some space between the two (Path). Subframes: arriving, cotheme, crossing, following, going, going_for, setting_off.
perception	Perception frame. FrameNet description: A Perceiver perceives a Phenomenon. Subframes: becoming_visible, hearing, seeing, spotting.
placement	Placement frame (Placing). FrameNet description: Generally without overall (translational) motion, an Agent places a Theme at a location, the Goal, which is profiled. Subframes: burying, hanging, holding_up, laying, leaning, putting, seating, saddling, setting, smearing, wrapping.
possession	Possession frame. FrameNet description: An Owner has (or lacks) a Possession. Subframes: having.
protection	Protection frame (Protecting). FrameNet description: Some Protection prevents a Danger from harming an Asset. Subframes: guarding.
removal	Removal frame (Removing). FrameNet description: An Agent causes a Theme to move away from a location, the Source. Subframes: cleaning, clearing, cutting_off, excavating, extracting, pulling_out, removing, tearing_out, wiping.
scrutiny	Scrutiny frame. FrameNet description: This frame concerns a Cognizer (a person or other intelligent being) paying close attention to something, the Ground, in order to discover and note its salient characteristics. Subframes: examining, finding_out, inspecting, searching, testing.
sleep_cycle	Sleep cycle frame (Cause_to_wake). FrameNet description: An Agent or Cause causes a Sleeper to transition from the Sleep_state to wakeful consciousness. Subframes: waking.

Tag	Description
speech	Speech frame (Statement). FrameNet description: This frame contains verbs and nouns that communicate the act of a Speaker to address a Message to some Addressee using language. Subframes: saying, shouting, singing, speaking, telling.
storage	Storage frame (Storing). FrameNet description: An Agent has placed a Theme in an accessible but somewhat out of the way Location for the purposes of maintaining it free from harm and illegitimate use while it is not being used. Subframes: guarding, keeping.
transfer	Transfer frame. FrameNet description: This frame involves a Donor transferring a Theme to a Recipient. Comment: in this version oft he tagset, transfer is understood more widely including events without either Donor, or Recipient, such as taking or leaving. Such events may feature additional participants, e.g. Location, Source. Subframes: catching, getting, giving, leaving, losing, taking, sacrificing, stealing, taking.
utility	Utility frame (Using). FrameNet description: An Agent manipulates an Instrument in order to achieve a Purpose. Subframes: using.