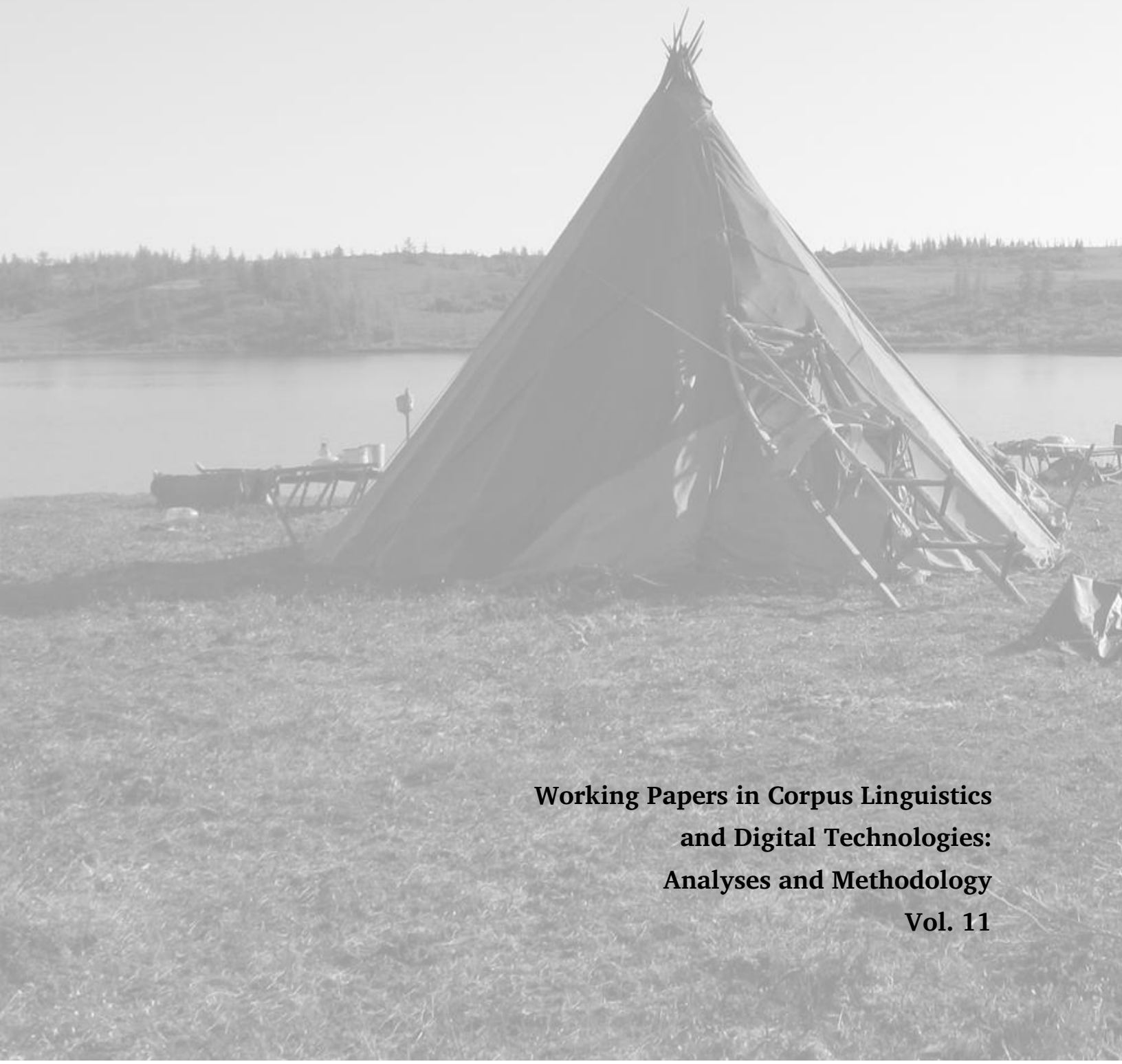


**Andrey Shluinsky & Alexandre Arkhipov**

## **User's Guide to INEL Enets Corpus**



**Working Papers in Corpus Linguistics  
and Digital Technologies:  
Analyses and Methodology  
Vol. 11**

**Andrey Shluinsky & Alexandre Arkhipov**

# **User's Guide to INEL Enets Corpus**



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

Vol. 11

Szeged – Hamburg

2025

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

**Vol. 11**

WPCL issues do not appear according to strict schedule.

© Copyrights of articles remain with the authors.

Vol. 11 (2025)

## **Editor-in-chief**

Kristin Bührig (Universität Hamburg)

## **Series editors**

Katalin Sipőcz (University of Szeged)

Elena A. Kryukova (Tomsk State Pedagogical University)

Sándor Szeverényi (University of Szeged)

Beáta Wagner-Nagy (Universität Hamburg)

## **Published by**

University of Szeged, Department of Finno-Ugric Studies

Egyetem utca 2. 6722 Szeged

Universität Hamburg, Zentrum für Sprachkorpora

Max-Brauer-Allee 60 22765 Hamburg

Published 2025

DOI 10.14232/wpcl.2025.11

ISSN 2677-0857

ISBN 978-963-688-062-0 (pdf)

## Contents

|        |   |    |
|--------|---|----|
| 1.     | Introduction .....                      | 7  |
| 1.1.   | Objective of the corpus .....           | 7  |
| 1.2.   | Enets language(s) .....                 | 7  |
| 1.2.1. | Description .....                       | 7  |
| 1.2.2. | Language Codes .....                    | 8  |
| 1.2.3. | Dialectal subdivisions .....            | 8  |
| 1.3.   | Archiving .....                         | 8  |
| 1.4.   | Citation .....                          | 9  |
| 1.5.   | Project members .....                   | 9  |
| 1.6.   | Acknowledgements .....                  | 10 |
| 2.     | The corpus .....                        | 15 |
| 2.1.   | The language(s) of the corpus .....     | 15 |
| 2.1.1. | Content .....                           | 15 |
| 2.1.2. | Annotations .....                       | 15 |
| 2.1.3. | Metadata .....                          | 15 |
| 2.2.   | Sources .....                           | 16 |
| 2.2.1. | Audio recordings .....                  | 16 |
| 2.2.2. | Transcripts / written sources .....     | 18 |
| 2.3.   | Content .....                           | 20 |
| 2.4.   | Corpus size .....                       | 20 |
| 2.5.   | Naming Conventions .....                | 21 |
| 2.5.1. | Name of the corpus .....                | 21 |
| 2.5.2. | Folder structure and filenames .....    | 21 |
| 2.5.3. | Names of texts (communications) .....   | 22 |
| 2.5.4. | Speaker codes .....                     | 22 |
| 2.5.5. | Abbreviations used in metadata .....    | 23 |
| 2.5.6. | Transliteration of Cyrillic names ..... | 25 |
| 2.6.   | Technical formats .....                 | 25 |
| 2.6.1. | Transcripts .....                       | 25 |

|        |   |    |
|--------|---|----|
| 2.6.2. | Metadata .....  | 26 |
| 2.6.3. | Media .....   | 26 |
| 2.6.4. | Other data .....  | 27 |
| 2.7.   | Workflow of the source files .....  | 27 |
| 2.7.1. | Transcripts .....   | 27 |
| 2.7.2. | Media .....   | 29 |
| 2.7.3. | Metadata .....  | 29 |
| 2.8.   | Metadata for the corpus .....   | 30 |
| 2.8.1. | General corpus metadata .....   | 30 |
| 2.8.2. | Text (“communication”) metadata .....   | 30 |
| 2.8.3. | Speaker metadata.....   | 32 |
| 3.     | Transcription and annotation .....  | 33 |
| 3.1.   | Tier layout.....  | 33 |
| 3.2.   | Transcription tiers.....  | 36 |
| 3.2.1. | Main transcription tiers (tx and ts).....                                       | 36 |
| 3.2.2. | Source transcription tiers (st and stl) .....                                   | 36 |
| 3.2.3. | Cyrillic orthography tier (tsc).....  | 37 |
| 3.2.4. | Transcription and orthographical conventions in the corpus.....                 | 37 |
|        | INEL transcription system .....   | 37 |
|        | Original transcriptions and their correspondence to the INEL transcription..... | 37 |
|        | Capitalization and punctuation .....  | 38 |
|        | Code-switching .....  | 38 |
| 3.3.   | Annotation tiers.....   | 38 |
| 3.3.1. | Reference (ref) .....   | 38 |
| 3.3.2. | Morpheme breaks (mb) .....  | 40 |
| 3.3.3. | Morphemes (lexical) (mp) .....  | 40 |
| 3.3.4. | Gloss (ge, gr).....   | 40 |
| 3.3.5. | Morphological category (mc) .....   | 40 |
| 3.3.6. | Part of speech (ps) .....   | 41 |
| 3.3.7. | Syntactic function (SyF).....   | 41 |
| 3.3.8. | Semantic role (SeR) .....   | 42 |

|  |     |
|--|-----|
| 3.3.9. Information status (IST).....   | 43  |
| 3.3.10. Borrowings (BOR, BOR-Phon, BOR-Morph).....   | 44  |
| 3.3.11. Code-switching (CS) .....  | 45  |
| 3.3.12. Geographical coordinates (geo).....  | 46  |
| 3.3.13. Free translation (fe, fr) .....  | 46  |
| 3.3.14. Original Russian, German or Hungarian translation (ltr, ltrg, lth respectively) .....  | 46  |
| 3.3.15. Notes (nt, nto) .....  | 47  |
| 3.4. Searching the corpus.....   | 47  |
| 3.4.1. Search with EXMARaLDA EXAKT.....  | 47  |
| 3.4.2. Online search in Tsakorpus.....   | 48  |
| References.....  | 53  |
| Appendix A1. Correspondence between Enets texts; Enets texts not included into the corpus..... | 57  |
| A1.1. Correspondence between published sources and texts in the corpus.....                    | 57  |
| A1.2. Correspondence of texts in Toolbox databases and texts in the corpus .....               | 68  |
| A1.3. Note on further texts in the corpus .....  | 85  |
| A1.4. Note on Enets texts not included into the corpus.....                                    | 85  |
| Appendix A2. INEL transcription for Enets .....  | 89  |
| Appendix A3. Transliteration .....   | 93  |
| Appendix A4. Enets speakers.....   | 95  |
| Appendix A5. Morpheme glossing labels (tiers ge, gr) and Tsakorpus grammar tags .....          | 99  |
| Appendix A6. Part-of-speech and morphological category tags (tier mc, ps).....                 | 118 |

## 1. Introduction<sup>1</sup>

### 1.1. Objective of the corpus

The present corpus of the Enets language(s) 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 based on published and unpublished data of both Enets lects (see more details in 2.2) and aims to bring together most transcribed Enets texts ever available. While a large number of Forest Enets texts are still not included into the current version of the corpus, for Tundra Enets this concerns only a handful of texts. The corpus makes possible typologically oriented corpus-based research on Enets and expands the documentation of the lesser described indigenous languages of Northern Eurasia.

### 1.2. Enets language(s)

#### 1.2.1. Description

Enets, which belongs to the Samoyedic group of the Uralic language family, is spoken in the western part of the Taymyr Peninsula in Central Siberia. It is most closely related to Nenets.

There are two Enets lects: **Tundra Enets (TE)** and **Forest Enets (FE)**, which can be considered mutually intelligible, but have a number of clear distinctions in lexicon, phonology, and morphology. They were traditionally presented as two dialects of the same language, but recently have been more and more frequently considered two different languages. A study by Khanina et al. (2018) describes the sociolinguistic and geographical history of the two Enets lects. The main phonological and lexical differences between them are summarized by Helimski (1984/2000). A more detailed account dealing also with grammatical features is presented by Khanina & Shluinsky (2022). It concludes that the history of the two Enets lects has included an initial divergence, followed by a secondary convergence, and then by a secondary divergence, as the latest regular interaction between the two groups has been broken off since the early 20<sup>th</sup> century. Speakers of TE and those of FE do not consider themselves belonging to the same ethnic group. The present corpus contains data of both TE and FE, but every text is consistently marked as belonging either to TE or to FE (those marked as belonging to both contain speech of different speakers speaking different lects).

TE speakers live in the village of Vorontsovo or nomadize together with Tundra Nenets in the Tukhard tundra. FE speakers live mainly in the village of Potapovo and in the town of Dudinka. Both lects are critically endangered, as (nearly) full command in Enets is kept by a maximum of 25–30 people overall; all of them belong to the elder generation and are bilingual in Russian (or trilingual in Russian and Tundra Nenets). The transfer of Enets between the generations is broken. Enets is not used on an everyday basis (see a sociolinguistic overview in Khanina & Shluinsky 2016 and Khanina et al. 2018).

The typological profile of Enets is rather typical for the Uralic languages, having mainly agglutinative morphology and head-final syntax, in particular, predominantly SOV word order. While TE is better represented in Uralic historical linguistics (cf. early Castrén’s (1854) data and Helimski’s (Ms.) unpublished dictionary), FE is better represented in public and in studies of grammar (in particular, Tereščenko 1966; Sorokina 2010; Siegl 2013; all available sources of Enets texts provide more FE texts than TE ones). The

---

<sup>1</sup> We are grateful to Olesya Khanina who carefully read a previous version of this document and whose criticisms helped us to make it significantly better. This document was also read and approved by Beáta Wagner-Nagy. Still, only us are responsible for remaining errors, inconsistencies, incorrect formulations and further shortcomings.

grammar sketch by Khanina & Shluinsky (2023) gives a grammatical summary of both Enets lects. Analytical decisions taken in morpheme-by-morpheme glossing provided in the present corpus mostly come from (Khanina & Shluinsky 2023), but at the same time the corpus follows general conventions of INEL corpora.

### 1.2.2. Language Codes

ISO-639-3 codes: **enf** for Forest Enets (FE), **enh** for Tundra Enets (TE)

Glottolog codes: **enet1250** for Enets in general, **fore1265** for Forest Enets, **tund1254** for Tundra Enets

### 1.2.3. Dialectal subdivisions

There is no established more elaborate dialectal classification than the distinction between TE (also referred to as Somatu or Mad(d)u) and FE (also referred to as Bay or Pe-Bay).

A special Nganasan-influenced variety of TE (henceforth Avam TE) used to be spoken in Avam tundra (the villages of Ust'-Avam and Volochanka); the last active speakers of this variety passed away by the 2000s. While there exist audio recordings of Avam TE, these recordings were not considered comprehensible by Tundra Enets consultants from Vorontsovo. Only a handful of Avam TE texts are transcribed (see Appendix A1). Since such transcriptions exist only in a draft form and the texts themselves are quite different from Vorontsovo/Tukhard TE, no Avam TE texts are included in the corpus. On the one hand, the peculiarities of Avam TE make these texts hardly analyzable based on TE background coming from the main body of TE available data. On the other hand, it makes these texts too different from other TE texts included in the corpus.

Differences in the speech of various FE resp. TE speakers (except for Avam TE) can be considered minor, although they are more significant in TE than in FE.

## 1.3. Archiving

The corpus comprises source media files (whenever available), annotated transcripts in *EXMARaLDA*<sup>2</sup> transcript formats and metadata descriptions in *EXMARaLDA* Coma format (see 2.6 and 2.8 for details).

A copy of transcripts in ELAN<sup>3</sup> 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 Tsakorporus online search platform.

The corpus is archived and published by the Research Data Repository of the University of Hamburg<sup>4</sup> under open-access conditions with Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0).<sup>5</sup>

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 MPEG2 or MPEG4 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.

<sup>2</sup> <http://exmaralda.org/en/>, last access: 07.06.2024.

<sup>3</sup> <https://tla.mpi.nl/tools/tla-tools/elan/>, last access: 12.06.2024.

<sup>4</sup> <https://www.fdr.uni-hamburg.de/communities/inel>, last access: 15.07.2024.

<sup>5</sup> <https://creativecommons.org/licenses/by-nc-sa/4.0/>, last access: 15.07.2024.



- The “lite” package does not include any sound, video or PDF files.

Besides the downloadable packages, the corpus is accessible online through Tsakorpus,<sup>6</sup> 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/EnetsCorpus/search>.

#### 1.4. Citation

The corpus is to be cited as follows:

Shluinsky, Andrey; Khanina, Olesya; Wagner-Nagy, Beáta. 2024. INEL Enets Corpus. Version 1.0. Publication date 2024-11-30. <https://hdl.handle.net/11022/0000-0007-FE1D-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 (i.e. principally morpheme-by-morpheme glosses) of included texts. Many other people contributed to the corpus. First of all, needless to say, this refers to Enets speakers who acted as storytellers (further referred to as speakers). 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.<sup>7</sup>

Additional reference to published (e.g. Sorokina & Bolina 2005: text 1, sentence 1) or manuscript archive source (e.g. Tereščenko archive; e.g. Urmanchieva Ms a) is welcome when citing sentences coming from there.

#### 1.5. Project members

##### Project summary information

The INEL Enets 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 Enets subproject spanned three years from January 2022 to December 2024. The research was carried out at the Institute for Finno-Ugric/Uralic Studies (IFUU) of the Hamburg University (UHH).

At the same time, the INEL Enets corpus includes FE and TE glossed texts databases compiled by Olesya Khanina and Andrey Shluinsky in 2009–2013. This work began within the ELDP project “Documentation of Enets: digitization and analysis of legacy field materials and fieldwork with last speakers” (2008–2011) conducted at Max Planck Institute for Evolutionary Anthropology<sup>8</sup> and afterwards was supported directly by the MPI-EVA.

##### Project leader

Beáta Wagner-Nagy

<sup>6</sup> <https://github.com/timarkh/tsakorpus>, last accessed: 17.11.2024.

<sup>7</sup> As Shluinsky was the main developer of the corpus in terms of INEL project, he has the main responsibility for remaining errors, inconsistencies and further shortcomings of the corpus. The same applies to analytical and representation solutions, except for those following the framework of the INEL project in general.

<sup>8</sup> <https://www.eva.mpg.de/linguistics/past-research-resources/documentation-and-description/documentation-of-enets/>, last access: 07.06.2024.

**Editors**

Beáta Wagner-Nagy, Alexandre Arkhipov

**Main corpus authors**

Olesya Khanina, University of Helsinki (preliminary glossing of transcribed texts, 2009–2013; collaboration as a consultant)

Andrey Shluinsky (preliminary glossing of transcribed texts, 2009–2013; main development of the corpus, July 2022 – September 2024)

Beáta Wagner-Nagy (preliminary glossing of the published texts, January – June 2022)

Contributions of Enets 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**

Timm Lehmberg (technical coordinator; until January 2023)

Aleksandr Riaposov

Elena Lazarenko

**Student assistants**

Natalia Kim (February – September 2024)

**1.6. Acknowledgements****Funding**

This corpus has been produced in 2022–2024 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.

Preliminary glossing work included into this corpus was supported by Endangered Languages Documentation Programme project “Documentation of Enets: digitization and analysis of legacy field materials and fieldwork with last speakers” (2008–2011, Bernard Comrie, Olesya Khanina, Andrey Shluinsky) and by Max Planck Institute for Evolutionary Anthropology. Within the mentioned ELDP project, most recordings were made, most of digitization and transcription work, as well as systematization of legacy data was conducted. A number of texts were recorded and transcribed in 2005 in a pilot fieldtrip supported by Max Planck Institute for Evolutionary Anthropology. At the same time, fieldwork in 2010 and some preliminary glossing work was also partially supported by the grant “Corpus of the languages of Russia” of the Branch of history and philology of the Russian Academy of sciences (PI Valentin Gusev, 2009–2011, hosted by Institute of Linguistics RAS).

A number of texts were recorded and/or transcribed in 2015–2017 within projects supported by the following grants:

- Russian Foundation for Humanities grant “Creation of multimedia dictionaries of Samoyedic languages” (#14-04-12019, 2014–2016, PI Julia Normanskaja, hosted by Institute of Linguistics RAS);
- Russian Foundation for Humanities grant “Language change in lects with no writing tradition (Altaic, Paleoasiatic and Uralic languages)” (#13-04-00416, 2013–2015, PI Evgeny Golovko, hosted by Institute for Linguistic Studies RAS);
- Russian Foundation for Humanities grant “Fieldwork on collecting spontaneous texts with the last Enets speakers” (#16-04-18023, 2016, PI Maria Ovsjannikova, hosted by Institute for Linguistic Studies RAS);
- Russian Science Foundation grant “Dynamics of language contact in the circumpolar region” (#17-18-01649, 2017–2021, PI Olesya Khanina, Institute of Linguistics RAS).

Data coming from the ELDP project, from the work supported by the MPI-EVA and from the mentioned projects supported by Russian Academy of sciences, Russian Foundation for Humanities and Russian Science Foundation are briefly referred to together as “ELDP + data”; see details in the section 2.2.

### **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 Institute for Linguistic Studies RAS (Russian: *Институт лингвистических исследований РАН*; abbreviated as ILI in the metadata) kindly provided access to manuscript archives of Natal'ya Mitrofanovna Tereščenko (1908–1987) and of Irina Petrovna Sorokina (1941–2021). Personally, we are grateful to Evgeny Golovko, Anna Urmanchieva and Marina Lyublinskaya. Special thanks go to Sofia Oskolskaya for her practical help, in particular but not exclusively in scanning the Enets part of Tereščenko's archive.

The Taymyr House of National Arts (Russian: *Таймырский дом народного творчества*; also known in English as Taymyr House of Folk Culture; abbreviated as TDNT in the metadata) kindly provided access to the archive of Kazimir Isidorovich (Kazis) Labanauskas (1942–2002), which contained both manuscripts and an extensive audio tapes collection. Some of these tapes had been already digitized by Vasilij Anatol'evich Zav`yalov, TDNT's technician, on an earlier Florian Siegl's request. Further audio recordings (done by Zoya Nikolaevna Bolina in 2007) were also provided by TDNT and converted by Zav`yalov from the minidisk format. Besides Vasilij Zav`yalov and Zoya Bolina, we personally thank the then director of the Taymyr House of National Arts Lyubov` Yur`evna Popova and the then head of its Department of folklore Larisa Dmitrievna Bettu.

The Dudinka branch of GTRK “Norilsk” (Russian: *Дудинское отделение ГТРК «Норильск»*; also known informally as “Taymyr radio”; abbreviated as GTRK in the metadata) kindly provided access to the Enets part of its extensive audio archive (Enets recordings done in 1990s – early 2000s by Nina Nikolaevna Bolina, then a radio journalist). We personally thank the then head of All-Russia State Television and Radio Broadcasting Company (*Всероссийская государственная телевизионная и радиовещательная компания*, ВГТРК/VGTRK) Vladimir Ivanovich Sharonov and the then head of the Dudinka branch of GTRK “Norilsk” Aleksandr Veniaminovich Moskvina, as well as Radik Irikovich Baty`rshin and Nina Dobrushina for their help in procuring the recordings.

The following colleagues kindly shared audio recordings from their private archives: Oksana E`duardovna Dobzhanskaya, Dar`ya Spiridonovna Bolina, Anna Urmanchieva and the late Irina Petrovna Sorokina. Viktor Nikolaevich Pal`chin kindly permitted to copy some manuscripts from his personal archive.

Larisa Leisiö and Valentin Gusev shared their audio recordings and transcripts done in 2010 in Päivölä (Finland; supported by a grant from the Academy of Finland, project 125225 (2008–2013), PI Larisa Leisiö).

Anna Urmanchieva kindly permitted to include into the corpus the TE text known as *Solda* she had edited and manually glossed (Urmanchieva Ms a; see more details in section 2.2).

Valentin Gusev shared his electronic archive with transcriptions of Enets texts (done by him and by other people).

Olga Kazakevich permitted to copy FE audio recordings, video recordings and draft transcriptions done by her, Leonid Zakharov and their students in 2011.

Finally, Egor Kashkin kindly provided fieldnotes from the personal archive of Ariadna Ivanovna Kuznetsova (1932–2015), which were an important source of metadata on Enets speakers available in 1970s.

Note that the fact that a piece of legacy data comes from a specific archive does not mean directly that it had been collected by the owner of this archive (see more details in section 2.2). In particular, Marina Lyublinskaya kindly permitted us to include in the corpus a text recorded and preliminarily transcribed by her, while the recording comes from Oksana Dobzhanskaya’s personal archive and the transcription comes from TDNT and Valentin Gusev’s archives.

Some data mentioned here are not included into the current version of the corpus. In particular, this concerns TE texts transcribed by Eugene Helimskiĭ coming from Valentin Gusev’s archive and FE texts coming from Olga Kazakevich’s collection (see more details in Appendix A1). Such data are not mentioned in section 2.2.

### **Storytelling,<sup>9</sup> text recording and analysis, processing the data**

First of all, we want to express our deepest gratitude to Enets speakers who generously shared their unique knowledge of Enets and made this corpus a tangible reality. They include speakers who told or wrote their stories, speakers who transcribed previously recorded stories orally<sup>10</sup> or in a written form and speakers who consulted us during long elicitation sessions answering our questions about Enets words and sentences (this was crucial for further analysis of transcribed texts). It is thanks to these individuals that the linguistic analysis represented in the corpus was possible, and the corpus authors express their deepest gratitude to them, for their wisdom, kindness, effort, and patience. While here list only those who worked directly with the authors of the corpus, we acknowledge to the same extent the contribution of those who generously worked with other researchers. All Enets speakers who contributed to the corpus through storytelling (except for those who unfortunately remain unidentified) are listed in Appendix A4. Sadly, we are mostly unaware of those who did oral transcription of most published or legacy texts (when we still have this information, it is mentioned in the section 2.2).

---

<sup>9</sup> We use the term storytelling broadly referring to producing a text of any genre for recording.

<sup>10</sup> By oral transcription, we mean listening to an original audio recording of text and producing its phrase-by-phrase slow and careful re-speaking (see Reinman 2010). Together with oral Russian translation, this kind of very hard and time-consuming work done by Enets speakers is the most crucial input into the corpus, besides producing the texts themselves.

Nikolaj Sergeevich Pal`chin† was the speaker recorded by Wagner-Nagy in 1994 in Potapovo, and Viktor (Vitalij) Nikolaevich Pal`chin did the oral transcription.

The following Enets speakers contributed to ELDP + data (in alphabetical order, but independently for FE and TE).

### ***Forest Enets***

Aleksandr Spiridonovich Bolin†: storytelling; elicitation sessions

Leonid Dmitrievich Bolin†: storytelling; oral transcription and Russian translation; elicitation sessions

Petr Nikolaevich Bolin: storytelling; elicitation sessions

Viktor (Vitalij) Nikolaevich Bolin: storytelling; oral transcription and Russian translation; elicitation sessions

Antonina Puyakovna Bolina†: storytelling

Lyubov` Kupriyanovna Bolina: storytelling

Nadezhda Konstantinovna Bolina: storytelling; oral transcription and Russian translation; elicitation sessions

Nina Nikolaevna Bolina: oral transcription and Russian translation

Tat`yana Nikolaevna Bolina: elicitation sessions

Zoya Nikolaevna Bolina: oral and written transcription and Russian translation; elicitation sessions

Nina Kupriyanovna Borisova†: storytelling; elicitation sessions

Nelli Alekseevna Chirimkina: storytelling

Ekaterina Spiridonovna Glibchenko: storytelling; oral transcription and Russian translation; elicitation sessions

Gennadij Afanas`evich Ivanov†: storytelling; oral transcription and Russian translation; elicitation sessions

Nikolaj Dmitrievich Ly`rmin†: storytelling

Galina Kirillovna Neznajkina: storytelling

Anatolij (Anton) Nikolaevich Pal`chin: storytelling; composing a text

Viktor (Vitalij) Nikolaevich Pal`chin: oral and written transcription and Russian translation; elicitation sessions

Svetlana Alekseevna Roslyakova†: storytelling; oral transcription and Russian translation; elicitation sessions

Lyubov` Nikolaevna Shhepeleva: oral transcription and Russian translation

Anatolij Maksimovich Silkin: storytelling

Daniil Alekseevich Silkin†: storytelling

Ivan Ivanovich Silkin†: composing texts and storytelling; oral transcription and Russian translation; elicitation sessions

Nikolaj Ivanovich Silkin: storytelling; oral transcription and Russian translation; elicitation sessions

Alevtina Spiridonovna Silkina: oral transcription and Russian translation

Zoya Kupriyanovna Tret`yakova†: storytelling

### ***Tundra Enets***

Marina Dyogolevna Beregovaya: storytelling

Irina Pajkovna Koshkaryova†: storytelling; oral transcription and Russian translation; elicitation sessions  
Lyudmila Pudidovna Novosyolova†: storytelling; oral transcription and Russian translation  
Valentina Puyakovna Nade`r: storytelling; oral transcription and Russian translation; elicitation sessions  
Semyon Dyogolevich Silkin: storytelling  
Sergej Kasovich Tuglakov†: storytelling  
Serne Urtomovna Tuglakova†: storytelling  
Vladimir Ayakovich Turutin: storytelling  
Zoya Ayakovna Ve`ngo (Turutina): storytelling; oral transcription and Russian translation  
Tat`yana Chanovna Yar: oral transcription and Russian translation; elicitation sessions

Zoya Nikolaevna Bolina and Viktor (Vitalij) Nikolaevich Pal`chin worked as collaborators of the ELDP project in 2008–2011; while Viktor Pal`chin transcribed a large amount of recordings with pen and paper, Zoya Bolina did some transcriptions with pen and paper and some in ELAN, and also made a number of audio recordings. In 2010, Zoya Bolina also assisted in the process of making some TE recordings. Both Viktor Pal`chin and Zoya Bolina also contributed extensively to oral transcription and Russian translation.

Besides, our special thanks go to those who spent many hours and days working with us (again in alphabetical order): Aleksandr Spiridonovich Bolin†, Leonid Dmitrievich Bolin†, Viktor (Vitalij) Nikolaevich Bolin, Nadezhda Konstantinovna Bolina, Nina Nikolaevna Bolina, Ekaterina Spiridonovna Glibchenko, Gennadij Afanas`evich Ivanov†, Irina Pajkovna Koshkaryova†, Valentina Puyakovna Nade`r, Lyudmila Pudidovna Novosyolova†, Svetlana Alekseevna Roslyakova†, Ivan Ivanovich Silkin†, Nikolaj Ivanovich Silkin, Alevtina Spiridonovna Silkina, Zoya Ayakovna Ve`ngo (Turutina), Tat`yana Chanovna Yar.

Next, we thank all those who contributed to text recording and analysis, as well as to technical processing of the data.

Recordings and transcriptions were done by Wagner-Nagy in 1994 jointly with Eugene Helimski† and under his supervision; Anna Urmanchieva also took part in some of these sessions.

Besides Khanina and Shluinsky, a big amount of ELDP+ data, namely audio recordings and draft transcriptions in ELAN in collaboration with Enets consultants, were done by Maria Ovsjannikova (in 2009, 2010, 2015, 2016), by Natalia Stoyanova (in 2010) and by Sergey Trubetskoy (in 2008). Maria Ovsjannikova kindly allowed us to include into the corpus the texts she recorded and/or transcribed in 2015–2016, which was done beyond the ELDP phase.

Important consultations on existing Enets speakers, their language proficiency and speaker metadata were given in different years by many people, but in particular by Valentin Gusev, Anna Urmanchieva, Florian Siegl, Dar`ya Spiridonovna Bolina, Zoya Nikolaevna Bolina, Oksana E`duardovna Dobzhanskaya, Ekaterina Spiridonovna Glibchenko, Viktor Urtomovich Yando†, Marina Nikolaevna Pal`china, Nina Nikolaevna Bolina, Viktor Pavlovich Krivonogov. Ekaterina Spiridonovna Glibchenko's help in speaker identification of FE legacy recordings was crucial for compiling the metadata.

Special thanks go to Vadim Alekseevich Silkin for his very detailed consultation in March 2008 on Forest Enets speakers' genealogical relationship that was helpful in compiling the corpus metadata.

Audio recordings coming from Dudinka branch of GTRK "Norilsk" were converted from the minidisk format by Peter Fröhlich, a technician at MPI-EVA. Peter Fröhlich also digitized analog video recordings

done in 2008. Audio recordings on reel-to-reel tapes coming from Sorokina’s personal archive were digitized by Sven Grawunder, at that time a researcher at MPI-EVA.

Anna Pazelskaya kindly helped with automatic recognition of (Labanauskas 2002) in PDF Transformer software from a scanned PDF file back in 2010.

One of previously published texts was translated into English by Jeremy Bradley.

Bernard Comrie’s support of ELD+ phase of our work goes far beyond funding provided by MPI-EVA and his formal capacity of the head of ELD+ project.

Dar`ya Spiridonovna Bolina kindly answered questions that arose from time to time in the process of development of the corpus in 2022–2024 and in particular helped with handwriting recognition of some written texts coming from Sorokina’s archive. Maria Amelina commented the situation with Tundra Enets speakers in Tukhard tundra.

We acknowledge the great effort of Enets speakers who contributed to published and legacy text collections and of all those who recorded and/or transcribed these texts (see more details in 2.2).

## **2. The corpus**

### **2.1. The language(s) of the corpus**

#### **2.1.1. Content**

The language of content in the corpus is mainly Enets, but audio recordings, especially those done in 2000s–2010s, contain instances of code-switching into Russian (part of them are left untranscribed in the corpus). Some TE recordings have some instances of code-switching into Tundra Nenets. Most texts are either in FE or in TE (FE texts being significantly more numerous, see 2.4), but some contain dialogues between a FE speaker and a TE speaker (the former interviewing the latter).

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** and **fr** 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**, 3.3.14). For texts transcribed as a part of ELD+ data, 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).

#### **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

In the context of the present corpus, it is useful to discuss the sources of audio recordings and the sources of written transcriptions separately. Some texts reached us only in the transcribed written form and thus have no audio recordings in the corpus. But all texts for which the audio recordings were accessible to us were time-aligned with the recordings, checked against them, and the transcriptions were corrected accordingly. Few texts happened to be independently transcribed by different people; in these cases, one version was taken as the primary one, but other available transcriptions were also consulted.

By ELDP+ data (ELDP+ collection, ELDP+ recordings, ELDP+ transcriptions) we mean, first, those coming from the ELDP project in 2008–2011 recorded and/or transcribed by Khanina and Shluinsky, by Zoya Bolina and Viktor Pal'chin who collaborated with the project and by Maria Ovsjannikova, Natalia Stoyanova and Sergey Trubetskoy who were fieldwork assistants of the project. Second, we mean earlier and later data (see 1.6 for details on funding) that continue the same collection both in terms of data organization and in terms of data accessibility. Namely, we mean recordings and transcriptions done by Khanina and Shluinsky in 2005, by Ovsjannikova and Shluinsky in 2015–2016 and by Khanina in 2017.

### 2.2.1. Audio recordings

The two main sources of audio recordings are, first, ELDP+ recordings done by Olesya Khanina, Andrey Shluinsky, Maria Ovsjannikova, Natalia Stoyanova, Sergey Trubetskoy in 2005–2017, as well as by Zoya Bolina in 2008–2010, and, second, archives. Besides, five song recordings in FE come from published compact disks (Bettü (ed.) 2006).

ELDP+ TE recordings were done in Vorontsovo (by Khanina and Shluinsky, 2008; by Khanina and Shluinsky, 2010, also with assistance of Zoya Bolina), in Tukhard and Tukhard tundra (by Shluinsky, 2009) and in Dudinka (by Khanina and Shluinsky, 2008; by Khanina, 2017). ELDP+ FE recordings were done mainly in Potapovo (by Khanina and Shluinsky, 2008; by Shluinsky and Trubetskoy, 2008; by Shluinsky and Ovsjannikova, 2009, 2015, 2016; by Shluinsky, Ovsjannikova and Stoyanova, 2010), but also in Dudinka (by Khanina and Shluinsky, 2008). Zoya Nikolaevna Bolina did her FE recordings in Dudinka in 2008–2010. Most of ELDP+ recordings were done in WAV format with the professional recorder Marantz PMD-660 accompanied with the external microphone AKG 1000s. Still, in some TE recordings from 2009 the external microphone was not used. Some FE recordings from 2010 were done in WAV format with the digital recorder Olympus VN-960PC. FE recordings from 2005 were done originally in MP3 format with consumer-grade MP3-recorders. Zoya Bolina's FE recordings from 2008–2010 were done with a cassette tape recorder and further digitized with tape recorder TASCAM PORTA 02MKII and digital recorder Marantz PMD-660. Some ELDP+ audio recordings are also accompanied by simultaneous video recordings; no video recordings from other sources are included into the corpus, except for one video recording coming from Leisiö and Gusev's archive.

Legacy audio recordings come from various archives. Some archive owners had obtained audio recordings done by other people. Sometimes different archives hold different copies of the same recordings. In the latter case, recordings of the best quality were chosen for transcription and/or for time-alignment with an existing transcription; the archive source of the copy which was actually used is indicated in the metadata to the specific text.



The Taymyr House of National Arts (TDNT) archive is first of all the source of tape recordings collection of Kazimir Isidorovich Labanauskas. Part of them were already digitized by 2008 by Vasilij Zav`yalov, a technician of the same institution, and part of them were digitized in 2008 by Olesya Khanina and Andrey Shluinsky with tape recorder TASCAM PORTA 02MKII and digital recorder Marantz PMD-660. The recordings from this source which make part of the corpus are as follows: (1) recordings of spontaneous speech in TE and in FE done by Labanauskas in 1970s–1990s, (2) recordings of oral transcription of texts in TE done by Labanauskas, (3) tape recordings of radio transmissions in FE and in TE done by Labanauskas from radio broadcasts (in the cases when the same recording was not available in the GTRK archive), (4) recordings in FE done in 1990s by other people, namely by Dar`ya Spiridonovna Bolina, by Viktor Nikolaevich Pal`chin and by an unidentified woman. The identification of the person who did the recording is sometimes based on indirect evidence and is not always certain (which is mentioned in metadata). Besides, audio recordings in FE done by Zoya Nikolaevna Bolina in 2007 on minidisks also come from the archive of the Taymyr House of National Arts; these recordings were converted from the minidisk format by Vasilij Zav`yalov in 2016.

The Dudinka branch of GTRK “Norilsk” (GTRK) shared the Enets part of its extensive audio archive of recordings once broadcasted on the radio. These recordings in FE and in TE were done by Nina Nikolaevna Bolina in 1990–2003; at the same time, according to Sorokina & Bolina’s (2005) metadata, some recordings might have been done by Dar`ya Spiridonovna Bolina in 1991. Nina Bolina used a professional cassette tape recorder “Reporter,” then when transmitted on radio, the sound was recorded onto reels, and later the reels were copied to minidisks (as a part of the archive modernization process). We obtained copies from these minidisks, and these copies were converted to WAV files by Peter Fröhlich at the MPI-EVA.

Audio tapes from Oksana Dobzhanskaya’s personal archive were digitized in 2008 by Olesya Khanina and Andrey Shluinsky with tape recorder TASCAM PORTA 02MKII and digital recorder Marantz PMD-660. These recordings in FE were mainly originally done in 1992 by Dobzhanskaya herself, but some recordings were done in 1990s by Viktor Nikolaevich Pal`chin, his wife Marina Nikolaevna Pal`china, Vera Bettu, Marina Lyublinskaya and an unidentified woman. Besides Dobzhanskaya’s own recordings, the identification of the person who did the recording is sometimes based on indirect evidence and is not always certain. Besides, Dobzhanskaya kindly permitted to copy an audio recording in FE in MP3 format she had done in 2009.

An audio tape from Dar`ya Spiridonovna Bolina’s personal archive contained a recording in FE presumably done by Dar`ya Bolina herself in 1990s. It was digitized in 2008 by Olesya Khanina and Andrey Shluinsky with tape recorder TASCAM PORTA 02MKII and digital recorder Marantz PMD-660.

Irina Petrovna Sorokina’s personal archive contained reel tapes with Sorokina’s own recordings in FE (some of them were surely done in 1969, while others were done later or might have been done later, but no later than in 1985). These recordings were professionally digitized by Sven Grawunder at the MPI-EVA.

Larisa Leisiö and Valentin Gusev’s digital archive contained mainly audio recordings in FE in WAV format from Zoya Nikolaevna Bolina done during her and Tat`yana Nikolaevna Bolina’s visit to Päivölä (Finland) in 2010 (this working session was conducted in terms of Academy of Finland’s project 125225, 2008–2013). For one of these recordings, a video recording in MOV format was available. Besides, it contained audio recordings in FE in MP3 format done by Kaur Mägi in Dudinka in 2006.

Wagner-Nagy's personal archive and Anna Urmanchieva's personal archive contained different digitized versions of the same Eugene Helimski's cassette tapes with recordings in FE coming from the fieldtrip in 1994 to Potapovo led by Helimski, where both Wagner-Nagy and Urmanchieva took part. While the versions coming from Wagner-Nagy's archive are of a better quality and are finally included into the corpus, mostly Urmanchieva's versions were earlier used for transcribing.

### 2.2.2. Transcripts / written sources

#### Published texts

Almost all published Enets texts are included into the corpus (exceptions are listed in Appendix A1). Namely, the following publications are included: extensive collections (Sorokina & Bolina 2005) and (Labanauskas 2002), as well as other publications: (Mikola 1967; 1980), (Pusztay 1978), (Sorokina 1982), (Gluxij & Susekov 1982), (Labanauskas 1992), (Urmanchieva 2008); and a short text published in a research paper (Helimski 1981/2000: 83). (Mikola 1967), (Sorokina 1982), (Gluxij & Susekov 1982) and (Labanauskas 1992) were reprinted in (Sorokina & Bolina 2005); three of five texts from (Labanauskas 1992) were reprinted in (Labanauskas 2002). While one of the published versions was chosen as the main source for further work, other versions were also consulted. The same concerns cases when there happen to be parallel transcriptions of the same recordings that were used for published texts, and cases when legacy manuscripts were also available for published texts. Table 27 in Appendix A1 shows the correspondence between published texts and texts in the corpus. It gives information on reprinted or legacy manuscript versions, parallel transcriptions of the same text and further alignment with audio recordings.

A long TE text known as *Solda* was originally written down and published in Russian by Dolgix (1961: 37–51), then retold in TE by Xolyu Nikolaevich Kaplin† and transcribed by Eugene Helimski in the 1970s and manually glossed by Anna Urmanchieva in the 1990s. While formally this text remains unpublished, it is available in a completely finished form and is known and widely cited by specialists in Samoyedic studies as (Urmanchieva Ms a).

#### Archive manuscripts

Natal'ya Mitrofanovna Tereščenko's archive is the most important source of unpublished texts with no audio recordings available.<sup>11</sup> It is preserved at the Institute for Linguistic Studies RAS and was systematized and described by a Russian Academy of Science Archive specialist Antonina Nikolaevna Anfert`eva†; further work with this archive has been coordinated by Marina Lyublinskaya. These FE texts were written down in 1962, and the details of the working process are unknown; sheets from the archive contain various handwritings. Still, Vasilij Fedorovich Ly`rmin† is frequently mentioned there as Tereščenko's assistant in recording and/or transcription.

Irina Petrovna Sorokina's archive is also preserved at the Institute for Linguistic Studies RAS (but is not yet systematized or described). Data copied from there contained handwritten draft transcriptions of some texts in FE published in (Sorokina & Bolina 2005), handwritten transcriptions of two unpublished texts in FE corresponding to audio recordings coming from Sorokina's personal archive, two unpublished texts in FE with no corresponding audio recordings available and two texts in FE written down by Dar`ya

<sup>11</sup> Although these texts were never published (except for a short text extract published in Tereščenko 1966: 436–437), examples taken from them are widely used in (Tereščenko 1973).

Spiridonovna Bolina and Galina Spiridonovna Bolina respectively (according to Dar`ya Bolina's identification of handwriting).

Kazimir Isidorovich Labanauskas' archive is preserved at the Taymyr House of National Arts (TDNT) and contains a number of Labanauskas' manuscripts with transcriptions of some texts in TE. While all of these texts were published in (Labanauskas 1992) and/or (Labanauskas 2002) in Cyrillic, manuscript transcriptions in Latin are more phonetically detailed and can be assumed to be closer to the original speech. These transcriptions were consulted in the process of analyzing the corresponding published texts. Besides, this archive contained a transcription of a text in FE recorded by Marina Lyublinskaya and preliminarily transcribed by her and Nadezhda Konstantinovna Bolina.

Larisa Leisiö and Valentin Gusev's archive contained unpublished draft transcriptions in MS Word documents produced in 2010 in collaboration with Zoya Nikolaevna Bolina and Tat`yana Nikolaevna Bolina. These are transcriptions of the above mentioned recordings in FE coming from Leisiö and Gusev's archive, mainly recorded from Zoya Bolina in 2010 in Päivölä, but also ones in FE recorded in 2006 by Kaur Mägi in Dudinka from Nina Kupriyanovna Borisova.

Manuscripts coming from Viktor Nikolaevich Pal`chin's personal archive contain his draft transcription of a FE song published both in (Labanauskas 2002) and (Sorokina & Bolina 2005), this transcription was also consulted.

The collection coming from Valentin Gusev's personal digital archive contained draft transcriptions of Enets texts. Some of these transcriptions were additional versions of FE and TE texts included in the corpus based on versions coming from other sources. Versions coming from Gusev's archive were also consulted.

Finally, a draft transcription of a FE song done by Eugene Helinski and Viktor Nikolaevich Pal`chin comes from Beáta Wagner-Nagy's personal archive (as a MS Word file).

### **ELDP + transcriptions**

ELDP + transcriptions and translations into Russian of both ELDP + and archive recordings included into the corpus were compiled in the following ways.

First, for most texts, draft transcriptions were done in collaboration with Enets speakers who provided an oral transcription and a Russian translation. This was mainly done directly in ELAN, but for some recordings (mostly those recorded and transcribed back in 2005) these transcriptions were done in MS Word (these transcriptions were later aligned with the audio recordings in the process of the compilation of the corpus). While Olesya Khanina and Andrey Shluinsky did such transcriptions for both FE and TE, Maria Ovsjannikova, Natalia Stoyanova and Sergey Trubetskoy conducted such work exclusively for FE. TE transcriptions were done mostly in 2008 and in 2010 in Vorontsovo, but in 2017, Khanina did some transcriptions in Dudinka. Most FE transcriptions were done in fieldtrips to Potapovo (in 2005, 2008, 2009, 2010, 2015, 2016), but a large part of this work was conducted in Dudinka (especially in 2010, but also in 2005, 2008, 2016), as well as during sessions in Moscow with Zoya Nikolaevna Bolina and Nina Nikolaevna Bolina in 2008–2009.

Second, for a significant part of FE texts, a draft transcription was done by Zoya Nikolaevna Bolina and Viktor Nikolaevich Pal`chin in their capacity as collaborators of the ELDP project in 2008–2011. While Viktor Pal`chin produced exclusively handwritten transcriptions, Zoya Bolina transcribed most recordings with pen and paper, but some others in ELAN. Additionally, Anton Nikolaevich Pal`chin and Ivan Ivanovich

Silkin provided handwritten versions of their texts. Handwritten transcriptions were aligned with the recordings at further stages of work by Shluinsky and Khanina.

Finally, for some short and simple texts, transcriptions were done by Shluinsky in 2022–2024.

The core part of ELDP + transcriptions were preliminarily glossed in Toolbox in 2009–2013; most of FE texts were glossed by Shluinsky, while Khanina glossed some FE texts and all TE texts from the Toolbox databases. Toolbox databases were imported into FLE<sub>x</sub> (see 2.7.1) and finalized there. Other texts (especially, but not exclusively those recorded and transcribed already in 2015–2016) were glossed directly in FLE<sub>x</sub> in 2022–2024.

The ELDP + collection is archived in ELAR archive: <http://elar.soas.ac.uk/deposit/0302>. (Currently the ELAR deposit contains contributions up to 2011, but later contributions will be added later.) Many but not all texts in this collection are part of the mentioned Toolbox databases. Given the fact that Toolbox databases were used by researchers and are referred to in some publications, Table 28 in Appendix 1 provides a correspondence between text abbreviations in the Toolbox databases and texts in the present corpus (as well as with files in the ELDP archive). Some texts of this collection have been published, and the information about these publications is mentioned in the metadata.

### 2.3. Content

The corpus contains texts of various genres, which are broadly classified (following the INEL project conventions) as folklore, narrative (monologues that are neither folklore texts nor songs), conversation and song. The songs category also includes a poem. The only text directly translated from Russian into FE in a written form is classified as translation apart from the other genres.

### 2.4. Corpus size

Table 1 contains the summary of the number of speakers, number of texts (“communications,” in EXMARaLDA terms) in total and specified by genre, number of sentences and number of tokens, separately for FE and TE and in total. Texts where a FE speaker interacts with TE speakers are classified here as TE texts.

*Table 1. Corpus size*

|                             |               | FE               | TE            | Total          |
|-----------------------------|---------------|------------------|---------------|----------------|
| Speakers                    |               | 57 <sup>12</sup> | 25            | 82             |
| Texts<br>(“Communications”) | Folklore      | 134              | 18            | 152            |
|                             | Narratives    | 337              | 93            | 430            |
|                             | Conversations | 47               | 24            | 71             |
|                             | Songs         | 23               | 2             | 25             |
|                             | Translations  | 1                | 0             | 1              |
|                             | Total         | 541              | 137           | 678            |
| <b>Sentences</b>            |               | <b>41,396</b>    | <b>12,737</b> | <b>54,133</b>  |
| <b>Tokens</b>               |               | <b>173,379</b>   | <b>45,331</b> | <b>218,710</b> |

<sup>12</sup> This number does not include upto 11 unidentified speakers of FE; some of these might possibly coincide with one another or with some identified speakers.

## 2.5. Naming Conventions

### 2.5.1. Name of the corpus

The name of the corpus is INEL Enets Corpus.

### 2.5.2. Folder structure and filenames

The entire corpus is contained in the folder “enets” which has the following files and subfolders.

Folders with text transcripts, organized by genre:

- “flk” (folklore texts)
- “nar” (narrative texts)
- “song” (songs)
- “conv” (conversations)
- “transl” (a text translated into Enets 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 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)
- (if available) sound file with the recording in WAV format (\*.wav) [“standard” and “video” packages] or MP3 format (\*.mp3) [“mp3” package]
- (if available) video file with the simultaneous video recording in MPEG2 (\*.mpg) or MPEG4 (\*.mp4) format [“video” package]
- (if available) 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”]
- (if available) legacy versions of transcripts originally glossed in Toolbox and converted into ELAN format, as well as texts aligned and transcribed in ELAN (\*\_eaf.xml) [all packages except “lite”]
- (if available; only for legacy oral transcription recordings of texts for which there is no original audio recording available) oral transcription audio recordings in WAV format (\*.wav) [“video” package]

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.

Video files as well as legacy oral transcription sound recordings are only included in the “video” package.

In the “lite” package of the corpus, neither sound nor video files are included.

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 3.3.7–3.3.10)

- o "gloss-panel-enets.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:

- "enets.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); year of recording; short title; genre abbreviation. These components are joined by underscore ("\_").

If the year of recording is only approximate or altogether unknown, a placeholder character "X" is used to fill the missing digits (e.g., "199X"). In the text metadata, more details on the date of recording can be specified if known (e.g., "early 1990s" instead of "199X").

The short title as part of a text name is a (possibly shortened) version of the English title, spelled without spaces, hyphens or other non-letter characters, with all initial capitals.

For ELDP + data, as well as for texts with no original title, both Russian and English titles are given by the corpus developers. If a text taken from a secondary source has an original Russian title, this title is usually kept as is, and the English title is its reasonably close translation. German and Hungarian original titles were reasonably closely translated into Russian and into English.

The genre abbreviations used are listed in 2.5.5.

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

**Text name:** PNS\_1994\_FormerLife\_nar

**Speaker code:** PNS (i.e. Pal`chin, Nikolaj Sergeevich)

**Year of recording:** 1994

**Short title:** FormerLife (i.e. "Former life")

**Genre:** nar (narrative)

A number of texts have the same short title, but are distinguished by numbers ("1", "2" etc.). First, this may refer to different parts of texts related to the same topic with presumed same event of recording: e.g. BLD\_IGA\_2016\_Fishing1\_conv and BLD\_IGA\_2016\_Fishing2\_conv both refer to "Conversation about fishing" between the same two speakers recorded the same day in the same place, but with a pause in between. Second, this may refer to different texts on the same topic (the same folklore or narrative story, description of the same object) recorded from the same speaker on different occasions: e.g., BVN\_1993\_Witcher1\_flk and BVN\_1996\_Witcher2\_flk are different versions of the same folk tale performed by the same speaker, but recorded by different researchers. Metadata contain explicit information on whether different parts or different versions are meant.

### 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. aforementioned "PNS" stands for Pal`chin, Nikolaj Sergeevich (Пальчин, Николай

Сепреевич). Still, letter combinations Ch, Sh, Yu are used for the corresponding transliterated letters Ч, Ш, Ю, e.g. “ChNP” (not “CNP”) for Chardu, Nina Petrovna (Чарду, Нина Петровна). If a code is already assigned to a different speaker, additional letters are used from one or more of the name parts, e.g. “BNK” for Bolina, Nadezhda Konstantinovna (Болина, Надежда Константиновна) and “BNKu” for Borisova, Nina Kupriyanovna (Борисова, Нина Куприяновна). In the rare cases where speakers have identical family names, first names and patronymics, a fourth capital letter is used for the code of one of them referring to their domicile, e.g. “SIIP” referring to Silkin, Ivan Ivanovich (Силкин, Иван Иванович) who lived in Potapovo, in contrast to “SII” referring to Silkin, Ivan Ivanovich (Силкин, Иван Иванович) who lived in Vorontsovo.

Appendix A4 contains the full list of Enets speakers who contributed to the corpus along with their codes.

For a number of texts, the speakers are unidentified; the placeholder “NN” stands for the speaker’s name in such a case. Still, presumably different unidentified speakers are distinguished with digits (NN1, NN2 etc.).

## 2.5.5. Abbreviations used in metadata

### Contributors

The following codes are used for people who are not Enets speakers but who contributed to recording, digitizing, transcribing, translating, glossing and editing the texts included into the corpus, or mentioned in the metadata for some other reasons. This list also includes contributors who are Enets speakers and who were involved into activities other than composing texts, storytelling and oral transcription (their codes are the same as their speakers’ codes). In square brackets, the standard transliteration (see Appendix A3) is provided for those whose variants of names used in English publications are different. In parentheses, variants of names are provided for those who have ones.

AAV – Arkhipov [Arxipov], Alexandre (Aleksandr Vladimirovich)

BDS – Bolina, Dar`ya Spiridonovna

BeV – Bettu, Vera

BrJ – Bradley, Jeremy

BNN – Bolina, Nina Nikolaevna

BZN – Bolina, Zoya Nikolaevna

DOE – Dobzhanskaya, Oksana E`duardovna

FP – Fröhlich, Peter

GrS – Grawunder, Sven

GVY – Gusev, Valentin (Valentin Yur`evich)

GYA – Gluxij, Yaroslav Andreevich†

HE – Helimski [Xelimskij], Eugene (Eugen, Evgenij Arnol`dovich) †

KAI – Kuznetsova [Kuznecova], Ariadna Ivanovna†

KhO – Khanina [Xanina], Olesya (Ol`ga Vladimirovna)

KiN – Kim, Natalia

LE – Lazarenko, Elena

LK – Labanauskas, Kazimir Isidorovich (Kazis)†  
 LL – Leisiö (Afonina), Larisa (Larisa Gennadievna)  
 LM – Lyublinskaya, Marina Dmitirievna  
 LVF – Ly`rmin, Vasilij Fedorovich†  
 MK – Mägi, Kaur  
 MT – Mikola, Tibor†  
 OM – Ovsjannikova [Ovsyannikova], Maria (Mariya Aleksandrovna)  
 PJ – Pusztay, János  
 PMN – Pal`china, Marina Nikolaevna  
 PVN – Pal`chin, Viktor Nikolaevich  
 RA – Riaposov [Ryaposov], Aleksandr  
 SF – Siegl, Florian  
 ShA – Shluinsky [Shluinskij], Andrey (Andrej Boleslavovich)  
 SIP – Sorokina, Irina Petrovna†  
 StN – Stojnova [Stojnova], Natalia (Natal`ya Markovna)  
 SVA – Susekov, Vasilij Andreevich†  
 TeNM – Tereščenko [Tereshhenko], Natal`ya Mitrofanovna†  
 TrS – Trubetskoy [Trubeckoj], Sergey (Sergej Anatol`evich)  
 UA – Urmanchieva, Anna (Anna Yur`evna)  
 WNB – Wagner-Nagy, Beáta  
 ZV – Zav`yalov, Vasilij Anatol`evich

### **Institutions**

The following codes are used for institutions whose archive data are used in the corpus:

GTRK – Dudinka branch of GTRK “Norilsk” (Дудинское отделение ГТРК «Норильск»; also known informally as “Taymyr radio” or “Dudinka radio”)  
 ILI – Institute for Linguistic Studies RAS (Institut lingvisticheskix issledovanij RAN / Институт лингвистических исследований РАН)  
 TDNT – Taymyr House of National Arts (Tajmy`rskij Dom Narodnogo Tvorchestva / Таймырский дом народного творчества, also known informally as “Glavny`j chum / Главный чум” and also known in English as Taymyr House of Folk Culture)

### **Publications**

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

GS1982 – Gluxij & Susekov 1982  
 L1992 – Labanauskas 1992  
 L2002 – Labanauskas 2002  
 M1967 – Mikola 1967  
 M1980 – Mikola 1980



P1978 – Pusztay 1978

SB2005 – Sorokina & Bolina 2005

U2008 – Urmanchieva 2008

## **Genres**

The following abbreviations are used for text genres:

flk – folklore texts

nar – narrative texts

song – texts of songs

conv – conversations

transl – texts translated into Enets from Russian

### **2.5.6. Transliteration of Cyrillic names**

In the metadata fields referring to personal names and placenames (see 2.8), a romanized spelling is used alongside their Cyrillic spelling according to the Russian orthography.

The following placenames use (a variant of) traditional English spelling:

Dudinka

Leningrad (historical name of St. Petersburg)

Norilsk

St. Petersburg

Taymyr

Yenisey

All personal names and most placenames in the metadata are transliterated following the GOST 7.79–2000 System B transliteration standard (GOST 2001; see Appendix A3, Table 32).

Elsewhere, e.g. in text titles, English glosses and free translation, English-style romanization is used (see Appendix A3, Table 33).

## **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.

### **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”<sup>13</sup>) for representation of spoken data, and at the same time a TEI<sup>14</sup> 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**”) mono or, more rarely, stereo, with 44.1 or 48 kHz sampling frequency and 16 bit depth. It is the native format of most ELDP+ recordings done in 2008–2017, while others have been either digitized from analog recordings or converted from other formats (see more details in sections 2.2 and 2.7.2). Audio recorded in 2005 has lower sampling frequency (16 or 24 kHz). Some texts recorded in 1994 were digitized at a higher sampling frequency (96 kHz), with 16 or 24 bit depth.

---

<sup>13</sup> <https://www.iso.org/standard/37338.html>, last access: 05.11.2024

<sup>14</sup> <https://tei-c.org/>, last access: 05.11.2024

For texts with available video recordings, video files are provided in MPEG2 or MPEG4 formats (file extensions “.mpg” and “.mp4”).

- MPEG2 files from 2008 were digitized from analog video cassettes (Video: MPEG2, 768x576, 25 fps, max. bit rate 8192 kbps, avg. bit rate ca. 3000 kbps; Audio: mp2, stereo, 48 kHz, 128 kbps)
- MPEG2 files from 2009–2010 are natively digital recordings (Video: MPEG2, 1024x576, 25 fps, max. bit rate 8192 kbps, avg. bit rate ca. 3300 kbps / 720x576, 25 fps, max. bit rate 8192 kbps, avg. bit rate ca. 2600 kbps; Audio: mp2, stereo, 48 kHz, 128 kbps)
- MPEG4 files were converted from original MTS files (Video: H264, 720x576, 25 fps, avg. bit rate ca. 640 kbps 584–746kbps; Audio: AAC, stereo, 48kHz, ca. 240kbps)

MP3 versions of all sound files are also provided as a light-weight option (44.1kHz, 192kbps).

#### 2.6.4. Other data

For all texts, original source(s) of their transcription are available.

PDF format (“\*.pdf”) corresponds to original Word documents, to scans of original publications or manuscripts and to excerpts from editorial layouts.

ELAN EAF format (“\*\_eaf.xml”) corresponds to glossed texts imported from Toolbox databases (technically, converted from interlinear plain text files to EAF format) and to texts aligned and transcribed in ELAN. They keep their legacy filenames with the addition of “\_eaf” suffix and use the “\*.xml” extension. This allows an ELAN user to select the entire corpus folder as a search domain without including the legacy transcripts. In order to open legacy transcript files with ELAN, they should be renamed to have the “\*.eaf” extension.

For some texts digitized recordings of their oral transcription coming from Labanauskas’ tapes in TDNT archive are available in WAV format (“\*.wav”). In contrast to the original audio recordings, such sound files are linked in the enets.coma (main metadata file) as “attached files”, not as “recordings”.

### 2.7. Workflow of the source files

#### 2.7.1. Transcripts

While the main part of analysis of all texts was done in the *SIL Fieldworks Language Explorer* (FLEX)<sup>15</sup> and all the transcripts were exported to the EXB format via the FLEX’s Flextext XML format, previous workflow was different depending on the source of the text.

Most published sources were imported into FLEX via plain text SFM (SIL Standard Format, used e.g. in Toolbox<sup>16</sup>). Later, some of these texts were exported into ELAN in order to align them with the available audio recordings and then imported back to FLEX.

For some other published texts, their original Latin and/or Cyrillic transcription lines (**st** and **stl** correspondingly) and the original Russian translation line (**ltr**) were copied from the original files into text files (for Labanauskas 2002, an automatically recognized text from a scanned PDF file was used, which was done in PDF Transformer back in 2010). The same was done with texts preliminarily transcribed in Word format. For all texts from archive manuscripts and for a minority of published texts, **st** / **stl** and **ltr** lines were typed manually into SFM plain text files (for Leisiö and Gusev’s transcripts they were simply copied

---

<sup>15</sup> <https://software.sil.org/fieldworks/>, last access: 12.06.2024.

<sup>16</sup> <http://www.fieldlinguisttoolbox.org/ToolboxReferenceManual.pdf>, last access: 04.11.2024

from MS Word files). Tiers containing draft versions of the phonological transcription (**ts** tier) and edited Russian translation (**fr** tier) were produced manually at the SFM plain text stage (and corrected later in FLEEx). Texts with no available audio recordings were then imported directly into FLEEx from the SFM plain text format. Texts with available audio recordings were first imported to ELAN, time-aligned with the sound file and then exported to FLEEx via Flextext XML.

As mentioned in section 2.2, ELDP+ draft transcriptions and Russian translations were either done in ELAN or preliminarily done with pen and paper. In the latter case, transcription and translation were manually typed in ELAN (though Latin transcription was typed in instead of the Cyrillic transcription). For texts previously glossed in Toolbox, it was done back in 2009–2013; for texts glossed directly in FLEEx it was done directly before glossing in 2023–2024.

As mentioned in section 2.2, most of ELDP+ transcriptions included into the corpus were glossed in Toolbox<sup>17</sup> in 2009–2013 after being time-aligned in ELAN. Later, Toolbox interlinear texts were again imported into ELAN and then into FLEEx. Importing glossed texts into FLEEx was done with a set of custom XSLT and Python scripts (see Riaposov et al. In preparation) which allowed to keep existing Toolbox interlinear analysis and merge it with the independently filled FLEEx lexicon. During and after the import, numerous changes were made to transcription, glosses, part of speech tags and speaker codes in order to harmonize the two datasets.

The following tiers originally come from Toolbox interlinear in this case: original draft Latin transcription line (tier **stl**), original draft Russian translation line (tier **ltr**), INEL transcription (**ts**), edited Russian and English translation lines (tiers **fr** and **fe**), occasional notes done in the process of transcribing (**nto**), and most morpheme-level tiers (**mp**, **ge**, **gr**, **mc**). Most of these tiers (except **stl**, **ltr**) were substantially changed compared to Toolbox source. Morpheme breaks tier (**mb**) and the part-of-speech tier (**ps**) were created in FLEEx.

ELAN transcriptions of texts that were not previously glossed in Toolbox were first preliminarily edited in ELAN (both technically and in order to check the transcription) and directly exported from ELAN to FLEEx. In this case the following tiers were created in ELAN: original draft Latin transcription line (tier **stl**), original draft Russian translation line (tier **ltr**), INEL transcription (**ts**), edited Russian translation line (tier **fr**), occasional notes done in the process of transcribing (**nto**) and sometimes developers' notes (tier **nt**).

For all transcripts except for those imported from Toolbox, the morphological analysis (interlinear glossing) was done in FLEEx; for transcripts imported from Toolbox, the morphological analysis was checked and edited in FLEEx. In the process of import from Toolbox, all the morpheme-level tiers were created or edited (**mb**, **mp**, **ge**, **gr**, **mc**), as well as the part-of-speech tier (**ps**). INEL transcription (tiers **ts** and **tx**) was also edited in FLEEx, as the work on unified phonological transcription checking (see section 3.2.4) was technically conducted in FLEEx lexicon. The **BOR** and **geo** tiers were also filled directly from the FLEEx lexicon. Together with the glossing, Russian translation (**fr**), English translation (**fe**) and notes (**nt**) were provided and/or edited. Separate FLEEx databases (“projects”) were maintained for FE and TE. Several texts primarily focused on TE but also containing speech in FE were analyzed in the TE FLEEx database.

Glossed texts were exported from FLEEx as Flextext XML and converted to EXMARaLDA EXB format. During this conversion, the **ref** tier was created which combines text code and sentence numbering (see

---

<sup>17</sup> <https://software.sil.org/toolbox/>, last access: 12.06.2024.

below). There are also some changes to the morpheme-level tiers concerning the representation of zero morphs (see 3.3.2–3.3.4).

After that, all further editing was done in the EXMARaLDA Partitur-Editor.

### 2.7.2. Media

As mentioned in section 2.2, audio recordings used in the corpus come from various sources.

Recordings made in Linear PCM WAV format are kept in their original format. The great majority of these were done on a professional recorder Marantz PMD-660 equipped with an external microphone AKG 1000S. Still, in some TE recordings from 2009 the external microphone was not used, and some FE recordings from 2010 were done with the digital recorder Olympus VN-960PC.

Recordings done in 2005, 2006 and 2009 in MP3 format were converted into PCM WAV format with Goldwave software.<sup>18</sup>

Some analog recordings from cassette tapes were digitized with a tape recorder TASCAM PORTA 02MKII and a digital recorder Marantz PMD-660. Technical parameters of those recordings which were originally analog, but were available to us already in the digitized form are at this point unknown. Neither are we aware of the technical details of digitizing of reel-to-reel tapes and of the conversion of Minidisk recordings into PCM WAV format.

Video recordings from 2008 were digitized from analog video cassettes into MPEG2 format; video recordings from 2009–2010 were originally done in MPEG2 format. In both cases, MPEG2 video files were edited in the VideoPad Video Editor,<sup>19</sup> so that the audio tracks were inserted there from the simultaneous audio recordings stored as WAV files. Video recordings from 2015–2016 were originally done in MTS format; a video recording from 2010 was available in MOV format. These recordings were edited with the Shotcut<sup>20</sup> software in order to add audio tracks from the simultaneous audio recordings stored as WAV files and to convert them into MP4 format.

### 2.7.3. Metadata

Text and speaker metadata were manually systematized in EXMARaLDA Coma and come from the following sources:

- Original metadata from published sources and from legacy manuscripts and legacy recordings
- Metadata maintained by Khanina, Shluinsky, Ovsjannikova, Stoyanova and Trubetskoy in 2008–2017 in XML format
- Legacy recordings' version analysis done by Khanina in 2008–2009
- Speaker metadata collected by Khanina and Shluinsky in handwritten form in 2008–2010, as well as Shluinsky's sporadic handwritten notes in fieldtrips and handwritten genealogical schemas done by Vadim Alekseevich Silkin during his consultation
- Legacy handwritten notes from 1974–1975 done by Ariadna Ivanovna Kuznetsova
- Speaker metadata provided in (Barbolina & Kudryakova 2013) and speaker metadata mentioned in (Amelina 2020)
- Facts mentioned in the Enets texts themselves

---

<sup>18</sup> <https://goldwave.com/>, last access: 12.06.2024.

<sup>19</sup> <https://www.nchsoftware.com/vepad/index.html>, last access: 10.07.2024.

<sup>20</sup> <https://www.shotcut.org/>, last access: 12.06.2024.

- Speaker identification of legacy recordings by Enets speakers, mostly by Ekaterina Spiridonovna Glibchenko
- Handwriting identification of some legacy manuscripts by Dar`ya Spiridonovna Bolina
- Occasional facts known from elsewhere

In cases where metadata from different sources contradicted each other, the information from the source judged to be more reliable was kept. Still, in some cases the metadata include alternatives with further references to existing sources.

## 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 mainly biographical data, but sometimes also basic data on language repertoires.

Personal names and place names are given in their romanized spelling alongside Cyrillic spelling according to Russian orthography. See 2.5.6 on the use of transliteration in these fields.

### 2.8.1. General corpus metadata

The general metadata about the whole corpus include the corpus name (“INEL Enets 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:**

- **0a. Title:** Short title (in English)
- **0b. Title (RU):** Short title (in Russian)
- **1. Genre:** Abbreviation of the genre of the text (flk = folklore, nar = narrative, song = song, transl = translation, conv=conversation).
- **2a. Recorded by:** Abbreviation of the person by whom the text was recorded (see 2.5.5)
- **2b. Date of recording:** Here the date of recording is given.
- **3. Dialect:** Enets lect used by the speaker(s) is given here (Forest, Tundra, both, or Forest with Tundra influence).
- **4. Speakers:** Code(s) of the speaker(s)
- **5a. Transcribed by:** Code(s) of the person(s) who did the transcription
- **5b. Date 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. Digitized by:** Code(s) of the person (or institution when specific person is not known) who did the digitizing (or converting from Minidisks) of original recordings (left empty for texts with no recording available, marked “Originally digital” for texts originally recorded in WAV or MP3 formats)

- **6c. Time-aligned by:** Code(s) of the person(s) who time-aligned the transcription
- **6d. Has audio:** Marked “Yes” if a sound recording is available
- **6e. Has video:** Marked “Yes” if a video recording is available
- **7a. Translation into Russian:** Code(s) of the person(s) who did the first available translation into Russian (in most cases, the same as person(s) who did the transcription).
- **7b. Translation into Russian edited:** Code(s) of the person(s) who edited the Russian translation.
- **7c. Translation into English:** Code(s) of the person(s) who did the translation into English or edited it.
- **7d-e. Translation into German / Hungarian:** Code of the person who did the translation in question (filled for texts coming from sources published in German or Hungarian respectively).
- **8. Glossed by:** Code(s) of the person(s) who did the morphological glossing
- **9a-e. Annotation SeR / SyF / IST / BOR\_CS / geo:** Codes of the person(s) who annotated the respective tiers

**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), except for those recorded in 2010 in Päivölä (Finland) where FE speakers were invited for a short working visit
- **Region:** The great majority of the texts originate from Taymyr (including Norilsk which is not formally a part of Taymyr region, but an enclave in it; currently, Taymyr region is part of the Krasnoyarsk Krai). Exceptions are texts recorded in Päivölä (Finland) in 2010 and texts recorded in 1960s–1970s in Leningrad (nowadays St. Petersburg)
- **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(s) of the text (*enf* – Forest Enets, or *enh* – Tundra Enets, or both).

**Setting:** In this section some information about legacy sources and existing publications is given.

- **1a. Archive (sound):** Code(s) of the owner(s) of the archive from which the recording comes (further details, such as names of the file where the recording is cut from can be mentioned); empty for texts with no recording available
- **1b. Archive (written):** Code(s) of the owner(s) of the archive for texts for which unpublished transcriptions are available (further details, such as names of the file or folder/sheets of paper archive can be mentioned); empty for texts available only in published form
- **2a. Published in:** If the text has been published, the publication reference is provided here (for most texts, in a shortened form, see 2.5.5)
- **2b. Published in (bibtex):** If the text has been published, the BiBTex key of the corresponding entry in the INEL Bibliography<sup>21</sup> is given here
- **3. Cyrillic orthography:** Marked “Yes” if a line in Cyrillic orthography (tier **tsc**) 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 (.exs) are linked here to the text description; the latter is needed for searching the corpus.

<sup>21</sup> <https://inel.corpora.uni-hamburg.de/portal/kataloge/bib/>; <https://doi.org/10.25592/uuhfdm.731>

**Attached file(s):** All other relevant files are linked to the text description here if available: PDFs with original publication(s) and/or scanned manuscripts, EAF files with original transcriptions in ELAN, video recordings, WAV files with Labanauskas' oral transcriptions.

### 2.8.3. Speaker metadata

Metadata of speaker(s) include, on the one hand, biographical information of the speaker and, on the other hand, information on their sociolinguistic background. The level of detail is determined by the available information. 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, but in special cases in parentheses also maiden / married names, unofficially used alternate names, as well as place of residence for a speaker who has the same name as a different speaker)

**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 (as well as former married names for those married for the second time)
- **2a-b. Given name (EN, RU):** For those who have alternate names widely used unofficially, these names are given in brackets
- **3a-b. Patronymic (EN, RU)**
- **4. Also known as:** Unofficial names are given here. These can be traditional names following pre-Christian tradition or nicknames.

**Basic biographic data:** Here basic biographical data of the speaker are provided. Note that settlement names can refer not always strictly to the settlement itself, but also to tundra camps around this settlement.

- **1. Place of birth**
- **1a Place of birth (LatLng):** Geographic coordinates (latitude, longitude) of place of birth
- **2. Region:** (always Taymyr)
- **3. Country:** (always Russia)
- **4. Date of birth**
- **5. Date of death**
- **6. Grown up in / former residences:** If former residences prior to the work with the linguist are known and differ significantly from place of birth, they are mentioned here
- **7. Domicile:** The current (i.e. at the time of the recording) place of residence of the speaker if known
- **7a. Domicile (LatLng):** Geographic coordinates (latitude, longitude) of domicile

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

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



**Family:** Here information about the ethnicity of the speaker (in the majority of cases, Enets) 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**
- **2-3. Ethnicity of mother / Name of mother**
- **4-5. Ethnicity of father / Name of father**
- **6-7. Ethnicity of husband/wife / Name of husband/wife**
- **8. Ethnicity of grandparents**
- **9. Clan:** Name of the Enets traditional clan the speaker belongs to is given here if known

#### Language documentation activities

- **Consultant of:** Here the linguists or other researchers with whom the speaker worked are mentioned. If known, this information concerns also work beyond the texts included in the corpus.

**Languages:** Here the language repertoire of the speaker is given; Russian (most commonly L2) is not mentioned.

#### L1 (First language)

- **Language code:** Here the ISO code is given (usually *enf* – Forest Enets or *enh* – Tundra Enets, but for some speakers *yrk* – Tundra Nenets).
- **1-2. First language / Dialect:** Normally, Enets / Forest Enets or Tundra Enets.

#### L2 (Second language)

- **Language code:** Here the ISO code is given (*yrk* – Tundra Nenets for some speakers; *enf* – Forest Enets or *enh* – Tundra Enets for those who have Tundra Nenets as L1).
- **1-2. Second language / Dialect:** Enets (Forest or Tundra), or Nenets (Tundra).

### 3. Transcription and annotation

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

#### 3.1. Tier layout

Table 2. Overview of the annotation tiers

| Tier label | Tier full name                  | Description  | Unit     | Optionality |
|------------|---------------------------------|--|----------|-------------|
| ref        | Reference                       | Text ID + sentence number<br>Text ID + speaker code + sentence number (for texts with multiple speakers) | sentence | obligatory  |
| st         | Source transcription (Cyrillic) | Original transcription / orthography of the source if written in Cyrillic alphabet                       | sentence | optional    |

| <b>Tier label</b> | <b>Tier full name</b>        | <b>Description</b>   | <b>Unit</b>           | <b>Optionality</b>                         |
|-------------------|------------------------------|--|-----------------------|--|
| <b>stl</b>        | Source transcription (Latin) | Original transcription of the source if written in Latin alphabet, or original Cyrillic transcription / orthography converted into Latin | sentence              | optional                                   |
| <b>ts</b>         | Text (sentence)              | Main phonological transcription for glossing and annotation  | sentence              | obligatory                                 |
| <b>tsc</b>        | Text (sentence) (Cyrillic)   | Main transcription rendered in Cyrillic orthography  | sentence              | optional                                   |
| <b>tx</b>         | Text (word)                  | Main phonological transcription segmented by words   | word                  | obligatory                                 |
| <b>mb</b>         | Morpheme breaks              | Morpheme breakdown of words (hyphen-separated morphemes for each word)   | morph                 | obligatory                                 |
| <b>mp</b>         | Morphemes (lexical)          | Lexical representation of morphemes (see notes in 3.3.3)   | morph                 | obligatory                                 |
| <b>ge</b>         | Gloss (English)              | Morphological glosses (with lexical glosses in English)  | morph                 | obligatory                                 |
| <b>gr</b>         | Gloss (Russian)              | Morphological glosses (with lexical glosses in Russian)  | morph                 | obligatory                                 |
| <b>mc</b>         | Morphological category       | Morphological category/part of speech for each morpheme  | morph                 | obligatory                                 |
| <b>ps</b>         | Part of speech               | Part of speech for each word   | word                  | obligatory                                 |
| <b>SeR</b>        | Semantic role                | Semantic (thematic) roles of NPs   | word / group of words | optional                                   |
| <b>SyF</b>        | Syntactic function           | Syntactic functions of predicates and arguments, as well as for subordinate clauses  | word / group of words | optional                                   |
| <b>IST</b>        | Information status           | Information status of NPs (given/new/accessible)   | word                  | optional (always empty in current version) |
| <b>BOR</b>        | Borrowing                    | Borrowings (source language and borrowing type)  | word                  | optional                                   |
| <b>BOR-Phon</b>   | Borrowing phonology          | Phonological adaptations in borrowings   | word                  | optional (always empty in current version) |
| <b>BOR-Morph</b>  | Borrowing morphology         | Morphological adaptations in borrowings  | word                  | optional (always empty in current version) |
| <b>CS</b>         | Code switching               | Code switching and calques (source language and type)  | group of words        | optional                                   |
| <b>geo</b>        | Geographical coordinates     | Coordinates of named places and geographical objects   | word                  | optional                                   |

| Tier label | Tier full name                   | Description   | Unit     | Optionality |
|------------|----------------------------------|---|----------|-------------|
| fr         | Free translation (Russian)       | Free Russian translation  | sentence | obligatory  |
| fe         | Free translation (English)       | Free English translation  | sentence | obligatory  |
| ltr        | 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    |
| ltg        | Original translation (German)    | Original German translation, as provided in the source (for sources published in German)  | sentence | optional    |
| lth        | Original translation (Hungarian) | Original Hungarian translation, as provided in the source (for sources published in Hungarian)  | sentence | optional    |
| nt         | Notes                            | Notes from corpus developers (in English)   | sentence | optional    |
| nto        | Notes (original)                 | Notes from the source / notes made in the process of draft transcription  | sentence | optional    |

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

Figure 1. A sample transcript fragment showing the tier layout

|     |  |         |                         |                    |                   |       |         |                            |                   |
|-----|--|---------|-------------------------|--------------------|-------------------|-------|---------|----------------------------|-------------------|
| ref | BTN_1969_Guzzler_flk.009 (009)   |         |                         |                    |                   |       |         |                            |                   |
| st  | Эсаза сойза лубхаза мие, эза ан' сойза мальчазуда мя.                  |         |                         |                    |                   |       |         |                            |                   |
| stl | Esađa sojđa lubxađa m'ije, eđa an' sojđa mal'čazoda m'a.               |         |                         |                    |                   |       |         |                            |                   |
| ts  | Eseza sojza lubaxazoda toza, eeza an'i sojza mal'čazoda miʔe.          |         |                         |                    |                   |       |         |                            |                   |
| tx  | Eseza  | sojza   | lubaxazoda              | toza,              | eeza              | an'i  | sojza   | mal'čazoda                 | miʔe.             |
| mb  | ese-za   | sojza   | lubaxa-zo-da            | toza               | ee-za             | an'i  | sojza   | mal'ča-zo-da               | miʔe              |
| mp  | ese-za   | sojza   | lubaxa-zo-da            | toza               | ee-za             | an'i  | sojza   | mal'ča-zo-da               | mis               |
| ge  | father-NOM.SG.3SG  | good    | cloth-DST.SG-OBL.SG.3SG | bring(pfv).[3SG.S] | mother-NOM.SG.3SG | again | good    | fur.coat-DST.SG-OBL.SG.3SG | give(pfv).[3SG.S] |
| gr  | отец-NOM.SG.3SG  | хороший | ткань-DST.SG-OBL.SG.3SG | принести.[3SG.S]   | мать-NOM.SG.3SG   | опять | хороший | малица-DST.SG-OBL.SG.3SG   | дать.[3SG.S]      |
| mc  | n(d)-n:(poss)  | adj     | n(d)-n:dst-n:(poss)     | v(d(irr)).[v:pn]   | n(d)-n:(poss)     | ptcl  | adj     | n(d)-n:dst-n:(poss)        | v(avs).[v:pn]     |
| ps  | n  | adj     | n                       | v                  | N                 | ptcl  | adj     | n                          | v                 |
| BOR |  |         | RUS:cult                |                    |                   |       |         | NEN:cult                   |                   |
| fe  | His father gave him a good shirt, his mother gave him a good fur coat. |         |                         |                    |                   |       |         |                            |                   |
| fr  | Отец ему хорошую рубаху дал, мать хорошую малицу ему дала.             |         |                         |                    |                   |       |         |                            |                   |
| ltr | Отец хорошую рубаху дал, мать хорошую малицу-ему дала.                 |         |                         |                    |                   |       |         |                            |                   |

See 3.4.2 (Table 12 and Table 15) for the correspondence between annotation tiers and search fields in the online version provided via the Tsakorpus platform.

### 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 **tx** has the same content 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 and alternatives in transcription, unintelligible fragments, false starts and non-speech sounds largely follow general INEL conventions described in (Arkhipov 2020: 12–17). Still, due to the fact that the Enets corpus imports a large number of texts transcribed and glossed independently of the INEL project, these conventions are not always followed consistently. The main differences are as follows.

- All kinds of uncertainties are usually marked informally in notes tiers (**nt** and/or **nto**).
- There are no cases of alternative transcriptions in the Enets corpus.
- Ellipsis character surrounded with double round brackets, ((...)) sequences, is used in the Enets corpus not only for unintelligible fragments, but also for non-speech events, untranscribed code-switching and sentences said by recording linguists (in Enets or in Russian). In the cases where ((...)) marks something else than unintelligible fragments this is usually explained in notes tiers (**nt** and/or **nto**). Non-speech sounds (such as cough or laugh) are never labelled in the **tx** tier (but are sometimes commented in notes tiers).
- Most false starts smaller than a word are not transcribed but marked as unintelligible ((...)). False starts with completed words are not enclosed brackets (and, generally no distinction is made between false starts with completed words, other kinds of speech repetitions and other kinds of disfluencies).
- Mismatches between original texts and translation are not usually marked in any form. Free translation tiers (**fe** and **fr**) are aimed to be as close to the original as possible, but at the same time to express the relevant sense in standard English and Russian; sometimes information missing in the source sentence but important for interpretation can be added in parentheses.

#### 3.2.2. Source transcription tiers (st and stl)

For the texts coming from a source using Cyrillic transcription / orthography, the source transcription tier (**st**) contains the original Cyrillic version of the text (otherwise it is empty); in this case the **stl** tier may contain this original phonetic transcription converted into Latin alphabet. An example for these two tiers in this configuration may be found in Figure 1.

For the texts coming from a source using Latin transcription (including those transcribed in ELAN), the **st** tier is always empty, and the **stl** tier contains the original transcription. In the case of original transcription done with pen and paper in Cyrillic, it is not typed, and the **stl** tier provides a latinized version of this transcription.

### 3.2.3. Cyrillic orthography tier (tsc)

The tier **tsc** contains, for some texts, the version of the **ts** tier represented in the Cyrillic orthography proposed for FE in 2019. In this tier, redundancies, false starts and marks of incomprehensible parts are usually omitted.

### 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. The choice of symbols follows the general conventions adopted in the INEL project. In particular,

- *č* (not the IPA *tʃ* and not the diachronically oriented *t'* or *ṭ*) is used for the voiceless dental affricate
- *š* (not the IPA *ʃ* and not the diachronically oriented *s'* or *ʂ*) is used for the alveolar sibilant
- *d', n', l'* (not the IPA *dʲ* or *j*, *nʲ* or *ɲ*, *lʲ* or *ɭ* and not *ǎ, ǎ́, ǎ̂*) are used for palatalized or palatal consonants
- *z* (not *ǎ, ǎ́, ǎ̂* or *ʒ*) is used for the dental or alveolar voiced fricative
- *ɛ* (not *ǎ*) is used for the FE mid-open front vowel

We distinguish between mid-open and mid-close vowels both in FE and TE, using *ɔ* (not *o*) for the mid-open one and *o* (not *ô*) for the mid-close one. For the whole set of phonemes and corresponding characters used in the transcription see Appendix A2. The project transcription is represented in tiers **ts** (Text (Sentence)) and **tx** (Text (Word)).

#### Original transcriptions and their correspondence to the INEL transcription

As seen from section 2.2, the corpus includes texts coming from very different sources. Helimski (1981/2003) and Urmanchieva (Ms a; 2008) use a phonological account presented in a Latin transcription. Mikola (1967; 1980), Puszta (1978) and Labanauskas in his archive manuscripts use Latin non-phonemic transcriptions following FUT. Gluxij & Susekov (1982) provide a phonetic transcription based on Cyrillic and presumably following the transcription system developed by Andrei Dulson. Sorokina & Bolina (2005) present their texts in a non-phonemic Cyrillic orthography (which is presumed to follow Tereščenko's (1986) orthographical proposals, but in fact is simplified); the same applies to manuscripts coming from Tereščenko's and Sorokina's archives. Labanauskas (1992; 2002) also uses a simplified version of Cyrillic orthography. Draft transcriptions of Leisiö and Gusev are also in Cyrillic script, but the logic of their writing is quite different. In Toolbox databases, before the import to FLEx we used an IPA transcription which aimed to be phonological, but transcriptions in ELAN files were not phonological neither exact, although also used IPA.

The phonological transcription used in the main tiers (**ts** and **tx**) is finally provided and/or verified manually based on the following sources of information on the phonological shape of morphemes.

For FE, the main source of such data are extensive audio recordings of lexicon elicitation collected in 2015–2016 and in 2019 by Andrey Shluinsky and Maria Ovsjannikova; the main consultants were Leonid Dmitrievich Bolin', Nadezhda Konstantinovna Bolina, Nikolaj Ivanovich Silkin and Gennadij Afanas'evich Ivanov', but Svetlana Alekseevna Roslyakova' and Zoya Nikolaevna Bolina also contributed. Beside these recordings, some instances of lexical items were also audially checked in the recordings included into the

corpus. Existing dictionaries (Sorokina & Bolina 2009 and FE part of Helimski (Ms) (based on the data collected by Anna Urmanchieva in 1995–1996), and existing spellings of lexical items in written sources were consulted, too. The FE dictionary (Shluinsky In press) was compiled in the process of verifying the phonological shape of items attested in the corpus.

For TE, the main source of phonological transcription was Helimski's (Ms) unpublished TE dictionary. Lexical items were also auditorily checked in the recordings included into the corpus. Existing spellings of these items in written sources were consulted.

For texts coming from a source with a Cyrillic orthography, this orthography is provided in the **st** tier (Source transcription). For some such texts, the **stl** tier (Source transcription (Latin)) contains results of an automatic conversion into Latin alphabet, further processed manually to obtain the transcription used in the main tiers. For texts coming from a source with a Latin orthography, it is provided in the **stl** tier (Source transcription (Latin)). In particular, in texts with transcriptions coming from ELDP+ data the **stl** tier contains the earliest draft of the transcription.

Table 31 in Appendix A2 summarizes the main correspondences of the transcription used in this corpus to some other transcriptions and/or orthographies used for Enets. Importantly, except for Helimski's system (also used in Urmanchieva's work) and for the orthography proposed for FE in 2019, other writing systems are not phonologically consistent and thus the correspondences are not strict.

### **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. At the same time, other punctuation marks are the same as in the sources and are not always uniform throughout the corpus.

The same applies to the sentence segmentation. For the texts coming from secondary sources, the sentences breaks are usually kept the same as in the source. For texts with transcriptions coming from ELDP+ data, the sentences are more similar to minimal discourse units and are based on intonation and pauses.

### **Code-switching**

Code-switching insertions in Russian are given in Cyrillic in Russian orthography in **ts** and **tx** tiers. Code-switching insertions in Tundra Nenets are given in a draft transcription using the INEL transcription system.

## **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 “PNS\_1980\_DyooAndStump\_flk.001” and “PNS\_1980\_DyooAndStump\_flk.002” respectively:

|            |   |   |
|------------|---|---|
| <b>ref</b> | PNS_1980_DyooAndStump_flk.001 (001.001) | PNS_1980_DyooAndStump_flk.002 (001.002) |
|------------|---|---|

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.)

|                |  |  |  |
|----------------|--|--|--|
| <b>ref-BAS</b> | BAP_BNN_1997_Interview1_conv.<br>BNN.001 (001) |  |  |
| <b>ref-BAP</b> |  | BAP_BNN_1997_Interview1_conv.<br>BAP.001 (002) | BAP_BNN_1997_Interview1_conv.<br>BAP.002 (003) |

### 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 (lexical) (mp)

The underlying morphemes tier (**mp**) shows the lexical representation of the morphs, both stems and affixes, which appear in the **mb** tier. It follows the FLEx lexicon which, importantly, differs for FE and TE (see section 2.7).

For both FE and TE, the lexical representation is given following a unified phonological account elaborated in the process of the corpus development. For FE it is significantly based on elicited lexical data and for TE to a significant extent follows Helimski's (Ms.) unpublished dictionary (see 3.2.4 for details), but even in the latter case the phonological representation in the corpus is not equal to the one provided by Helimski. As both Enets lects have a widespread phonological variation, for many lexical entries, still different variants are used in the transcription. In such cases the selection of the allomorph which appears in the **mp** tier is arbitrary.

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.<sup>22</sup> For the list of abbreviations used see Appendix A5.

Glosses for all morphemes within a word are separated by hyphens. Glosses for zero morphs are given in square brackets preceded by a dot (e.g. ".[3SG]").

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").

Morphemes with unknown meaning are glossed with two percent signs ("%"). In case of uncertainty (both in the meaning of the lexical item and/or in its phonological shape), the gloss is preceded with a single percent sign (e.g. "%whistle").

For an example 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 40 in Appendix A6 shows the tags used for parts of speech and inflectional categories. For inflectional affixes the pattern "x:a" is used, where x stands for part of speech, to which an affix can be attached, and a stands for the category of this affix. Derivational processes are marked as "x>y", x and y being the tags for part of speech. Elements with unknown meaning are marked with two percent signs ("%").

---

<sup>22</sup> <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>, last access: 07.06.2024.



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, e.g. the outcome of derivational processes is marked here. The list of possible parts of speech can be found in Appendix A6 (Table 39). Words with unknown part of speech are marked with two percent signs ("%%").

For an example see Figure 1.

### 3.3.7. Syntactic function (SyF)<sup>23</sup>

The annotation scheme used in the syntactic function tier was developed by Beáta Wagner-Nagy and Sándor Szeverényi (Wagner-Nagy et al. 2018: 24ff.) who also made it available for the INEL project.

In the Syntactic function tier (**SyF**), 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 3. Tags for syntactic functions

| Tag                        | Description                      |
|----------------------------|----------------------------------|
| <b>Main arguments</b>      |                                  |
| S                          | subject                          |
| O                          | direct object                    |
| <b>Predicate</b>           |                                  |
| v:pred                     | verbal predicate                 |
| n:pred                     | nominal predicate                |
| adj:pred                   | attributive/adjectival predicate |
| pro:pred                   | pronominal predicate             |
| ptcl:pred                  | particle predicate               |
| cop                        | copula                           |
| <b>Subordinate clauses</b> |                                  |
| 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

<sup>23</sup> Syntactic functions are only annotated in a few texts in the current version of the corpus.

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 (<.>).

*Table 4. Tags for referent expressions*

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

### 3.3.8. Semantic role (SeR)<sup>24</sup>

The annotation of semantic (thematic) roles is given in tier labelled **SeR**. It is based on GRAID principles (Haig & Schnell 2014) with some further developments by Beáta Wagner-Nagy and Sándor Szeverényi (Wagner-Nagy & Szeverényi 2018: 21ff.), further adapted for the current project.

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

*Table 5. Tags for semantic roles*

| Tag        | Description  |
|------------|--|
| <b>A</b>   | Agent: volitional initiator of the action, the participant is causing the action or is responsible for something happening.  |
| <b>B</b>   | Beneficiary: entity for whose benefit the action is performed.   |
| <b>Com</b> | Comitative: an entity which is involved in the situation in the same way as some other participant (e.g. co-agent, co-patient, etc.).                              |
| <b>Cau</b> | Cause: entity (mostly non-human) that causes an event without volition.  |
| <b>E</b>   | Experiencer: living entity that undergoes a sensory, cognitive, or emotional experience (first argument of the verbs of emotion, volition, cognition, perception). |

<sup>24</sup> Semantic roles are only annotated in a few texts in the current version of the corpus.

| Tag         | Description  |
|-------------|--|
| <b>G</b>    | Goal: location or entity towards which the movement happens.   |
| <b>Ins</b>  | Instrument: entity used to perform an action, which does not control the situation.  |
| <b>L</b>    | Location: location or entity where an event takes place or where something is located (i.e. stative location).   |
| <b>P</b>    | Patient: entity, which undergoes physical changes, is created or destroyed.  |
| <b>Path</b> | Path: entity or location along or through which the movement takes place.  |
| <b>Poss</b> | Possessor: entity which owns something.  |
| <b>R</b>    | Recipient: animate recipient of transfer or addressee of verb of speech  |
| <b>So</b>   | Source: location or entity from which the movement originates.   |
| <b>St</b>   | Stimulus: entity perceived by experiencer.   |
| <b>Th</b>   | 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.). |
| <b>Time</b> | 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 3.3.7).

### 3.3.9. Information status (IST)<sup>25</sup>

The Information status tier (IST) contains the annotation of information status. 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 (2018: 28ff.) and made available by them. 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 [ $\pm$  discourse-old] and [ $\pm$  hearer-old]:

Table 6. Parameters for determining information status

|                    | <b>+discourse-old</b> | <b>–discourse-old</b> |
|--------------------|-----------------------|-----------------------|
| <b>+hearer-old</b> | given                 | accessible            |
| <b>–hearer-old</b> | —                     | 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:

<sup>25</sup> Information status is not annotated in the current version of the corpus.

Table 7. Basic tags for annotating information status

| Tag                         | Description   |
|-----------------------------|---|
| <b>Given referents</b>      |   |
| giv-active                  | given and active referent (i.e. mentioned in the current or last sentence)  |
| giv-inactive                | given and inactive referent (i.e. mentioned before the last sentence)   |
| <b>Accessible referents</b> |   |
| accs-sit                    | referent, accessible through the situation (e.g. having breakfast: “Give me <u>the butter</u> , please.”)   |
| accs-aggr                   | referent, accessible through the aggregation of other referents (e.g. “Once upon a time, a king had a wife and two children. <u>They</u> lived happily.”) |
| accs-inf                    | referent, accessible through inference, e.g. part-whole relations (e.g. “We had a turkey for thanksgiving. I ate its <u>wings</u> .”)                     |
| accs-gen                    | referent, accessible through general knowledge (e.g. “ <u>The president of the U.S.</u> travelled to Cuba.”)  |
| <b>New referents</b>        |   |
| new                         | new referent  |

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

The Borrowing tier (**BOR**) contains the annotation of borrowed lexical items. Both the origin of the item in question and the type of borrowing is annotated. The tags are made up as follows: <LANGUAGE:class>. The annotation is implemented already in the FLEx lexicon and automatically exported to EXMARaLDA. For Enets, Russian (RUS), Tundra Nenets (NEN) and marginally – Evenki (EV) and Dolgan (DOLG) borrowings are tagged; Forest Enets (FE) borrowings are marked in TE texts and, vice versa, Tundra Enets (TE) borrowings are marked in FE texts.

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

Table 8. 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**<sup>26</sup> contains the annotation of phonological processes in borrowing. The tag set is the following:

Table 9. Annotation panel for phonological processes in borrowings

| Tag              | Description                |
|------------------|----------------------------|
| <b>Deletions</b> |                            |
| inCdel           | initial consonant deletion |

<sup>26</sup> The tier **BOR-Phon** is not annotated in the current version of the corpus.

|                      |                                     |
|----------------------|-------------------------------------|
| inVdel               | initial vowel deletion (aphaeresis) |
| medCdel              | medial consonant deletion           |
| medVdel              | medial vowel deletion (syncope)     |
| finCdel              | final consonant deletion            |
| finVdel              | final vowel deletion (apocope)      |
| <b>Insertions</b>    |                                     |
| inVins               | initial vowel insertion             |
| medVins              | medial vowel insertion              |
| finVins              | final vowel insertion               |
| <b>Substitutions</b> |                                     |
| Csub                 | consonant substitution              |
| Vsub                 | vowel substitution                  |
| <b>Other</b>         |                                     |
| lenition             | lenition (weakening)                |
| fortition            | fortition (strengthening)           |

The tier **BOR-Morph**<sup>27</sup> contains the annotation of morphological processes in borrowing. The tags are made up as follows: <Strategy:Inflection>. The tag set is the following:

Table 10. Tags for annotating morphological processes in borrowings

| Tag  | Description  |
|--|--|
| <b>Adaptation strategies</b>                       |  |
| 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)   |
| <b>Further inflection (in the matrix language)</b> |  |
| :bare  | no inflection  |
| :infl  | further inflection   |

### 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 language is mostly Russian (RUS), sometimes also Tundra Nenets (NEN). The tag set for the type of code-switch is the following:

Table 11. Tags for annotating code-switching

| Tag                                     | Description  |
|---|--|
| <b>Sentence-external code-switching</b> |  |
| :ext                                    | languages change at sentence (clause, utterance) borders |

<sup>27</sup> The tier **BOR-Morph** is not annotated in the current version of the corpus.

| 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                             | a single word is inserted, distinguishing between subtypes is problematic |

In the current version of the corpus, some instances of code-switching are not transcribed and not annotated; these are marked with ((...)) in **tx** tier. Sentence-external code-switching into Russian is annotated in **CS** tier when transcribed, but sentence-internal code-switching is mostly left unannotated (although can be seen as insertions of Russian words and expressions in **tx** and **ts** tiers). Code-switching into Tundra Nenets is mentioned in **nt** tier. Non-transcribed code-switching into Russian is also usually mentioned in **nt** tier.

### 3.3.12. Geographical coordinates (geo)

For some placenames encountered in the corpus, geographical coordinates in “latitude, longitude” format are provided in the **geo** tier, e.g. “68.683965, 86.276724” (68°41'02.3"N 86°16'36.2"E) for Potapovo.

### 3.3.13. Free translation (fe, fr)

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 do NOT reflect morphological and syntactical properties of the Enets original, but aim to express as clearly as possible the sense of the Enets original in a more or less grammatical and standard form. The translations loosely follow the common guidelines presented in Arkhipov (2020: Ch. 3).

When a part of the sentence is not transcribed or cannot be translated, the ellipsis character ... is typically used:

Figure 2. Incomplete translations

|            |                                       |
|------------|---------------------------------------|
| <b>ref</b> | LND_SNI_2005_MyWife_nar.LND.054 (071) |
| <b>ts</b>  | Etot, d'čriš ((...)).                 |
| <b>fe</b>  | Well, it was deep ..."                |
| <b>fr</b>  | Это, оно глубокое было ..."           |

Entirely unintelligible or otherwise untranscribed utterances have “(…).” in the free translation tiers.

### 3.3.14. Original Russian, German or Hungarian translation (ltr, ltg, lth respectively)

The tiers **ltr**, **ltg** and **lth** respectively contain the Russian, German or Hungarian respectively original translation of texts included into the corpus as provided in the source (a including draft translation written down during a session of oral transcription). German and Hungarian translations are available for a minority of texts coming from published sources and are more or less literary translations. Russian translations come from different sources and vary widely from literary translations (more or less precise or rather different from the original version) to word-by-word translation.

### 3.3.15. Notes (nt, nto)

The Notes tier (**nt**) contains corpus compilers' notes related to the sentence. The notes begin with the indication of who made the note (abbreviation as listed in section 2.5.5) in square brackets, followed by a colon.

The original notes tier (**nto**) contains notes that appeared on previous stages of processing the texts, including some of those available in original publications.

## 3.4. Searching the corpus

### 3.4.1. Search with EXMARaLDA EXAKT

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

EXAKT offers the following possibilities:

- Search in any one of transcription and annotation tiers using regular expressions
- Display search results in a concordance view including left and right context (**tx** tier)
- Display speaker and communication metadata and filter search results according to the metadata
- Display annotations overlapping with the annotations on the primary search tier and filter search results according to their content
- Add columns for custom user analyses
- View the fragment of the transcript containing the selected search result and listen to the corresponding sound
- Open the complete transcript containing the selected search result in EXMARaLDA Partitur Editor
- Copy search results as CSV

### Local and remote search

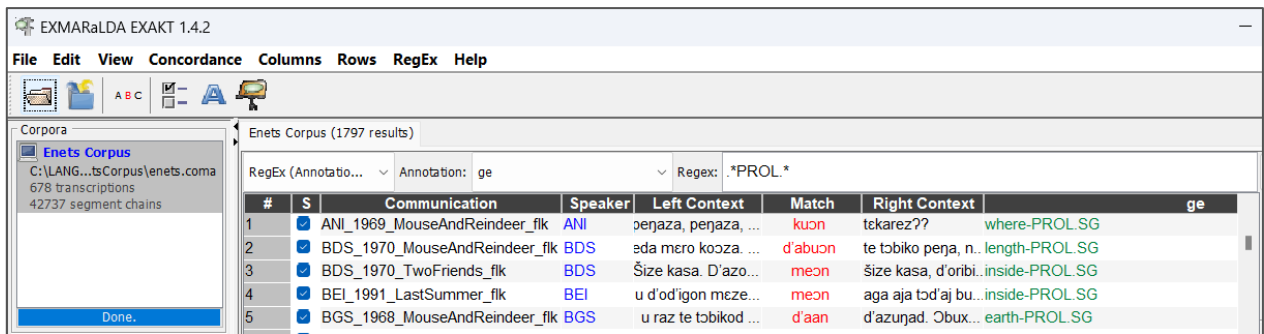
In order to perform a search on the downloaded corpus files locally, the main metadata file (**enets.coma**) should be opened with “File > Open corpus” command. (Creating a word list is optional.)

It is also possible to perform a remote search on the corpus without downloading it to the local computer. Use the command “File > Open remote corpus > Catalog lookup”, specify the following Catalog URL: <https://inel.corpora.uni-hamburg.de/remote/corpora.xml> and click “Open”. Choose the desired corpus and confirm. <https://s3-uhh.lzs.uni-hamburg.de/gwiss-inel-corpora/remote/enets-1.0/enets.coma> will appear as the Corpus URL. Check the “Anonymous login” option and press “OK”. (Creating a word list is optional.)

### Performing a search in EXAKT

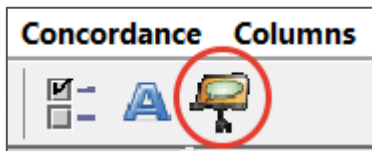
- One of the tiers should be selected in the main concordance window: either one of the annotation tiers (recommended; use “RegEx (Annotations)”); select any of tiers except **tx** under “Annotation”) or the transcription tier (**tx**; use “RegEx (Transcription)”).
- A search expression (interpreted as a regular expression) 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, Appendix A5 and Appendix A6 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 3. 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 4. EXAKT: “Open Partitur” button



Please refer to EXMARaLDA manuals<sup>28</sup> for further details on using EXAKT and Partitur Editor.

### 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/EnetsCorpus/search>. The interface of the 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
- Export 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 “Enets” and the search expression(s) entered in one or more corresponding fields.

Table 12. Tsakorpus search fields and EXMARaLDA tiers: main transcription and word-/morph-level annotation

| Tsakorpus search field | Corresponding tier in EXMARaLDA |
|------------------------|---------------------------------|
| Word                   | <b>tx</b>                       |

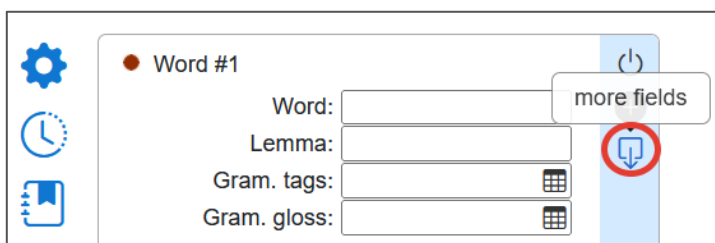
<sup>28</sup> <https://exmaralda.org/en/quickstart-documents/>, last accessed: 15.11.2024.



|                     |   |
|---------------------|---|
| Lemma               | <b>mp</b> (stem)  |
| Gram. tags          | <b>ps</b> ; grammar tags generated from grammatical glosses ( <b>ge</b> , <b>gr</b> ) |
| Gram. gloss         | grammatical (i.e. affix) glosses ( <b>ge</b> , <b>gr</b> )                            |
| Lexical gloss (en)* | lexical (i.e. stem) glosses ( <b>ge</b> )   |
| Lexical gloss (ru)* | lexical (i.e. stem) glosses ( <b>gr</b> )   |
| Morph. slot*        | <b>mc</b>   |
| Semantic role*      | <b>SeR</b>  |
| Syntactic function* | <b>SyF</b>  |
| Inform. status*     | <b>IST</b>  |
| Borrowing*          | <b>BOR</b>  |
| Bor. phonetics*     | <b>BOR-Phon</b>   |
| Bor. morphology*    | <b>BOR-Morph</b>  |
| Code-switching*     | <b>CS</b>   |
| Geogr. coordinates* | <b>geo</b>  |

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

Figure 5. Tsakorpus interface: Show more fields



Please refer to sections 3.2–3.3, Appendix A5 and Appendix A6 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).

Table 13. Stems and affixes in Tsakorpus

| EXMARaLDA tier | Word            | Stem     | Search field              | Affix  | Search field                              |
|----------------|-----------------|----------|---------------------------|--------|---|
| tx             | texiz           |          |                           |        |   |
| mb             | te-xiz          | te       |                           | xiz    | Gram. gloss: ABL.PL{xiz}                  |
| mp             | te-xiti         | te       | Lemma: te                 | xiti   | Gram. gloss: ABL.PL{_xiti}                |
| ge             | reindeer-ABL.PL | reindeer | Lex. gloss (en): reindeer | ABL.PL | Gram. gloss: ABL.PL<br>Gram. tags: abl,pl |
| gr             | олень-ABL.PL    | олень    | Lex. gloss (ru): олень    | ABL.PL | Gram. gloss: ABL.PL<br>Gram. tags: abl,pl |
| ps             | n               |          | Gram. tags: n             |        |   |

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

The affixes can be found by searching for the complete gloss (e.g. “ABL.PL”) in the **Gram. gloss** field, or with corresponding grammar tags (e.g. “abl,pl”) 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: “ABL.PL{xiz}”.

In case there exist more than one underlying form of an affix in **mp** tier (e.g. two dialectal variants), a particular underlying form can be specified in curly braces with underscore in the **Gram. gloss** field: “ABL.PL{\_xiti}”.

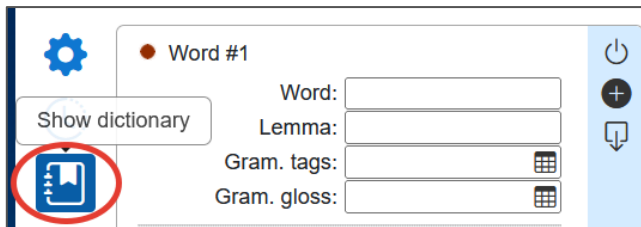
Note that some stems may have a glossing label similar to a grammatical gloss, e.g. NEG for “negative verb”. In such cases, this gloss should be entered into **Lex. gloss (en)** or **Lex. gloss (ru)** field, since it will not be treated as a grammatical gloss by Tsakorpus. It will however get a grammar tag (see next section) and can be found by searching for this tag. Such glosses are marked as “lexical” in Comment columns in Appendix A5.

Table 14. Stems and affixes in Tsakorpus: stems with grammatical glosses

| EXMARaLDA tier | Word      | Stem | Search field                                 | Affix | Search field                                     |
|----------------|-----------|------|--|-------|--|
| tx             | n'iezo?   |      |  |       |  |
| mb             | n'ie-zo?  | n'ie |  | zo?   | Gram. gloss: 1SG.S{zo?}                          |
| mp             | i-zo?     | i    | Lemma: i                                     | zo?   | Gram. gloss: 1SG.S {_zo?}                        |
| ge             | NEG-1SG.S | NEG  | Lex. gloss (en): NEG<br>Gram. tags: neg,negv | 1SG.S | Gram. gloss: 1SG.S<br>Gram. tags: pn1,pnsg,subjc |
| gr             | NEG-1SG.S | NEG  | Lex. gloss (ru): NEG<br>Gram. tags: neg,negv | 1SG.S | Gram. gloss: 1SG.S<br>Gram. tags: pn1,pnsg,subjc |
| ps             | aux       |      | Gram. tags: aux                              |       |  |

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 6. 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 7. Tsakorpus interface: Show list of values

The screenshot shows a search interface with the following elements:

- On the left, three icons: a gear (settings), a clock (history), and a book (dictionary).
- At the top, a red dot followed by "Word #1".
- Four search fields: "Word:", "Lemma:", "Gram. tags:", and "Gram. gloss:". The "Gram. tags:" and "Gram. gloss:" fields have a red circle with a grid icon next to them.
- On the right, a vertical toolbar with icons for power, add, and download.

### 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 Appendix A5.

- 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”, “ADJZ” => “adjz”. 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 Appendix A5. 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, “HORT1” will get tags “hort, hort1” and “HORT2” will get tags “hort, hort2”. Therefore each of them separately can be found with their distinctive tags (resp. “hort1” and “hort2”), while searching for “hort” will find both of them. Please refer to Appendix A5 for complete lists of tags.
- When a gloss consists of multiple components, such as “ABL.SG” or “3DU.MD”, each of them is usually translated into a tag, e.g. “ABL.SG” => “abl” (ablative), “sg” (singular); “PROL.PL” => “prol” (prolative), “pl” (plural); “3DU.MD” => “pn3” (3 person), “pndu” (dual number), “md” (medial conjugation). A search for tag “pn3” will return all words with any of glosses “3SG”, “3DU”, “3PL”, “3DU.MD”, etc.; a search for tag “abl” will return all words with any of glosses “ABL.SG”, “ABL.PL”.
- 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 “OBL.SG” will not find a zero morph in “house.[OBL.SG]”. Such wordforms can only be found with corresponding grammar tags, e.g. “obl,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 Appendix A5) 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”.

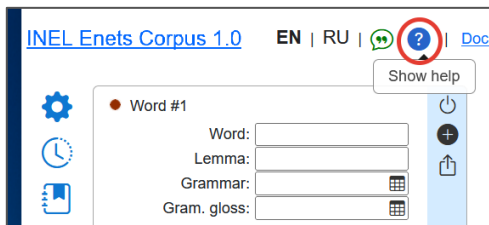
To search in one of the sentence-level annotation tiers, the search expression should be entered into “Word” field, and “Language/tier” field set to one of the following:

Table 15. Tsakorpus search fields and EXMARaLDA tiers: sentence-level annotation

| Language/tier label         | Corresponding tier in EXMARaLDA |
|-----------------------------|---------------------------------|
| Enets orthography (Cyr.)    | <b>tsc</b>                      |
| Enets source (Cyr.)         | <b>st</b>                       |
| Enets source (Lat.)         | <b>stl</b>                      |
| Russian translation         | <b>fr</b>                       |
| English translation         | <b>fe</b>                       |
| Orig. Russian translation   | <b>ltr</b>                      |
| Orig. German translation    | <b>ltg</b>                      |
| Orig. Hungarian translation | <b>lth</b>                      |
| Notes (Eng.)                | <b>nt</b>                       |
| Orig. notes                 | <b>nto</b>                      |
| Sentence ID                 | <b>ref</b>                      |

For further details please refer to Tsakorpus online help.

Figure 8. Tsakorpus interface: Show help



## References

- Amelina, Maria K. 2020. From small-scale multilingualism to “the big shift” to Tundra Nenets: Linguistic ideologies and language shift dynamics in Tukhard tundra and the Lower Yenisei area (20th — the beginning of the 21st century). *Ural-Altaic studies* 1(36). 7–48.  
[https://ural-altai.ru/userfiles/files/publications/UAI-36\\_Amelina%20\(2\).pdf](https://ural-altai.ru/userfiles/files/publications/UAI-36_Amelina%20(2).pdf)
- 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>
- 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>
- Barbolina, Anna A. & Kurdryakova, Nina S. (eds.). 2013. *Skaziteli Tajmyra* [Narrators of Taymyr]. Dudinka: Tajmyrskij Dom Narodnogo Tvorčestva. [https://www.tdnt.org/pdf\\_poli.php?id=37&t=publ](https://www.tdnt.org/pdf_poli.php?id=37&t=publ)
- Bettu, Larisa D. (ed.), 2006. *Fol'klornoe nasledie narodov Tajmyra* [Folklore legacy of Taymyr peoples]. 2 CDs. Dudinka. <https://www.tdnt.org/cd.php?t=cd>
- Bolina, Dar'ja S. (translator). 1995. *Luka pazduj ede baza (Evangelie ot Luki – fragmenty)* [Luka pazduj ede baza (Gospel of Luke – fragments)]. Stockholm: Institute for Bible Translation.
- Bolina, Dar'ja S. 2015. *Direda boduj' dja: Živaja tundra* [Direda boduj d'a: Live tundra]. St. Petesburg: Almaz-Graf. (A textbook for knidegarden.)
- Bolina, Dar'ja S. 2019. *Ènčuu' bukvar' (Èneckij bukvar')* [Enets primer]. St. Petersburg: Almaz-Graf.
- Bolina, Dar'ja S. & Rosljakova, Svetlana A. 2015. *Kajaku' bine: Lučik solnca* [Kajaku bine: A solar light]. St. Petesburg: Almaz-Graf. (A textbook for kindegarden.)
- Bolina, Zoja N. 2012. *Èneckij kartinnyj slovar'* [Enets picture dictionary]. Dudinka: Tajmyrskij dom narodnogo tvorčestva. [https://www.tdnt.org/pdf\\_poli.php?id=29&t=publ](https://www.tdnt.org/pdf_poli.php?id=29&t=publ)
- Bolina, Zoja N. 2014a. *Èzzuuj: Sled narty* [Èzzuuj: A sledge's trace]. Dudinka: Tajmyrskij dom narodnogo tvorčestva. [https://www.tdnt.org/pdf\\_poli.php?id=25&t=publ](https://www.tdnt.org/pdf_poli.php?id=25&t=publ)
- Bolina, Zoja N. 2014b. *Pesni rodnoj zemli: Samodejatel'nyj èneckij xudožnik Ivan Silkin* [Songs of home place: Amateur Enets painter Ivan Silkin]. Dudinka: Tajmyrskij dom narodnogo tvorčestva. [https://www.tdnt.org/pdf\\_poli.php?id=23&t=publ](https://www.tdnt.org/pdf_poli.php?id=23&t=publ)
- Brykina, Maria & Gusev, Valentin & Szeverényi, Sandor & 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>
- Castrén, M. Alexander. 1854. *Grammatik der samojedischen Sprachen*. St. Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.
- Čarušin, Evgenij. 2015. *Sama" nèon eza d'jori" (Razzkazy o životnyx)* [Sama? neon eza d'ori? (Stories about animals)]. Translated into Enets by Galina S. Bolina. St. Petesburg: Almaz-Graf.

- Dolgix, Boris O. 1961. *Mifologičeskie skazki i istoričeskie predanija èncev* [Mythological tales and histories of Enets people]. Moscow: Izdatel'stvo AN SSSR.
- 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>
- Gluxij, Jaroslav A. & Susekov, Vasilij A. 1982. Èneckie teksty [Enets texts]. In: Porotova, Tel'mina I. (ed.), *Skazki narodov Sibirskogo Severa* [Tales of peoples of Siberian North], 144–155. Tomsk: TGPU.
- 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](https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/2036/file/Kapitel6_07.pdf)
- 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](https://multicast.aspra.uni-bamberg.de/data/pubs/graid/Haig+Schnell2014_GRAID-manual_v7.0.pdf)
- Helimski, Eugene. 1981/2000. Ètimologičeskie zametki po èneckoj onomastike [Etymological notes on Enets onomastics]. *Sovetskoe Finno-Ugrovedenie* 17(2). 119–130. <https://doi.org/10.3176/lu.1981.2.07>. Reprinted in: Helimski 2000: 82–93.
- Helimski, Eugene. 1984/2000. Ustanovlenie dialektnoj prinadležnosti èneckix materialov [Establishing the dialectal affiliation of Enets data]. In: Veenker, Wolfgang (ed.), *Dialectologia Uralica*. Materialien des erstes Internationales Symposions zur Dialektologie der uralischen Sprachen in Hamburg, 38–43. Wiesbaden: Otto Harrassowitz. Reprinted in: Helimski 2000: 56–59.
- Helimski, Eugene. 2000. *Komparativistika, uralistika* [Comparative studies, Uralic studies]. Moscow: Jazyki slavjanskix kul'tur.
- Helimski, Eugene. Ms. Èneckij slovar' [Enets dictionary]. <https://www.slm.uni-hamburg.de/ifuu/download/helimski-enzisch.zip>.
- Khanina, Olesya. 2008. Odin tekst na tundrovom dialekte èneckogo jazyka I ego lingvističeskij razbor [A text in Tundra Enets language and its linguistic analysis]. *Naučnyj vestnik Jamalo-Neneckogo avtonomnogo okruga* 6(58). 139–147.
- Khanina, Olesya & Koryakov, Yuri & Shluinsky, Andrey. 2018. Enets in space and time: A case-study in linguistic geography. *Finnisch-Ugrische Mitteilungen* 42. 109–135.

- Khanina, Olesya & Shluinsky, Andrey. 2016. Èneckij jazyk [Enets]. In: Mixal'čenko, Vida Ju. (ed.), *Jazyk i obščestvo: Ènciklopedija* [Language and the society: An encyclopaedia], 588–593. Moscow: Azbukovnik. [https://iling-ran.ru/library/sociolingva/2016\\_jazyk\\_i\\_obschestvo.pdf](https://iling-ran.ru/library/sociolingva/2016_jazyk_i_obschestvo.pdf)
- Khanina, Olesya & Shluinsky, Andrey. 2022. Forest Enets and Tundra Enets: How different are they and why? In: Kaheinen, Kaisla & Leisiö, Larisa & Erkkilä, Riku & Qiu, Toivo E. H. (eds.), *Hämeenmaalta Jamalille. Kirja Tapani Salmiselle* [From Hämeenmaa to Jamal. The Festschrift for Tapani Salminen], 181–200. Helsinki: University of Helsinki Helda Open Books. <https://doi.org/10.31885/9789515180858>
- Khanina, Olesya & Shluinsky, Andrey. 2023. Forest and Tundra Enets. In Abondolo, Daniel & Valijärvi, Riita-Liisa (eds.), *The Uralic Languages*, 793–852. London – New York: Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781315625096-18>
- [Labanauskas, Kazimir I.] 1992. *Fol'klor narodov Tajmyra* [Folklore of Taymyr's peoples.] Vol. 1. Èneckij fol'klor [Enets folklore]. Dudinka: Tajmyrskij okružnoj centr narodnogo tvorčestva.
- Labanauskas, Kazimir I. 2002. *Rodnoe slovo: Èneckie pesni, skazki, istoričeskie predanija, tradicionnye rasskazy, mify* [Mother tongue: Enets songs, tales, legends, traditional stories, myths]. St. Petersburg: Prosveščenie.<sup>29</sup>
- Mikola, Tibor. 1967. Enzische Sprachmaterialien. *Acta Linguistica Academiae Scientiarum Hungaricae* 17. 59–74.
- Mikola, Tibor. 1980. Enyec és nganaszan nyelvi adalékok [Enets and Nganasan linguistic data]. *Nyelvtudományi Közlemények* 82. 223–236.
- Prokof'ev, Georgij N. Èneckij (enisejsko-samoediskij dialect) [Enets (Yenisey-Samoyedic dialect)]. In Prokof'ev, Georgij N. (ed.). *Jazyki i pis'mennost' samojedskix i finno-ugorskix narodov* [Languages and written systems of Samoyedic and Finno-Ugric peoples], 75–90. Moscow: Učpedgiz.
- Pusztay, János. 1978. Kleines Jenissej-Samojedisches Material. *Finnisch-Ugrische Mitteilungen* 2. 3–33.
- Reiman, D. Will. 2010. Basic oral language documentation. *Language Documentation & Conservation* 4. 254–268.
- Riaposov, Aleksandr & Arkhipov, Alexandre & Lazarenko, Elena. In preparation. Importing complete interlinear glossed texts into SIL FLEx.
- Shluinsky, Andrey. In press. *Forest Enets dictionary*. Szeged: University of Szeged – Department of Finno-Ugrian Philology, Department of Altaic Studies. (Studia Uralo-Altaica 58.)
- Shluinsky, Andrey; Khanina, Olesya; Ovsjannikova, Maria. 2021. Lesnoj èneckij jazyk: Kto na prostorax arktičeskoj tundry govorit na nastojaščem jazyke [Forest Enets: Who speaks the “true language” in the Arctic tundra]. <https://postnauka.org/longreads/156201>
- Siegl, Florian. 2006–2008. Item "Stories in FE" in collection "Enets and Forest Nenets". The Language Archive. <https://hdl.handle.net/1839/00-0000-0000-0008-6E7A-B>

<sup>29</sup> This publication includes musical notation for published songs. It is not mentioned there, but according to drafts available in TDNT archive, it was done by A.V. Koryukov.

- Siegl, Florian. 2013a. *Materials on Forest Enets, an indigenous language of Northern Siberia*. Helsinki: Société Finno-Ougrienne. <https://www.sgr.fi/sust/sust267/sust267.pdf>.
- Sorokina, Irina P. 1982. Èneckie teksty [Enets texts]. In: Porotova, Tel'mina I. (ed.), *Skazki narodov Sibirskogo Severa* [Tales of peoples of Siberian North] 4, 155–159. Tomsk: TGPU.
- Sorokina, Irina P. 2010. *Èneckij jazyk* [Enets]. St. Petersburg: Nauka.
- Sorokina, Irina P. & Bolina, Dar'ja S. 2005. *Èneckie teksty* [Enets texts]. St. Petersburg: Nauka.
- Sorokina, Irina P. & Bolina, Dar'ja S. 2009. *Èneckij slovar'* [Enets dictionary]. St. Petersburg: Nauka.
- Tereščenko, Natal'ja M. 1966. Èneckij jazyk. [Enets.] In Lytkin, Vasilij E. & Majtinskaja, Klara E. (eds.), *Jazyki narodov SSSR: finno-ugorskie i samodijskie jazyki* [Languages of the USSR: Finno-Ugric and Samoyedic languages], 438–457. Moscow: Nauka.
- Tereščenko, Natal'ja M. 1973. *Sintaksis samodijskix jazykov* [Syntax of Samoyedic languages]. Leningrad: Nauka.
- Tereščenko, Natal'ja M. 1986. Alfavit èneckogo jazyka [Enets alphabet]. In Skorik, Petr Ja. (ed.), *Paleoaziatskie jazyki* [Paleoasiatic languages], 50–52. Leningrad: Nauka.
- Urmanchieva, Anna Ju. 2008. Èneckij jazyk [Enets]. In Kibrik, Aleksandr E., et al. (eds.), *Malye jazyki: Suščestvovanie na grani* [Minor languages: Borderline existence] 2, 162–191. Moscow: Moscow State University.
- Urmanchieva, Anna. Ms a. Solda. A glossed text in Tundra Enets.
- Urmanchieva, Anna. Ms b. Otpusk [Vacation]. A glossed text in Forest Enets.  
<https://www.philol.msu.ru/~languedoc/corpus/enets/enets-01-otpusk.xhtml>.
- 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>.



## Appendix A1. Correspondence between Enets texts; Enets texts not included into the corpus

### A1.1. Correspondence between published sources and texts in the corpus

#### Forest Enets

Table 16. Sorokina & Bolina 2005

| Text number | Text code in corpus <sup>30</sup>  | Other versions  | Audio recording or other legacy sources <sup>31</sup>   | Comment  |
|-------------|------------------------------------|---|---|--|
| 1           | PNS_1969_DyoyManyStories_flk       |   | SIP's legacy manuscript (notebook 9, pp. 11–18)   | SB2005 version was mainly used   |
| 2           | PNS_1969_DyoyCheatingGuests_flk    | Reprints Sorokina 1982: Дѐа, 155–157  |   | SB2005 version was mainly used   |
| 3           | PNS_1980_DyoyAndStump_flk          | Reprints in a simplified orthography GS1982 Д'о?а (5-я часть), 150–151          |   | SB2005 version was mainly used   |
| 4           | PNS_1980_DyoyCheatingGuests_flk    | Reprints in a simplified orthography GS1982 Д'о?а (часть 6-я), 151–152          |   | SB2005 version was mainly used   |
| 5           | PNS_1969_DyoyPretendedToBeDead_flk |   |   |  |
| 6           | PNS_1980_DyoyCheatingShaman_flk    | Reprints in a simplified orthography GS1982 Д'о?а, 148–150                      |   | SB2005 version was mainly used   |
| 7           | PNS_1980_DyoyCheatingOldCouple_flk | Reprints in a simplified orthography GS1982 Д'о?а (7-я часть), 153–154          |   | SB2005 version was mainly used   |
| 8           | PNS_1990_DyoyAndInak_flk           | Independently transcribed and published in L2002: Дѐа, 78–92 (a longer version) | GTRK archive audio recording (KhO&ShA's working file nsp900419_Djoa_rad.wav)  | For the first part SB2005 version was used, but retranscribed according to the audio recording. The second part was transcribed by ShA and SIIP (KhO&ShA's working file nsp900419_Djoa_rad.eaf) In L2002, the text continues beyond the recording, the continuation is included into the corpus as PNS_199X_DyoyCheatingPeople_flk |
| 9           | PNS_1969_DyoyAndPike_flk           |   |   |  |
| 10          | PNS_1969_Moreo_flk                 |   | SIP's legacy manuscript (notebook 9, pp. 3–6)<br>SIP's archive audio recording (KhO&ShA's working file Sorokina_6_B_Moreo_Olasne.wav) | Only the first part of the text is published in SB2005; it is retranscribed according to the audio recording. For the second part, SIP's legacy manuscript transcription (notebook 9, pp. 39ff) was used. The remainder of the source audio recording remains untranscribed  |

<sup>30</sup> See 2.5.3 on its structure. Note that the first part of a code refers to the speaker, also see Appendix A4.

<sup>31</sup> If no source of an audio recording is mentioned, no audio recording corresponding to the text was accessible for us and the text is included into the corpus only in its written form. This situation means either that such a recording existed, but we could not find it, or that the text originally was written by a speaker or dictated to a linguist with no audio recording in any form.

| Text number | Text code in corpus <sup>30</sup>                           | Other versions  | Audio recording or other legacy sources <sup>31</sup>  | Comment  |
|-------------|---|---|--|--|
| 11          | ChNP_1974_Mosquito_flk                                      |   | SIP's archive audio recording (KhO&ShA's working file Sorokina_5_B_PojavilisjKomary1.wav)                        | Retranscribed according to the audio recording   |
| 12          | PNS_1990_MoreoAndWitch1_flk and PNS_1990_MoreoAndWitch2_flk | Independently transcribed by ShA and SIIP (nsp900817_Olasne_1_rad.eaf and nsp900817_Olasne_2_rad.eaf) and then glossed by ShA in Toolbox (NSP900817_OLA1, NSP900817_OLA2) | GTRK archive audio recording (KhO&ShA's working files nsp900817_Olasne_1_rad.wav and nsp900817_Olasne_2_rad.wav) | Toolbox version was mainly used  |
| 13          | BSP_1991_WitchAndEnetsWoman_flk                             |   | TDNT archive audio recording (KhO&ShA's working file spb91_Olasne_tdnt.wav)                                      | Retranscribed according to available audio recording; the published text differs significantly from it |
| 14          | YKD_1969_Witch_flk  |   | SIP's legacy manuscript (notebook 9, pp. 6–10) with a longer version   | SB2005 version was mainly used for the published part, but the continuation follows the manuscript     |
| 15          | PNS_1991_Witch_flk  |   |  |  |
| 16          | BSP_1991_MakingVeins_flk                                    |   |  |  |
| 17          | YKD_1969_War_flk  |   |  |  |
| 18          | BSP_1991_ColorfulFace_flk                                   |   |  |  |
| 19          | BSP_1991_Tungus_flk   |   |  |  |
| 20          | BEI_1991_LastSummer_flk                                     |   |  |  |
| 21          | PNS_1991_TundraPeople_flk                                   |   |  |  |
| 22          | BVN_1991_FreaksCape1_flk                                    | Independently transcribed by ShA and BZN (vnb9111_UrodlivjMys_rad.eaf) and then glossed by ShA in Toolbox (VNB9111_UROD)  | GTRK archive audio recording (KhO&ShA's working file vnb9111_UrodlivjMys_rad.wav)                                | Toolbox version was mainly used  |
| 23          | PNS_1991_TwoShamans_flk                                     |   |  |  |
| 24          | SDA_1969_TwoBrothers_flk                                    |   |  |  |
| 25          | BSP_1991_Loon_flk   |   |  |  |
| 26          | SDA_1969_TwoReindeer_flk                                    |   |  |  |
| 27          | SDA_1969_BirdAndMouse_flk                                   |   |  |  |
| 28          | PLN_1985_Partridge_flk                                      |   |  |  |
| 29          | PLN_1985_PartridgeAndPike_flk                               |   |  |  |
| 30          | PLN_1985_SturgeonAndPike_flk                                |   |  |  |
| 31          | PLN_1985_TwoBears_flk                                       |   |  |  |
| 32          | BTN_1969_Guzzler_flk  |   | SIP's archive audio recording (KhO&ShA's working file Sorokina_5_A_Obzhora.wav)                                  | Retranscribed according to the audio recording   |

| Text number | Text code in corpus <sup>30</sup> | Other versions  | Audio recording or other legacy sources <sup>31</sup>  | Comment   |
|-------------|-----------------------------------|---|--|---|
| 33          | BNN_1969_Cuckoo_flk               |   |  |   |
| 34          | BViN_1969_Cuckoo_flk              |   |  |   |
| 35          | ANI_1969_Cuckoo_flk               |   | SIP's archive audio recording (KhO&ShA's working file Sorokina_5_B_Kukushka3.wav)                      | Retranscribed according to the audio recording  |
| 36          | BYuS_1969_Cuckoo_flk              |   |  |   |
| 37          | SDA_1969_RedMoss_flk              |   |  |   |
| 38          | SMN_1990_Cuckoo_flk               | Independently transcribed by ShA and BZN (mns900522_MyshkaKukushka_rad.eaf) and then glossed by ShA in Toolbox (MNS900522_MYSHKU) | GTRK (KhO&ShA's working file mns900522_MyshkaKukushka_rad.wav)   | Toolbox version was mainly used.<br>The recording from GTRK had one more text that is included into the corpus separately as SMN_1990_Mouse_flk |
| 39          | SDA_1969_MouseAndReindeer_flk     |   |  |   |
| 40          | BViN_1969_MouseAndReindeer_flk    |   |  |   |
| 41          | ANI_1969_MouseAndReindeer_flk     |   | SIP's archive audio recording (KhO&ShA's working file Sorokina_5_A_OlenjImyshka.wav)                   | Retranscribed according to the audio recording  |
| 42          | SDA_1969_ZakharAndReindeer_flk    |   |  |   |
| 43          | SDA_1974_Earlier_flk              |   |  |   |
| 44          | SDA_1969_OldMan_flk               |   |  |   |
| 45          | PNS_1990_TwoBrothers_flk          |   | GTRK archive audio recording (KhO&ShA's working file nsp900810_MD60-1_2_MD1-8_9_NeskoljkoSkazok_2.wav) | Retranscribed according to the audio recording  |
| 46          | PLN_1985_Stoat_flk                |   |  |   |
| 47          | PLN_1985_HareAndWolverine_flk     |   |  |   |
| 48          | BViN_1969_HareAndWolverine_flk    |   |  |   |
| 49          | PLN_1985_FishAndCrow_flk          |   |  |   |
| 50          | SDA_1974_SturgeonSkin_flk         |   |  |   |
| 51          | BSP_1992_Dog_flk                  |   |  |   |
| 52          | PNS_1990_Reindeer_flk             |   | GTRK archive audio recording (KhO&ShA's working file nsp900810_MD60-1_2_MD1-8_9_NeskoljkoSkazok_3.wav) | Retranscribed according to the audio recording  |
| 53          | SDA_1969_PoorMan_flk              |   |  |   |
| 54          | NN2_1974_OldStory_nar             | Independently transcribed by ShA and BViN (xxx70s80s_DavnoEtoBylo_ips.eaf)  | SIP's archive audio recording (KhO&ShA's working file xxx70s80s_DavnoEtoBylo.wav)                      | Both versions were equally used.<br>In SB2005, the text is attributed to BVN, but by mistake.   |
| 55          | BND_1966_Turutin_nar              | Reprints in Cyrillics M1967: text 3   |  | M1967 version was mainly used   |

| Text number | Text code in corpus <sup>30</sup> | Other versions  | Audio recording or other legacy sources <sup>31</sup>   | Comment   |
|-------------|-----------------------------------|---|---|---|
| 56          | BND_1966_Ngela_flk                | Reprints in Cyrillics M1967: text 2   |   | M1967 version was mainly used   |
| 57          | PNS_1991_TripleJump_flk           |   |   |   |
| 58-1        | BYuS_1969_Bear_flk                |   |   |   |
| 58-2        | BYuS_1969_Hare_flk                |   |   |   |
| 58-3        | BYuS_1969_Hunter_flk              |   |   |   |
| 58-4        | BYuS_1969_OldWoman_flk            |   |   |   |
| 58-5        | BYuS_1969_Tent_flk                |   |   |   |
| 59          | SDA_1974_Ivan_flk                 |   |   |   |
| 60          | BSP_1991_EnetsAndCzar_flk         |   |   |   |
| 61          | PNS_1990_EnetsAndRussian1_flk     | Independently transcribed and published in L2002: Онай энче, роса энче, 92–95; the text was glossed by ShA in Toolbox following L2002 transcription (nsp900810_EneclRUsskij_rad.eaf, NSP900810_ENE) | GTRK archive audio recording (KhO&ShA's working file nsp900810_EneclRUsskij_rad.wav)  | Toolbox version based on L2002 version was mainly used, but SB2005 version was also consulted   |
| 62          | SDA_1969_LittleChildSong_song     |   | SIP's archive audio recording (KhO&ShA's working file Sorokina_5_B_PesnjaRebenka.wav)   | Retranscribed according to the audio recording, but the difference is rather minor<br>In L2002:63, a version of this song is published based on a different recording, which is a part of SMN_BNN_199X_Interview_conv text included into the corpus |
| 63          | PVN_1992_ReindeerDriverSong_song  | Also published in L2002 Тэа понида бари, 64   | TDNT written legacy data (KhO&ShA's working folder Chernaja_Papka_File11_Songs, p. 3–4,9)   | As the text is composed by the speaker, the versions are equivalent. A later recording done by KhO and ShA is synchronized in the corpus (cut from KhO&ShA's working file vn050922_Legenda.mp3)   |
| 64          | PNS_1991_OneWhoCameFromSky_song   | Independently transcribed and published in L2002: Ҳаҳаз той, 65–66  | TDNT archive audio recording (KhO&ShA's working file nsp70s90s_PrishedshijEN_tdnt.wav)<br>PVN's archive manuscript (KhO&ShA's working folder Tetradj_VNPalchina, from 35 to 44) | Retranscribed according to the audio recording using all available sources  |
| 65          | BMA_1969_HowIGotMarried_nar       |   |   |   |
| 66          | SDA_1974_HowIVisitedTundra_nar    |   |   |   |
| 67          | PNS_1969_Fire_nar                 |   |   |   |

| Text number | Text code in corpus <sup>30</sup> | Other versions   | Audio recording or other legacy sources <sup>31</sup>   | Comment   |
|-------------|-----------------------------------|--|---|---|
| 68          | NN1_1969_Potapovo_nar             |  | SIP's archive audio recording of the end of the text (KhO&ShA's working file Sorokina_5_B_Vpotapovo_Okonchanie.wav) | The end is retranscribed according to the audio recording.<br>In SB2005, the text is attributed to BYuS, but a woman's voice is heard in the recording (maybe SDA, but not surely).<br>Another short recording related to this text comes from SIP's archive audio recording (KhO&ShA's working file Sorokina_5_A_Vpotapovo_Popytka.wav); this is presumably indeed from BYuS, but this recording is not at all the whole text. |
| 69          | BND_1966_MyLife_nar               | Reprints in Cyrillics M1967: text 1  |   | M1967 version was mainly used   |
| 70          | BND_1966_PartridgeNest_nar        | Reprints in Cyrillics M1967: text 4  |   | M1967 version was mainly used   |
| 71          | BVN_1969_HowWeLived_nar           |  |   |   |
| 72          | ChNP_1980_Reindeer_nar            | Reprints in a simplified orthography GS1982 Кун'и те? пон'да? ..., 146–147         |   | SB2005 version was mainly used  |
| 73          | IKV_1980_ReindeerInTundra_nar     | Reprints in a simplified orthography GS1982 Кун'и те? д'ир'и бадуй д'аХан, 144–146 |   | SB2005 version was mainly used  |
| 74          | IKV_1980_ReindeerSkin_nar         | Reprints in a simplified orthography GS1982 Кун'и? онай эн'ч'о?... , 154–155       |   | SB2005 version was mainly used  |
| 75          | BVN_1969_ReindeersFormerly_nar    |  |   |   |
| 76          | PAS_1969_HowCatchFish_nar         |  |   |   |
| 77          | BNN_1974_MyLife_nar               |  |   |   |
| 78          | PNS_1991_HowlFished_nar           |  |   |   |
| 79          | YKD_1969_Hungry_nar               |  |   |   |
| 80          | ChNP_1969_Volgograd_transl        |  |   |   |

Table 17. Labanauskas 2002

| Text title and pages      | Text code in corpus                      | Other versions  | Audio recording or other legacy sources  | Comment  |
|---------------------------|--|---|--|--|
| Онай энче бари,<br>63, 75 | A part of<br>SMN_BNN_199X_Interview_conv | Independently transcribed as a small part of a longer recording by OM/ShA and GES (mns_nnb90s_Interview_tdnt.eaf) and then glossed by ShA in Toolbox (MNS_NNB90S_INT) | TDNT archive audio recording (KhO&ShA's working file mns_nnb90s_Interview_tdnt.wav)<br>Only for the song TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File11_Songs, p.1,7) | Toolbox version was mainly used.<br>In SB2005: text 62, a version of this song is published based on a different recording (included into the corpus as SDA_1969_LittleChildSong_song) |

| Text title and pages   | Text code in corpus              | Other versions   | Audio recording or other legacy sources  | Comment  |
|--|----------------------------------|--|--|--|
| Бака, 63–64, 75  | PVN_1992_Priluki_song            | Reprinted in a different orthography in (Shluinsky et al. 2021)  | Sound recording comes from (Bettu (ed.) 2006) CDs.<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File11_Songs, p.2,8)                                    | The song also exists in various other versions of audio recordings (both legacy and ELDP+). Only one of them is included into the corpus as BAS_2005_Priluki_song  |
| Тэа понида бари, 64, 76  | PVN_1992_ReindeerDriverSong_song | Also published in SB2005: text 63  | TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File11_Songs, p. 3–4,9)   | As the text is composed by the speaker, the versions are equivalent. A later recording done by KhO and ShA is synchronized in the corpus (cut from KhO&ShA's working file vn050922_Legenda.mp3)  |
| Каяку, 64–65, 76   | PVN_1992_Sun_song                |  | Sound recording comes from (Bettu (ed.) 2006) CDs.<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File11_Songs, p.5–6,10)                                 | The song also exists in some other versions of audio recordings  |
| Ѓахаз той, 65–66, 76   | PNS_1991_OneWhoCameFromSky_song  | Independently transcribed and published in SB2005: text 64   | TDNT archive audio recording (KhO&ShA's working file nsp70s90s_PrishedshijEN_tdnt.wav)<br>PVN's archive manuscript (KhO&ShA' working folder Tetradj_VNPalchina, from 35 to 44) | Retranscribed according to the audio recording using all available sources   |
| Оласнэ, 67–74, 77  | PNS_1992_WitchSongVersion_flk    | Independently preliminarily transcribed by ShA and BNK (nsp92_Pesnja_Olasne_od.eaf)  | DOE's archive audio recording (KhO&ShA's working file NSPalchin1992byOD.wav)   | Retranscribed according to the audio recording using both sources  |
| Дѐа, 78–92, Beginning (up to words Кабид недоць а", дирѐдоць, p. 81) | PNS_1990_DyoeAndInak_flk         | Independently transcribed and published in SB2005: text 8 (a shorter version).<br>The remainder of the available recording was independently transcribed in ELAN by ShA and SIIP (KhO&ShA's working file nsp900419_Djoa_rad.eaf) | GTRK archive audio recording (KhO&ShA's working file nsp900419_Djoa_rad.wav)   | SB2005 and ELAN versions were mainly used.<br>In L2002, the text is continued further, the continuation is included into the corpus separately as PNS_199X_DyoeCheatingPeople_flk (see next line)  |
| Дѐа, 78–92, End (starting from words Дѐор манѐ, p. 81)               | PNS_199X_DyoeCheatingPeople_flk  |  |  | The beginning of this text is included into the corpus as PNS_1990_DyoeAndInak_flk based on independent transcription of the available recording (see previous line).<br>In contrast, the source audio recording for this end of the published text is unknown |
| Онай энче, роса энче, 92–95  | PNS_1990_EnetsAndRussian1_flk    | Independently transcribed and published in SB2005: text 61   | GTRK archive audio recording (KhO&ShA's working file nsp900810_EnecIRUsskij_rad.wav)   | The text was glossed by ShA in Toolbox using L2002 transcription (nsp900810_EnecIRUsskij_rad.eaf, NSP900810_ENE).<br>Toolbox version based on L2002 version was mainly used, but SB2005 version was also consulted   |

| Text title and pages           | Text code in corpus                  | Other versions  | Audio recording or other legacy sources  | Comment  |
|--------------------------------|--------------------------------------|---|--|--|
| Онэй тазобэ, ося тазобэ, 95–99 | PNS_199X_EnetsShamanEvenkiShaman_flk |   | DOE's archive audio recording (KhO&ShA's working file nsp90s_2Shamana_od.wav)  | The text was glossed by ShA in Toolbox using L2002 transcription (nsp90s_2Shamana_od.eaf, NSP90S_DVA). Toolbox version was used  |
| Нэху каса, 100–103             | BVN_1992_ThreeBrothers_flk           |   |  | The text was glossed by ShA in Toolbox using L2002 transcription (with no recording available; RODNSL_TRBR). Toolbox version was used  |
| Баручи соз, 104–108            | BVN_1992_FreaksCape2_flk             | Independently transcribed by ShA and RSA (vnb92_BaruchiSoz_od.eaf) and then glossed by ShA in Toolbox (VNB92_BAR) | DOE's archive audio recording (KhO&ShA's working file vnb92_BaruchiSoz_od.wav)   | Toolbox version was used.  |
| Нэху нэ, 108–110               | BMN_1994_ThreeWomen_flk              | Reprinted in (Siegl 2013: 496–501)  | TDNT archive audio recording (KhO&ShA's working file TDNT_3_F_MNBolina_3Zhenschiny.wav). GTRK archive audio recording (KhO&ShA's working file mnb9404_MD59-6_MD1-12_Slepajaljubovj.wav).<br>Since the GTRK recording is of a better quality, but lacks first sentences, these sentences were cut from the TDNT recording and added to the GTRK recording (the result is KhO&ShA's working file mnb9404_TriZhenschiny_rad_tdnt.wav) | The text was glossed by ShA in Toolbox using L2002 transcription (mnb9404_TriZhenschiny_rad_tdnt.eaf, MNB9404_TRI). Toolbox version was used                                   |
| Нэла"а Мугади, 137–139         | TNM_1984_Ngela_flk                   |   |  | Presumably, the text is based on a TE recording, but transcribed by a FE speaker (maybe PVN) and thus changed mainly into FE, but there is no direct evidence for its history. |

Table 18. Mikola 1967

| Text number | Text code in corpus        | Other versions                            | Audio recording or other legacy sources | Comment                       |
|-------------|----------------------------|---|---|-------------------------------|
| 1           | BND_1966_MyLife_nar        | Reprinted in Cyrillics in SB2005: text 69 |   | M1967 version was mainly used |
| 2           | BND_1966_Ngela_flk         | Reprinted in Cyrillics in SB2005: text 56 |   | M1967 version was mainly used |
| 3           | BND_1966_Turutin_nar       | Reprinted in Cyrillics in SB2005: text 55 |   | M1967 version was mainly used |
| 4           | BND_1966_PartridgeNest_nar | Reprinted in Cyrillics in SB2005: text 70 |   | M1967 version was mainly used |

Table 19. Pusztay 1978

| Text number | Text code in corpus   | Other versions | Audio recording or other legacy sources | Comment |
|-------------|-----------------------|----------------|---|---------|
| 1           | BDS_1970_Reindeer_nar |                |   |         |

|   |                               |  |  |  |
|---|-------------------------------|--|--|--|
| 2 | BDS_1970_TwoPeople_flk        |  |  |  |
| 3 | BDS_1970_TwoFriends_flk       |  |  |  |
| 4 | BDS_1970_Lazy_flk             |  |  |  |
| 5 | BDS_1970_MouseAndReindeer_flk |  |  |  |
| 6 | BDS_1970_Conversation_conv    |  |  |  |
| 7 | BDS_1970_Seasons_nar          |  |  |  |

*Table 20. Mikola 1980*

| <b>Text number</b> | <b>Text code in corpus</b>    | <b>Other versions</b> | <b>Audio recording or other legacy sources</b> | <b>Comment</b> |
|--------------------|-------------------------------|-----------------------|--|----------------|
| 1                  | BGS_1968_TwoFriends_flk       |                       |  |                |
| 2                  | BGS_1968_MouseAndReindeer_flk |                       |  |                |

*Table 21. Gluxij & Susekov 1982*

| <b>Text title and pages</b>            | <b>Text code in corpus</b>         | <b>Other versions</b>                                    | <b>Audio recording or other legacy sources</b> | <b>Comment</b>                 |
|--|------------------------------------|--|--|--------------------------------|
| Кун'и те? д'ир'и бадуй д'аХан, 144–146 | IKV_1980_ReindeerInTundra_nar      | Reprinted in a simplified orthography in SB2005: text 73 |  | SB2005 version was mainly used |
| Кун'и те? пон'да? ..., 146–147         | ChNP_1980_Reindeer_nar             | Reprinted in a simplified orthography in SB2005: text 72 |  | SB2005 version was mainly used |
| Д'о?а, 148–150                         | PNS_1980_DyooCheatingShaman_flk    | Reprinted in a simplified orthography in SB2005: text 6  |  | SB2005 version was mainly used |
| Д'о?а (5-я часть), 150–151             | PNS_1980_DyooAndStump_flk          | Reprinted in a simplified orthography in SB2005: text 3  |  | SB2005 version was mainly used |
| Д'о?а (часть 6-я), 151–152             | PNS_1980_DyooCheatingGuests_flk    | Reprinted in a simplified orthography in SB2005: text 4  |  | SB2005 version was mainly used |
| Д'о?а (7-я часть), 153–154             | PNS_1980_DyooCheatingOldCouple_flk | Reprinted in a simplified orthography in SB2005: text 7  |  | SB2005 version was mainly used |
| Кун'и? онай эн'ч'о?..., 154–155        | IKV_1980_ReindeerSkin_nar          | Reprinted in a simplified orthography in SB2005: text 74 |  | SB2005 version was mainly used |

*Table 22. Sorokina 1982*

| <b>Text title and pages</b> | <b>Text code in corpus</b> | <b>Other versions</b>       | <b>Audio recording or other legacy sources</b> | <b>Comment</b>                 |
|-----------------------------|----------------------------|-----------------------------|--|--------------------------------|
| Дѐа, 155–157                | PNS_1969_Dyoo2_flk         | Reprinted in SB2005: text 2 |  | SB2005 version was mainly used |



Table 23. *Urmanchieva 2008*

| Text title and pages | Text code in corpus   | Other versions | Audio recording or other legacy sources   | Comment |
|----------------------|-----------------------|----------------|---|---------|
| Ведьмак, 164–191     | BVN_1993_Witcher1_flk |                | GTRK archive audio recording (KhO&ShA's working file vnb930317_MD59-3_MD2-4_Olasne.wav) |         |

## Tundra Enets

Table 24. *Sorokina & Bolina 2005*

| Text number | Text code in corpus       | Other versions   | Audio recording or other legacy sources   | Comment   |
|-------------|---------------------------|--|---|---|
| 81          | TDM_1977_Dia_flk          |  |   |   |
| 82          | SNS_1977_TwoBrothers_flk  |  |   |   |
| 83          | KTN_1977_Devil_flk        |  |   |   |
| 84          | KXN_1978_ThreeShamans_flk | Reprinted from L1992 Нэхү' тазобэ 5–10; an edited version was reprinted in L 2002 Цизе тазобэ, 116–121 | TDNT archive audio recording (KhO&ShA's working file TDNTKass11_Ne_klassif_St2.wav)<br>TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NexuqTazobe.wav)   | Retranscribed according to the audio recording  |
| 85          | PiNS_1977_Mouse_flk       | Reprinted from Sorokina 1982: 158–159  |   | SB2005 version was used   |
| 86          | PiNS_1977_Hare_nar        |  |   |   |
| 87          | KXN_1977_ThreeSons_flk    |  |   |   |
| 88          | SII_1973_SevenGirls_flk   | Reprinted from L1992, Сэ"о кати, 11–12; an edited version was reprinted in L2002 Сэ"о кати, 133–136    | TDNT archive audio recordings with oral transcription (KhO&ShA's working files TDNTKas5_Nauch_St1_1_LodosezaSeoKati_1.wav, TDNTKas5_Nauch_St1_2_LodosezaSeoKati_2.wav, TDNTKass11_Ne_klassif_St1_LodosezaSeoKatiza.wav).<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 47–52) | The source line follows L2002, but L1992 and LK's manuscript and oral transcription were also actively used |
| 89          | TTP_1973_OldWomansSon_nar | Reprinted from L1992, Меню"о" нию, 17–18   |   | SB2005 version was mainly used, but L1992 version was also consulted  |
| 90          | TTP_1973_OldMan_nar       | Reprinted from L1992, Бахыо, 15–16   |   | SB2005 version was mainly used, but L1992 version was also consulted  |
| 91          | KXN_1977_Flood_nar        |  |   |   |
| 92          | KXN_1977_IceFloe_nar      |  |   |   |

| Text number | Text code in corpus            | Other versions  | Audio recording or other legacy sources  | Comment  |
|-------------|--------------------------------|---|--|--|
| 93          | KXN_1977_HuntingPolarFoxes_nar |   |  |  |
| 94          | KXN_1977_MakingSkis_nar        |   |  |  |
| 95          | KTN_1977_Stove_nar             |   |  |  |
| 96          | KXN_1977_PolarFox_nar          |   |  |  |
| 97          | KTX_1977_Bear_nar              |   |  |  |
| 98          | TNM_1977_EvilReindeer_nar      |   |  |  |
| 99          | SOI_1977_MyMarriage_nar        |   |  |  |
| 100         | TNM_1977_OldDays_nar           |   |  |  |
| 101         | KDX_1977_Vorontsovo_nar        |   |  |  |
| 102         | SII_1973_Taimen_nar            | Reprinted from L1992, <i>Наро кәре</i> , 13–14; an edited version was reprinted in L2002 <i>Наро кәре</i> , 136–137 | TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NaroKare.wav)<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 31–34) | The source line follows L2002, but L1992, LK's manuscript and oral transcription were also actively used |

Table 25. *Labanauskas 1992*

| Text title and pages | Text code in corpus       | Other versions   | Audio recording or other legacy sources   | Comment   |
|----------------------|---------------------------|--|---|---|
| Нэхү' тазобэ 5–10    | KXN_1978_ThreeShamans_flk | Reprinted in SB2005: text 84; an edited version was reprinted in L 2002 <i>Ҷизэ тазобэ</i> , 116–121 | TDNT archive audio recording (KhO&ShA's working file TDNTKass11_Ne_klassif_St2.wav)<br>TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NexuqTazobe.wav)   | Retranscribed according to the audio recording  |
| Сэ"о кати, 11–12     | SII_1973_SevenGirls_flk   | Reprinted in SB2005: text 88; an edited version was reprinted in L2002 <i>Сэ"о кати</i> , 133–136    | TDNT archive audio recordings with oral transcription (KhO&ShA's working files TDNTKas5_Nauch_St1_1_LodosezaSeoKati_1.wav, TDNTKas5_Nauch_St1_2_LodosezaSeoKati_2.wav, TDNTKass11_Ne_klassif_St1_LodosezaSeoKatiza.wav).<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 47–52) | The source line follows L2002, but L1992 and LK's manuscript and oral transcription were also actively used |
| Наро кәре, 13–14     | SII_1973_Taimen_nar       | Reprinted in SB2005: text 102; an edited version was reprinted in L2002 <i>Наро кәре</i> , 136–137   | TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NaroKare.wav)<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 31–34)  | The source line follows L2002, but L1992 and LK's manuscript and oral transcription were also actively used |

| Text title and pages | Text code in corpus       | Other versions               | Audio recording or other legacy sources | Comment  |
|----------------------|---------------------------|------------------------------|---|--|
| Бахуо, 15–16         | TTP_1973_OldMan_nar       | Reprinted as SB2005: text 90 |   | SB2005 version was mainly used, but L1992 version was also consulted |
| Меню”о” нии, 17–18   | TTP_1973_OldWomansSon_nar | Reprinted as SB2005: text 89 |   | SB2005 version was mainly used, but L1992 version was also consulted |

Table 26. Labanauskas 2002

| Text title and pages    | Text code in corpus       | Other versions   | Audio recording or other legacy sources   | Comment   |
|-------------------------|---------------------------|--|---|---|
| Нэхү түзэго, 110–116    | SII_1973_ThreeTuzagos_flk |  | TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas5_Nauch_St2_NexuqTuzago.wav)<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 35–46; Zelenaja_Papka, p.13–20)  | The source line follows L2002, but LK's manuscript and oral transcription were also actively used           |
| Үизэ тазобэ, 116–121    | KXN_1978_ThreeShamans_flk | Reprints in edited version L1992 Нэхү' тазобэ 5–10, which is also reprinted in SB2005: text 84 | TDNT archive audio recording (KhO&ShA's working file TDNTKass11_Ne_klassif_St2.wav)<br>TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NexuqTazobe.wav)   | Retranscribed according to the audio recording  |
| Бунэля тиарега, 122–131 | SII_1973_Bunelja_flk      |  | TDNT archive audio recordings with oral transcription (KhO&ShA's working files TDNTKas1_Nauch_St1_Bunelja1.wav, TDNTKas1_Nauch_St2_Bunelja2.wav, TDNTKas24_Folk_St1_Bunelja.wav).<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 1–23)   | The source line follows L2002, but LK's manuscript and oral transcription were also actively used           |
| Аседа бахуо, 131–133    | SII_1973_Aseda_nar        |  | TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_Aseda.wav)<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 25–30)   | The source line follows L2002, but LK's manuscript and oral transcription were also actively used           |
| Сэ”о кати, 133–136      | SII_1973_SevenGirls_flk   | Reprints in edited version L1992 Сэ”о кати, 11–12, which is also reprinted in SB2005: text 88  | TDNT archive audio recordings with oral transcription (KhO&ShA's working files TDNTKas5_Nauch_St1_1_LodosezaSeoKati_1.wav, TDNTKas5_Nauch_St1_2_LodosezaSeoKati_2.wav, TDNTKass11_Ne_klassif_St1_LodosezaSeoKatiza.wav).<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 47–52) | The source line follows L2002, but L1992 and LK's manuscript and oral transcription were also actively used |

| Text title and pages | Text code in corpus | Other versions   | Audio recording or other legacy sources  | Comment   |
|----------------------|---------------------|--|--|---|
| Hapo kape, 136–137   | SII_1973_Taimen_nar | Reprints in edited version L1992 Hapo kape, 13–14, which is also reprinted in SB2005: text 102 | TDNT archive audio recording with oral transcription (KhO&ShA's working file TDNTKas1_Nauch_St2_NaroKare.wav)<br>TDNT written legacy data (KhO&ShA' working folder Chernaja_Papka_File13_Skazki, p. 31–34) | The source line follows L2002, but L1992 and LK's manuscript and oral transcription were also actively used |

### ***Urmanchieva Ms a***

Sentence splits and numeration in the version widely accessible to researchers did not coincide with those referred to in (Helimski Ms.). The source PDF version contained in the corpus has been edited to make the sentence numbers fully equivalent to those referenced in Helimski (Ms.), and the sentence numbering in the corpus also follows it. No other changes were made to the original source file.

## **A1.2. Correspondence of texts in Toolbox databases and texts in the corpus**

### **Forest Enets**

*Table 27. Correspondence of Forest Enets texts in Toolbox databases and texts in the corpus*

| Text code in Toolbox | Files name                   | Text code in corpus                               | Audio recording source | Comment |
|----------------------|------------------------------|---|------------------------|---------|
| AM050924_DYO         | am050924_Dyoo                | SAM_2005_CheatingOldMan_flk                       | ELDP+                  |         |
| AM080814_VER         | am080814_Frazy_NaVertolete   | SAM_2008_IHaveComeFromTundra_nar                  | ELDP+                  |         |
| AN090721_OTE         | an090721_Otec                | PAN_2009_MyFather_nar                             | ELDP+                  |         |
| AN090803_OXO         | an090803_OxotaVChislax       | PAN_2009_Hunting_nar                              | ELDP+                  |         |
| AP_NNB970724_INT1    | ap_nnb970724_Interview_1_rad | BAP_BNN_1997_Interview1_conv                      | GTRK                   |         |
| AP_NNB970724_INT2    | ap_nnb970724_Interview_2_rad | BAP_BNN_1997_Interview2_conv                      | GTRK                   |         |
| AP050924_SVA         | ap050924_Svadba              | BAP_2005_Marriage_nar                             | ELDP+                  |         |
| AS_AM_ASS050925_RAZ1 | as_am_ass050925_Razgovor1    | BAS_SAM_SAS_2005_ConversationWhoSpeaksEnets1_conv | ELDP+                  |         |
| AS_AM_ASS050925_RAZ2 | as_am_ass050925_Razgovor2    | BAS_SAM_SAS_2005_ConversationFamily_conv          | ELDP+                  |         |
| AS_AM_ASS050925_RAZ4 | as_am_ass050925_Razgovor4    | BAS_SAM_SAS_2005_ConversationWhoSpeaksEnets2_conv | ELDP+                  |         |
| AS_AM_ES050925_NGA   | as_am_es050925_Nnganasany    | BAS_SAM_GES_2005_ConversationNnganasans_conv      | ELDP+                  |         |
| AS_AM_ES050925_ROD   | as_am_es050925_Roda          | BAS_SAM_GES_2005_ConversationEnetsClans_conv      | ELDP+                  |         |
| AS_AM050925_OLE      | as_am050925_Olenj            | SAM_BAS_2005_ReindeerAndMouse_flk                 | ELDP+                  |         |

| Text code in Toolbox | Files name                              | Text code in corpus                   | Audio recording source | Comment   |
|----------------------|---|---------------------------------------|------------------------|---|
| AS_DA080402_SHKU     | as_da080402_ExaliNaShkure               | BAS_SDnA_2008_HowWeRodeOnASkin_nar    | ELDP+                  |   |
| AS_NI_GA100713_RYB5  | as_ni_ga100713_Razgovory_RazdelkaRyby_5 | BAS_SNI_IGA_2010_CuttingFish6_conv    | ELDP+                  |   |
| AS_NI100713_RAZ      | as_ni100713_RazdelajuRybu               | BAS_SNI_2010_CuttingFish1_conv        | ELDP+                  |   |
| AS_NI100713_RYB1     | as_ni100713_Razgovory_RazdelkaRyby_1    | BAS_SNI_2010_CuttingFish2_conv        | ELDP+                  |   |
| AS_NI100713_RYB2     | as_ni100713_Razgovory_RazdelkaRyby_2    | BAS_SNI_2010_CuttingFish3_conv        | ELDP+                  |   |
| AS_NI100713_RYB3     | as_ni100713_Razgovory_RazdelkaRyby_4    | BAS_SNI_2010_CuttingFish4_conv        | ELDP+                  |   |
| AS_NI100713_RYB4     | as_ni100713_Razgovory_RazdelkaRyby_5    | BAS_SNI_2010_CuttingFish5_conv        | ELDP+                  |   |
| AS_TN69_RAZ          | as_tn69_Razgovor_ips                    | BAS_BTN_1969_Conversation_conv        | SIP                    |   |
| AS080324_SKA         | as080324_Skazka                         | BAS_2008_LegendOfBear_flk             | ELDP+                  |   |
| AS080826_VNU         | as080826_Vnuchka                        | BAS_2008_MyGranddaughter_nar          | ELDP+                  |   |
| AS090714_MOR         | as090714_Moroshka                       | BAS_2009_Cloudberries_nar             | ELDP+                  |   |
| AS090715_KAKPLE      | as090715_KakPletutMaut                  | BAS_2009_MakingLasso_nar              | ELDP+                  |   |
| AS090715_POLZ        | as090715_KakPolzujutsaMautom            | BAS_2009_UsingLasso_nar               | ELDP+                  |   |
| AS100703_JAZ         | as100703_KtoNaKakixJazykax              | BAS_2010_LocalLanguages_nar           | ELDP+                  |   |
| AS100703_VNU         | as100703_SVnuchkoj                      | BAS_2010_ChatingWithGranddaughter_nar | ELDP+                  |   |
| AS100709_CHU         | as100709_KakPostavitjChum               | BAS_2010_InstallingATent_nar          | ELDP+                  |   |
| AS100709_KRAS        | as100709_KakNaKranoeOzero               | BAS_2010_RedLake_nar                  | ELDP+                  |   |
| AS100709_SET         | as100709_KakPostavitjSetj               | BAS_2010_InstallingANet_nar           | ELDP+                  |   |
| AS100709_TIM         | as100709_Timka                          | BAS_2010_LazyTimka_flk                | ELDP+                  |   |
| AS100709_ZHI         | as100709_ZhiznjVTundre                  | BAS_2010_LifeInTundra_nar             | ELDP+                  |   |
| AS100711_CHI         | as100711_ChinitSeti                     | BAS_2010_RepairingNets_nar            | ELDP+                  |   |
| AS100713_KLEJ        | as100713_RybijKlejRazdelyvaet           | BAS_2010_FishGlue_nar                 | ELDP+                  |   |
| AS100715_IDOL        | as100715_IdolUMateri                    | BAS_2010_MothersIdol_nar              | ELDP+                  |   |
| AS100715_SHAM        | as100715_ChtoZnaetProShamanov           | BAS_2010_Shamans_nar                  | ELDP+                  |   |
| ASP_NNB90S_INT       | asp_nnb90s_Interview_tdn                | PAS_BNN_1994_Interview2_conv          | TDNT                   |   |
| ASP_NNB940905_INT    | asp_nnb940905_Interview_rad             | PAS_BNN_BIE_1994_Interview1_conv      | GTRK                   |   |
| ASS_VMS_NNB90S_INT   | ass_vms_nnb90s_Interview_tdn            | SAS_SVM_NNB_199X_Interview_conv       | TDNT                   |   |
| DA_AS_SA080403_SUM   | da_as_sa080403_SumasshedshijStarik      | SDnA_BAS_RSA_2008_InsaneMan_flk       | ELDP+                  |   |
| DS9111_DVE           | ds9111_DveSkazki_rad                    | BDS_1991_TwoTales_flk                 | GTRK                   | First part is independently transcribed and published in (Siegl 2013: 481–486). |
| EDB_NNB910207_VOS    | edb_nnb910207_Vospominanija_rad         | BED_BNN_1991_MyLife_nar               | GTRK                   |   |

| Text code in Toolbox | Files name                        | Text code in corpus             | Audio recording source | Comment  |
|----------------------|-----------------------------------|---------------------------------|------------------------|--|
| EIB_NNB90S00S_INT1   | eib_nnb90s00s_Interview_1_tdnt    | BEI_BNN_200X_Interview1_conv    | TDNT                   |  |
| EIB_NNB90S00S_INT2   | eib_nnb90s00s_Interview_2_tdnt    | BEI_BNN_200X_Interview2_conv    | TDNT                   |  |
| ES080401_TJU         | es_ap080401_Tjulj                 | GES_BAP_2008_Tulle_nar          | ELDP+                  |  |
| ES090722_UL          | es090722_StaruxaUljjanaVBolnice   | GES_2009_InAHospital_nar        | ELDP+                  |  |
| ES100702_OBY         | es100712_EneckieObychai           | GES_2010_EnetsCustoms_nar       | ELDP+                  |  |
| ES100702_ZJA         | es100702_SkazkaOZjatjax           | GES_2010_Sons-in-law_flk        | ELDP+                  |  |
| ES100712_BERE        | es100712_ObychaiBeremennostj      | GES_2010_CustomsPregnancy_nar   | ELDP+                  |  |
| ES100715_MAL         | es100715_SkazkaOMaljchike         | GES_2010_PoorBoy_flk            | ELDP+                  |  |
| GA090713_VI          | ga090713_Videnie                  | IGA_2009_Ghost_nar              | ELDP+                  |  |
| GA090721_KUR         | ga090721_Kuropatka                | IGA_2009_Partridge_nar          | ELDP+                  |  |
| GA100629_RYB         | ga100629_Rybalka                  | IGA_2010_Fishing_nar            | ELDP+                  |  |
| GA100708_TUN         | ga100708_TundraOleni              | IGA_2010_ReindeerInTundra_nar   | ELDP+                  |  |
| IE_NNB90S_INT        | ie_nnb90s_Interview_rad           | BIE_BNN_199X_Interview_conv     | GTRK                   |  |
| II100712_TAB         | ii100712_TabakKonchalsjaVTundre   | SIIP_2010_TobaccoInTundra_nar   | ELDP+                  |  |
| II100713_PROZ        | ii100713_Prozvishcha              | SIIP_2010_Nicknames_nar         | ELDP+                  |  |
| KDJA_XXX69_RAZ       | kdja_xxx69_RazgovorSDevochkoj_ips | YKD_NN3_1969_Conversation_conv  | SIP                    | SIP's legacy manuscript (notebook 22, p. 4–5) became available after the text was independently transcribed and glossed; it was still consulted at the later stage |
| KDJA69_DIK           | kdja69_Dikij_ips                  | YKD_1969_WildReindeer_nar       | SIP                    | SIP's legacy manuscript (notebook 22, p. 2–3) became available after the text was independently transcribed and glossed; it was still consulted at the later stage |
| LD_VINB080818_VESCH  | ld_vinb080818_ShamanskjeVeschi    | BLD_BViN_2008_ShamansThings_nar | ELDP+                  |  |
| LD050925_JASOB       | ld050925_JaschikDljaSobole        | BLD_2005_SableTrap_nar          | ELDP+                  |  |
| LD050925_OLEN_1      | ld050925_OleniZabolevali_1        | BLD_2005_IllReindeer1_nar       | ELDP+                  |  |
| LD050925_OLEN_2      | ld050925_OleniZabolevali_2        | BLD_2005_IllReindeer2_nar       | ELDP+                  |  |
| LD050925_SOB         | ld050925_SluchajSSobakoj          | BLD_2005_DogsAteHareCubs_nar    | ELDP+                  |  |
| LD080325_DY1         | ld080325_StarikIDyrjavjyKamenj_1  | BLD_2008_StoneWithHole1_nar     | ELDP+                  |  |
| LD080325_DY2         | ld080325_StarikIDyrjavjyKamenj_2  | BLD_2008_StoneWithHole2_nar     | ELDP+                  |  |
| LD080402_OTSH        | ld080402_Otshelnik                | BLD_2008_Hermit_nar             | ELDP+                  |  |
| LD080406_TUN         | ld080406_StarikTungus             | BLD_2008_EvenkiOldMan_nar       | ELDP+                  |  |
| LD080406_VOD         | ld080406_Vodovorot                | BLD_2008_Whirlpool_nar          | ELDP+                  |  |

| Text code in Toolbox | Files name                     | Text code in corpus                  | Audio recording source | Comment |
|----------------------|--------------------------------|--------------------------------------|------------------------|---------|
| LD080815_JAS         | ld080815_Jasnovidjaschij       | BLD_2008_Clairvoyant_nar             | ELDP+                  |         |
| LD080815_PRI1        | ld080815_Prizrak_1             | BLD_2008_Ghost1_nar                  | ELDP+                  |         |
| LD080815_PRI2        | ld080815_Prizrak_2             | BLD_2008_Ghost2_nar                  | ELDP+                  |         |
| LD080816_CEP         | ld080816_SobakaCepIjalas       | BLD_2008_BadDog_nar                  | ELDP+                  |         |
| LD080816_GAG         | ld080816_ElGagar               | BLD_2008_IAteloona_nar               | ELDP+                  |         |
| LD080816_KOP         | ld080816_Kopytka               | BLD_2008_Necrobacteriosis_nar        | ELDP+                  |         |
| LD080816_MED         | ld080816_MedvedjSeti           | BLD_2008_BearCheckingNets_nar        | ELDP+                  |         |
| LD080816_SAL         | ld080816_MedvezhjeSalo         | BLD_2008_BearMeat_nar                | ELDP+                  |         |
| LD080816_SOB         | ld080816_MalenkajaSobaka       | BLD_2008_LittleDog_nar               | ELDP+                  |         |
| LD080816_STAR        | ld080816_StaruxaSobaki         | BLD_2008_WomanDogs_nar               | ELDP+                  |         |
| LD080817_DOZH        | ld080817_Dozhdj                | BLD_2008_NoRain_nar                  | ELDP+                  |         |
| LD080817_LOD         | ld080817_UgnaliLodku           | BLD_2008_SomeoneTookBoat_nar         | ELDP+                  |         |
| LD080817_MALK        | ld080817_Maljki                | BLD_2008_Fry_nar                     | ELDP+                  |         |
| LD080817_MALOZ       | ld080817_MalenjkoeOzero        | BLD_2008_LittleLake_nar              | ELDP+                  |         |
| LD080817_MED         | ld080817_OleniMedvedj          | BLD_2008_BearCame_nar                | ELDP+                  |         |
| LD080817_OGO         | ld080817_Ogonj                 | BLD_2008_Fire_nar                    | ELDP+                  |         |
| LD080817_PESOST      | ld080817_PeschanyjOstrov       | BLD_2008_SandyIsland_nar             | ELDP+                  |         |
| LD080817_SLED        | ld080817_SobachijSled          | BLD_2008_DogFootprint_nar            | ELDP+                  |         |
| LD080817_SOP         | ld080817_Sopka                 | BLD_2008_CapeCoveredWithForest_nar   | ELDP+                  |         |
| LD080818_DOR         | ld080818_DorogaNaNikolskoe     | BLD_2008_NikolskoeWay_nar            | ELDP+                  |         |
| LD080818_LEB         | ld080818_Lebedi                | BLD_2008_Swans_nar                   | ELDP+                  |         |
| LD080818_MEN         | ld080818_MenjaetsaNaOzere      | BLD_2008_ChangingLandscape_nar       | ELDP+                  |         |
| LD080818_NEM         | ld080818_NemcyZemljanka        | BLD_2008_GermansEarthHouse_nar       | ELDP+                  |         |
| LD080818_OTKA        | ld080818_OtkarmIvali           | BLD_2008_HowWeFedASwan_nar           | ELDP+                  |         |
| LD080818_OZE         | ld080818_OzeralReki            | BLD_2008_LakesAndRivers_nar          | ELDP+                  |         |
| LD080818_PODL        | ld080818_PodlednyeSeti         | BLD_2008_UnderIceNets_nar            | ELDP+                  |         |
| LD080818_SJUZIM      | ld080818_Sjudazimoj            | BLD_2008_ComingToTheLakeInWinter_nar | ELDP+                  |         |
| LD080819_BLES        | ld080819_RybachiliSBlesnoj     | BLD_2008_FishingWithAMinnow_nar      | ELDP+                  |         |
| LD080819_CHUMSTO     | ld080819_ChumStojal            | BLD_2008_MyTentHere_nar              | ELDP+                  |         |
| LD080819_JAG         | ld080819_ChtoSJagodami         | BLD_2008_PreservingBerries_nar       | ELDP+                  |         |
| LD080819_MOR         | ld080819_KakZagotovitjMoroshku | BLD_2008_PreservingCloudberries_nar  | ELDP+                  |         |

| Text code in Toolbox | Files name                    | Text code in corpus                 | Audio recording source | Comment                                      |
|----------------------|-------------------------------|-------------------------------------|------------------------|--|
| LD080819_MYSHGEN     | ld080819_MyshPoGenke          | BLD_2008_GenaAndMouse_nar           | ELDP+                  |  |
| LD080819_OLEN        | ld080819_OlenevodstvoProdukty | BLD_2008_FoodToReindeerHerders_nar  | ELDP+                  |  |
| LD080819_OTME        | ld080819_OtmetkaNajtiSeti     | BLD_2008_MarkForFindingNets_nar     | ELDP+                  | Published in (Khanina & Shluinsky 2023: 849) |
| LD080819_PASTRAS     | ld080819_PastjRastaschili     | BLD_2008_PeopleRemovedTrap_nar      | ELDP+                  |  |
| LD080819_PURG        | ld080819_PurgjByli            | BLD_2008_Snowstorms_nar             | ELDP+                  |  |
| LD080819_RYB         | ld080819_RybyMolodezh         | BLD_BViN_2008_FishYoungPeople_nar   | ELDP+                  |  |
| LD080819_RYBVNU      | ld080819_RybjiVnutrennosti    | BLD_2008_FishEntrails_nar           | ELDP+                  |  |
| LD080819_SDEBOL      | ld080819_SdelatjBolok         | BLD_2008_IWantToMakeABolok_nar      | ELDP+                  |  |
| LD080819_SESMAT      | ld080819_SestraMateri         | BLD_2008_MyMothersSister_nar        | ELDP+                  |  |
| LD080819_SETXOD      | ld080819_SetjXodunomXodila    | BLD_2008_ShakingNet_nar             | ELDP+                  |  |
| LD080819_UT          | ld080819_Utki                 | BLD_2008_Ducks_nar                  | ELDP+                  |  |
| LD080819_UTSID       | ld080818_UtkaSidela           | BLD_2008_DuckSatHere_nar            | ELDP+                  |  |
| LD080819_VIDGLU      | ld080819_VidelGluxarja        | BLD_2008_ISawCapercaillie_nar       | ELDP+                  |  |
| LD080819_VIDPTI      | ld080819_VidelPticu           | BLD_2008_ISawBird_nar               | ELDP+                  |  |
| LD080819_VODCVE      | ld080819_VodaCvetet           | BLD_2008_AlgalBloom_nar             | ELDP+                  |  |
| LD080819_ZAG         | ld080819_ZagotavlivaliRybu    | BLD_2008_PreservingFish_nar         | ELDP+                  |  |
| LD080819_ZAPDRO      | ld080819_ZapasDrov            | BLD_2008_FirewoodSupply_nar         | ELDP+                  |  |
| LD080819_ZHITUN      | ld080819_ZhiznjVTundre        | BLD_2008_LifeInTundra_nar           | ELDP+                  |  |
| LD080825_PCHE        | ld080825_Pchely               | BLD_2008_Bees_nar                   | ELDP+                  |  |
| LD080826_DVESOB      | ld080826_DveSobaki            | BLD_2008_TwoDogs_nar                | ELDP+                  |  |
| LD080826_ISPU        | ld080826_IspugalsjaMedvedej   | BLD_2008_IFrightenedBears_nar       | ELDP+                  |  |
| LD080826_STAOL       | ld080826_StarikOleni          | BLD_2008_OldManAndReindeer_nar      | ELDP+                  |  |
| LD080826_TRIMED      | ld080826_TriMedvedja          | BLD_2008_ThreeBears_nar             | ELDP+                  |  |
| LD090716_CHUM        | ld090716_KakSobiratChum       | BLD_2009_InstallingATent_nar        | ELDP+                  |  |
| LD090718_NEV         | ld090718_KakNevodjat          | BLD_2009_FishingWithTrotline_nar    | ELDP+                  |  |
| LD090718_ROS         | ld090718_Rosomaxa             | BLD_2009_Glutton_nar                | ELDP+                  |  |
| LD090718_XOD         | ld090718_XodRyby              | BLD_2009_FishGoingToSpawning_nar    | ELDP+                  |  |
| LD090721_GUBBRASP    | ld090721_GubaRaspuxla         | BLD_2009_MyCousinsLipWasSwollen_nar | ELDP+                  |  |
| LD090721_MED         | ld090721_Medvedica            | BLD_2009_SheBear_nar                | ELDP+                  |  |
| LD090726_UPR         | ld090726_Uprijazh             | BLD_2009_Harness_nar                | ELDP+                  |  |



| Text code in Toolbox | Files name                   | Text code in corpus                       | Audio recording source | Comment |
|----------------------|------------------------------|---|------------------------|---------|
| LD090726_VJJU        | ld090726_VjjuniNazhivka      | BLD_2009_WaterWormsForBait_nar            | ELDP+                  |         |
| LD090727_OLZHI       | ld090727_OleniZhirnajaGrudj  | BLD_2009_ReindeerWithFatBeast_nar         | ELDP+                  |         |
| LD090727_SHES        | ld090727_KakDelalShesty      | BLD_2009_MakingPoles_nar                  | ELDP+                  |         |
| LD090729_BED         | ld090729_BednyjNaSokakax     | BLD_2009_PoorManHarnessingDogs_nar        | ELDP+                  |         |
| LD090729_DVE         | ld090729_DveDorogiNazad      | BLD_2009_TwoWaysBack_nar                  | ELDP+                  |         |
| LD090729_JAG         | ld090729_OstalisBezJagelja   | BLD_2009_HowReindeerStayedWithoutMoss_nar | ELDP+                  |         |
| LD090729_KAKID       | ld090729_KakIdtiNazad        | BLD_2009_HowToGoBack_nar                  | ELDP+                  |         |
| LD090729_ODNOR       | ld090729_OdnorogieOleni      | BLD_2009_OneAntleredReindeer_nar          | ELDP+                  |         |
| LD090729_PONSOB      | ld090729_PonjatlivajaSobaka  | BLD_2009_IntelligentDog_nar               | ELDP+                  |         |
| LD090729_REFOK       | ld090729_RechkaFokina        | BLD_2009_FokinaRiver_nar                  | ELDP+                  |         |
| LD090729_RUCH        | ld090729_RuchejOtOzeraFokina | BLD_2009_BrookBehindFokinaRiver_nar       | ELDP+                  |         |
| LD090729_SOBCHU      | ld090729_SobakiChuma         | BLD_2009_DogPlague_nar                    | ELDP+                  |         |
| LD090730_SPA         | ld090730_SpalnyjMeshok       | BLD_2009_SleepingBag_nar                  | ELDP+                  |         |
| LD090803_ELIZ        | ld090803_PokazalosElizlv     | BLD_2009_GhostSeenByElizaveta_nar         | ELDP+                  |         |
| LD090803_EXA         | ld090803_ExaliPoEniseju      | BLD_2009_HowWeRidAlongYenisey_nar         | ELDP+                  |         |
| LD090803_REB         | ld090803_RebjonokZaMedvedem  | BLD_2009_MyChildRanBehindPeople_nar       | ELDP+                  |         |
| LD090803_SPO         | ld090803_SpokojnyjGod        | BLD_2009_QuietYear_nar                    | ELDP+                  |         |
| LD100629_CHAG        | ld100629_ChagaChaj           | BLD_2010_ChagaTea_nar                     | ELDP+                  |         |
| LD100629_KAM         | ld100629_KamusyDljaStarika   | BLD_2010_OldMansKamuses_nar               | ELDP+                  |         |
| LD100629_KOMARY      | ld100629_Komary              | BLD_2010_Mosquitoes_nar                   | ELDP+                  |         |
| LD100629_KOMLICH     | ld100629_KomaryLichinki      | BLD_2010_MosquitoLarvae_nar               | ELDP+                  |         |
| LD100629_KOMMOSH     | ld100629_KomarlMoshka        | BLD_2010_MosquitoesAndMidges_nar          | ELDP+                  |         |
| LD100629_KUPMLAD     | ld100629_KupanieMladenca     | BLD_2010_BathingABaby_nar                 | ELDP+                  |         |
| LD100629_LAP         | ld100629_Lapnik              | BLD_2010_SpruceBranches_nar               | ELDP+                  |         |
| LD100629_MEDSOB      | ld100629_MedvedjSobaka       | BLD_2010_SheBearAndCubs_nar               | ELDP+                  |         |
| LD100629_NASHLI      | ld100629_NashliUtinojeGneздо | BLD_2010_WeFoundDuckNest_nar              | ELDP+                  |         |
| LD100629_PUT         | ld100629_PutjKOzeru          | BLD_2010_WayToLake_nar                    | ELDP+                  |         |
| LD100629_TRO         | ld100629_ZajachjaTropa       | BLD_2010_HarePath_nar                     | ELDP+                  |         |
| LD100629_UTGNEZ      | ld100629_UtkiGneздо          | BLD_2010_DuckNest_nar                     | ELDP+                  |         |
| LD100629_UTINJA      | ld100629_UtinojeJajco        | BLD_2010_DuckEgg_nar                      | ELDP+                  |         |
| LD100629_UTPERJAJ    | ld100629_UtkiPerenosjatJajca | BLD_2010_AnimalsAndTheirChildren_nar      | ELDP+                  |         |

| Text code in Toolbox | Files name                  | Text code in corpus                   | Audio recording source | Comment                              |
|----------------------|-----------------------------|---------------------------------------|------------------------|--------------------------------------|
| LD100629_VID         | ld100629_VidyRyby           | BLD_2010_FishSorts_nar                | ELDP+                  |                                      |
| LD100629_VIDKOM      | ld100629_VidyKomarov        | BLD_2010_MosquitoSorts_nar            | ELDP+                  |                                      |
| LD100629_VLOD        | ld100629_VLodke             | BLD_2010_InTheBoat_nar                | ELDP+                  |                                      |
| LD100629_ZCHAG       | ld100629_ZavarkaChaga       | BLD_2010_BrewingChaga_nar             | ELDP+                  |                                      |
| LD100629_ZHAR        | ld100629_ZharkaRyby         | BLD_2010_FryingFish_nar               | ELDP+                  |                                      |
| LD100629_ZIMKOM      | ld100629_ZimnijKomar        | BLD_2010_WinterMosquito_nar           | ELDP+                  |                                      |
| LD100630_BOR         | ld100630_BorodatyjMuzhik    | BLD_2010_BeardedMan_nar               | ELDP+                  |                                      |
| LD100630_MAUT        | ld100630_KakDelatjMaut      | BLD_2010_MakingLasso_nar              | ELDP+                  |                                      |
| LD100701_NIT         | ld100701_VygorajutNitki     | BLD_2010_ThreadsGetFaded_nar          | ELDP+                  |                                      |
| LD100701_OLUX        | ld100701_OleniUxodjat       | BLD_2010_ReindeerLeaveAway_nar        | ELDP+                  |                                      |
| LD100701_SMESH       | ld100701_SmeshnyeMedvedi    | BLD_2010_FunnyBears_nar               | ELDP+                  |                                      |
| LD100702_NAZAB       | ld100702_NaZaborOtMedvedja  | BLD_2010_ClimbingOnFence_nar          | ELDP+                  |                                      |
| LD100702_ROS         | ld100702_Rosomaxa           | BLD_2010_DirtyGlutton_nar             | ELDP+                  |                                      |
| LD100702_VER         | ld100702_VerevkaZaputalasj  | BLD_2010_HowRopeEntangled_nar         | ELDP+                  |                                      |
| LD100704_PROV        | ld100704_ProvalijsjaPodLjod | BLD_2010_IFellThroughIce_nar          | ELDP+                  |                                      |
| LD100705_GDEOL       | ld100705_GdeKakieOleni      | BLD_2010_SortsOfWildReindeer_nar      | ELDP+                  |                                      |
| LD100705_KAM         | ld100705_Kamenj             | BLD_2010_Stone_nar                    | ELDP+                  |                                      |
| LD100705_KRY         | ld100705_KrysinyjKorolj     | BLD_2010_RatsAndErmine_nar            | ELDP+                  |                                      |
| LD100705_PES         | ld100705_SledyPestsalLisy   | BLD_2010_PolarFoxAndFoxFootprints_nar | ELDP+                  |                                      |
| LD100705_SOB         | ld100705_UmnajaSobaka       | BLD_2010_CleverDog_nar                | ELDP+                  | Published in (Shluinsky et al. 2021) |
| LD100706_PES         | ld100706_Pesets             | BLD_2010_Fox_nar                      | ELDP+                  |                                      |
| LD100706_PRIM        | ld100706_Primanka           | BLD_2010_Bait_nar                     | ELDP+                  |                                      |
| LD100706_SEZO        | ld100706_SezonOxoty         | BLD_2010_HuntingSeason_nar            | ELDP+                  |                                      |
| LD100706_SILKUR      | ld100706_SilkiKuropatki     | BLD_2010_TrapsForPartridges_nar       | ELDP+                  |                                      |
| LD100706_SILZAJ      | ld100706_SilkiZajats        | BLD_2010_TrapsForHares_nar            | ELDP+                  |                                      |
| LD100706_ZHILY       | ld100706_Zhily              | BLD_2010_Tendons_nar                  | ELDP+                  |                                      |
| LD100711_DEZH        | ld100711_Dezhurstvo         | BLD_2010_MyVigil_nar                  | ELDP+                  |                                      |
| LD100711_KLE         | ld100711_KakDelatKley       | BLD_2010_MakingGlue_nar               | ELDP+                  |                                      |
| LD100711_RANXOR      | ld100711_KakRanjsheXoronili | BLD_2010_BuryingDeceasedFormerly_nar  | ELDP+                  |                                      |
| LD100711_RAZDE       | ld100711_RazdelkaRyby       | BLD_2010_CuttingFish_nar              | ELDP+                  |                                      |
| LD100712_CHER        | ld100712_Chervi             | BLD_2010_WaterWorms_nar               | ELDP+                  |                                      |

| Text code in Toolbox | Files name                      | Text code in corpus              | Audio recording source                       | Comment   |
|----------------------|---------------------------------|----------------------------------|--|---|
| LD100712_NELM        | ld100712_Nelma                  | BLD_2010_WhiteSalmon_nar         | ELDP+  |   |
| LD100712_OGR         | ld100712_OgromnajaRyba          | BLD_2010_HugeFish_nar            | ELDP+  |   |
| LD100712_OND         | ld100712_OndatrVSetjax          | BLD_2010_MuskratInTheNets_nar    | ELDP+  |   |
| LD100712_SHKUPES     | ld100712_ShkuraPesets           | BLD_2010_PolarFoxSkin_nar        | ELDP+  |   |
| LD100712_SHTA        | ld100712_ShtanyProzhgli         | BLD_2010_HowWeBurntTrousers_nar  | ELDP+  |   |
| LD100713_POKOL       | ld100713_Pokolka                | BLD_2010_PokolkaLake_nar         | ELDP+  |   |
| LD100713_POMGENP     | ld100713_PomeresjilosjGenke     | BLD_2010_SomeoneSeemedToGena_nar | ELDP+  |   |
| LD100713_ROG         | ld100713_Roga                   | BLD_2010_Antlers_nar             | ELDP+  |   |
| LD100713_RYBMED      | ld100713_RybakiMedvedj          | BLD_2010_FishermenAndBear_nar    | ELDP+  |   |
| LD100713_SHA         | ld100713_Shaman                 | BLD_2010_Shaman_nar              | ELDP+  |   |
| LD100713_UPR         | ld100713_PomeresjilasjUprjazhka | BLD_2010_HowSledgeSeemedToMe_nar | ELDP+  |   |
| LD100714_KUKU        | ld100714_Kukushka               | BLD_2010_Cuckoo_nar              | ELDP+  |   |
| LD100714_OV          | ld100714_Ovod                   | BLD_2010_ReindeerGadfly_nar      | ELDP+  |   |
| LD100715_DOOZER      | ld100715_KakDoytiDoOzera        | BLD_2010_HowToGetToLake_nar      | ELDP+  |   |
| LD100715_GUS         | ld100715_ProzvisheGus           | BLD_2010_NicknameGoose_nar       | ELDP+  |   |
| LD100715_KAP         | ld100715_Kapkany                | BLD_2010_Traps_nar               | ELDP+  |   |
| LD100715_MIR         | ld100715_Mirazh                 | BLD_2010_Mirage_nar              | ELDP+  |   |
| LD100715_OLENJA      | ld100715_ShkuraOlenjat          | BLD_2010_ReindeerCalfSkin_nar    | ELDP+  |   |
| LD100715_POR         | ld100715_Porne                  | BLD_2010_Witch_flk               | ELDP+  |   |
| LD100715_ROSOM       | ld100715_Rosomaxa               | BLD_2010_GluttonInATrap_nar      | ELDP+  |   |
| LD100715_SHKU        | ld100715_KakSkoblitShkury       | BLD_2010_ScrapingSkins_nar       | ELDP+  |   |
| LD100715_SORT        | ld100715_SortaRyby              | BLD_2010_FishVarieties_nar       | ELDP+  |   |
| LD100715_VOL         | ld100715_Volki                  | BLD_2010_Wolves_nar              | ELDP+  |   |
| MNB_NNB90S_INT       | mnbn90s_Interview_tdnt          | BMN_BNN_199X_Interview_conv      | TDNT   |   |
| MNB_NNB90S_PRI       | mnbn90s_Prимety_tdnt            | BMN_BNN_199X_FolkSigns_conv      | TDNT   |   |
| MNB9404_TRI          | mnbn9404_TriZhenschiny_rad_tdnt | BMN_1994_ThreeWomen_flk          | GTRK, but lacking there first sentences TDNT | Published as L2002: Нэxу нэ, 108–110 (reprinted in Siegl 2013: 496–501).<br>Glossing in Toolbox was done based on published L2002 transcription |

| Text code in Toolbox | Files name                           | Text code in corpus                              | Audio recording source | Comment  |
|----------------------|--------------------------------------|--|------------------------|--|
| MNS_NNB90S_INT       | mns_nnb90s_Interview_tdnt            | SMN_BNN_199X_Interview_conv                      | TDNT                   | A song that is a little fragment of this text is published as L2002: Онай энче бари, 63, 75 (but no other transcription of the whole recording is available).<br>The text SDA_1969_LittleChildSong_song is a version of the same song, but recorded from a different person. |
| MNS900522_MYSHKU     | mns900522_Kukushka_rad               | SMN_1990_Cuckoo_flk and<br>SMN_1990_Mouse_flk    | GTRK                   | In Toolbox, two texts mad a single text, cut into two text in the corpus.<br>The first one (SMN_1990_Cuckoo_flk) is published as SB2005: text 38. Transcribed for glossing independently   |
| ND_NI050923_ROZH     | nd_ni050923_Rozhdenija               | LND_SNI_2005_HowMyChildrenWereBorn_nar           | ELDP+                  |  |
| ND_NI050923_ZHEN     | nd_ni050923_Zhena                    | LND_SNI_2005_MyWife_nar                          | ELDP+                  |  |
| ND970718_ZHI         | nd970718_OZhizni_rad                 | LND_BNN_1997_MyLife_nar                          | GTRK                   |  |
| NI_ES090801_SHU      | ni_es090801_ShutlivyjRazgovor        | SNI_GES_2009_JokingConversation_conv             | ELDP+                  |  |
| NI_ES100712_PERE     | ni_es100712_Razgovor_Perekur         | SNI_GES_2010_ConversationSmoking_conv            | ELDP+                  |  |
| NI_GA100713_RYBNIZ   | ni_ga100713_Razgovor_Rybalka         | SNI_IGA_2010_ConversationFishing_conv            | ELDP+                  |  |
| NI_GA100714_RYBNIZ   | ni_ga100714_Razgovor_RybalkaNizovjja | SNI_IGA_2010_ConversationLowerPartOfYenisey_conv | ELDP+                  |  |
| NI080815_DOCH        | ni080815_StarshajaDochj              | SNI_2008_MyElderDaughter_nar                     | ELDP+                  |  |
| NI080815_EZSOB       | ni080815_EzditNaSobakax              | SNI_2008_HowIHarnessDogs_nar                     | ELDP+                  |  |
| NI080815_OSEN        | ni080815_OxotaOsenju                 | SNI_2008_HuntingInAutumn_nar                     | ELDP+                  |  |
| NI080815_OZKOM       | ni080815_OzeroKomary                 | SNI_2008_LakeAndMosquitoes_nar                   | ELDP+                  |  |
| NI080815_RAB         | ni080815_KakajaRabota                | SNI_2008_AvailableJobs_nar                       | ELDP+                  |  |
| NI080815_VOLTUN      | ni080815_VolkiVTundre                | SNI_2008_WolvesInTundra_nar                      | ELDP+                  |  |
| NI080817_DJO         | ni080817_Djoa                        | SNI_2008_Dyoa_flk                                | ELDP+                  |  |
| NI080823_SKAOZH      | ni080823_SkazkaIOzhizni              | SNI_2008_TaleAndRealLife_nar                     | ELDP+                  |  |
| NI080823_TOV1        | ni080823_DvaTovarischa_1             | SNI_2008_TwoFellows1_flk                         | ELDP+                  |  |
| NI080823_TOV2        | ni080823_DvaTovarischa_2             | SNI_2008_TwoFellows2_flk                         | ELDP+                  |  |
| NI090712_TAB         | ni090712_Tabak                       | SNI_2009_Tobacco_nar                             | ELDP+                  |  |
| NI090715_LOS         | ni090715_PojmalLosja                 | SNI_2009_HowICaughtAnElk_nar                     | ELDP+                  |  |
| NI090715_LYZH        | ni090715_LyzhiKamusami               | SNI_2009_RubbingSkisWithKamuses_nar              | ELDP+                  |  |
| NI090715_MAUT        | ni090715_KakDelatjMaut               | SNI_2009_MakingLasso_nar                         | ELDP+                  |  |

| Text code in Toolbox | Files name                       | Text code in corpus                        | Audio recording source | Comment |
|----------------------|----------------------------------|--|------------------------|---------|
| NI090719_ZOL         | ni090719_ZolotajaRybka           | SNI_2009_TalkingFish_flk                   | ELDP+                  |         |
| NI090727_GRI         | ni090727_GribyJagody             | SNI_2009_MushroomsAndBerries_nar           | ELDP+                  |         |
| NI100630_SNY         | ni100630_OSnax                   | SNI_2010_MyDreams_nar                      | ELDP+                  |         |
| NI100705_XODTUN      | ni100705_XodiliVTundru           | SNI_2010_HowWeWereInTundra_nar             | ELDP+                  |         |
| NI100705_ZHO         | ni100705_ChjaZhopa               | SNI_2010_LostCroup_nar                     | ELDP+                  |         |
| NI100706_NJU         | ni100706_NjukilzKory             | SNI_2010_TentCoversMadeOfBark_nar          | ELDP+                  |         |
| NI100713_IDO         | ni100713_ProisxozhdeniIdoly      | SNI_2010_Idols_nar                         | ELDP+                  |         |
| NI100713_OBY         | ni100713_ObychaiDveZheny         | SNI_2010_CustomOfPolygamy_nar              | ELDP+                  |         |
| NI100713_XOR         | ni100713_KakXoronili             | SNI_2010_BuryingDeceasedPeopleFormerly_nar | ELDP+                  |         |
| NI100714_SME         | ni100714_SmeshnojSluchaj         | SNI_2010_FunnyStory_flk                    | ELDP+                  |         |
| NI100715_NAKON       | ni100715_NakonechnikiXoreja      | SNI_2010_ReindeerPoleTips_nar              | ELDP+                  |         |
| NI100715_OLAS        | ni100715_KtoTakieOlasne          | SNI_2010_Witches_nar                       | ELDP+                  |         |
| NI100715_SLU         | ni100715_SluchaiMedvediVolki     | SNI_2010_StoriesBearsWolves_nar            | ELDP+                  |         |
| NK_LD080403_SVA      | nk_ld080403_Svadba               | BNK_BLD_2008_OurMarriage_nar               | ELDP+                  |         |
| NK080325_RYB         | nk080325_Ryba                    | BNK_2008_CuttingFish1_nar                  | ELDP+                  |         |
| NK080403_SK          | nk080403_Skazka                  | BNK_2008_TaleGirl_flk                      | ELDP+                  |         |
| NK080403_ZHI         | nk080403_OtecIBeglecy            | BNK_2008_MyFatherAndFugitives_nar          | ELDP+                  |         |
| NK080406_ARG         | nk080406_ArgishiMedvedj          | BNK_2008_ReindeerCaravanAndBear_nar        | ELDP+                  |         |
| NK080406_IVA         | nk080406_IstorijaIvanovyh        | BNK_2008_IvanovsHistory_nar                | ELDP+                  |         |
| NK080406_POZVA       | nk080406_KakPozvaliZhitjPalchiny | BNK_2008_HowILivedWithPalchins_nar         | ELDP+                  |         |
| NK080406_UBE         | nk080406_UbezahavshieOleni       | BNK_2008_LostReindeer_nar                  | ELDP+                  |         |
| NK080823_MYSH        | nk080823_Myshi                   | BNK_2008_Mice_nar                          | ELDP+                  |         |
| NK090723_RYB         | nk090723_RazdelkaRyby            | BNK_2009_CuttingFish2_nar                  | ELDP+                  |         |
| NK090729_OBU         | nk090729_Obuv                    | BNK_2009-Shoes_nar                         | ELDP+                  |         |
| NK090729_UNI1        | nk090729_KakShitUntajki_I_1      | BNK_2009_SewingShoes1_nar                  | ELDP+                  |         |
| NK090729_UNI2        | nk090729_KakShitUntajki_I_2      | BNK_2009_SewingShoes2_nar                  | ELDP+                  |         |
| NK090729_UNI3        | nk090729_KakShitUntajki_I_3      | BNK_2009_SewingShoes3_nar                  | ELDP+                  |         |
| NK090729_UNII1       | nk090729_KakShitUntajki_II_1     | BNK_2009_SewingShoes4_nar                  | ELDP+                  |         |
| NK090729_UNII2       | nk090729_KakShitUntajki_II_2     | BNK_2009_SewingShoes5_nar                  | ELDP+                  |         |
| NK090729_UNII3       | nk090729_KakShitUntajki_II_3     | BNK_2009_SewingShoes6_nar                  | ELDP+                  |         |
| NK090730_RODY        | nk090730_RodyObrjad              | BNK_2009_ChildbirthCustom_nar              | ELDP+                  |         |

| Text code in Toolbox | Files name                           | Text code in corpus                 | Audio recording source | Comment   |
|----------------------|--------------------------------------|-------------------------------------|------------------------|---|
| NK090803_DIK         | nk090803_UbilDikogo                  | BNK_2009_KilledWildReindeer_nar     | ELDP+                  |   |
| NK090803_EXPOEN      | nk090803_ExaliPoEniseju              | BNK_2009_HowWeRidAlongYenisey_nar   | ELDP+                  |   |
| NK090803_SBE         | nk090803_SbezhavshyeZakljuchjonnye   | BNK_2009_FugitivePrisoners_nar      | ELDP+                  |   |
| NK090803_UKRA        | nk090803_KakUkrali                   | BNK_2009_HowIWasStolen_nar          | ELDP+                  |   |
| NK100703_KRAS        | nk100703_KrasnajaZemlja              | BNK_2010_RedSoil_nar                | ELDP+                  |   |
| NK100703_LJU         | nk100703_VeshaliLjuljku              | BNK_2010_HangingCradle_nar          | ELDP+                  |   |
| NK100703_MOL         | nk100703_MaloMoloka                  | BNK_2010_LowMilkSupply_nar          | ELDP+                  |   |
| NK100703_PERE        | nk100703_KakPerevoziliDetej          | BNK_2010_TransportationOfBabies_nar | ELDP+                  |   |
| NK100703_PRIHER      | nk100703_PrimetyBeremennostj         | BNK_2010_FolkSignsPregnancy_nar     | ELDP+                  |   |
| NK100703_PRIK        | nk100703_Prikorm                     | BNK_2010_ComplementaryFoods_nar     | ELDP+                  |   |
| NK100703_PRINOG      | nk100703_PrimetyPolRebenkaNogi       | BNK_2010_FolkSignsBabysSex2_nar     | ELDP+                  |   |
| NK100703_PRIKPOL     | nk100703_PrimetyPolRebenka           | BNK_2010_FolkSignsBabysSex1_nar     | ELDP+                  |   |
| NK100703_RODY        | nk100703_RodyVTundre                 | BNK_2010_ChildbirthInTundra_nar     | ELDP+                  |   |
| NK100709_JUK         | nk100709_KakDelatjJukolu             | BNK_2010_MakingDryFish_nar          | ELDP+                  |   |
| NK100711_GNEZ        | nk100711_NaxodiliGnezdo              | BNK_2010_HowWeFoundNests_nar        | ELDP+                  |   |
| NK100711_UKR         | nk100711_UkryvatjBolokSnegom         | BNK_2010_CoveringBolokWithSnow_nar  | ELDP+                  |   |
| NKU_ZK_LKU080319_RAZ | nku_zk_lku080319_Razgovory           | BNKu_TZK_BLK_2008_Conversation_conv | ELDP+                  |   |
| NKU_ZKU080319_DIA    | nku_zku080319_Dialog                 | BNKu_TZK_2008_Conversation_conv     | ELDP+                  |   |
| NPCH_NNB970801       | npch_nnb970801_Interview_rad         | ChNP_BNN_1997_Interview_conv        | GTRK                   |   |
| NSB69_STIH           | nsb69_Stih_ips                       | BNS_1969_Poem_song                  | SIP                    |   |
| NSP70S80S_OLKOM      | nsp70s80s_MoreoOlasneKomary_ips      | PNS_19XX_WitchMosquitoes_flk        | SIP                    |   |
| NSP70S90S_MUZ        | nsp70s90s_MuzikalnyeInstrumenty_tdnt | PNS_19XX_MusicalInstruments_nar     | TDNT                   |   |
| NSP70S90S_OLAS       | nsp70s90s_Olasne_tdnt                | PNS_19XX_WitchAndGirl_flk           | TDNT                   |   |
| NSP70S90S_UKR        | nsp70s90s_Ukrashenija_tdnt           | PNS_19XX_Jewelry_nar                | TDNT                   |   |
| NSP70S90S_ZHEN       | nsp70s90s_Zhenitjba_tdnt             | PNS_19XX_TraditionalMarriage_nar    | TDNT                   |   |
| NSP900810_ENE        | nsp900810_EneciRussskij_rad          | PNS_1990_EnetsAndRussian1_flk       | GTRK                   | Independently transcribed and published in SB2005: text 61 and L2002: Онай энче, роса энче, 92–95.<br>Glossing in Toolbox was done based on L2002 published transcription |

| Text code in Toolbox | Files name                 | Text code in corpus                  | Audio recording source | Comment  |
|----------------------|----------------------------|--------------------------------------|------------------------|--|
| NSP900817_OLA1       | nsp900817_Olasne_1_rad     | PNS_1990_MoreoAndWitch1_flk          | GTRK                   | Published as SB2005: text 22 (as a single text together with PNS_1990_MoreoAndWitch2_flk).<br>Transcribed for glossing independently   |
| NSP900817_OLA2       | nsp900817_Olasne_2_rad     | PNS_1990_MoreoAndWitch2_flk          | GTRK                   | Published as SB2005: text 22 (as a single text together with PNS_1990_MoreoAndWitch1_flk).<br>Transcribed for glossing independently   |
| NSP90S_BYK           | nsp90s_Skazka_UbilBykov_au | PNS_1994_WildReindeerBulls_flk       | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible.<br>An independent earlier draft transcription done by GVV is available and was checked |
| NSP90S_DVA           | nsp90s_2Shamana_od         | PNS_199X_EnetsShamanEvenkiShaman_flk | DOE                    | Published as L2002:<br>Онэй тазобэ, ося тазобэ, 95–99.<br>Glossing in Toolbox was done based on published L2002 transcription  |
| NSP90S_DYOP1         | nsp90s_Djoa_Opisanie_1_au  | PNS_1994_HowDyoeLooks1_nar           | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible   |
| NSP90S_DYOP2         | nsp90s_Djoa_Opisanie_2_au  | PNS_1994_HowDyoeLooks2_nar           | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible   |
| NSP90S_ENRU          | nsp90s_EneclRusskij_od     | PNS_199X_EnetsAndRussian2_flk        | DOE                    |  |
| NSP90S_NOVOS         | nsp90s_Novosibirsk_au      | PNS_1994_Novosibirsk_nar             | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible   |
| NSP90s_RAN           | nsp90s_KakRansheZhili_au   | PNS_1994_FormerLife_nar              | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible   |
| NSP90S_SHAM          | nsp90s_Shaman_au           | PNS_1994_Shaman_nar                  | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible   |

| Text code in Toolbox   | Files name                       | Text code in corpus               | Audio recording source | Comment   |
|------------------------|----------------------------------|-----------------------------------|------------------------|---|
| NSP90S_UR              | nsp90s_Urody_au                  | PNS_1994_MythologicalFreaks_nar   | UA / WNB               | Transcription was done with the recording from UA's archive, but later a better quality recording from WNB's archive became accessible                            |
| NSP910302_ENRU         | nsp910302_EnetsIRusskij_tdnt     | PNS_1991_EnetsAndRussian3_flk     | TDNT                   |   |
| NSP92_DVA              | nsp92_DvaCheloveka_od            | PNS_1992_TwoMen_flk               | DOE                    |   |
| PN080320_CHA           | pn080320_Chajki                  | BPN_2008_Gulls_nar                | ELDP+                  |   |
| PN080320_SKAZ          | pn080320_SkazkiOtca              | BPN_2008_FathersTales_nar         | ELDP+                  |   |
| RODNSL_TRBR            | RodnoeSlovo_TriBrata             | BVN_1992_ThreeBrothers_flk        | -                      | Published as L2002: Hэxy кaкa, 100–103<br>Glossed in Toolbox using L2002 transcription (with no recording available)  |
| SA080331_DVE           | sa080331_DveZhenschiny           | RSA_2008_TwoWomen_flk             | ELDP+                  |   |
| SA080815_ISP           | sa080815_Ispugalisj              | RSA_2008_WeGotAfraidOfBear_nar    | ELDP+                  |   |
| SA080818_KOLYB         | sa080818_Pesnja_KolybeljnajaSynu | RSA_2008_Lullaby_song             | ELDP+                  |   |
| SA080820_KU            | sa080820_SkazkaKukushka          | RSA_2008_Cuckoo_flk               | ELDP+                  |   |
| SA080824_TONU          | sa080824_TonulaVBolote           | RSA_2008_IAlmostPerishedInBog_nar | ELDP+                  |   |
| SPB_EDB_KAB70S80S_POLY | spb_edb_kab70s80s_Polylogue_ips  | BSP_BED_BKA_19XX_HowWeLive_nar    | SIP                    |   |
| SPB_NNB910131_INT_NNB  | spb_nnb910131_Interview_rad      | BSP_BNN_1991_Interview_conv       | GTRK                   |   |
| SPB_XXX70S80S          | spb_xxx70s80s_DikieOleni_ips     | BSP_NN4_19XX_WildReindeer_conv    | SIP                    | SIP's legacy manuscript (notebook 22, p. 6–15) became available after the text was independently transcribed and glossed; it was still checked at the later stage |
| SPB70s80s_KULD         | spb70s80s_KukushkaLjdina_ips     | BSP_19XX_TwoTales_flk             | SIP                    |   |
| SPB70S80S_SHKU         | spb70s80s_Shkury_ips             | BSP_19XX_HowIHurtMyHand_nar       | SIP                    | SIP's legacy manuscript (notebook 22, p. 15) became available after the text was independently transcribed and glossed; it was still checked at the later stage   |
| SPB70S80S_VOJ          | spb70s80s_VVojnu_ips             | BSP_19XX_FormerlyAndNowadays_nar  | SIP                    |   |
| SPB90S_BAR             | spb90s_BaruchiSoz_tdnt           | BSP_1992_FreaksCape_flk           | TDNT                   |   |
| SPB90s_POR             | spb90s_Porne_od.                 | BSP_199X_Witch_flk                | DOE                    |   |
| VINB_NSB69_DIA         | vinb_nsb69_Dialog_ips            | BViN_BNS_1969_Conversation_conv   | SIP                    |   |



| Text code in Toolbox | Files name                      | Text code in corpus              | Audio recording source | Comment   |
|----------------------|---------------------------------|----------------------------------|------------------------|---|
| VINB090731_MYSH      | vinb090731_MyshkaOlenj          | BViN_2009_MouseAndReindeer_flk   | ELDP+                  |   |
| VNB900110_MES        | vnb900110_NazvanijaMesjacev_rad | BVN_1990_MonthNames_nar          | GTRK                   |   |
| VNB90S_KAKSVA        | vnb90s_KakSvatalis_dsb          | BVN_199X_Matchmaking2_nar        | BDS                    |   |
| VNB9111_UROD         | vnb9111_UrodlivijMys_rad        | BVN_1991_FreaksCape1_flk         | GTRK                   | Published as SB2005: text 22<br>Transcribed for glossing independently            |
| VNB92_BAR            | vnb92_BaruchiSoz_od             | BVN_1992_FreaksCape2_flk         | DOE                    | Published as L2002: Баручи соз, 104–108<br>Transcribed for glossing independently |
| VNB9306_KAKZHI       | vnb9306_KakZhili_rad            | BVN_1993_HowWeLived_nar          | GTRK                   |   |
| VNB930621_SZ         | vnb930621_SojuzZoloto_rad       | BVN_1993_UnionGold_nar           | GTRK                   |   |
| VNB940221_PRIR       | vnb940221_PrimetyVPriode_rad    | BVN_1994_FolkSigns1_nar          | GTRK                   |   |
| VNB950604_GRI        | vnb950604_GribyJagody_rad       | BVN_1995_MushroomsAndBerries_nar | GTRK                   |   |
| VNB950724_BRAT       | vnb950724_BratUrodec_rad        | BVN_1995_FreakBrother_flk        | GTRK                   |   |
| VNB960309_SVA        | vnb960309_Svatovstvo_rad        | BVN_1996_Matchmaking1_nar        | GTRK                   |   |
| VNB961119_PRIM       | vnb961119_Primety_rad           | BVN_1996_FolkSigns2_nar          | GTRK                   |   |
| XXX70S80S_DEV        | xxx70s80s_DevochkaVolk_ips      | NN5_19XX_LittleRedRidingHood_flk | SIP                    |   |

## Tundra Enets

Table 28. Correspondence of Tundra Enets texts in Toolbox databases and texts in the corpus

| Text code in Toolbox | Files name                            | Text code in corpus                      | Audio recording source | Comment |
|----------------------|---------------------------------------|--|------------------------|---------|
| APT_NNB90s00s_INT    | apt_nnb90s00s_Interview_tdnt          | TAP_BNN_199X_Interview_conv              | TDNT                   |         |
| DMT99_SKA            | dmt99_Skazka_rad                      | TDM_BNN_1999_WomanAndHerBrotherinlaw_flk | GTRK                   |         |
| ELS_NNB970514_INTW   | els_nnb970514_Interview_rad           | SEL_1997_Interview_conv                  | GTRK                   |         |
| GL_NNB990112_INTW    | gl_nnb990112_Interview_rad            | YGL_199X_Interview_conv                  | GTRK                   |         |
| IP_VP080715_IDO      | ip_vp080715_IdolKukla                 | KIP_NVP_2008_Idols_nar                   | ELDP+                  |         |
| IP_VP080715_PRKOR    | ip_vp080715_PrividenieKorjaga         | KIP_NVP_2008_GhostSnag_nar               | ELDP+                  |         |
| IP_VP080715_RAZREB   | ip_vp080715_Razgovor_RebjataNaRybalku | KIP_NVP_2008_PeopleWentFishing_conv      | ELDP+                  |         |
| IP_VP080715_ROZH     | ip_vp080715_RozhalaKolju              | KIP_NVP_2008_HowIBoreMySon_nar           | ELDP+                  |         |
| IP_VP080718_RAZKAK   | ip_vp080718_Razgovor_KakDomoj         | KIP_NVP_2008_GoingHome_conv              | ELDP+                  |         |
| IP_VP080718_RAZTUA   | ip_vp080718_Razgovor_Tualet           | KIP_NVP_2008_Toilet_conv                 | ELDP+                  |         |
| IP_VP080725_RAZRUG   | ip_vp080726_Razgovor_Rugajut          | KIP_NVP_2008_CriticizingLinguists_conv   | ELDP+                  |         |

| Text code in Toolbox | Files name                              | Text code in corpus                        | Audio recording source | Comment                     |
|----------------------|---|--|------------------------|-----------------------------|
| IP_VP080726_MED      | ip_vp080726_Medvedj                     | KIP_NVP_2008_Bear_nar                      | ELDP+                  | Published as (Khanina 2008) |
| IP_VP080726_PLA      | ip_vp080726_Plastilin                   | KIP_NVP_2008_ModellingClay_conv            | ELDP+                  |                             |
| IP_VP080726_RAZANA   | ip_vp080726_Razgovor_AnatPetr           | KIP_NVP_2008_WhoSpokeEnets_conv            | ELDP+                  |                             |
| IP_VP080726_RAZCHTO  | ip_vp080726_Razgovor_ChtoRasskazatj     | KIP_NVP_2008_HowThingsGo_conv              | ELDP+                  |                             |
| IP_VP080726_RAZDED   | ip_vp080726_Razgovor_DedLjagaj          | KIP_NVP_2008_LyagajOldMan_conv             | ELDP+                  |                             |
| IP_VP080726_RAZDVE   | ip_vp080726_Razgovor_DveNedeli          | KIP_NVP_2008_TwoWeeksHere_conv             | ELDP+                  |                             |
| IP_VP080726_RAZDYM   | ip_vp080726_Razgovor_DymPudidovna       | KIP_NVP_2008_Smoke_conv                    | ELDP+                  |                             |
| IP_VP080726_RAZMED   | ip_vp080726_Razgovor_Medvedj            | KIP_NVP_2008_Bears_conv                    | ELDP+                  |                             |
| IP_VP080726_RAZOLE   | ip_vp080726_Razgovor_OlesyalzVorontsova | KIP_NVP_2008_LyagajOldManInTheShop_conv    | ELDP+                  |                             |
| IP_VP080726_RAZSED   | ip_vp080726_Razgovor_SedojMagazin       | KIP_NVP_2008_Shop_conv                     | ELDP+                  |                             |
| IP_VP080726_VOL      | ip_vp080726_Volk                        | KIP_NVP_2008_Wolf_nar                      | ELDP+                  |                             |
| IP_VP100803_RAZRAZ   | ip_vp100803_Razgovor_Raznoe             | KIP_NVP_2010_CurrentLifeInTheFamily_conv   | ELDP+                  |                             |
| IP_VP100807_RAZSTA   | ip_vp100807_Razgovor_StaruxuVkljuchi    | KIP_NVP_2010_CurrentAffairs_conv           | ELDP+                  |                             |
| IP_VP100807_RAZTEB   | ip_vp100807_Razgovor_TebjaUvidela       | KIP_NVP_2010_IHaveSeenYou_conv             | ELDP+                  |                             |
| IP_VP100810_RAZHOR   | ip_vp100810_Razgovor_XoroshGovorit      | KIP_NVP_2010_HowDrunkPersonSpeaks_conv     | ELDP+                  |                             |
| IP_VP100810_RAZZRA   | ip_vp100810_Razgovor_ZrjaVstali         | KIP_NVP_2010_WeGotUpInVain_conv            | ELDP+                  |                             |
| IP080712_ZNA         | ip080714_BabkaZnakharka                 | KIP_NLP_2008_HealerWoman_nar               | ELDP+                  |                             |
| IP080714_KAM         | ip080714_Kamusa                         | KIP_2008_ProcessingKamuses_flk             | ELDP+                  |                             |
| IP080714_OCH         | ip080714_Ochischenie                    | KIP_2008_TraditionalCleansing_nar          | ELDP+                  |                             |
| IP080714_POTD        | ip080714_PoterjalasjDed                 | KIP_2008_HowIGotLostAndMyGrandfather_nar   | ELDP+                  |                             |
| IP080714_PRRU        | ip080714_PrividenieRuka                 | KIP_2008_GhostHand_nar                     | ELDP+                  |                             |
| IP080714_SHAM        | ip080714_TriShamanki                    | KIP_2008_ThreeShamanWomen_nar              | ELDP+                  |                             |
| IP080714IP_ZHA       | ip080714_ZhadnajaZhena                  | KIP_2008_GreedyWife_flk                    | ELDP+                  |                             |
| IP080716_BAB         | ip080716_Babka                          | KIP_2008_MyAunt_nar                        | ELDP+                  |                             |
| IP080716_EDA         | ip080716_EdaOleni                       | KIP_2008_TundraLife_nar                    | ELDP+                  |                             |
| IP080716_POTM        | ip080716_PoterjalasjMatj                | KIP_2008_HowIGotLost_nar                   | ELDP+                  |                             |
| IP080716_PROF1       | ip080716_ProfessorShamanskaja_1         | KIP_2008_ProfessorBuyingShamanClothes1_nar | ELDP+                  |                             |
| IP080716_PROF2       | ip080716_ProfessorShamanskaja_2         | KIP_2008_ProfessorBuyingShamanClothes2_nar | ELDP+                  |                             |
| IP080716_SEM         | ip080716_MojaSemja                      | KIP_2008_MyFamily_nar                      | ELDP+                  |                             |
| IP080716_SVA         | ip080716_SvadbaTroick                   | KIP_2008_JokeWedding_nar                   | ELDP+                  |                             |

| Text code in Toolbox | Files name                   | Text code in corpus                     | Audio recording source | Comment  |
|----------------------|------------------------------|---|------------------------|--|
| IP080716_ZAM         | ip080716_Zamuzh              | KIP_2008_MyMarriage_nar                 | ELDP+                  |  |
| IP080722_CHU         | ip080722_Chum                | KIP_2008_TraditionalTent_nar            | ELDP+                  |  |
| IP080722_RYB         | ip080722_Ryby                | KIP_2008_PregnancyAndRedFish_nar        | ELDP+                  | Published in (Khanina & Shluinsky 2023: 849–850) |
| IP080722_SHAJ        | ip080722_PrividenieShajtanka | KIP_2008_ShaytankaHillPhantom_nar       | ELDP+                  |  |
| IP080722_SIR         | ip080722_Sirota              | KIP_2008_OrphanGirl_flk                 | ELDP+                  |  |
| IP080722_UST         | ip080722_UstjAvam            | KIP_2008_VisitingUstAvam_nar            | ELDP+                  |  |
| MD090825_SHK         | md090825_Shkury              | BMD_2009_Skins_nar                      | ELDP+                  |  |
| MD090825_SOB         | md090825_Sobaki              | BMD_2009_Dogs_nar                       | ELDP+                  |  |
| MD090825_TAB         | md090825_Tabak               | BMD_2009_Tobacco_nar                    | ELDP+                  |  |
| MD090826_CHA         | md090826_ChajMuka            | BMD_2009_TeaFlatbread_nar               | ELDP+                  |  |
| MD090826_DET         | md090826_DetiVTundre         | BMD_2009_ChildrenInTundra_nar           | ELDP+                  |  |
| MD090826_GRI         | md090826_JagodyGriby         | BMD_2009_BerriesMushrooms_nar           | ELDP+                  |  |
| MD090826_JUK         | md090826_Jukola              | BMD_2009_DryFish_nar                    | ELDP+                  |  |
| MD090826_KAM         | md090826_Kamusa              | BMD_2009_Kamuses_nar                    | ELDP+                  |  |
| MD090826_MJA         | md090826_Mjaso               | BMD_2009_Meat_nar                       | ELDP+                  |  |
| MD090826_NAR         | md090826_NarvatjJagelj       | BMD_2009_CollectingMoss_nar             | ELDP+                  |  |
| MD090826_RYB         | md090826_Ryba                | BMD_2009_Fish_nar                       | ELDP+                  |  |
| MD090826_SVA         | md090826_Svadba              | BMD_2009_MarriageInTundra_nar           | ELDP+                  |  |
| PLP810508KIL         | plp810508_OZhizni_tdnt       | PLP_1981_MyLife_nar                     | TDNT                   |  |
| RD_NNB981221_BES     | rd_nnb981221_Beseda_rad      | SRD_199X_Interview_conv                 | GTRK                   |  |
| SD080808_DET         | sd080808_DetstvoSmertjOtca   | SSD_2008_MyFathersDeath_nar             | ELDP+                  |  |
| SD080808_OSEB        | sd080808_OSebe               | SSD_2008_AboutMyself_nar                | ELDP+                  |  |
| SD080808_OTE         | sd080808_OtecDed             | SSD_2008_MyFatherAndHisBrother_nar      | ELDP+                  |  |
| SD080808_ZAOL        | sd080808_ZaOlenjami          | SSD_2008_LostReindeer_nar               | ELDP+                  |  |
| SD080809_DOLE        | sd080809_DetiOleni           | SSD_2008_HowMySonsLookedForReindeer_nar | ELDP+                  |  |
| SD080809_PUJ         | sd080809_Pujaku              | SSD_2008_MyUnclePuyaku_nar              | ELDP+                  |  |
| SD080809_STA         | sd080809_StarshijBrat        | SSD_2008_MyElderBrother_nar             | ELDP+                  |  |
| SK090828_MOR         | sk090828_Moroshka            | TSK_2009_Fishing_nar                    | ELDP+                  |  |
| SK090828_OHO         | sk090828_KakXodiliNaOxotu    | TSK_2009_HowWeWentHunting_nar           | ELDP+                  |  |
| SK090828_OLE         | sk090828_OleniVGodu          | TSK_2009_ReindeerYearCycle_nar          | ELDP+                  |  |

| Text code in Toolbox | Files name                          | Text code in corpus               | Audio recording source | Comment |
|----------------------|-------------------------------------|-----------------------------------|------------------------|---------|
| SK090828_STA         | sk090828_StarikIdol                 | TSK_2009_OldMenAndIdol_nar        | ELDP+                  |         |
| SK090828_TAL         | sk090828_Taljniki                   | TSK_2009_KnifeForFirelighter_nar  | ELDP+                  |         |
| SK090828_VOR         | sk090828_VoroncovoDrova             | TSK_2009_FirewoodInVorontsovo_nar | ELDP+                  |         |
| SK090828_XOD         | sk090828_XodiliSTovarischem         | TSK_2009_WildReindeer_nar         | ELDP+                  |         |
| SUT090816_OZH        | sut090816_OZhizni                   | TSU_2009_MyLife_nar               | ELDP+                  |         |
| SUT090816_PER        | sut090816_KakPereshliizVorontsova   | TSU_2009_LeavingVorontsovo_nar    | ELDP+                  |         |
| SUT090817_BOK        | sut090817_KakShjutBokari            | TSU_2009_SewingBoots_nar          | ELDP+                  |         |
| SUT090817_LEP        | sut090817_LepeshkiVTundre           | TSU_2009_FlatbreadInTundra_nar    | ELDP+                  |         |
| SUT090817_NEV        | sut090817_KakRybachiliNevodom       | TSU_2009_FishingWithSeine_nar     | ELDP+                  |         |
| SUT090817_SHK        | sut090817_KakVShkole                | TSU_2009_SchoolYears_nar          | ELDP+                  |         |
| SUT090817_TAL        | sut090817_TalnikVVorontsovo         | TSU_2009_FirewoodInVorontsovo_nar | ELDP+                  |         |
| VA080723_KUR         | va080723_PochemuKuriti              | TVA_2008_HowIStartedSmoking_nar   | ELDP+                  |         |
| VA080723_MOR         | va080723_MoraChaj                   | TVA_2008_HerbTea_nar              | ELDP+                  |         |
| VA080723_SAM         | va080723_Samarskij                  | TVA_2008_MyFatherInKolkhoz_nar    | ELDP+                  |         |
| VA080723_SOV         | va080723_SovhozSevernoy             | TVA_2008_SovkhozSevernoy_nar      | ELDP+                  |         |
| VA080723_STA         | va080723_StarikXabaj                | TVA_2008_KhabajOldMan_nar         | ELDP+                  |         |
| VP080730_JUK         | vp080730_Jukola                     | NVP_2008_MakingDryFish_nar        | ELDP+                  |         |
| VP100810_LES         | vp100810_KakDelatjLeski             | NVP_2010_MakingFlatbread_nar      | ELDP+                  |         |
| ZA_VA080723_ZHU1     | za080723_Zhukovskij_1               | TZA_TVA_2008_Phantom1_nar         | ELDP+                  |         |
| ZA_VA080723_ZHU2     | za080723_Zhukovskij_2               | TZA_TVA_2008_Phantom2_nar         | ELDP+                  |         |
| ZA_ZN100810_OSEBE    | za_zn100810_OSebe                   | TZA_BZN_2010_MyLife_nar           | ELDP+                  |         |
| ZA_ZN100810_SKA0     | za_zn100810_SkazkaOSebe             | TZA_BZN_2010_Fool_flk             | ELDP+                  |         |
| ZA_ZN100811_PESOTL   | za_zn100811_Pesnja_OtLicaOtca       | TZA_BZN_2010_MyFathersSong1_song  | ELDP+                  |         |
| ZA080722LI_ELI       | za080723_EliMylo                    | TZA_2008_RussianWoman_nar         | ELDP+                  |         |
| ZA080723_DET         | za080723_Detstvo                    | TZA_2008_MyChildhood_nar          | ELDP+                  |         |
| ZA080723_SKA         | za080723_Skazka                     | TZA_2008_Yompu_flk                | ELDP+                  |         |
| ZA080723_ZHI         | za080723_Zhiznj                     | TZA_2008_MyLife_nar               | ELDP+                  |         |
| ZA100811_PESOTE      | za_zn100811_Pesnja_OtecEgoTovarisch | TZA_2010_MyFathersSong2_song      | ELDP+                  |         |

### A1.3. Note on further texts in the corpus

All other texts included into the corpus which are not mentioned in tables in Appendix A1 (A1.2) have never been published previously nor widely accessible to linguists in unpublished form. The only exception is the short text published in (Tereščenko 1966: 436–437), which is an extract from the text coming from Tereščenko’s archive included into the corpus as BIA\_1962\_HuntingWildReindeer\_nar.

### A1.4. Note on Enets texts not included into the corpus

#### Forest Enets

##### *Published texts*

Most published Forest Enets texts were included into the corpus.<sup>32</sup> The exceptions are as follows.

First and most significant, at this stage the corpus includes no published written texts composed and published by their authors, as well as no published translations from Russian (except for a single text that makes part of Sorokina & Bolina’s (2005) collection). Importantly, while such kinds of texts do not exist for Tundra Enets, such publications are available in Forest Enets, and the list below is not necessarily exhaustive. Zoya Nikolaevna Bolina published a collection of her selected journalistic texts (2014a), more of them having been published for many years in the local newspaper “Taymyr”<sup>33</sup>; further short written texts accompany Zoya Bolina’s picture dictionary (2012) and a collection of Ivan Ivanovich Silkin’s paintings (2014b). Dar`ya Spiridonovna Bolina included numerous written texts, original and translated, in her pedagogical textbooks (D. Bolina 2015; 2019; Bolina & Rosljakova 2015). A Forest Enets translation of the Gospel of Luke was also published by Dar`ya Bolina (1995), and a collection of Forest Enets translations of stories by a Soviet writer Evgenij Charushin was published by Galina Spiridonovna Bolina (Čarušin 2015).

Second, five spoken Forest Enets texts and two written Forest Enets text are published in (Siegl 2013: 462–492, 502–509). These texts are part of Florian Siegl’s collection, and for them source audio files are available (Siegl 2006–2008, see below). These texts should eventually be processed together with the media files.

Third, a very short Forest Enets text fragment recorded back in 1934 is published in (Prokof’ev 1937: 90). This text fragment was consciously not included into the corpus due to following reasons. On the one hand, its phonology is more archaic and represents a different variety of FE than all other accessible FE texts, and at the same time is too small to provide a relevant comparative background to other texts which make part of the corpus. On the other hand, it is assumed to be a part of a significantly larger Enets text collection from Prokof’ev’s archive (see below) and is to be potentially to be processed as its part.

---

<sup>32</sup> Only coherent texts are meant here, not isolated sentences.

<sup>33</sup> <https://taimyr24.ru/gazeta/>, last access 14.06.2024

Finally, Urmanchieva (Ms b) provided a manually glossed Forest Enets text recorded from Zoya Nikolaevna Bolina. The original audio file corresponding to this text has not been found so far, but is expected to be preserved somewhere and thus the text is postponed.

### ***Unpublished texts***

Significantly more unpublished transcribed Forest Enets texts exist, which were not included into this corpus.

First, an extensive further ELDP+ collection of transcribed Forest Enets texts was not included into the corpus only because limited time resources. The great majority of them were recorded in 2005–2016, but several texts are transcriptions of legacy recordings (see details in section 2.2). 172 texts are preliminarily transcribed in ELAN (several of these ELAN files are based on draft handwritten transcriptions of Zoya Bolina or Viktor Pal'chin), while 121 texts are transcribed so far only in a handwritten form by Zoya Bolina or Viktor Pal'chin. Even more Forest Enets ELDP+ recordings done in 2005–2016 unfortunately remain untranscribed so far.

Second, a collection of recordings of more than 50 Forest Enets texts recorded in 2006–2008 is deposited by Florian Siegl at The Language Archive (TLA)<sup>34</sup> (Siegl 2006–2008), several of them having been published in (Siegl 2013), as mentioned above. These texts were recorded from the same generation of speakers as ELDP+ collection (and the majority were recorded exactly from the same speakers). According to Siegl (2013: 462), even more transcribed texts are at his disposal, but are not archived by any institution. The same is true for his transcriptions of any texts other than those published in (Siegl 2013), as mentioned above.

Third, a collection of 44 handwritten paper sheets of Forest Enets texts written down in 1930s by Georgij N. Prokof'ev (1897–1942) is kept as a part of his and his wife Ekaterina Dmitrievna Prokof'eva's archive preserved at the Peter the Great Museum of Anthropology and Ethnography (the Kunstkamera) in St. Petersburg.<sup>35</sup> So far, the archive is not published, and the access to it is restricted.

Fourth, a collection of a dozen of Forest Enets texts (some of them are original, some are translated from Russian) in a paper manuscript form collected in late 1970s – early 1980s by Yaroslav Andreevich Gluxij (1941–2014) and Vasilij Andreevich Susekov (passed away in 2022) is a part of the archive of the Department of Siberian Indigenous Languages at the Tomsk State Pedagogical University. So far, the archive is not published, and the access to it is restricted.

Fifth, a collection of Forest Enets texts was recorded by Olga Kazakevich, Leonid Zakharov and their team in Potapovo in 2011, and for 6 of them draft transcriptions are available. These texts were recorded from the same speakers as ELDP+ collection.

Sixth, a number of Forest Enets texts were transcribed by Valentin Gusev (p.c.) in 1995–1996. Some of these transcriptions are available in electronic form in the collection of draft transcriptions of Enets texts from Gusev's personal archive. Most of these transcriptions were additional versions of FE texts included into the corpus based on versions coming from other sources, but one of them is an unfinished draft transcription of an FE text with no

---

<sup>34</sup> Which is an integral part of the Max Planck Institute for Psycholinguistics in Nijmegen, see <https://archive.mpi.nl/tla/>, last access 17.06.2024

<sup>35</sup> The archive inventory is accessible here: [https://www.kunstkamera.ru/files/Archive\\_KK/opisi/f\\_6\\_op\\_1\\_polevye\\_materialy\\_i\\_nauchnye\\_trudy.pdf](https://www.kunstkamera.ru/files/Archive_KK/opisi/f_6_op_1_polevye_materialy_i_nauchnye_trudy.pdf), last access 17.06.2024

source audio recording available. It remains unclear where the original manuscript versions are stored now, and it's possible that more transcribed FE texts are available there.

Finally, a number of Forest Enets audio and video recordings which are not transcribed, but are translated into Russian were done by Nikolay Pluzhnikov in 2022 in terms of “Cyrillica” center’s project supported by president grants foundation and are deposited at Youtube<sup>36</sup>.

### **Tundra Enets**

There are no published or widely accessible Tundra Enets texts of any kind that are not included into the corpus.

As mentioned in 1.2.3, one can distinguish between Vorontsovo / Tukhard TE (the variety mainly considered in all sources dealing with TE, including the present corpus) and Avam TE (the Nganasan-influenced variety used to be spoken in Avam tundra).

#### ***Vorontsovo / Tukhard TE***

Several Tundra Enets texts in a paper manuscript form collected in late 1970s – early 1980s by Gluxij and Susekov are a part of the archive of the chair of the languages of Siberia at the Tomsk State Pedagogical University (see above). So far, the archive is not published, and the access to it is restricted.

Several Tundra Enets texts were written down in Vorontsovo in 1970s by Eugene Helimski. Some of them are further retold versions of texts published in Russian in (Dolgix 1961), like *Solda* (Urmanchieva Ms a), but others are not. While original paper manuscripts form collected are stored in the personal archive of Anna Urmanchieva (p.c.), electronic versions of some or all of them are available in electronic form in the collection of draft transcriptions of Enets texts from Gusev’s personal archive. Unfortunately, the latter transcriptions became available to us too late to fit including them into the current version of the corpus.<sup>37</sup>

Several Tundra Enets recordings from Xolyu Nikolaevich Kaplin† done by Kazimir Isidorovich Labanauskas in 1970s in Vorontsovo were preliminarily transcribed by in mid-2000s Valentin Gusev and Ekaterina Subobteevna Kostyorkina†, a heritage speaker Tundra Enets whose main language was Nganasan. Unfortunately, since both the quality of available recordings is not perfect, and the available transcription and translation are rather imprecise, including these texts into the corpus was not considered reasonable.

Finally, some Vorontsovo / Tukhard TE ELDP+ recordings done in 2008–2010 by Khanina and Shluinsky unfortunately remain untranscribed so far; the same applies for a legacy Vorontsovo / Tukhard TE recording coming from GTRK archive.

#### ***Avam TE***

So far neither the present corpus nor any other source has properly analyzed Avam TE texts.

---

<sup>36</sup> [https://www.youtube.com/watch?v=wLox\\_2CEz-M](https://www.youtube.com/watch?v=wLox_2CEz-M), <https://www.youtube.com/watch?v=BC9YXVCbaTc>, <https://www.youtube.com/watch?v=qVuXigtup1U>, last access 25.07.2024

<sup>37</sup> While except for (Urmanchieva Ms a) none of these texts is edited and analyzed, according to Anna Urmanchieva (p.c.), lexical items and examples from these texts are assumed to be included in (Helimski Ms.).

For an Avam TE text recorded from Den`chude Nyujteevich Mirnyx† by Eugene Helimski in 1992 in Volochanka, a draft transcription was done in 2017 in ELAN by Olesya Khanina and Nina Dentumeevna Chunanchar.

Several Avam TE recordings from Kurumaku Fyodorovich Tuglakov† done by Kazimir Isidorovich Labanauskas in 1970s in Vorontsovo were preliminarily transcribed by in mid-2000s Valentin Gusev and Ekaterina Subobteevna Kostyorkina†.

There are more legacy Avam TE recordings available that unfortunately remain untranscribed so far. One of them was recorded from Kurumaku Fyodorovich Tuglakov in 2001 by Oksana Dobzhanskaya and comes from her personal archive. Another one was recorded from an unidentified male speaker in 1977 by Galina Grigor`evna Grigor`eva and comes from the archive of the Institute of Russian Literature (the Pushkin House, Russian: *Пушкинский дом*). More recordings from Kurumaku Fyodorovich Tuglakov done by Anna Urmanchieva and Aleksandr Pevnov in 2009 are preserved by Urmanchieva herself.



## Appendix A2. INEL transcription for Enets

Table 29. INEL Enets transcription characters

| INEL transcription | Description   | Unicode Character Name <sup>38</sup>                                |
|--------------------|---|---|
| <b>Vowels</b>      |   |   |
| a                  | open central unrounded vowel  | LATIN SMALL LETTER A (U+0061)                                       |
| e                  | close-mid front unrounded vowel   | LATIN SMALL LETTER E (U+0065)                                       |
| ɛ <sup>39</sup>    | open-mid front unrounded vowel  | LATIN SMALL LETTER OPEN E (U+025B)                                  |
| i                  | close front unrounded vowel   | LATIN SMALL LETTER I (U+0069)                                       |
| o                  | close-mid back rounded vowel  | LATIN SMALL LETTER O (U+006F)                                       |
| ɔ                  | open-mid back rounded long vowel  | LATIN SMALL LETTER OPEN O (U+0254)                                  |
| u                  | close back rounded vowel  | LATIN SMALL LETTER U (U+0075)                                       |
| <b>Consonants</b>  |   |   |
| b                  | voiced bilabial stop  | LATIN SMALL LETTER B (U+0062)                                       |
| č                  | voiceless alveolo-palatal or palatal affricate                          | LATIN SMALL LETTER C WITH CARON (U+010D)                            |
| d                  | voiced alveolar stop  | LATIN SMALL LETTER D (U+0064)                                       |
| d'                 | voiced palatalized alveolar stop or voiced palatal stop                 | LATIN SMALL LETTER D (U+0064) + MODIFIER LETTER APOSTROPHE (U+02BC) |
| g                  | voiced velar stop   | LATIN SMALL LETTER G (U+0067)                                       |
| j                  | palatal approximant   | LATIN SMALL LETTER J (U+006A)                                       |
| k                  | voiceless velar stop  | LATIN SMALL LETTER K (U+006B)                                       |
| l                  | alveolar lateral approximant  | LATIN SMALL LETTER L (U+006C)                                       |
| l'                 | alveolar palatalized lateral approximant or palatal lateral approximant | LATIN SMALL LETTER L (U+006C) + MODIFIER LETTER APOSTROPHE (U+02BC) |
| m                  | bilabial nasal  | LATIN SMALL LETTER M (U+006D)                                       |
| n                  | alveolar nasal  | LATIN SMALL LETTER N (U+006E)                                       |
| n'                 | palatalized alveolar nasal or palatal nasal                             | LATIN SMALL LETTER N (U+006E) + MODIFIER LETTER APOSTROPHE (U+02BC) |
| ŋ                  | velar nasal   | LATIN SMALL LETTER ENG (U+014B)                                     |
| p                  | voiceless bilabial stop   | LATIN SMALL LETTER P (U+0070)                                       |
| r                  | voiced alveolar trill   | LATIN SMALL LETTER R (U+0072)                                       |
| s                  | voiceless dental or alveolar fricative                                  | LATIN SMALL LETTER S (U+0073)                                       |

<sup>38</sup> Capital letters appearing sentence-initially and in proper names are not listed.

<sup>39</sup> Available as a phoneme in FE only, not in TE.

| INEL transcription | Description  | Unicode Character Name <sup>38</sup>     |
|--------------------|--|--|
| š                  | voiceless postalveolar fricative or voiceless palatalized alveolar fricative | LATIN SMALL LETTER S WITH CARON (U+0161) |
| t                  | voiceless alveolar stop  | LATIN SMALL LETTER T (U+0074)            |
| x                  | voiceless or voiced velar or uvular fricative                                | LATIN SMALL LETTER X (U+0078)            |
| z                  | voiced dental or alveolar fricative  | LATIN SMALL LETTER Z (U+007A)            |
| ʔ                  | glottal stop   | LATIN LETTER GLOTTAL STOP (U+0294)       |

Table 30. Phonemes which occur only in loanwords (from Russian or Tundra Nenets)

| INEL transcription | Description   | Unicode Character Name <sup>40</sup>     |
|--------------------|---|--|
| c                  | voiceless alveolar affricate  | LATIN SMALL LETTER C (U+0063)            |
| f                  | voiceless labiodental fricative   | LATIN SMALL LETTER F (U+0066)            |
| v                  | voiced labiodental fricative  | LATIN SMALL LETTER V (U+0076)            |
| w                  | labiovelar approximant  | LATIN SMALL LETTER W (U+0077)            |
| ž                  | voiced postalveolar fricative   | LATIN SMALL LETTER Z WITH CARON (U+017E) |
| '                  | indicates palatalization of Russian or Nenets consonants that have no phonemic palatalized counterparts in native Enets morphemes | MODIFIER LETTER APOSTROPHE (U+02BC)      |

As the transcription system of the INEL project has some differences in graphic representation of Enets phonemes with other existing transcriptions and orthographies, the following table illustrates possible correspondences with some of them.

Table 31. Comparative transcription / orthography chart

| INEL transcription | Main realizations (IPA) | Khanina & Shluinsky 2023 | Helinski Ms. | Mikola 1967 | Mikola 1980 | Pusztay 1978 | Siegl 2013 | Gluxij & Susekov 1982 | Tereščenko 1986 | Orthography 2019 <sup>41</sup> | Other Cyrillic writing systems |
|--------------------|-------------------------|--------------------------|--------------|-------------|-------------|--------------|------------|-----------------------|-----------------|--------------------------------|--------------------------------|
| a                  | a                       | a                        | a            | a           | a           | a            | a          | a                     | а, я*           | а, я*                          | а, я*                          |
| b                  | b, b <sup>j</sup> , p   | b                        | b            | b, б        | b           | b            | b, б       | б, б'                 | б               | б                              | б                              |
| č                  | tʃ ~ c ~ t'             | tʃ                       | č            | č, č        | č           | č            | č          | ч'                    | ч**             | ч**                            | ч**; Тб*                       |
| d                  | d, t                    | d                        | d            | d           | d           | d            | d          | д                     | д               | д                              | д                              |

<sup>40</sup> Capital letters appearing sentence-initially and in proper names are not listed.

<sup>41</sup> Established as a result of a week-long seminar in Dudinka in October 2019 held by Dar'ya Spiridonovna Bolina, Nina Nikolaevna Bolina, Zoya Nikolaevna Bolina, Ekaterina Spiridonovna Glibchenko, Anna Ivanovna Saktobergenova, Andrey Shluinsky and Evgeniya Nikolaevna Stoly'pina.

| INEL transcription | Main realizations (IPA)                | Khanina & Shluinsky 2023 | Helinski Ms.       | Mikola 1967 | Mikola 1980 | Pusztay 1978 | Siegl 2013 | Gluxij & Susekov 1982 | Tereščenko 1986     | Orthography 2019 <sup>41</sup> | Other Cyrillic writing systems |
|--------------------|--|--------------------------|--------------------|-------------|-------------|--------------|------------|-----------------------|---------------------|--------------------------------|--------------------------------|
| d'                 | d <sup>j</sup> ~ ʃ                     | d <sup>j</sup>           | d                  | d           | d           | d'           | d'         | Д'                    | Дб <sup>*42</sup>   | Дб*                            | Дб*                            |
| e                  | e                                      | e                        | e, diphthongs      | e           | e, ɛ        | e, ɛ         | e          | е, ɛ                  | э, е*               | э, е*, ы, и <sup>43</sup>      | э, е; ы, и*                    |
| ɛ                  | ɛ                                      | ɛ                        | e <sup>44</sup>    | ɛ           | ɛ           | e, a, ä      | ä          | е                     | ɛ                   | ɛ, я                           | ɛ; я; э; ё                     |
| g                  | g, g <sup>j</sup> , k                  | g                        | g                  | g           | g           | g            | g          | г                     | г                   | г                              | г                              |
| i                  | i                                      | i                        | i                  | i, i̯       | i, i̯       | i, i̯        | i          | и, ы                  | и, ы <sup>*45</sup> | и, ы*                          | и, ы*                          |
| j                  | j                                      | j                        | j                  | j           | j           | j            | j          | й, й̆                 | й*                  | й*                             | й*                             |
| k                  | k, k <sup>j</sup>                      | k                        | k                  | k, k̆       | K           | k            | k          | к, к'                 | к                   | к                              | к                              |
| l                  | l                                      | l                        | l                  | l           | l           | l, ɫ         | l          | л                     | л                   | л                              | л                              |
| l'                 | l <sup>j</sup> ~ ʎ                     | l <sup>j</sup>           | l̆                 | l̆          | l̆          | l'           | l'         | л'                    | ль*                 | ль*                            | ль*                            |
| m                  | m, m <sup>j</sup>                      | m                        | m                  | m, m̆       | m           | m, m'        | m          | м, м'                 | м                   | м                              | м                              |
| n                  | n                                      | n                        | n                  | n           | n           | n            | n          | н                     | н                   | н                              | н                              |
| n'                 | n <sup>j</sup> ~ ɲ                     | n <sup>j</sup>           | ñ                  | ñ           | ñ           | n'           | ñ          | н'                    | нь*                 | нь*                            | нь*                            |
| ŋ                  | ŋ, ŋ <sup>j</sup>                      | ŋ                        | ŋ                  | ŋ           | ŋ           | ŋ            | ŋ          | Н̆                    | Н̆                  | Н̆                             | Н̆, Н̆, Н̆, Н̆ <sup>46</sup>   |
| o                  | o                                      | o                        | ô, o <sup>47</sup> | o           | Q           | o            | o          | Q                     | ó                   | ô, ё*, у, ю                    | о, ё; е; у, ю <sup>*48</sup>   |
| ɔ                  | ɔ                                      | ɔ                        | o                  | o           | o           | o            | o          | o                     | o, ё*               | o, ё*                          | o, ё; е*                       |
| p                  | p, p <sup>j</sup>                      | p                        | p                  | p, p̆       | p           | p, p'        | p          | п, п'                 | п                   | п                              | п                              |
| r                  | r, r <sup>j</sup>                      | r                        | r                  | r, r̆       | r           | r, r'        | r, r̆      | р, р'                 | р                   | р                              | р                              |
| s                  | θ ~ s, θ <sup>j</sup> ~ s <sup>j</sup> | s                        | s                  | s           | s           | s, s'        | s          | θ                     | с                   | с                              | с                              |
| š                  | ʃ, ʃ <sup>j</sup>                      | ʃ                        | ś                  | ś           | š           | š            | š          | ш, ш'                 | ς**                 | ш, щ**                         | ς; ш; щ**                      |
| t                  | t                                      | t                        | t                  | t           | t           | t            | t          | т                     | т                   | т                              | т                              |
| u                  | u                                      | u                        | u                  | u           | u           | u, w         | u          | у                     | у, ю*               | у, ю*                          | у, ю*; в                       |
| x                  | x, x <sup>j</sup> ~ χ ~ ɣ              | x                        | x                  | h           | h           | x            | x          | х, γ                  | х                   | х                              | х                              |
| z                  | ð ~ z, ð <sup>j</sup> ~ z <sup>j</sup> | z                        | ð                  | z           | ð           | ʒ            | đ          | џ                     | ʒ                   | з                              | з                              |
| ʔ                  | ʔ                                      | ʔ                        | ʔ                  | '           | '           | ʔ            | ʔ          | ʔ                     | ' , "49             | "                              | ' , "                          |

\* All Cyrillic writing systems (except for (Gluxij & Susekov 1982) and other papers by Gluxij and/or Susekov) involve the following conventions (mainly typical for Cyrillic orthographies):

<sup>42</sup> Tereščenko (1986) remains implicit if d', l', n' and š are analyzed as separate phonemes.

<sup>43</sup> Orthographical conventions from 2019 include that single close-mid and close vowels are not distinguished in non-first syllables and the symbols for close vowels are always used in this position.

<sup>44</sup> A mentioned, the phoneme ɛ is absent in TE; for FE, Helinski (Ms.) uses an archaic representation where e usually corresponds to modern standard ɛ and diphthongs correspond to modern standard e.

<sup>45</sup> Tereščenko's (1986) description implies these are two distinct phonemes.

<sup>46</sup> The variation of technically different Cyrillic symbols for the velar ŋ that appear in different sources has no significance, since all of them are perceived as graphic variations of the same grapheme.

<sup>47</sup> Helinski (Ms) always uses o in non-first syllables where the distinction between o and ɔ is neutralized.

<sup>48</sup> The symbol e sporadically appears for ɔ and o as a result of influence of the Russian orthography where the distinction between letters e and ё is not obligatory.

<sup>49</sup> Tereščenko (1986) proposed two different symbols for two morphophonemic glottal stops (following the conventions of Tundra Nenets orthography). These two symbols are used less systematically in Enets Cyrillic writing systems.

- *e, ě, ю, я* are used word-initially for sequences of *j + a, o* (and *ɔ*), *u, a* and after the palatalized consonants *d', l', n'*
- *u* is used word-initially and after the palatalized consonants *d', l', n'*, while after other consonants *u* or *ы* can be used<sup>50</sup>
- *ŭ* is used for *j* word-finally or before a consonant
- *Ѧ, Ѧь, Ѧь* are used word-finally or before a consonants (including before a word-final glottal stop)

\*\* All Cyrillic writing systems (except for the one used in Leisiö and Gusev's draft transcription) sporadically or systematically use the redundant symbol *Ѧ* after the symbols *ч, щ* and *ç* at the end of a word.

As mentioned, except for (Urmanchieva Ms a) and Helimski (1981/2000), which follow Helimski's (Ms) phonological transcription, all other secondary sources used in the corpus are not phonologically consistent (see 3.2.4 for more details). Because of this, significantly more correspondences than presented in the table are possible at the level of specific phonemes. Since Mikola (1967; 1980), Pusztay (1978) and Gluxij & Susekov (1982) aim at a detailed phonetic, not phonological, transcription, other symbols than presented in the table marginally appear in their publications.

---

<sup>50</sup> Systematized in the orthographical conventions from 2019, inconsistent in preceding writing systems.

### Appendix A3. Transliteration

All personal names and most placenames in respective metadata fields are transliterated from Cyrillic into Latin alphabet following the transliteration standard GOST 7.79 System B (published as GOST 2001).

Elsewhere, e.g. in text titles, English glosses (**ge** tier) and free translations (**fe** tier), English-style romanization is used.

*Table 32. INEL transliteration of Russian letters (GOST 7.79 System B)*

| Cyrillic |   | Roman |     | Notes   |
|----------|---|-------|-----|---|
| А        | а | A     | a   |   |
| Б        | б | B     | b   |   |
| В        | в | V     | v   |   |
| Г        | г | G     | g   |   |
| Д        | д | D     | d   |   |
| Е        | е | E     | e   |   |
| Ё        | ё | Yo    | yo  |   |
| Ж        | ж | Zh    | zh  |   |
| З        | з | Z     | z   |   |
| И        | и | I     | i   |   |
| Й        | й | J     | j   |   |
| К        | к | K     | k   |   |
| Л        | л | L     | l   |   |
| М        | м | M     | m   |   |
| Н        | н | N     | n   |   |
| О        | о | O     | o   |   |
| П        | п | P     | p   |   |
| Р        | р | R     | r   |   |
| С        | с | S     | s   |   |
| Т        | т | T     | t   |   |
| У        | у | U     | u   |   |
| Ф        | ф | F     | f   |   |
| Х        | х | X     | x   |   |
| Ц        | ц | C     | c   | INEL uses c everywhere (cz is recommended in GOST when not before i, e, y, j) |
| Ч        | ч | Ch    | ch  |   |
| Ш        | ш | Sh    | sh  |   |
| Щ        | щ | Shh   | shh |   |
| Ъ        | ъ | ``    | ``  | double grave accent   |
| Ы        | ы | Y`    | y`  |   |
| Ь        | ь | `     | `   | grave accent  |
| Э        | э | E`    | e`  |   |
| Ю        | ю | Yu    | yu  |   |
| Я        | я | Ya    | ya  |   |

Table 33. INEL Russian to English romanization

| Cyrillic |     | English |          | Note   |
|----------|-----|---------|----------|--|
| А        | а   | A       | a        |  |
| Б        | б   | B       | b        |  |
| В        | в   | V       | v        |  |
| Г        | г   | G       | g        |  |
| Д        | д   | D       | d        |  |
| Е        | е   | Ye / e  | ye / e   | ye word-initially, after a vowel or ъ; e elsewhere<br>E.g. Енисей > Yenisey, Большое > Bolshoye        |
| Ё        | ё   | Yo      | yo       |  |
| Ж        | ж   | Zh      | zh       |  |
| З        | з   | Z       | z        |  |
| И        | и   | I       | i        |  |
|          | -ий |         | -y / -iy | y in Russian endings, e.g. Самарский > Samarsky<br>iy in Enets or Nenets names, e.g. Волебий > Volebiy |
| Й        | й   | Y       | y        |  |
| К        | к   | K       | k        |  |
| Л        | л   | L       | l        |  |
| М        | м   | M       | m        |  |
| Н        | н   | N       | n        |  |
| О        | о   | O       | o        |  |
| П        | п   | P       | p        |  |
| Р        | р   | R       | r        |  |
| С        | с   | S       | s        |  |
| Т        | т   | T       | t        |  |
| У        | у   | U       | u        |  |
| Ф        | ф   | F       | f        |  |
| Х        | х   | Kh      | kh       |  |
| Ц        | ц   | Ts      | ts       |  |
| Ч        | ч   | Ch      | ch       |  |
| Ш        | ш   | Sh      | sh       |  |
| Щ        | щ   | Shch    | shch     |  |
| –        | ъ   | –       | omitted  |  |
| Ы        | ы   | Y       | y        |  |
| –        | -ый | –       | -y / -iy | y in Russian endings, e.g. Малый > Maly<br>iy in Enets or Nenets names, e.g. Ласый > Lasiy             |
| –        | ь   | –       | omitted  | E.g. Татьяна > Tatyana, Ванька > Vanka   |
| –        | -ье | –       | -ye      | E.g. Афанасьевич > Afanasyevich  |
| –        | -ьо | –       | -yo      | E.g. Бульон > Bulyon   |
| Э        | э   | E       | e        |  |
| Ю        | ю   | Yu      | yu       |  |
| Я        | я   | Ya      | ya       |  |

#### Appendix A4. Enets speakers

Appendix A4 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 and those whose name we still do not know.

Codes used in the Toolbox databases and draft ELAN transcriptions are also included into this table. Maiden names are mentioned in parentheses for married women, as well as widely used alternate names. More detailed information about the speakers is provided in speaker metadata.

In this Appendix, transliteration as in the metadata is used (see 2.5.6).

*Table 34. Forest Enets speakers*

| Name                                  | Main residence place <sup>51</sup>  | Code in the corpus | Code in Toolbox | Comments  |
|---------------------------------------|-------------------------------------|--------------------|-----------------|---|
| Ashlyapkina, Evdokiya Alekseevna      | Voroncovo                           | AEA                | aea             |   |
| Ashlyapkina, Nina Il'ichna            | Potapovo                            | ANI                | -               | The speaker's patronymic is not provided in (Sorokina & Bolina 2005), but is reported by Dar'ya S. Bolina |
| Bolin, Aleksandr Spiridonovich        | Potapovo                            | BAS                | as              |   |
| Bolin, Innokentij Egorovich           | Potapovo                            | BIE                | ie              |   |
| Bolin, Ivan Andreevich                | Potapovo                            | BIA                | -               |   |
| Bolin, Kupriyan Andreevich            | Potapovo                            | BKA                | kab             |   |
| Bolin, Leonid Dmitrievich             | Potapovo                            | BLD                | ld              |   |
| Bolin, Nikolaj Dmitrievich            | Dudinka                             | BND                | -               |   |
| Bolin, Petr Nikolaevich               | Dudinka                             | BPN                | pn              |   |
| Bolin, Petr Spiridonovich             | Malaya Xeta                         | BPS                | -               |   |
| Bolin, Spiridon Petrovich             | Potapovo                            | BSP                | spb             |   |
| Bolin, Viktor (Vitalij) Nikolaevich   | Potapovo, Dudinka                   | BViN               | vinb            |   |
| Bolin, Yuriy Spiridonovich            | Potapovo                            | BYuS               | -               |   |
| Bolina (Silkina), Antonina Puyakovna  | Potapovo                            | BAP                | ap              | The first language of this speaker was Tundra Enets   |
| Bolina, Dar'ya Spiridonovna           | Dudinka, St. Petersburg / Leningrad | BDS                | ds              |   |
| Bolina (Yamkina), Elizaveta Danilovna | Potapovo                            | BED                | edb             |   |
| Bolina (Silkina), Elizaveta Ivanovna  | Potapovo                            | BEI                | eib             |   |
| Bolina, Galina Spiridonovna           | Dudinka, St. Petersburg / Leningrad | BGS                | -               |   |
| Bolina, Lyubov` Kupriyanovna          | Dudinka                             | BLK                | lku             |   |
| Bolina, Mariya Andreevna              | Potapovo                            | BMA                | -               |   |

<sup>51</sup> Settlements names also refer to tundra places nearby.

| Name  | Main residence place <sup>51</sup> | Code in the corpus | Code in Toolbox | Comments   |
|---|------------------------------------|--------------------|-----------------|--|
| Bolina, Mariya Nikolaevna                   | Dudinka                            | BMND               | mn              |  |
| Bolina (Pal`china), Mariya Nikolaevna       | Karaul                             | BMN                | mnb             |  |
| Bolina (Ivanova), Nadezhda Konstantinovna   | Potapovo                           | BNK                | nk              |  |
| Bolina, Nina Nikolaevna                     | Dudinka                            | BNN                | nnb             |  |
| Bolina, Tat`yana Nikolaevna                 | Dudinka                            | BTN                | tn              |  |
| Bolina (Silkina), Vera Nikolaevna           | Potapovo                           | BVN                | vnb             |  |
| Bolina, Zoya Nikolaevna                     | Dudinka                            | BZN                | zn              |  |
| Borisova (Bolina), Nina Kupriyanovna        | Sakha; Dudinka                     | BNKu               | nku             |  |
| Chardu (Bolina), Nina Petrovna              | Potapovo                           | ChNP               | npch            |  |
| Chirimkina (Silkina), Nelli Alekseevna      | Norilsk                            | ChNA               | na              |  |
| Glibchenko (Bolina), Ekaterina Spiridonovna | Potapovo                           | GES                | es              |  |
| Ivanov, Gennadij Afanas`evich               | Potapovo                           | IGA                | ga              |  |
| Ivanov, Konstantin Vasil`evich              | Potapovo                           | IKV                | -               |  |
| Ly`rmin, Nikolaj Dmitrievich                | Potapovo                           | LND                | nd              |  |
| Ly`rmin, Vasilij Fedorovich                 | Potapovo                           | LVF                | -               | The first lanugage of this speaker was Tundra Nenets   |
| Muravskaya (Bolina), Natal`ya Spiridonovna  | Dudinka                            | BNS                | nsb             |  |
| Neznajkina (Yamkina), Galina Kirillovna     | Potapovo                           | NGK                | gk              |  |
| Pal`chin, Aleksey Sergeevich                | Potapovo                           | PAS                | asp             |  |
| Pal`chin, Anatolij (Anton) Nikolaevich      | Potapovo                           | PAN                | an              |  |
| Pal`chin, Nikolaj Sergeevich                | Potapovo                           | PNS                | nsp             |  |
| Pal`chin, Viktor (Vitalij) Nikolaevich      | Potapovo, Dudinka                  | PVN                | vn              |  |
| Pal`chin, Yakov Nikolaevich                 | Potapovo                           | PYN                | -               |  |
| Pal`china, Lidiya Nikolaevna                | Nosok                              | PLN                | -               | In (Sorokina & Bolina 2005), by mistake "Bolina, Lidiya Nikolaevna"; the correct information is reported by Dar`ya S. Bolina |
| Roslyakova (Silkina), Svetlana Alekseevna   | Potapovo                           | RSA                | sa              |  |
| Shhepeleva (Bolina), Lyubov` Nikolaevna     | Dudinka                            | ShLN               | ln              |  |
| Silkin, Anatolij Maksimovich                | Potapovo                           | SAM                | am              |  |
| Silkin, Daniil Alekseevich                  | Potapovo                           | SDnA               | da              |  |
| Silkin, Ivan Ivanovich                      | Potapovo                           | SIIP               | ii              | There is a TE speaker with an identical name (SII)   |
| Silkin, Nikolaj Ivanovich                   | Potapovo                           | SNI                | ni              |  |
| Silkina (Bolina), Alevtina Spiridonovna     | Potapovo                           | SAS                | ass             |  |
| Silkina (Bolina) Marija Nikolaevna          | Potapovo                           | SMN                | mns             |  |
| Silkina, Valentina Maksimovna               | Potapovo                           | SVM                | vms             |  |
| Stoly`pina (Silkina), Ekaterina Nikolaevna  | Potapovo                           | SEN                | -               |  |
| Suslova (Silkina), Dar`ya Alekseevna        | Potapovo, (?) Nosok                | SDA                | -               |  |
| Tret`yakova (Bolina), Zoya Kupriyanovna     | Xatanga                            | TZK                | zku             |  |



| Name                                       | Main residence place <sup>51</sup> | Code in the corpus | Code in Toolbox | Comments |
|--|------------------------------------|--------------------|-----------------|----------|
| Tuglakova (Ashlyapkina), Mariya Alekseevna | Voroncovo                          | TMA                | mat             |          |
| Yamkin, Kirill Danilovich                  | Potapovo                           | YKD                | kdja            |          |
| unidentified speaker                       |                                    | NN1, NN2...        | xxx             |          |

Table 35. *Tundra Enets speakers*

| Name                                     | Main residence place | Code in the corpus | Code in Toolbox | Comments   |
|--|----------------------|--------------------|-----------------|--|
| Beregovaya (Silkina), Marina Dyogolevna  | Tuxard               | BMD                | md              |  |
| Kaplin, Dyojga Xol'evich                 | Voroncovo            | KDX                | -               |  |
| Koshkaryova (Tuglakova), Irina Pajkovna  | Voroncovo            | KIP                | ip              |  |
| Kaplin, Tediku Nikolaevich               | Voroncovo            | KTN                | -               |  |
| Kaplin, Xolyu Nikolaevich                | Voroncovo            | KXN                | -               |  |
| Novosyolova (Pil'ko), Lyudmila Pudidovna | Voroncovo            | NLP                | lp              |  |
| Nade`r (Silkina), Valentina Puyakovna    | Voroncovo            | NVP                | vp              |  |
| Pil'ko, Nouko Semenovich                 | Voroncovo            | PNS                | -               |  |
| Pil'ko, Pudido Lil'evich                 | Voroncovo            | PLP                | plp             |  |
| Silkina (Pil'ko), Ekaterina Lil'evna     | Tuxard               | SEL                | els             |  |
| Silkin, Ivan Ivanovich                   | Voroncovo            | SII                | -               |  |
| Silkin, Neechi Stepanovich               | Voroncovo            | SNS                | -               |  |
| Silkin, Roman Dyogolevich                | Tuxard               | SRD                | rd              |  |
| Silkin, Semyon Dyogolevich               | Tuxard               | SSD                | sd              |  |
| Silkina, O.I.                            | Voroncovo            | SOI                |                 | No further metadata are provided in (Sorokina & Bolina 2005)   |
| Tuglakov, Dyojga Maksimovich             | Voroncovo            | TDM                | dmt             |  |
| Tuglakov, Kaso Tanulovich                | Tuxard               | TKT                | ktt             |  |
| Tuglakov, Nikolaj Maksimovich            | Voroncovo            | TNM                | -               |  |
| Tuglakov, Sergej Kasovich                | Tuxard               | TSK                | sk              |  |
| Tuglakova, Antonina Puyalevna            | Norilsk              | TAP                | apt             |  |
| Tuglakova (Yando), Serne Urtomovna       | Tuxard               | TSU                | sut             | The first language of this speaker was Tundra Nenets   |
| Turutin, T.P.                            | Voroncovo            | TTP                | -               | No further metadata are provided in (Labanauskas 1992)   |
| Turutin, Vladimir Ayakovich              | Karepovsk            | TVA                | va              |  |
| Ve`ngo (Turutina), Zoya Ayakovna         | Karepovsk, Voroncovo | TZA                | za              |  |
| Yar (Pil'ko), Guli Lil'evna              | Voroncovo            | YGL                | gl              |  |
| Yar, Tat'yana (Lyudmila) Chanovna        | Voroncovo            | YTCh               | tc              | No texts from this speaker are included into the corpus; she contributed extensively to oral transcription |



## Appendix A5. Morpheme glossing labels (tiers **ge**, **gr**) and Tsakorpus grammar tags

The following tables list the glossing labels used in tiers **ge**, **gr** and corresponding grammar tags for use in Tsakorpus online search. Table 36 gives the listing by category while Table 37 is sorted alphabetically, without division into categories.

Table 36. List of morpheme glossing labels by category

| Gloss                     | Description       | Tsakorpus grammar tags | Tsakorpus category             | Comment  |
|---------------------------|-------------------|------------------------|--------------------------------|--|
| <b>Person and number</b>  |                   |                        |                                |  |
| 1SG                       | 1 person singular | pn1, pns               | persnum-person, persnum-number | only in combinations with personal pronouns, e.g. "PRO.1SG.NOM", in verbal subject cross-reference markers, e.g. "1SG.S", and in possessive markers, e.g. "NOM.SG.1SG" |
| 1DU                       | 1 person dual     | pn1, pdu               | persnum-person, persnum-number |  |
| 1PL                       | 1 person plural   | pn1, pnpl              | persnum-person, persnum-number |  |
| 2SG                       | 2 person singular | pn2, pns               | persnum-person, persnum-number |  |
| 2DU                       | 2 person dual     | pn2, pdu               | persnum-person, persnum-number |  |
| 2PL                       | 2 person plural   | pn2, pnpl              | persnum-person, persnum-number |  |
| 3SG                       | 3 person singular | pn3, pns               | persnum-person, persnum-number |  |
| 3DU                       | 3 person dual     | pn3, pdu               | persnum-person, persnum-number |  |
| 3PL                       | 3 person plural   | pn3, pnpl              | persnum-person, persnum-number |  |
| <b>Personal pronouns</b>  |                   |                        |                                |  |
| PRO                       | personal pronoun  | pers                   |                                | lexical, only in combinations with person and number and case, e.g. "PRO.1SG.NOM"  |
| <b>Nominal categories</b> |                   |                        |                                |  |
| <b>Number</b>             |                   |                        |                                |  |
| SG                        | singular number   | sg                     | n-num                          | only in combinations with case, e.g. "LAT.SG"  |
| DU                        | dual number       | du                     | n-num                          |  |
| PL                        | plural number     | pl                     | n-num                          |  |
| <b>Case</b>               |                   |                        |                                |  |
| ABL                       | ablative case     | abl                    | n-case                         | only in combinations with number or w, e.g. "ABL.SG" or with personal pronouns, e.g. "PRO.1SG.ABL"   |
| ACC                       | accusative case   | acc                    | n-case                         | only in combinations with personal pronouns, e.g. "PRO.1PL.ACC"  |

| Gloss                           | Description                    | Tsakorpus grammar tags | Tsakorpus category    | Comment  |
|---------------------------------|--------------------------------|------------------------|-----------------------|--|
| GEN                             | genitive case                  | gen                    | n-case                | only in combinations with personal pronouns, e.g. "PRO.1PL.GEN"  |
| LAT                             | lative case                    | lat                    | n-case                | only in combinations with number, e.g. "LAT.SG" or with personal pronouns, e.g. "PRO.1SG.LAT"                          |
| LAT.SG2                         | lative case, singular number 2 | lat, sg, latsg2        | n-case, n-num, n-case |  |
| LOC                             | locative case                  | loc                    | n-case                | only in combinations with number, e.g. "LOC.SG" or with personal pronouns, e.g. "PRO.1SG.LOC"                          |
| NOM                             | nominative case                | nom                    | n-case                | only in combinations with number and other categories, e.g. "NOM.SG.3DU" or with personal pronouns, e.g. "PRO.1SG.NOM" |
| OBL                             | oblique case                   | obl                    | n-case                | only in combinations with number, e.g. "OBL.SG.1SG"  |
| PROL                            | prolative case                 | prol                   | n-case                | only in combinations with number, e.g. "PROL.SG" or with personal pronouns, e.g. "PRO.1SG.PROL"                        |
| <b>Case of adverbials</b>       |                                |                        |                       |  |
| ABL.ADV                         | ablative case of adverbials    | abl, advcase           | n-case, n-misc        |  |
| LAT.ADV                         | lative case of adverbials      | lat, advcase           | n-case, n-misc        |  |
| LOC.ADV                         | locative case of adverbials    | loc, advcase           | n-case, n-misc        |  |
| NOM.ADV                         | nominative case of adverbials  | nom, advcase           | n-case, n-misc        |  |
| <b>Possession</b>               |                                |                        |                       |  |
|                                 | possessive                     | poss                   | n-poss                | only as a tag; is added to combinations of case and person-number markers, e.g. "OBL.SG.3DU"                           |
| <b>Other nominal categories</b> |                                |                        |                       |  |
| CAR.ADV                         | caritive adverbializer         | car, caradv            | deriv-misc, n-misc    |  |
| CMP                             | comparative                    | cmp                    | n-misc                |  |
| DST                             | destinative form               | dst                    | n-misc                | only in combinations with number, e.g. "DST.SG"  |
| ESS                             | essive case                    | ess                    | n-misc                |  |
| ESS2                            | essive case 2                  | ess2                   | n-misc                |  |
| ESS3                            | essive case 3                  | ess3                   | n-misc                |  |
| %ESS2                           | possibly an essive case marker | ess2, uncert           | n-misc, misc          |  |
| KPL                             | kinship plural                 | kpl                    | n-misc                |  |
|                                 | similative (any)               | simil                  | deriv-misc            | only as a tag  |
| SIMIL                           | similative (nominal)           | simil, similn          | deriv-misc, n-misc    |  |
| SOC                             | sociative                      | soc                    | n-misc                |  |

| Gloss                         | Description   | Tsakorpus<br>grammar tags | Tsakorpus category                        | Comment  |
|-------------------------------|---|---------------------------|---|--|
| SOC2                          | sociative 2   | soc2                      | n-misc                                    |  |
| VOC                           | vocative form   | voc                       | n-misc                                    |  |
| <b>Verbal categories</b>      |   |                           |   |  |
| <b>Conjugation type</b>       |   |                           |   |  |
| S                             | subjective conjugation                                  | subj                      | persnum-conj                              | only in combinations with person and number, e.g. "3DU.S"                          |
| SG.O                          | objective conjugation, singular object                  | objc, objsg               | persnum-conj, object-number               | only in combinations with person and number, e.g. "2SG.SG.O"                       |
| DU.O                          | objective conjugation, dual object                      | objc, objdu               | persnum-conj, object-number               |  |
| NSG.O                         | objective conjugation, nonsingular object               | objc, objnsg              | persnum-conj, object-number               | only in combinations with person and number, e.g. "3DU.NSG.O"                      |
| PL.O                          | objective conjugation, plural object                    | objc, objpl               | persnum-conj, object-number               |  |
| S/SG.O                        | subjective conj. or objective conj., nonsingular object | subj, objc, objsg         | persnum-conj, persnum-conj, object-number | only in combinations with person and number, e.g. "2DU.S/SG.O"                     |
| MD                            | middle conjugation                                      | md                        | persnum-conj                              |  |
| <b>Cross-reference series</b> |   |                           |   |  |
| PRM                           | presumptive cross-reference series                      | prm                       | persnum-series                            | only in combinations with person and number and conjugation type, e.g. "3DU.S.PRM" |
| PST                           | past tense cross-reference series                       | pst                       | persnum-series                            | only in combinations with person and number and conjugation type, e.g. "3DU.S.PST" |
| <b>TAM categories</b>         |   |                           |   |  |
| AUD                           | auditive  | aud                       | v-tam                                     |  |
| FUT                           | future tense  | fut                       | v-tam                                     |  |
| HAB                           | habitual  | hab                       | v-tam                                     |  |
| HAB2                          | habitual 2  | hab2                      | v-tam                                     |  |
|                               | hortative (any)   | hort                      | v-tam                                     | only as a tag  |
| HORT1                         | hortative 1   | hort, hort1               | v-tam, v-tam                              |  |
| HORT2                         | hortative 2   | hort, hort2               | v-tam, v-tam                              |  |
| HYPOT                         | hypothetical  | hypot                     | v-tam                                     |  |
| IMP                           | imperative  | imp                       | v-tam                                     | only in combinations with person and number, e.g. "IMP.2SG.S"                      |
| IMP.MLD                       | mild imperative   | imp, mld                  | v-tam, v-tam                              |  |
| INTER                         | interrogative   | inter                     | v-tam                                     |  |
| INTER2                        | interrogative 2   | inter2                    | v-tam                                     |  |
| JUSS                          | jussive   | juss                      | v-tam                                     |  |

| Gloss                   | Description                      | Tsakorpus grammar tags | Tsakorpus category     | Comment                              |
|-------------------------|----------------------------------|------------------------|------------------------|--------------------------------------|
| JUSS.NEG                | negative jussive                 | juss, neg              | v-tam, neg             |                                      |
|                         | necessitative (any)              | nec                    | v-tam                  | only as a tag                        |
| NEC1                    | necessitative 1                  | nec, nec1              | v-tam, v-tam           |                                      |
| NEC2                    | necessitative 2                  | nec, nec2              | v-tam, v-tam           |                                      |
| PRF                     | perfect                          | prf                    | v-tam                  |                                      |
| PROB                    | probabilitive                    | prob                   | v-tam                  |                                      |
| SBJV                    | subjunctive                      | sbjv                   | v-tam                  |                                      |
| SBJV2                   | subjunctive 2                    | sbjv2                  | v-tam                  |                                      |
| SIMIL.FUT               | similative, future tense         | simil, similfut        | deriv-misc, v-tam      |                                      |
| SIMIL.PRS               | similative, present tense        | simil, similprs        | deriv-misc, v-tam      |                                      |
| SIMIL.PST               | similative, past tense           | simil, similpst        | deriv-misc, v-tam      |                                      |
| (ipfv)                  | imperfective verb stem           | ipfv                   | v-asp                  | in lexical glosses                   |
| (pfv)                   | perfective verb stem             | pfv                    | v-asp                  | in lexical glosses                   |
| (pfv/ipfv)              | biaspectual verb stem            | ipfv, pfv              | v-asp, v-asp           | in lexical glosses                   |
| <b>Non-finite forms</b> |                                  |                        |                        |                                      |
|                         | conditional converb (any)        | cvb, cond              | v-nfin, v-nfin         | only as a tag                        |
| COND1                   | conditional converb 1            | cvb, cond, cond1       | v-nfin, v-nfin, v-nfin |                                      |
| COND2                   | conditional converb 2            | cvb, cond, cond2       | v-nfin, v-nfin, v-nfin |                                      |
| COND.IRR                | irreal converb                   | cvb, cond, irr         | v-nfin, v-nfin, v-nfin |                                      |
| CNG                     | connegative                      | cng                    | v-nfin                 |                                      |
| CVB                     | converb                          | cvb                    | v-nfin                 | only in combinations (see next rows) |
| CVB.IMM                 | immediate converb                | cvb, imm               | v-nfin, v-nfin         |                                      |
| CVB.SIM                 | simultaneous converb             | cvb, sim               | v-nfin, v-nfin         |                                      |
| INF                     | infinitive (general converb)     | cvb, inf               | v-nfin, v-nfin         |                                      |
| NMLZ                    | nominalization                   | nmlz                   | v-nfin                 | only in combinations (see next rows) |
| NMLZ.ANT                | anterior nominalization          | nmlz, nmlzant          | v-nfin, v-nfin         |                                      |
| NMLZ.ANT2               | anterior nominalization 2        | nmlz, nmlzant2         | v-nfin, v-nfin         |                                      |
| NMLZ.SIM                | simultaneous nominalization      | nmlz, nmlzsim          | v-nfin, v-nfin         |                                      |
| %NMLZ                   | possibly a nominalization marker | nmlz, uncert           | v-nfin, misc           |                                      |
| PTCP                    | participle                       | ptcp                   | v-nfin                 | only in combinations (see next rows) |
| PTCP.ABES               | abessive participle              | ptcp, ptcpabes         | v-nfin, v-nfin         |                                      |
| PTCP.ANT                | anterior participle              | ptcp, ptcpant          | v-nfin, v-nfin         |                                      |
| PTCP.ANT2               | anterior participle 2            | ptcp, ptcpant2         | v-nfin, v-nfin         |                                      |
| PTCP.PASS               | passive participle               | ptcp, ptcppass         | v-nfin, v-nfin         |                                      |
| PTCP.POST               | posterior participle             | ptcp, ptcppost         | v-nfin, v-nfin         |                                      |
| PTCP.POST2              | posterior participle 2           | ptcp, ptcppost2        | v-nfin, v-nfin         |                                      |
| PTCP.PRF                | perfect participle               | ptcp, ptcpprf          | v-nfin, v-nfin         |                                      |
| PTCP.SIM                | simultaneous participle          | ptcp, ptcpsim          | v-nfin, v-nfin         |                                      |
| SUP                     | supine                           | sup                    | v-nfin                 |                                      |

| Gloss                     | Description                          | Tsakorpus<br>grammar tags | Tsakorpus category  | Comment       |
|---------------------------|--------------------------------------|---------------------------|---------------------|---------------|
| <b>Nominal derivation</b> |                                      |                           |                     |               |
| ACTN                      | action nominal                       | actn                      | deriv-n             |               |
| AGN                       | agent nominal                        | agn                       | deriv-n             |               |
| AUG                       | augmentative                         | aug                       | deriv-n             |               |
|                           | diminutive (any)                     | dim                       | deriv-n             | only as a tag |
| DIM1                      | diminutive 1                         | dim, dim1                 | deriv-n, deriv-n    |               |
| DIM2                      | diminutive 2                         | dim, dim2                 | deriv-n, deriv-n    |               |
| DIM3                      | diminutive 3                         | dim, dim3                 | deriv-n, deriv-n    |               |
| DIM4                      | diminutive 4                         | dim, dim4                 | deriv-n, deriv-n    |               |
| DYA                       | dyadic (connective-reciprocal)       | dya                       | deriv-n             |               |
|                           | instrumental nominal (any)           | insn                      | deriv-n             | only as a tag |
| INSN1                     | instrumental nominal 1               | insn, insn1               | deriv-n, deriv-n    |               |
| INSN2                     | instrumental nominal 2               | insn, insn2               | deriv-n, deriv-n    |               |
|                           | location nominal (any)               | locn                      | deriv-n             | only as a tag |
| LOCN1                     | location nominal 1                   | locn, locn1               | deriv-n, deriv-n    |               |
| LOCN2                     | location nominal 2                   | locn, locn2               | deriv-n, deriv-n    |               |
| NMNPST                    | nominal past                         | nmnpst                    | deriv-n             |               |
| PEJ                       | pejorative                           | pej                       | deriv-n             |               |
| POOR                      | 'poor/unfortunate/deceased'          | poor                      | deriv-n             |               |
| <b>Verbal derivation</b>  |                                      |                           |                     |               |
| ATTNV                     | attenuative (verbal)                 | attnv                     | deriv-v             |               |
|                           | caritive (any)                       | car                       | deriv-misc          | only as a tag |
| CAR                       | caritive (verbal)                    | car, carv                 | deriv-misc, deriv-v |               |
| CNTV                      | continuative                         | cntv                      | deriv-v             |               |
| DISC                      | discontinuative                      | disc                      | deriv-v             |               |
| DISTR                     | distributive                         | distr                     | deriv-v             |               |
| DRV1                      | unspecified derivation 1             | drv1                      | deriv-v             |               |
| DRV2                      | unspecified derivation 2             | drv2                      | deriv-v             |               |
| DRV3                      | unspecified derivation 3             | drv3                      | deriv-v             |               |
| DRV4                      | unspecified derivation 4             | drv4                      | deriv-v             |               |
| DRV5                      | unspecified derivation 5             | drv5                      | deriv-v             |               |
| %DRV                      | possibly an (unspecified) derivation | drv, uncert               | deriv-v, misc       |               |
| DUR                       | durative                             | dur                       | deriv-v             |               |
| DUR2                      | durative 2                           | dur2                      | deriv-v             |               |
| FREQ                      | frequentative                        | freq                      | deriv-v             |               |
| INCH                      | inchoative                           | inch                      | deriv-v             |               |
| INCH2                     | inchoative 2                         | inch2                     | deriv-v             |               |
| INCH3                     | inchoative 3                         | inch3                     | deriv-v             |               |
| MULT1                     | multiplicative 1                     | mult1                     | deriv-v             |               |
| MULT2                     | multiplicative 2                     | mult2                     | deriv-v             |               |
| MULT3                     | multiplicative 3                     | mult2                     | deriv-v             |               |

| Gloss                | Description                   | Tsakorpus grammar tags | Tsakorpus category     | Comment   |
|----------------------|-------------------------------|------------------------|------------------------|---|
| PASS                 | passive                       | pass                   | deriv-v                |   |
| RES1                 | resultative 1                 | res1                   | deriv-v                |   |
| RES2                 | resultative 2                 | res2                   | deriv-v                |   |
|                      | transitivizer (any)           | tr                     | deriv-v                | only as a tag                                     |
| TR1                  | transitivizer 1               | tr, tr1                | deriv-v, deriv-v       |   |
| TR2                  | transitivizer 2               | tr, tr2                | deriv-v, deriv-v       |   |
| TR3                  | transitivizer 3               | tr, tr3                | deriv-v, deriv-v       |   |
| TR4                  | transitivizer 4               | tr, tr4                | deriv-v, deriv-v       |   |
| TR5                  | transitivizer 5               | tr, tr5                | deriv-v, deriv-v       |   |
| TR6                  | transitivizer 6               | tr, tr6                | deriv-v, deriv-v       |   |
|                      | translative (any)             | trl                    | deriv-v                | only as a tag                                     |
| TRL1                 | translative 1                 | trl, trl1              | deriv-v, deriv-v       |   |
| TRL2                 | translative 2                 | trl, trl2              | deriv-v, deriv-v       |   |
| TRL3                 | translative 3                 | trl, trl3              | deriv-v, deriv-v       |   |
|                      | verbalizer (any)              | vblz                   | deriv-v                | only as a tag                                     |
| VBLZ1                | verbalizer 1                  | vblz, vblz1            | deriv-v, deriv-v       |   |
| VBLZ2                | verbalizer 2                  | vblz, vblz2            | deriv-v, deriv-v       |   |
| VBLZ3                | verbalizer 3                  | vblz, vblz3            | deriv-v, deriv-v       |   |
| VBLZ4                | verbalizer 4                  | vblz, vblz4            | deriv-v, deriv-v       |   |
| VBLZ5                | verbalizer 5                  | vblz, vblz5            | deriv-v, deriv-v       |   |
| VBLZ6                | verbalizer 6                  | vblz, vblz6            | deriv-v, deriv-v       |   |
| <b>Negation</b>      |                               |                        |                        |   |
| NEG                  | negative verb                 | neg, negv              | neg, neg               | lexical   |
| EMPH.NEG             | emphatic negative verb        | neg, negemph           | neg, neg               | lexical   |
| NEG.ADJ              | adjective negation            | neg, negadj            | neg, neg               | lexical   |
| <b>Miscellaneous</b> |                               |                        |                        |   |
| ADJZ                 | adjectivizer                  | adjz                   | deriv-misc             |   |
| ADJZ2                | adjectivizer 2                | adjz2                  | deriv-misc             |   |
| %ADJZ2               | possibly an adjectivizer      | adjz2, uncert          | deriv-misc, misc       |   |
|                      | adverbializer (any)           | advz                   | deriv-misc             | only as a tag                                     |
| ADVZ1                | adverbializer 1               | advz1                  | deriv-misc             |   |
| ADVZ2                | adverbializer 2               | advz2                  | deriv-misc             |   |
| ADVZ3                | adverbializer 3               | advz3                  | deriv-misc             |   |
| ADV.EMPH             | emphatic adverbial derivation | advz, emph             | deriv-misc, disc       |   |
|                      | attenuative (any)             | attn                   | deriv-misc             | only as a tag                                     |
| ATTN1                | attenuative 1 (adjectival)    | attn, attn1            | deriv-misc, deriv-misc |   |
| ATTN2                | attenuative 2 (adjectival)    | attn, attn2            | deriv-misc, deriv-misc |   |
| CONC                 | concessive particle           | conc                   | misc                   | lexical   |
| EMPH                 | emphatic marker               | emph                   | disc                   | only in combinations, e.g. "EMPH.NEG", "ADV.EMPH" |
|                      | exclamative (any)             | excl                   | disc                   | only as a tag                                     |



| Gloss | Description          | Tsakorpus<br>grammar tags | Tsakorpus category | Comment                            |
|-------|----------------------|---------------------------|--------------------|------------------------------------|
| EXCL1 | exclamative 1        | excl, excl1               | disc, disc         | lexical                            |
| EXCL2 | exclamative 2        | excl, excl2               | disc, disc         | lexical                            |
| INSV  | insistive            | insv                      | deriv-misc         |                                    |
| LIM   | limitative           | lim                       | deriv-misc         |                                    |
| LIM2  | limitative 2         | lim2                      | deriv-misc         |                                    |
| ORD   | ordinal numeral      | ord                       | deriv-misc         |                                    |
| PLC   | placeholder          | plc                       | misc               | lexical                            |
| SEL   | selective            | sel                       | deriv-misc         |                                    |
| TOP   | topical marker       | top                       | deriv-misc         |                                    |
| %     | uncertain categories | uncert                    | misc               | as part of a gloss, e.g.<br>"%DRV" |
| %%    | unknown morph        | unkn                      | misc               |                                    |

Table 37. Alphabetical list of morpheme glossing labels

| Gloss    | Description                   | Tsakorpus grammar tags | Tsakorpus category             | Comment  |
|----------|-------------------------------|------------------------|--------------------------------|--|
| 1DU      | 1 person dual                 | pn1, pndu              | persnum-person, persnum-number | only in combinations with personal pronouns, e.g. "PRO.1SG.NOM", in verbal subject cross-reference markers, e.g. "1SG.S", and in possessive markers, e.g. "NOM.SG.1SG" |
| 1PL      | 1 person plural               | pn1, pnpl              | persnum-person, persnum-number |  |
| 1SG      | 1 person singular             | pn1, pnsq              | persnum-person, persnum-number |  |
| 2DU      | 2 person dual                 | pn2, pndu              | persnum-person, persnum-number |  |
| 2PL      | 2 person plural               | pn2, pnpl              | persnum-person, persnum-number |  |
| 2SG      | 2 person singular             | pn2, pnsq              | persnum-person, persnum-number |  |
| 3DU      | 3 person dual                 | pn3, pndu              | persnum-person, persnum-number |  |
| 3PL      | 3 person plural               | pn3, pnpl              | persnum-person, persnum-number |  |
| 3SG      | 3 person singular             | pn3, pnsq              | persnum-person, persnum-number |  |
| ABL      | ablative case                 | abl                    | n-case                         | only in combinations with number, e.g. "ABL.SG" or with personal pronouns, e.g. "PRO.1SG.ABL"  |
| ABL.ADV  | ablative case of adverbials   | abl, advcase           | n-case, n-misc                 |  |
| ACC      | accusative case               | acc                    | n-case                         | only in combinations with personal pronouns, e.g. "PRO.1PL.ACC"  |
| ACTN     | action nominal                | actn                   | deriv-n                        |  |
| ADJZ     | adjectivizer                  | adjz                   | deriv-misc                     |  |
| ADJZ2    | adjectivizer 2                | adjz2                  | deriv-misc                     |  |
| ADV.EMPH | emphatic adverbial derivation | advz, emph             | deriv-misc, disc               |  |
| ADVZ1    | adverbializer 1               | advz1                  | deriv-misc                     |  |
| ADVZ2    | adverbializer 2               | advz2                  | deriv-misc                     |  |
| ADVZ3    | adverbializer 3               | advz3                  | deriv-misc                     |  |
| AGN      | agent nominal                 | agn                    | deriv-n                        |  |
| ATTN1    | attenuative 1 (adjectival)    | attn, attn1            | deriv-misc, deriv-misc         |  |
| ATTN2    | attenuative 2 (adjectival)    | attn, attn2            | deriv-misc, deriv-misc         |  |
| ATTNV    | attenuative (verbal)          | attnv                  | deriv-v                        |  |
| AUD      | auditive                      | aud                    | v-tam                          |  |
| AUG      | augmentative                  | aug                    | deriv-n                        |  |
| CAR      | caritive (verbal)             | car, carv              | deriv-misc, deriv-v            |  |
| CAR.ADV  | caritive adverbializer        | car, caradv            | deriv-misc, n-misc             |  |
| CMP      | comparative                   | cmp                    | n-misc                         |  |
| CNG      | connegative                   | cng                    | v-nfin                         |  |

| Gloss    | Description                        | Tsakorpus grammar tags | Tsakorpus category          | Comment   |
|----------|------------------------------------|------------------------|-----------------------------|---|
| CNTV     | continuative                       | cntv                   | deriv-v                     |   |
| CONC     | concessive particle                | conc                   | misc                        | lexical   |
| COND.IRR | irreal converb                     | cvb, cond, irr         | v-nfin, v-nfin, v-nfin      |   |
| COND1    | conditional converb 1              | cvb, cond, cond1       | v-nfin, v-nfin, v-nfin      |   |
| COND2    | conditional converb 2              | cvb, cond, cond2       | v-nfin, v-nfin, v-nfin      |   |
| CVB      | converb                            | cvb                    | v-nfin                      | only in combinations (see next rows)                            |
| CVB.IMM  | immediate converb                  | cvb, imm               | v-nfin, v-nfin              |   |
| CVB.SIM  | simultaneous converb               | cvb, sim               | v-nfin, v-nfin              |   |
| DIM1     | diminutive 1                       | dim, dim1              | deriv-n, deriv-n            |   |
| DIM2     | diminutive 2                       | dim, dim2              | deriv-n, deriv-n            |   |
| DIM3     | diminutive 3                       | dim, dim3              | deriv-n, deriv-n            |   |
| DIM4     | diminutive 4                       | dim, dim4              | deriv-n, deriv-n            |   |
| DISC     | discontinuative                    | disc                   | deriv-v                     |   |
| DISTR    | distributive                       | distr                  | deriv-v                     |   |
| DRV1     | unspecified derivation 1           | drv1                   | deriv-v                     |   |
| DRV2     | unspecified derivation 2           | drv2                   | deriv-v                     |   |
| DRV3     | unspecified derivation 3           | drv3                   | deriv-v                     |   |
| DRV4     | unspecified derivation 4           | drv4                   | deriv-v                     |   |
| DRV5     | unspecified derivation 5           | drv5                   | deriv-v                     |   |
| DST      | destinative form                   | dst                    | n-misc                      | only in combinations with number, e.g. "DST.SG"                 |
| DU       | dual number                        | du                     | n-num                       |   |
| DU.O     | objective conjugation, dual object | objc, objdu            | persnum-conj, object-number |   |
| DUR      | durative                           | dur                    | deriv-v                     |   |
| DUR2     | durative 2                         | dur2                   | deriv-v                     |   |
| DYA      | dyadic (connective-reciprocal)     | dya                    | deriv-n                     |   |
| EMPH     | emphatic marker                    | emph                   | disc                        | only in combinations, e.g. "EMPH.NEG", "ADV.EMPH"               |
| EMPH.NEG | emphatic negative verb             | neg, negemph           | neg, neg                    | lexical   |
| ESS      | essive case                        | ess                    | n-misc                      |   |
| ESS2     | essive case 2                      | ess2                   | n-misc                      |   |
| ESS3     | essive case 3                      | ess3                   | n-misc                      |   |
| EXCL1    | exclamative 1                      | excl, excl1            | disc, disc                  | lexical   |
| EXCL2    | exclamative 2                      | excl, excl2            | disc, disc                  | lexical   |
| FREQ     | frequentative                      | freq                   | deriv-v                     |   |
| FUT      | future tense                       | fut                    | v-tam                       |   |
| GEN      | genitive case                      | gen                    | n-case                      | only in combinations with personal pronouns, e.g. "PRO.1PL.GEN" |
| HAB      | habitual                           | hab                    | v-tam                       |   |

| Gloss    | Description                    | Tsakorpus grammar tags | Tsakorpus category    | Comment   |
|----------|--------------------------------|------------------------|-----------------------|---|
| HAB2     | habitual 2                     | hab2                   | v-tam                 |   |
| HORT1    | hortative 1                    | hort, hort1            | v-tam, v-tam          |   |
| HORT2    | hortative 2                    | hort, hort2            | v-tam, v-tam          |   |
| HYPOT    | hypothetical                   | hypot                  | v-tam                 |   |
| IMP      | imperative                     | imp                    | v-tam                 | only in combinations with person and number, e.g. "IMP.2SG.S"                                 |
| IMP.MLD  | mild imperative                | imp, mld               | v-tam, v-tam          |   |
| INCH     | inchoative                     | inch                   | deriv-v               |   |
| INCH2    | inchoative 2                   | inch2                  | deriv-v               |   |
| INCH3    | inchoative 3                   | inch3                  | deriv-v               |   |
| INF      | infinitive (general converb)   | cvb, inf               | v-nfin, v-nfin        |   |
| INSN1    | instrumental nominal 1         | insn, insn1            | deriv-n, deriv-n      |   |
| INSN2    | instrumental nominal 2         | insn, insn2            | deriv-n, deriv-n      |   |
| INSV     | insistive                      | insv                   | deriv-misc            |   |
| INTER    | interrogative                  | inter                  | v-tam                 |   |
| INTER2   | interrogative 2                | inter2                 | v-tam                 |   |
| JUSS     | jussive                        | juss                   | v-tam                 |   |
| JUSS.NEG | negative jussive               | juss, neg              | v-tam, neg            |   |
| KPL      | kinship plural                 | kpl                    | n-misc                |   |
| LAT      | lative case                    | lat                    | n-case                | only in combinations with number, e.g. "LAT.SG" or with personal pronouns, e.g. "PRO.1SG.LAT" |
| LAT.ADV  | lative case of adverbials      | lat, advcase           | n-case, n-misc        |   |
| LAT.SG2  | lative case, singular number 2 | lat, sg, latsg2        | n-case, n-num, n-case |   |
| LIM      | limitative                     | lim                    | deriv-misc            |   |
| LIM2     | limitative 2                   | lim2                   | deriv-misc            |   |
| LOC      | locative case                  | loc                    | n-case                | only in combinations with number, e.g. "LOC.SG" or with personal pronouns, e.g. "PRO.1SG.LOC" |
| LOC.ADV  | locative case of adverbials    | loc, advcase           | n-case, n-misc        |   |
| LOCN1    | location nominal 1             | locn, locn1            | deriv-n, deriv-n      |   |
| LOCN2    | location nominal 2             | locn, locn2            | deriv-n, deriv-n      |   |
| MD       | middle conjugation             | md                     | persnum-conj          |   |
| MULT1    | multiplicative 1               | mult1                  | deriv-v               |   |
| MULT2    | multiplicative 2               | mult2                  | deriv-v               |   |
| MULT3    | multiplicative 3               | mult2                  | deriv-v               |   |
| NEC1     | necessitative 1                | nec, nec1              | v-tam, v-tam          |   |
| NEC2     | necessitative 2                | nec, nec2              | v-tam, v-tam          |   |
| NEG      | negative verb                  | neg, negv              | neg, neg              | lexical   |
| NEG.ADJ  | adjective negation             | neg, negadj            | neg, neg              | lexical   |

| Gloss     | Description                               | Tsakorpus grammar tags | Tsakorpus category          | Comment   |
|-----------|---|------------------------|-----------------------------|---|
| NMLZ      | nominalization                            | nmlz                   | v-nfin                      | only in combinations (see next rows)  |
| NMLZ.ANT  | anterior nominalization                   | nmlz, nmlzant          | v-nfin, v-nfin              |   |
| NMLZ.ANT2 | anterior nominalization 2                 | nmlz, nmlzant2         | v-nfin, v-nfin              |   |
| NMLZ.SIM  | simultaneous nominalization               | nmlz, nmlzsim          | v-nfin, v-nfin              |   |
| NMNPST    | nominal past                              | nmnpst                 | deriv-n                     |   |
| NOM       | nominative case                           | nom                    | n-case                      | only in combinations with number and other categories, e.g. "NOM.SG.3DU"                        |
| NOM.ADV   | nominative case of adverbials             | nom, advcase           | n-case, n-misc              |   |
| NSG.O     | objective conjugation, nonsingular object | objc, objnsg           | persnum-conj, object-number | only in combinations with person and number, e.g. "3DU.NSG.O"                                   |
| OBL       | oblique case                              | obl                    | n-case                      | only in combinations with number, e.g. "OBL.SG.1SG"   |
| ORD       | ordinal numeral                           | ord                    | deriv-misc                  |   |
| PASS      | passive                                   | pass                   | deriv-v                     |   |
| PEJ       | pejorative                                | pej                    | deriv-n                     |   |
| PL        | plural number                             | pl                     | n-num                       |   |
| PL.O      | objective conjugation, plural object      | objc, objpl            | persnum-conj, object-number |   |
| PLC       | placeholder                               | plc                    | misc                        | lexical   |
| POOR      | 'poor/unfortunate/deceased'               | poor                   | deriv-n                     |   |
| PRF       | perfect                                   | prf                    | v-tam                       |   |
| PRM       | presumptive cross-reference series        | prm                    | persnum-series              | only in combinations with person and number and conjugation type, e.g. "3DU.S.PRM"              |
| PRO       | personal pronoun                          | pers                   |                             | lexical, only in combinations with person and number and case, e.g. "PRO.1SG.NOM"               |
| PROB      | probabilitive                             | prob                   | v-tam                       |   |
| PROL      | prolative case                            | prol                   | n-case                      | only in combinations with number, e.g. "PROL.SG" or with personal pronouns, e.g. "PRO.1SG.PROL" |
| PST       | past tense cross-reference series         | pst                    | persnum-series              | only in combinations with person and number and conjugation type, e.g. "3DU.S.PST"              |
| PTCP      | participle                                | ptcp                   | v-nfin                      | only in combinations (see next rows)  |
| PTCP.ABES | abessive participle                       | ptcp, ptcpabes         | v-nfin, v-nfin              |   |
| PTCP.ANT  | anterior participle                       | ptcp, ptcpant          | v-nfin, v-nfin              |   |
| PTCP.ANT2 | anterior participle 2                     | ptcp, ptcpant2         | v-nfin, v-nfin              |   |

| Gloss      | Description   | Tsakorpus grammar tags | Tsakorpus category                        | Comment  |
|------------|---|------------------------|---|--|
| PTCP.PASS  | passive participle                                      | ptcp, ptcpPASS         | v-nfin, v-nfin                            |  |
| PTCP.POST  | posterior participle                                    | ptcp, ptcpPOST         | v-nfin, v-nfin                            |  |
| PTCP.POST2 | posterior participle 2                                  | ptcp, ptcpPOST2        | v-nfin, v-nfin                            |  |
| PTCP.PRF   | perfect participle                                      | ptcp, ptcpPRF          | v-nfin, v-nfin                            |  |
| PTCP.SIM   | simultaneous participle                                 | ptcp, ptcpSIM          | v-nfin, v-nfin                            |  |
| RES1       | resultative 1   | res1                   | deriv-v                                   |  |
| RES2       | resultative 2   | res2                   | deriv-v                                   |  |
| S          | subjective conjugation                                  | subjC                  | persnum-conj                              | only in combinations with person and number, e.g. "3DU.S"      |
| S/SG.O     | subjective conj. or objective conj., nonsingular object | subjC, objC, objSG     | persnum-conj, persnum-conj, object-number | only in combinations with person and number, e.g. "2DU.S/SG.O" |
| SBJV       | subjunctive   | sbjv                   | v-tam                                     |  |
| SBJV2      | subjunctive 2   | sbjv2                  | v-tam                                     |  |
| SEL        | selective   | sel                    | deriv-misc                                |  |
| SG         | singular number   | sg                     | n-num                                     | only in combinations with case, e.g. "LAT.SG"                  |
| SG.O       | objective conjugation, singular object                  | objC, objSG            | persnum-conj, object-number               | only in combinations with person and number, e.g. "2SG.SG.O"   |
| SIMIL      | similative (nominal)                                    | simil, similN          | deriv-misc, n-misc                        |  |
| SIMIL.FUT  | similative, future tense                                | simil, similFUT        | deriv-misc, v-tam                         |  |
| SIMIL.PRS  | similative, present tense                               | simil, similPRS        | deriv-misc, v-tam                         |  |
| SIMIL.PST  | similative, past tense                                  | simil, similPST        | deriv-misc, v-tam                         |  |
| SOC        | sociative   | soc                    | n-misc                                    |  |
| SOC2       | sociative 2   | soc2                   | n-misc                                    |  |
| SUP        | supine  | sup                    | v-nfin                                    |  |
| TOP        | topical marker  | top                    | deriv-misc                                |  |
| TR1        | transitivizer 1   | tr, tr1                | deriv-v, deriv-v                          |  |
| TR2        | transitivizer 2   | tr, tr2                | deriv-v, deriv-v                          |  |
| TR3        | transitivizer 3   | tr, tr3                | deriv-v, deriv-v                          |  |
| TR4        | transitivizer 4   | tr, tr4                | deriv-v, deriv-v                          |  |
| TR5        | transitivizer 5   | tr, tr5                | deriv-v, deriv-v                          |  |
| TR6        | transitivizer 6   | tr, tr6                | deriv-v, deriv-v                          |  |
| TRL1       | translative 1   | trl, trl1              | deriv-v, deriv-v                          |  |
| TRL2       | translative 2   | trl, trl2              | deriv-v, deriv-v                          |  |
| TRL3       | translative 3   | trl, trl3              | deriv-v, deriv-v                          |  |
| VBLZ1      | verbalizer 1  | vblz, vblz1            | deriv-v, deriv-v                          |  |
| VBLZ2      | verbalizer 2  | vblz, vblz2            | deriv-v, deriv-v                          |  |
| VBLZ3      | verbalizer 3  | vblz, vblz3            | deriv-v, deriv-v                          |  |
| VBLZ4      | verbalizer 4  | vblz, vblz4            | deriv-v, deriv-v                          |  |
| VBLZ5      | verbalizer 5  | vblz, vblz5            | deriv-v, deriv-v                          |  |

| Gloss      | Description                          | Tsakorpus<br>grammar tags | Tsakorpus category | Comment                            |
|------------|--------------------------------------|---------------------------|--------------------|------------------------------------|
| VBLZ6      | verbalizer 6                         | vblz, vblz6               | deriv-v, deriv-v   |                                    |
| VOC        | vocative form                        | voc                       | n-misc             |                                    |
| (ipfv)     | imperfective verb stem               | ipfv                      | v-asp              | in lexical glosses                 |
| (pfv)      | perfective verb stem                 | pfv                       | v-asp              | in lexical glosses                 |
| (pfv/ipfv) | biaspectual verb stem                | ipfv, pfv                 | v-asp, v-asp       | in lexical glosses                 |
| %          | uncertain categories                 | uncert                    | misc               | as part of a gloss, e.g.<br>"%DRV" |
| %ADJZ2     | possibly an adjectivizer             | adjz2, uncert             | deriv-misc, misc   |                                    |
| %DRV       | possibly an (unspecified) derivation | drv, uncert               | deriv-v, misc      |                                    |
| %ESS2      | possibly an essive case marker       | ess2, uncert              | n-misc, misc       |                                    |
| %NMLZ      | possibly a nominalization marker     | nmlz, uncert              | v-nfin, misc       |                                    |
| %%         | unknown morph                        | unkn                      | misc               |                                    |

Table 38. Alphabetical list of Tsakorpus grammar tags

| Tsakorpus grammar tag | Tsakorpus category | Gloss   | Description                | Comment   |
|-----------------------|--------------------|---|----------------------------|---|
| abl                   | n-case             | ABL.SG, ABL.PL, ABL.ADV; PRO.1SG.ABL, ...               | ablative case              | only in combinations with number, e.g. "ABL.SG" or with personal pronouns, e.g. "PRO.1SG.ABL" |
| acc                   | n-case             | PRO.1SG.ACC, ...  | accusative case            | only in combinations with personal pronouns, e.g. "PRO.1PL.ACC"                               |
| actn                  | deriv-n            | ACTN  | action nominal             |   |
| adj                   | part of speech     |   | adjective                  |   |
| adjz                  | deriv-misc         | ADJZ  | adjectivizer               |   |
| adjz2                 | deriv-misc         | ADJZ2, %ADJZ2   | adjectivizer 2             |   |
| adv                   | part of speech     |   | adverb                     |   |
| advcase               | n-misc             | ABL.ADV, LAT.ADV, LOC.ADV, NOM.ADV                      | case of adverbials         |   |
| advz                  | deriv-misc         | ADVZ1, ADVZ2, ADVZ3, ADV.EMPH                           | adverbializer (any)        |   |
| advz1                 | deriv-misc         | ADVZ1   | adverbializer 1            |   |
| advz2                 | deriv-misc         | ADVZ2   | adverbializer 2            |   |
| advz3                 | deriv-misc         | ADVZ3   | adverbializer 3            |   |
| agn                   | deriv-n            | AGN   | agent nominal              |   |
| attn                  | deriv-misc         | ATTN1, ATTN2  | attenuative (any)          |   |
| attn1                 | deriv-misc         | ATTN1   | attenuative 1 (adjectival) |   |
| attn2                 | deriv-misc         | ATTN2   | attenuative 2 (adjectival) |   |
| attnv                 | deriv-v            | ATTNV   | attenuative (verbal)       |   |
| aud                   | v-tam              | AUD   | auditive                   |   |
| aug                   | deriv-n            | AUG   | augmentative               |   |
| aux                   | part of speech     |   | auxiliary verb             |   |
| car                   | deriv-misc         | CAR, CAR.ADV  | caritive (any)             |   |
| caradv                | n-misc             | CAR.ADV   | caritive adverbializer     |   |
| carv                  | deriv-v            | CAR   | caritive (verbal)          |   |
| cmp                   | n-misc             | CMP   | comparative                |   |
| cng                   | v-nfin             | CNG   | connegative                |   |
| cntv                  | deriv-v            | CNTV  | continuative               |   |
| conc                  | misc               | CONC  | concessive particle        | lexical   |
| cond                  | v-nfin             | COND1, COND2, COND.IRR                                  | conditional converb (any)  |   |
| cond1                 | v-nfin             | COND1   | conditional converb 1      |   |
| cond2                 | v-nfin             | COND2   | conditional converb 2      |   |
| conj                  | part of speech     |   | conjunction                |   |
| cvb                   | v-nfin             | COND1, COND2, COND.IRR, CNG, CVB, CVB.IMM, CVB.SIM, INF | converb                    |   |



| <b>Tsakorpus grammar tag</b> | <b>Tsakorpus category</b> | <b>Gloss</b>            | <b>Description</b>             | <b>Comment</b>  |
|------------------------------|---------------------------|-------------------------|--------------------------------|---|
| dem                          | part of speech            |                         | demonstrative pronoun          |   |
| dim                          | deriv-n                   | DIM1, DIM2, DIM3, DIM4  | diminutive (any)               |   |
| dim1                         | deriv-n                   | DIM1                    | diminutive 1                   |   |
| dim2                         | deriv-n                   | DIM2                    | diminutive 2                   |   |
| dim3                         | deriv-n                   | DIM3                    | diminutive 3                   |   |
| dim4                         | deriv-n                   | DIM4                    | diminutive 4                   |   |
| disc                         | deriv-v                   | DISC                    | discontinuative                |   |
| distr                        | deriv-v                   | DISTR                   | distributive                   |   |
| drv                          | deriv-v                   | %DRV                    | an (unspecified) derivation    |   |
| drv1                         | deriv-v                   | DRV1                    | unspecified derivation 1       |   |
| drv2                         | deriv-v                   | DRV2                    | unspecified derivation 2       |   |
| drv3                         | deriv-v                   | DRV3                    | unspecified derivation 3       |   |
| drv4                         | deriv-v                   | DRV4                    | unspecified derivation 4       |   |
| drv5                         | deriv-v                   | DRV5                    | unspecified derivation 5       |   |
| dst                          | n-misc                    | DST.SG, DST.PL          | destinative form               |   |
| du                           | n-num                     | DU, NOM.DU, OBL.DU      | dual number                    |   |
| dur                          | deriv-v                   | DUR                     | durative                       |   |
| dur2                         | deriv-v                   | DUR2                    | durative 2                     |   |
| dya                          | deriv-n                   | DYA                     | dyadic (connective-reciprocal) |   |
| emph                         | disc                      | ADV.EMPH, EMPH.NEG      | emphatic marker                |   |
| ess                          | n-misc                    | ESS                     | essive case                    |   |
| ess2                         | n-misc                    | ESS2, %ESS2             | essive case 2                  |   |
| ess3                         | n-misc                    | ESS3                    | essive case 3                  |   |
| excl                         | disc                      | EXCL1, EXCL2            | exclamative (any)              | lexical   |
| excl1                        | disc                      | EXCL1                   | exclamative 1                  | lexical   |
| excl2                        | disc                      | EXCL2                   | exclamative 2                  | lexical   |
| freq                         | deriv-v                   | FREQ                    | frequentative                  |   |
| fut                          | v-tam                     | FUT                     | future tense                   |   |
| gen                          | n-case                    | PRO.1SG.GEN, ...        | genitive case                  | only in combinations with personal pronouns, e.g. "PRO.1PL.GEN" |
| hab                          | v-tam                     | HAB                     | habitual                       |   |
| hab2                         | v-tam                     | HAB2                    | habitual 2                     |   |
| hort                         | v-tam                     | HORT1, HORT2            | hortative (any)                |   |
| hort1                        | v-tam                     | HORT1                   | hortative 1                    |   |
| hort2                        | v-tam                     | HORT2                   | hortative 2                    |   |
| hypot                        | v-tam                     | HYPOT                   | hypothetical                   |   |
| imm                          | v-nfin                    | CVB.IMM                 | immediate converb              |   |
| imp                          | v-tam                     | IMP.2SG.S, ...; IMP.MLD | imperative                     | only in combinations with person and number, e.g. "IMP.2SG.S"   |

| <b>Tsakorpus grammar tag</b> | <b>Tsakorpus category</b> | <b>Gloss</b>                                       | <b>Description</b>             | <b>Comment</b>  |
|------------------------------|---------------------------|--|--------------------------------|---|
| inch                         | deriv-v                   | INCH   | inchoative                     |   |
| inch2                        | deriv-v                   | INCH2  | inchoative 2                   |   |
| inch3                        | deriv-v                   | INCH3  | inchoative 3                   |   |
| inf                          | v-nfin                    | INF  | infinitive (general converb)   |   |
| insn                         | deriv-n                   | INSN1, INSN2                                       | instrumental nominal (any)     |   |
| insn1                        | deriv-n                   | INSN1  | instrumental nominal 1         |   |
| insn2                        | deriv-n                   | INSN2  | instrumental nominal 2         |   |
| insv                         | deriv-misc                | INSV   | insistive                      |   |
| inter                        | v-tam                     | INTER  | interrogative                  |   |
| inter2                       | v-tam                     | INTER2   | interrogative 2                |   |
| intj                         | part of speech            |  | Interjection                   |   |
| interrog                     | part of speech            |  | Interrogative pronoun          |   |
| ipfv                         | v-asp                     | (ipfv), (pfv/ipfv)                                 | imperfective verb stem         |   |
| irr                          | v-nfin                    | COND.IRR   | irreal converb                 |   |
| juss                         | v-tam                     | JUSS, JUSS.NEG                                     | jussive                        |   |
| kpl                          | n-misc                    | KPL  | kinship plural                 |   |
| lat                          | n-case                    | LAT.SG, LAT.SG2, LAT.PL, LAT.ADV; PRO.1SG.LAT, ... | lative case                    | only in combinations with number, e.g. "LAT.SG" or with personal pronouns, e.g. "PRO.1SG.LAT" |
| latsg2                       | n-case                    | LAT.SG2  | lative case, singular number 2 |   |
| lim                          | deriv-misc                | LIM  | limitative                     |   |
| lim2                         | deriv-misc                | LIM2   | limitative 2                   |   |
| loc                          | n-case                    | LOC.SG, LOC.PL, LOC.ADV; PRO.1SG.LOC, ...          | locative case                  | only in combinations with number, e.g. "LOC.SG" or with personal pronouns, e.g. "PRO.1SG.LOC" |
| locn                         | deriv-n                   | LOCN1, LOCN2                                       | location nominal (any)         |   |
| locn1                        | deriv-n                   | LOCN1  | location nominal 1             |   |
| locn2                        | deriv-n                   | LOCN2  | location nominal 2             |   |
| md                           | persnum-conj              | MD   | middle conjugation             |   |
| mld                          | v-tam                     | IMP.MLD  | mild imperative                |   |
| mult1                        | deriv-v                   | MULT1  | multiplicative 1               |   |
| mult2                        | deriv-v                   | MULT2  | multiplicative 2               |   |
| mult2                        | deriv-v                   | MULT3  | multiplicative 3               |   |
| n                            | part of speech            |  | noun                           |   |
| nec                          | v-tam                     | NEC1, NEC2   | necessitative (any)            |   |
| nec1                         | v-tam                     | NEC1   | necessitative 1                |   |
| nec2                         | v-tam                     | NEC2   | necessitative 2                |   |
| neg                          | neg                       | EMPH.NEG, NEG, NEG.ADJ, JUSS.NEG                   | negation                       |   |
| negadj                       | neg                       | NEG.ADJ  | adjective negation             | lexical   |
| negemph                      | neg                       | EMPH.NEG   | emphatic negative verb         | lexical   |
| negv                         | neg                       | NEG  | negative verb                  | lexical   |

| Tsakorpus grammar tag | Tsakorpus category | Gloss   | Description                 | Comment   |
|-----------------------|--------------------|---|-----------------------------|---|
| nmlz                  | v-nfin             | NMLZ.ANT, NMLZ.ANT2, NMLZ.SIM, %NMLZ  | nominalization              |   |
| nmlzant               | v-nfin             | NMLZ.ANT  | anterior nominalization     |   |
| nmlzant2              | v-nfin             | NMLZ.ANT2   | anterior nominalization 2   |   |
| nmlzsim               | v-nfin             | NMLZ.SIM  | simultaneous nominalization |   |
| nmnpst                | deriv-n            | NMNPST  | nominal past                |   |
| nom                   | n-case             | NOM.SG, NOM.DU, NOM.PL;<br>NOM.SG.1SG, ...;<br>PRO.1SG.NOM, ...;<br>NOM.ADV                     | nominative case             | only in combinations with number and other categories, e.g. "NOM.SG.3DU"  |
| npr                   | part of speech     |   | proper noun                 |   |
| num                   | part of speech     |   | numeral                     |   |
| objc                  | persnum-conj       | 1SG.NSG.O, ...;<br>1SG.SG.O, ...;<br>2DU.S/SG.O, ...  | objective conjugation       | only in combinations with person and number, e.g. "3DU.NSG.O"   |
| objdu                 | object-number      | DU.O  | dual object                 |   |
| objnsg                | object-number      | 1SG.NSG.O, ...  | nonsingular object          | only in combinations with person and number, e.g. "3DU.NSG.O"   |
| objpl                 | object-number      | PL.O  | plural object               |   |
| objsg                 | object-number      | 1SG.SG.O, ...;<br>2DU.S/SG.O, ...   | singular object             | only in combinations with person and number, e.g. "2SG.SG.O"  |
| obl                   | n-case             | OBL.SG, OBL.SG.1SG, ...   | oblique case                | only in combinations with number, e.g. "OBL.SG.1SG"   |
| ord                   | deriv-misc         | ORD   | ordinal numeral             |   |
| pass                  | deriv-v            | PASS  | passive                     |   |
| pej                   | deriv-n            | PEJ   | pejorative                  |   |
| pers                  | part of speech     |   | personal pronoun            |   |
| pfv                   | v-asp              | (pfv), (pfv/ipfv)   | perfective verb stem        |   |
| pl                    | n-num              | PL, ABL.PL, DST.PL, LAT.PL, LOC.PL, NOM.PL, OBL.PL, PROL.PL;<br>NOM.PL.3SG, ...;<br>PL.1SG, ... | plural number               |   |
| plc                   | misc               | PLC   | placeholder                 | lexical   |
| pn1                   | persnum-person     | 1SG, 1DU, 1PL   | 1 person                    | in personal pronouns, e.g. "PRO.1SG.NOM", in verbal subject cross-reference markers, e.g. "1SG.S", and in possessive markers, e.g. "NOM.SG.1SG" |
| pn2                   | persnum-person     | 2SG, 2DU, 2PL   | 2 person                    |   |
| pn3                   | persnum-person     | 3SG, 3DU, 3PL   | 3 person                    |   |
| pndu                  | persnum-number     | 1DU, 2DU, 3DU   | dual number                 |   |
| pnpl                  | persnum-number     | 1PL, 2PL, 3PL   | plural number               |   |

| Tsakorpus grammar tag | Tsakorpus category | Gloss   | Description                        | Comment  |
|-----------------------|--------------------|---|------------------------------------|--|
| pns                   | persnum-number     | 1SG, 2SG, 3SG   | singular number                    |  |
| poor                  | deriv-n            | POOR  | 'poor/unfortunate/deceased'        |  |
| poss                  | n-poss             | OBL.SG.1SG, ...   | possessive                         | in combinations of case and person-number markers  |
| pp                    | part of speech     |   | postposition                       |  |
| prf                   | v-tam              | PRF   | perfect                            |  |
| prm                   | persnum-series     | 1SG.S.PRM, ...  | presumptive cross-reference series | only in combinations with person and number and conjugation type, e.g. "3DU.S.PRM"               |
| pro                   | part of speech     |   | pronoun                            |  |
| prob                  | v-tam              | PROB  | probabilitive                      |  |
| prol                  | n-case             | PROL.SG, PROL.PL;<br>PRO.1SG.PROL, ...  | prolative case                     | only in combinations with number, e.g. "PROL.SG", or with personal pronouns, e.g. "PRO.1SG.PROL" |
| pst                   | persnum-series     | 1SG.S.PST, ...  | past tense cross-reference series  | only in combinations with person and number and conjugation type, e.g. "3DU.S.PST"               |
| ptcl                  | part of speech     |   | particle                           |  |
| ptcp                  | v-nfin             | PTCP.ABES, PTCP.ANT,<br>PTCP.ANT2,<br>PTCP.PASS,<br>PTCP.POST,<br>PTCP.POST2,<br>PTCP.PRF, PTCP.SIM | participle                         |  |
| ptcpabes              | v-nfin             | PTCP.ABES   | abessive participle                |  |
| ptcpant               | v-nfin             | PTCP.ANT  | anterior participle                |  |
| ptcpant2              | v-nfin             | PTCP.ANT2   | anterior participle 2              |  |
| ptcppass              | v-nfin             | PTCP.PASS   | passive participle                 |  |
| ptcppost              | v-nfin             | PTCP.POST   | posterior participle               |  |
| ptcppost2             | v-nfin             | PTCP.POST2  | posterior participle 2             |  |
| ptcpprf               | v-nfin             | PTCP.PRF  | perfect participle                 |  |
| ptcpsim               | v-nfin             | PTCP.SIM  | simultaneous participle            |  |
| quant                 | part of speech     |   | quantifier                         |  |
| reln                  | part of speech     |   | relational noun                    |  |
| res1                  | deriv-v            | RES1  | resultative 1                      |  |
| res2                  | deriv-v            | RES2  | resultative 2                      |  |
| sbjv                  | v-tam              | SBJV  | subjunctive                        |  |
| sbjv2                 | v-tam              | SBJV2   | subjunctive 2                      |  |
| sel                   | deriv-misc         | SEL   | selective                          |  |
| sg                    | n-num              | ABL.SG, DST.SG,<br>LAT.SG, LAT.SG2,<br>LOC.SG, NOM.SG,<br>OBL.SG, PROL.SG                           | singular number                    | only in combinations with case, e.g. "LAT.SG"  |
| sim                   | v-nfin             | CVB.SIM   | simultaneous converb               |  |

| <b>Tsakorpus<br/>grammar tag</b> | <b>Tsakorpus<br/>category</b> | <b>Gloss</b>                                | <b>Description</b>        | <b>Comment</b>   |
|----------------------------------|-------------------------------|---|---------------------------|--|
| simil                            | deriv-misc                    | SIMIL                                       | similative (any)          |  |
| similfut                         | v-tam                         | SIMIL.FUT                                   | similative, future tense  |  |
| similn                           | n-misc                        | SIMIL                                       | similative (nominal)      |  |
| similprs                         | v-tam                         | SIMIL.PRS                                   | similative, present tense |  |
| similpst                         | v-tam                         | SIMIL.PST                                   | similative, past tense    |  |
| soc                              | n-misc                        | SOC   | sociative                 |  |
| soc2                             | n-misc                        | SOC2  | sociative 2               |  |
| subjc                            | persnum-conj                  | 1SG.S, ...; 1PL.S/SG.O,<br>...              | subjective conjugation    | only in combinations with person<br>and number, e.g. "3DU.S" |
| sup                              | v-nfin                        | SUP   | supine                    |  |
| top                              | deriv-misc                    | TOP   | topical marker            |  |
| tr                               | deriv-v                       | TR1, TR2, TR3, TR4,<br>TR5, TR6             | transitivizer (any)       |  |
| tr1                              | deriv-v                       | TR1   | transitivizer 1           |  |
| tr2                              | deriv-v                       | TR2   | transitivizer 2           |  |
| tr3                              | deriv-v                       | TR3   | transitivizer 3           |  |
| tr4                              | deriv-v                       | TR4   | transitivizer 4           |  |
| tr5                              | deriv-v                       | TR5   | transitivizer 5           |  |
| tr6                              | deriv-v                       | TR6   | transitivizer 6           |  |
| trl                              | deriv-v                       | TRL1, TRL2, TRL3                            | translative (any)         |  |
| trl1                             | deriv-v                       | TRL1  | translative 1             |  |
| trl2                             | deriv-v                       | TRL2  | translative 2             |  |
| trl3                             | deriv-v                       | TRL3  | translative 3             |  |
| uncert                           | misc                          | %ADJZ2, %DRV,<br>%ESS2, %NMLZ               | uncertain categories      |  |
| unkn                             | misc                          | %%  | unknown morph             |  |
| v                                | part of speech                |   | verb                      |  |
| vblz                             | deriv-v                       | VBLZ1, VBLZ2, VBLZ3,<br>VBLZ4, VBLZ5, VBLZ6 | verbalizer (any)          |  |
| vblz1                            | deriv-v                       | VBLZ1                                       | verbalizer 1              |  |
| vblz2                            | deriv-v                       | VBLZ2                                       | verbalizer 2              |  |
| vblz3                            | deriv-v                       | VBLZ3                                       | verbalizer 3              |  |
| vblz4                            | deriv-v                       | VBLZ4                                       | verbalizer 4              |  |
| vblz5                            | deriv-v                       | VBLZ5                                       | verbalizer 5              |  |
| vblz6                            | deriv-v                       | VBLZ6                                       | verbalizer 6              |  |
| voc                              | n-misc                        | VOC   | vocative form             |  |

## Appendix A6. Part-of-speech and morphological category tags (tier mc, ps)

Table 39. Part-of-speech tags (tiers mc, ps)

| Label    | Description           |
|----------|-----------------------|
| adj      | adjective             |
| adv      | adverb                |
| aux      | auxiliary             |
| conj     | conjunction           |
| dem      | demonstrative pronoun |
| intj     | interjection          |
| interrog | interrogative pronoun |
| n        | noun                  |
| npr      | proper noun           |
| num      | numeral               |
| pers     | personal pronoun      |
| pp       | postposition          |
| pro      | pronoun               |
| ptcl     | particle              |
| quant    | quantifier            |
| reln     | relational noun       |
| v        | verb                  |
| %%       | unknown               |

Table 40. Morphological categories tags (tier mc)

| Label     | Description  |
|-----------|--|
| n(avd)    | noun of the voiced alternating inflectional class                      |
| n(avd(l)) | noun of the voiced alternating inflectional class with basic l-stem    |
| n(avd(n)) | noun of the voiced alternating inflectional class with basic n-stem    |
| n(avd(o)) | noun of the voiced alternating inflectional class with basic o-stem    |
| n(avd(r)) | noun of the voiced alternating inflectional class with basic r-stem    |
| n(avd(u)) | noun of the voiced alternating inflectional class with basic u-stem    |
| n(avd(z)) | noun of the voiced alternating inflectional class with basic z-stem    |
| n(avs)    | noun of the voiceless alternating inflectional class                   |
| n(avs(s)) | noun of the voiceless alternating inflectional class with basic s-stem |

| Label         | Description   |
|---------------|---|
| n(avs(z))     | noun of the voiceless alternating inflectional class with basic z-stem        |
| n(d)          | noun of the default inflectional class  |
| n:case        | nominal case marker   |
| n:dst         | nominal destinative marker  |
| n:num         | nominal number marker   |
| n:pn          | nominal person-number cross-reference marker                                  |
| n:poss        | nominal possessive marker (case is expressed cumulatively)                    |
| reln:case     | relational noun case marker   |
| v(avd)        | verb of the voiced alternating inflectional class                             |
| v(avd(n))     | verb of the voiced alternating inflectional class with basic n-stem           |
| v(avd(n,irr)) | irregular verb of the voiced alternating inflectional class with basic n-stem |
| v(avd(o))     | verb of the voiced alternating inflectional class with basic o-stem           |
| v(avd(o,irr)) | irregular verb of the voiced alternating inflectional class with basic o-stem |
| v(avd(r))     | verb of the voiced alternating inflectional class with basic r-stem           |
| v(avs)        | verb of the voiceless alternating inflectional class                          |
| v(d)          | verb of the default inflectional class  |
| v(d(d))       | verb of the default inflectional class taking d-allomorphs                    |
| v(d(irr))     | irregular verb of the default inflectional class                              |
| v(d(z))       | verb of the default inflectional class taking z-allomorphs                    |
| v:conj        | verbal marker of middle conjugation or non-singular object                    |
| v:nfin        | verbal marker of uninflected non-finite forms                                 |
| v:nfinp       | verbal marker of non-finite forms taking possessive markers                   |
| v:pn          | verbal person-number marker   |
| v:poss        | verbal possessive marker (in non-finite forms)                                |
| v:TAM         | verbal tense-mood marker  |
| v:TAM.pn      | verbal tense-mood marker and person-number marker                             |
| %%            | unknown   |