

Instructions for the user of Pavlik Morozov Corpus

1st February 2019

The parallel text corpus of Research Unit for Volgaic Languages in University of Turku includes translations of a youth novel *Pavlik Morozov* (1947) by Vitali Gubarev organized in equivalent sentences. The corpus has fourteen translations and the original Russian text, altogether fifteen texts. The division into equivalent sentences was originally made for the Russian original and then introduced to the whole corpus. The division was made mechanically and may thus not always be appropriate.

Notable is that the translation into Chuvash is not original, but made afterwards to complete the corpus, because the original text was not found when collecting the material. The translation is made by Eduard Fomin in 2005. The Erzya translation exists in two versions, of which the first is a translation made for the corpus and the second is the later on added original translation from 1953, which was not found earlier. The first version of the Erzya translation made for the corpus was composed by Svetlana Saadat (née Gausheva). This translation was later corrected by Nina Agafonova. The translations were made between 2018 and 2010. The Erzya translation taken to the corpus is the one by Agafonova.

The original Russian text was taken to the corpus from the books with Khanty and Mansi translations. These books also included the original text.

BIBLIOGRAPHIC DATA

The bibliographic data about the original texts is listed below.

Original Russian text:

Gubarev = Губарев, Виталий 1947: *Павлик Морозов*. Государственное Учебно-педагогическое издательство Министерства просвещения РСФСР, Москва, Ленинград.

Finnish translation:

Gubarev, Vitali 1953: *Pavlik Morozov*. Suomentanut M. Liukkonen. Koululaisen kirjasto. Karjalais-suomalaisen SNT:n valtion kustannusliike, Petroskoi.

Erzya translation:

Gubarev = Губарев, Виталий 1953: *Павлик Морозов*. Перевод с русского П. Шмелёвой. Мордовской книжной издательства, Саранск.

Moksha translation:

Gubarev = Губарев, Виталий 1953: *Павлик Морозов*. Рузоннеста ётафтозе Н. М. Неякин. Мордовский книжной издательства, Саранск.

Meadow Mari translation:

Gubarev = Губарев, Виталий 1959: *Павлик Морозов*. Рушла гыц И. М. Смирнов кусарен. Книгам лукмо марий издательство, Йошкар-Ола.

Hill Mari translation:

Gubarev = Губарев, Виталий 1960: *Павлик Морозов*. Арк. Каюшков сӓрен. Книгам лыкшы мары издательство, Йошкар-Ола.

Komi translation:

Gubarev = Губарев, Виталий 1950: *Павлик Морозов*. Перевод Н. Денисовалӧн. Школьной библиотека. Коми Государственной Издательство, Сыктывкар.

Permian Komi translation:

Gubarev = Губарев, Виталий 1957: *Павлик Морозов*. Переводсӧ керис З. А. Тетюева. Коми-Пермяцкӧй книжнӧй издательство, Кудымкар.

Udmurt translation:

Gubarev = Губарев, Виталий 1953: *Павлик Морозов*. Редактор перевода Л. Осотова. Удмуртской книжной издательство, Ижевск.

Khanty translation:

Gubarev = Губарев, Виталий 1952: *Павлик Морозов*. Ханты ясынга К. Н. Дабин ханшис. Государственное Учебно-педагогическое издательство Министерства просвещения РСФСР Ленинградское отделение, Ленинград, Москва.

Mansi translation:

Gubarev = Губарев, Виталий 1953: *Павлик Морозов*. Н. М. Садомин русь латынгныл маньси латнгыг толмасьлас. Государственное Учебно-педагогическое издательство Министерства просвещения РСФСР, Москва, Ленинград.

Hungarian translation (Budapest):

Gubarjov, Vitalij 1952: *Pavlik Morozov*. Fordította F. Kemény Márta. Ifjúsági Könyvkiadó; Budapest.

Hungarian translation (Bucharest):

Gubarjev, Vitalij 1952: *Pavlik Morozov*. Fordította Kézdi Lenke. Az imsz. kv. Ifjúsági Könyvkiadója, Bukarest.

Tatar translation:

Gubarev = Губарев, Виталий 1953: *Павлик Морозов*. С. Фәйзуллина тәржемәсе. Урта яшьтәге балалар өчен. Татгосиздат, Казань.

REFERENCES

The abbreviation of the corpus is PM, which in the references is followed by the English abbreviation of the language of the translation. The English can be replaced by other language as well, if it is more practical. When referring to the Erzya texts, the year of the translation must also be mentioned. The English abbreviations are the following:

Finnish = Fin

Erzya = Erz1953 or Erz2010

Moksha = Mok

Meadow or Eastern Mari = MaE
Hill or Western = MaW
Komi = Zyr
Permian Komi = Per
Udmurt = Udm
Khanty = Kha
Mansi = Man
Hungarian (Budapest) = HBp
Hungarian (Bucharest) = HBk
Tatar = Tat
Chuvash = Chu
Russian = Rus

When referring to an equivalent sentence, the reference must contain the abbreviation of the corpus, the language abbreviation of the translation and the number of the equivalent sentence, which is found in the text. Structure of the whole reference is NAME OF THE CORPUS/LANGUAGE:THE EQUIVALENT SENTENCE(S). For instance, the equivalent sentences 125–129 of the Khanty translation in the Pavlik Morozov corpus are referred by the marking PM/Kha:125–129.

The users of the parallel text corpora can also mention in the description of their research material that the corpora have been composed by the research Unit for Volgaic Languages in University of Turku (Volgan alueen kielten tutkimusyksikkö, Turun yliopisto).

ABOUT THE TRANSLATIONS NOT USED IN THE CORPUS

The Estonian version was scanned for the corpus, but the translation turned out to be so free, that it was not possible to divide it into the equivalent sentences. Therefore it was not included into the corpus. The Nenets version was ordered from Saint Petersburg and scanned, but it has neither been converted with the optical character recognition nor divided into the equivalent sentences.