

The fruits of the project VepKar became visible in Wikipedia and how YouTube helps make multilingual subtitles for Wikimedia Commons

<http://dictorpus.krc.karelia.ru>

Andrew Krizhanovsky



Open corpus of Veps and Karelian languages (VepKar)

**Представление материалов научного проекта
ВепКар в Википедии и о том, как YouTube помогает
делать многоязычные субтитры для Викисклада**

Wiki conference 2018. Saint Petersburg, Russia

1. Veps language corpus (2009-2016)

Open source-code, open data, open, open, open, open....

Open corpus!

2. Open corpus of Veps and Karelian languages (2016 - ...)

Veps language corpus* (1)

Karelian Research Centre of the Russian Academy of Sciences (KarRC RAS)

- Institute of Language, Literature and History of KarRC RAS
- Institute of Applied Mathematical Research of KarRC RAS



Online: <http://vepsian.krc.karelia.ru>

*) N.G. Zaitseva, M.M. Filatova, N.L. Shibanova, A.A. Krizhanovsky. Veps corpora // International scientific conference "Corpus linguistics". Saint Petersburg, 2015. Pp. 202-212.

**1. Veps language corpus
(2009-2016), pure PHP, MySQL**

**2. Open corpus of Veps and
Karelian languages (2016 - ...),
Laravel, MySQL**

Artwork of our designer at Commons



Natalya Vdovitsyna
[[User:Natty610]]

[[commons:Category:Corpus linguistics]]

[[commons:Category:VepKar]]



open corpus of Veps
and Karelian languages

Video in native language

Project's deliverables

Popular science & public relation:

- YouTube,
- Social networks (vk.com, facebook),
- Wikipedia



Camera, tripod, audio recording by dictaphone

Refresh Sequencer Use Backdrop

01:31+15

00:10 00:20 00:30 00:40 00:50 01:00 01:10 01:20 01:30 01:40 01:50 01:60 02:10 02:20 02:30



- Video editor:**
1. Merge audio and video.
 2. Combine clips.
 3. Remove errors.

Refresh Sequencer Use Backdrop

01:31+15

01:12 01:13 01:14 01:15 01:16 01:17 01:18 01:19 01:20 01:21 01:22 01:23 01:24 01:25 01:26 01:27 01:28 01:29 01:30 01:31 01:32 01:33 01:34 01:35 01:36

Channel: 0

Final Length: 00:00:00
Playhead: 2706
Frame Offset 652:1766
Frame Still 0:0

▼ Sound

Vepkar 5 F X

../.audio/VepkarVeps_1...

Pack Caching

Mono

Draw Waveform

Volume: 0.000

Pitch: 1.000

Pan: 0.000

YouTube Creator Studio

Manage subtitles and closed captions

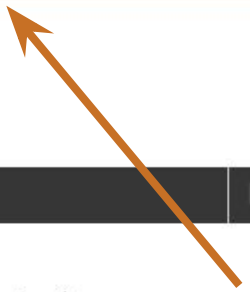
Add new subtitles or CC

Veps language and VepKar corpus by Nina Zaitseva, 2018

English	Published
Russian	Published

Search 189 other languages

- Faroese
- Fijian
- Filipino
- Finnish** Published
- French
- French (Belgium)
- French (Canada)
- French (France)
- French (Switzerland)



Too few



Type subtitle here then press Enter +

- 0:01.1 I'm Nina Zaitseva, and I'm speaking now
- 0:05.1 in Veps language.
- 0:05.4 Veps language is a part of the Baltic-Finnish branch of the Finno-Ugric language family. ✕
- 0:12.0 +
- 0:12.1 Veps live in three regions: in Karelia, in the
- 0:22.7 Leningrad and Vologda regions

Veps language and VepKar corpus by Nina Zaitseva, 2018

Veps language is a part of the Baltic-Finnish branch of the Finno-Ugric language family.

0:07 / 2:29

0:02 0:03 0:04 0:05 0:06 0:07 0:08 0:09 0:10 0:11

Nina Zaitseva, and I'm speaking in Veps language.

Veps language is a part of the Baltic-Finnish branch of the Finno-Ugric language family.

Pause video while typing

*Create subtitles
in selected language
in YouTube Creator
Studio*

View published subtitles and CC: Russian
Subtitles/CC source: IAMR KarRC RAS

Download YouTube subtitles in SRT format

Edit

Actions ▾

- Rename
- Delete
- Download
- Original format
- .vtt
- .srt**
- .sbv

Григорьевна Зайцева и моя
зучит на вепском языке.

ий язык входит в прибалтийско-
ю ветвь финно-угорской
ой семьи.

живут в трех регионах: в
и, в Ленинградской и
дской областях.



0:22.8 в районе Межозерья между Онежским,
0:28.4 Ладожским и Белым озерами.

→ ↻ https://commons.wikimedia.org 110% ... ☆ >>

English AKA MBG talk preference

timedtext discussion create wat

WIKIMEDIA COMMONS

Create new translation or edit existing

Select language and press the **Go** button

vep - vepsän kel' Go

- vec - vèneto
- vep - vepsän kel'**
- vi - Tiếng Việt
- vls - West-Vlams
- vmf - Mainfränkisch
- vo - Volapük
- vot - Vad'da
- vro - Võro
- wa - walon
- war - Winaray
- wo - Wolof
- wuu - 吴语

Search Wikimedia Cor

Go Search

Main page Welcome Community portal Village pump Help center

Upload file Recent changes Latest files Random file Contact us

Wikimedia Commons timedtext

Upload video,
then create
timedtext, copy
subtitles text.

Huge list of
languages to
create subtitles



VepKar videos at Commons and WP



Nina Zaitseva / Нина Григорьевна Зайцева

File:Veps language and VepKar corpus by Nina Zaitseva 2018.webm

Subtitles at Commons: Veps, English, Russian.

Wikipedia:

- Veps language
- Вепсский язык
- Зайцева, Нина Григорьевна



Boiko
Tatyana

Татьяна
Петровна
Бойко

File:Livvi-Karelian language and VepKar corpus by Tatiana Boiko.ogv

Subtitles at Commons:
Livvi-Karelian, English, Russian.

Wikipedia:

- Livvi-Karelian language
- Ливвиковское наречие

Conclusion



The study was supported by the Russian Foundation for Basic Research, grant No. 18-012-00117, “Problems of construction of text corpus of minor peoples of Russia in the case of Open corpus of Veps and Karelian languages”.

Проблемы создания корпусов языков малочисленных народов России на примере Открытого корпуса вепсского и карельского языков.

Positive campaigning & Popular science

Open software

Dictorpus at GitHub

<https://github.com/componavt/dictorpus>

Open Video

YouTube, Wikimedia commons

[Commons:Category:VepKar](https://commons.wikimedia.org/wiki/Category:VepKar)

Open data

Linguistics data has open license
(CC-BY)

<http://dictorpus.krc.karelia.ru/en/dumps>

Social network

Vkontakte, [Facebook](#)

<https://vk.com/karrc>,

Karelian Research Centre of RAS, CC BY 4.0

