

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

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)

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

[[commons:Category:Corpus linguistics]]

[[commons:Category:VepKar]]



Natalya Vdovitsyna
[[User:Natty610]]



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

Photo of camera, tripod, audio recording, clips.

The top-left panel shows a multi-track timeline. Tracks 1 through 7 contain audio waveforms. Track 3 contains a video strip labeled 'VepKar_Veps_two.003: //..VepKar_Veps_two.mo'. Track 1 contains a video strip labeled 'VepKar_Veps_one.001: //..VepKar_Veps_one.mov | 4976'. A green playhead is positioned at 01:31+15. The timeline axis is marked from 00:10 to 02:30.



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

The bottom-left panel shows a single track timeline. It contains a merged strip labeled 'VepKar_Veps_two.003: //..VepKar_Veps_two.mov | 4460'. The strip is a long blue rectangle with a teal audio waveform on top. A green playhead is positioned at 01:31+15. The timeline axis is marked from 01:12 to 01:36.

The properties panel on the right shows settings for an audio strip. It includes a 'Strip' section with 'Final Length: 00:00:00:00', 'Playhead: 2706', 'Frame Offset 652:1766', and 'Frame Still 0:0'. The 'Sound' section shows a volume of 5, a file path '://.Audio/VepkarVeps_1...', and options for 'Pack', 'Caching', 'Mono', and 'Draw Waveform'. The 'Volume' is set to 0.000, 'Pitch' to 1.000, and 'Pan' to 0.000.

YouTube Creator Studio

Manage subtitles and closed captions

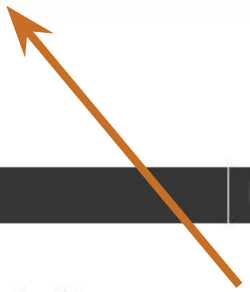
Add new subtitles or CC

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-
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

rch

Search Wikimedia Cor

Go Search

rigate

Main page
Welcome
Community portal
Village pump
Help center

ticipate

Upload file
Recent changes
Latest files
Random file
Contact us

le

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

Wikimedia Commons timedtext

Upload video,
then create
timedtext, copy
subtitles text.

Huge list of
languages to
create subtitles



VepKar videos at Commons and WP

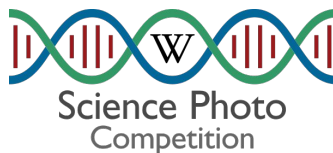


File:Veps language and VepKar corpus by Nina Zaitseva 2018.webm

Subtitles at Commons: Veps, English, Russian.

Wikipedia:

- Veps language
- Вепсский язык
- Зайцева, Нина Григорьевна



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



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>

Karelian Research Centre of RAS

