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

Natalya Vdovitsyna
[[User:Natty610]]

[[commons:Category:Corpus linguistics]]

[[commons:Category:VepKar]]

Vep Kar

OTRS

open corpus of Veps and Karelian languages

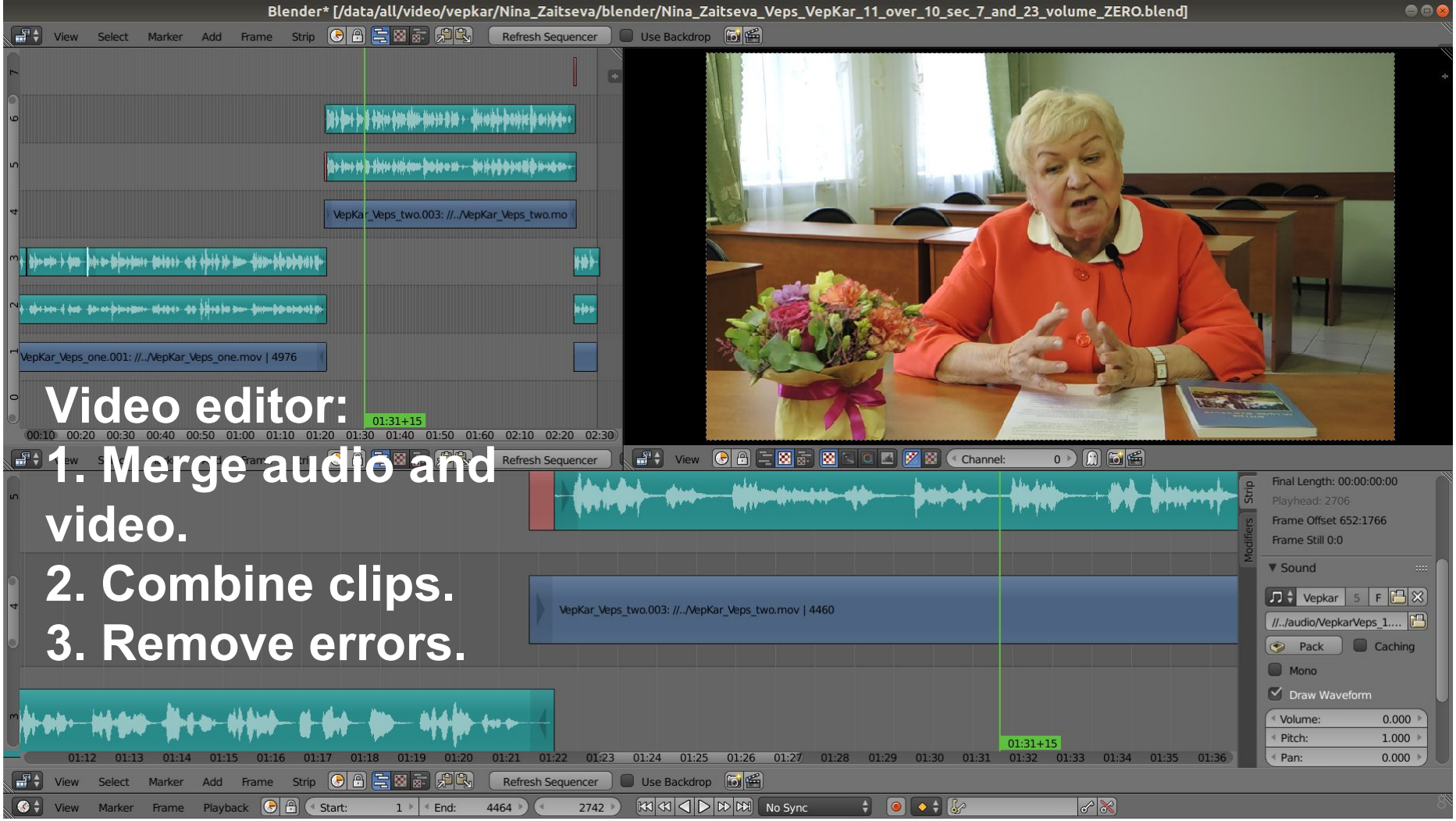
Video in native language

Project’s deliverables

Popular science & public relation:

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

Photo of camera, tripod, audio recording, clips.


Blender* [/data/all/video/vepkar/Nina_Zaitseva/blender/Nina_Zaitseva_Veps_VepKar_11_over_10_sec_7_and_23_volume_ZERO.blend]
View Select Marker Add Frame Strip
Refresh Sequencer
Use Backdrop
VepKar_Veps_two.003: //../VepKar_Veps_two.mo
VepKar_Veps_one.001: //../VepKar_Veps_one.mov | 4976
01:31+15
Video editor:
1. Merge audio and video.
2. Combine clips.
3. Remove errors.
Channel: 0
VepKar_Veps_two.003: //../VepKar_Veps_two.mov | 4460
Final Length: 00:00:00:00
Playhead: 2706
Frame Offset 652:1766
Frame Still 0:0
Sound
Vepkar
//../audio/VepkarVeps_1....
Pack
Caching
Mono
Draw Waveform
Volume: 0.000
Pitch: 1.000
Pan: 0.000
Strip
Modifiers
View Marker Frame Playback
Start: 1
End: 4464
2742
No Sync

Secure | https://www.youtube.com/timedtext_video?v=huKvkscI6CQ

YouTube Creator Studio

Manage subtitles and closed captions

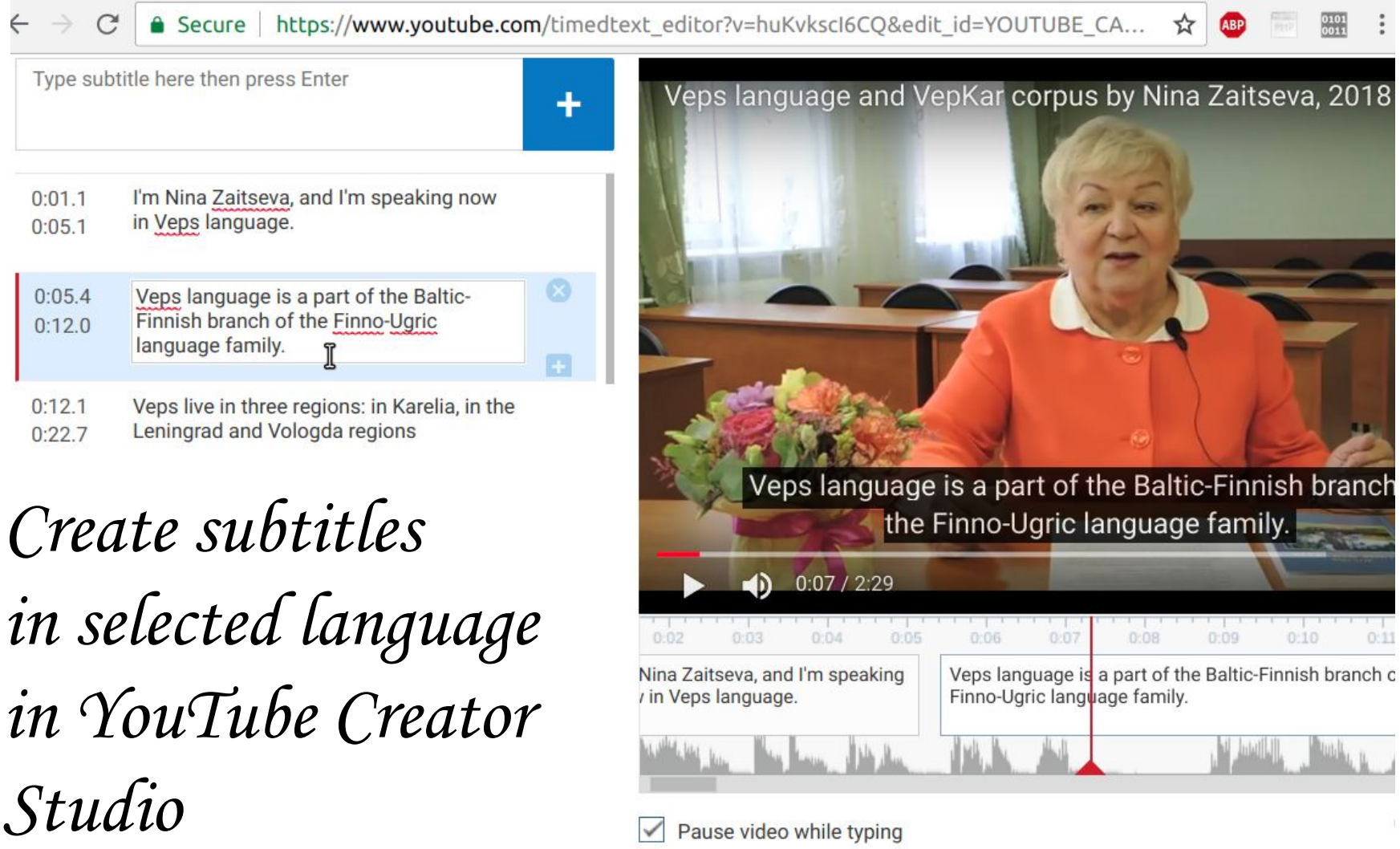
Add new subtitles or CC

English	Published
Russian	Published

Search 189 other languages

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

Too few



Create subtitles in selected language in YouTube Creator Studio

YouTube RU

Search

Info & Settings | Enhancements | Audio | End screen & Annotations | Cards | Subtitles/CC

View published subtitles and CC: Russian

Subtitles/CC source: IAMR KarRC RAS

Edit

Actions

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

Download YouTube subtitles in SRT format

Григорьевна Зайцева, и моя
зучит на вепсском 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.

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

Science Photo Competition

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 data

Linguistics data has open license (CC-BY)

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

Open Video

YouTube, Wikimedia commons

Commons:Category:VepKar

Karelian Research Centre of RAS