Self-supervised pretraining for low resource languages in Speech processing

Julien Hauret & Nicolas Dufour

Algorithms for speech and natural language processing

MVA 2020-2021

ENS Paris-Saclay

March 25th, 2021

I. Presentation of the task

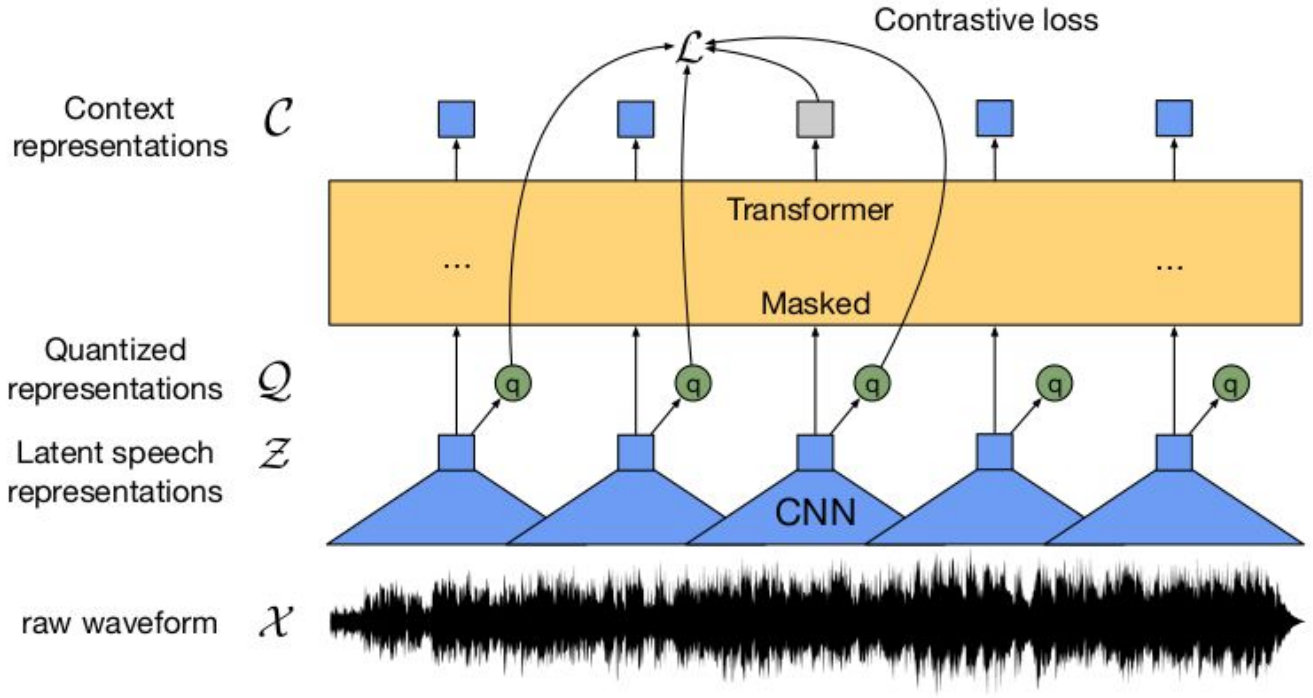
Phoneme recognition

ʒ a d ɔ ʁ l ə k a m ɑ̃ b ɛ ʁ t

II. Models And metrics

January 22th, 2021

Wav2Vec2 2.0 architecture



CTC Loss

For an input, like speech

Predict a sequence of tokens

h e e ϵ l ϵ l l o o !

Merge repeats, drop ϵ

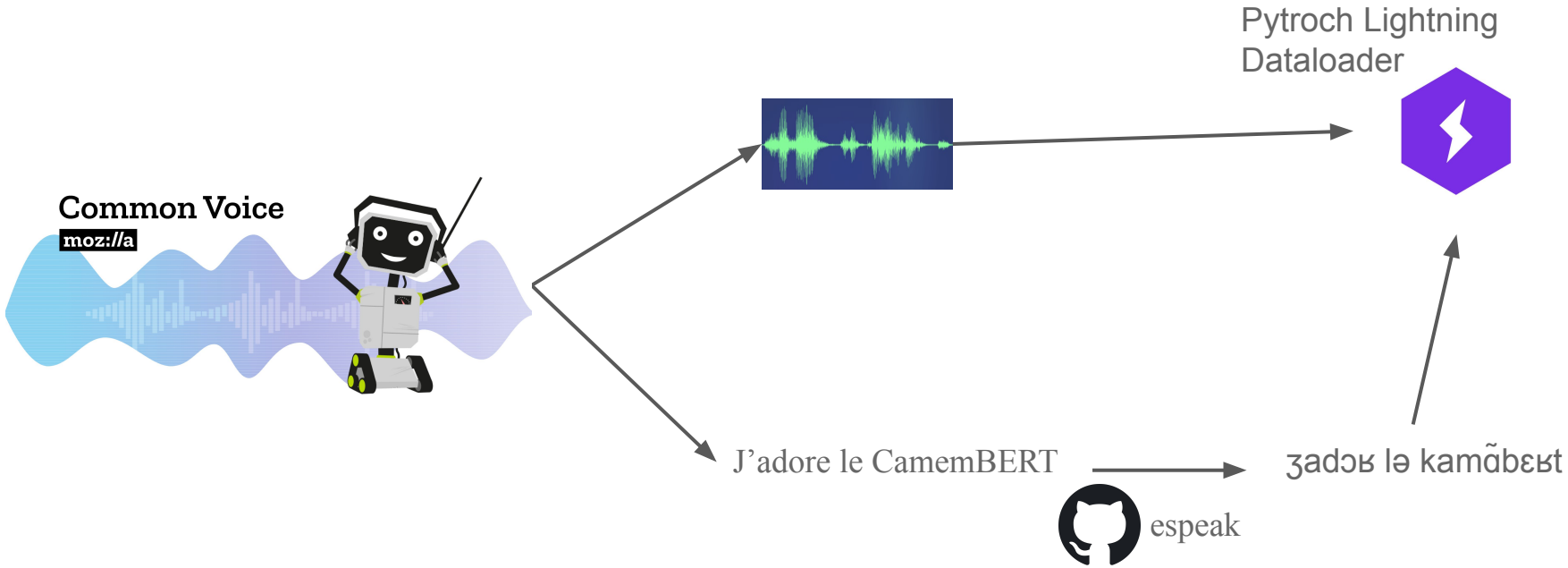
h e l l o !

Final output

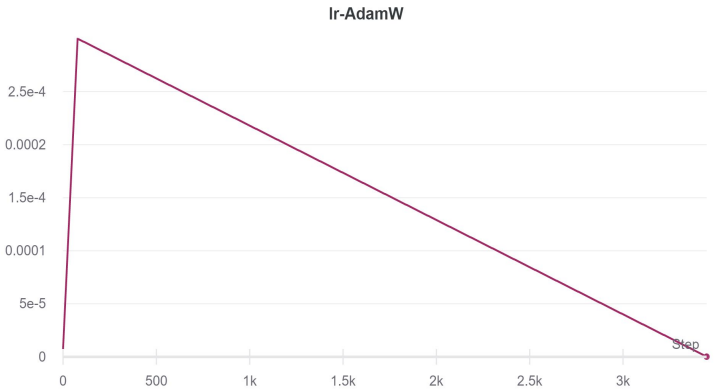
h e l l o !

Picture take from https://distill.pub/2017/ctc/

Data preprocessing



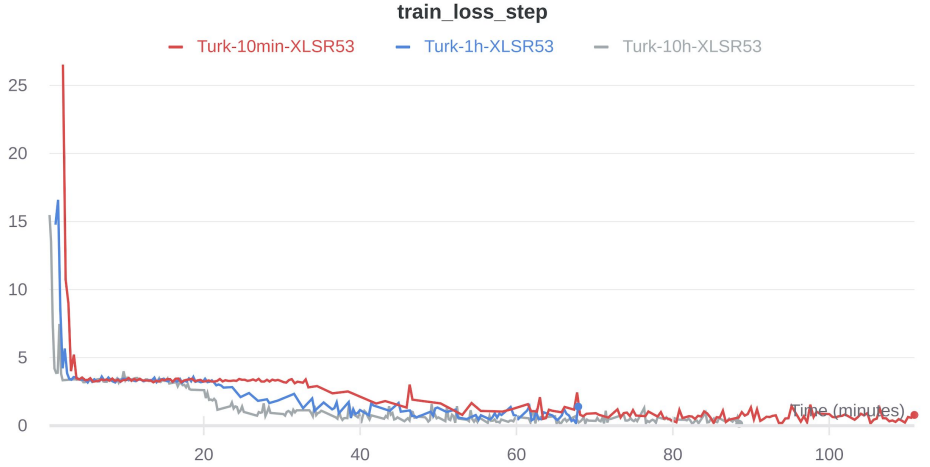
Training procedure

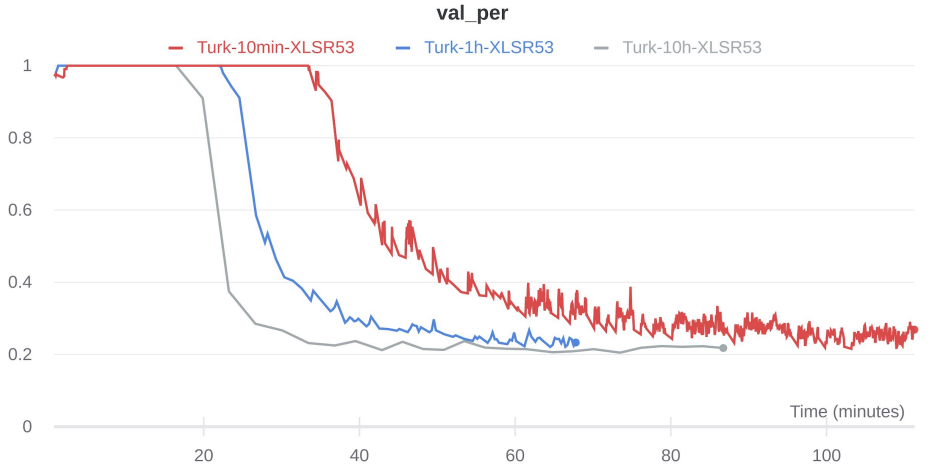


Learning Schedule

- AdamW with max lr of 3e-4
- Model selection on val PER
- 30 epochs

Training curves :

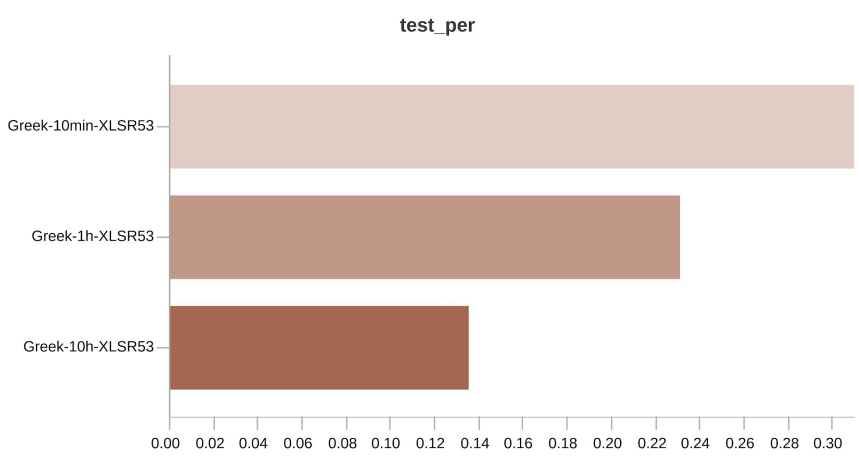




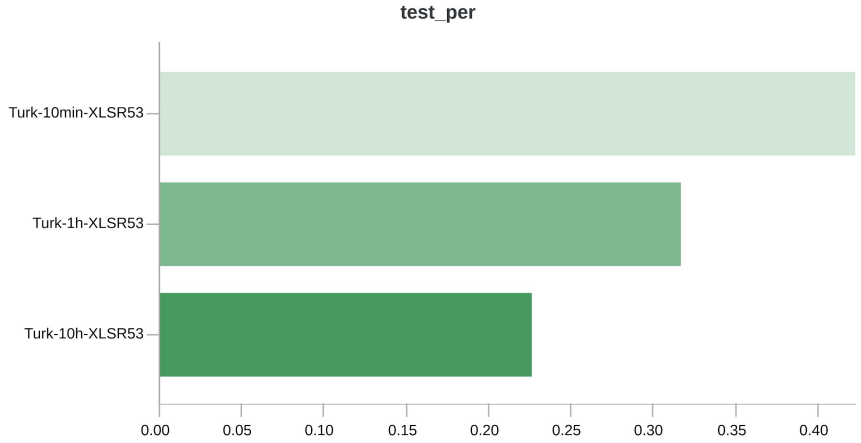
III. Experiments

January 22th, 2021

Impact of data quantity : low proportion in XLSR-53

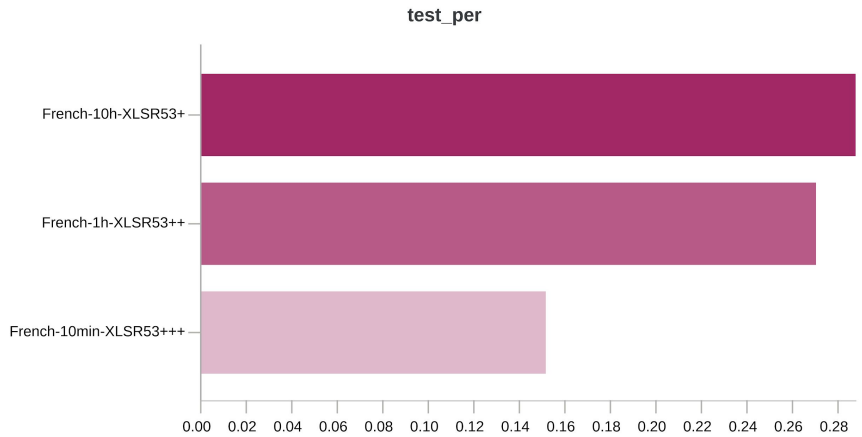


Greek

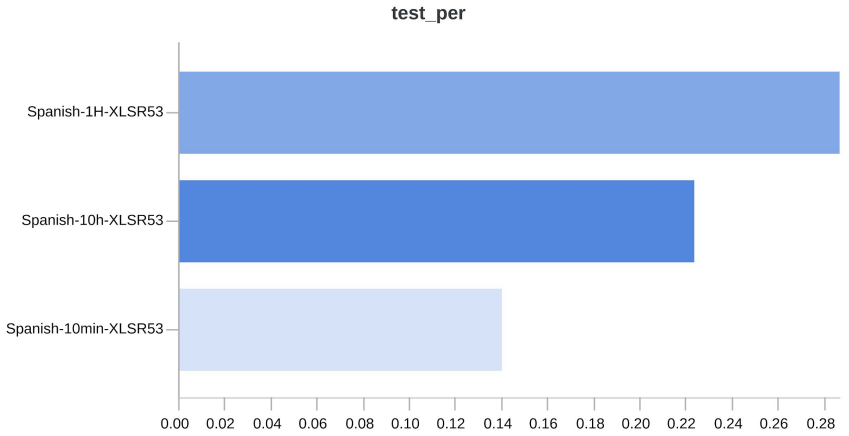


Turk

Impact of data quantity : high proportion in XLSR-53

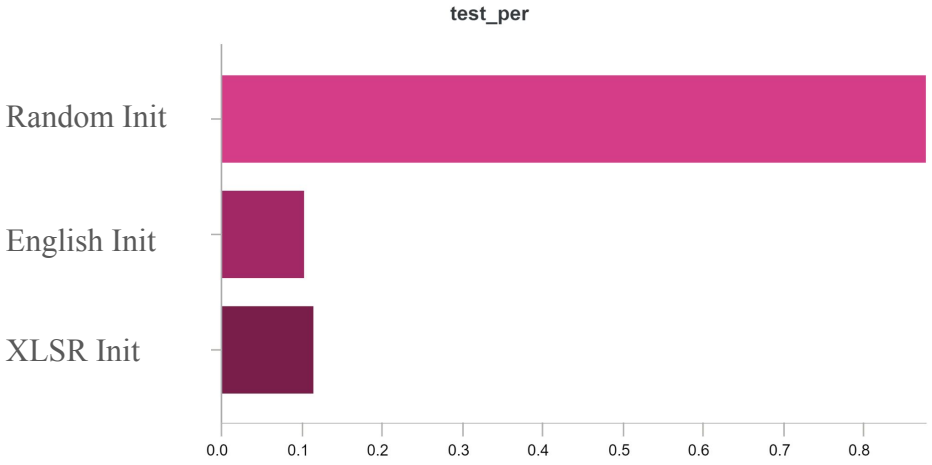


French

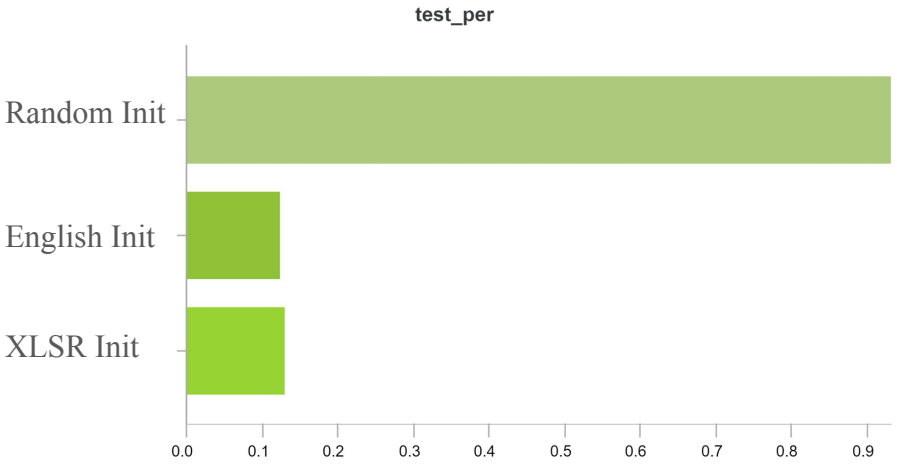


Spanish

What's the best pretrained model?

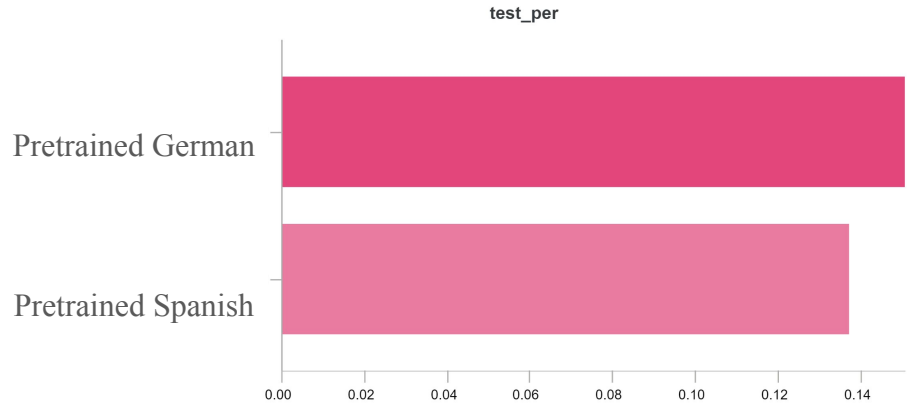


Czech training



Greek training

Does it matter on which language we pretrain?



Portuguese training

Dutch training

école
normale
supérieure
paris–saclay

Conclusion

Conclusion

We've train multiple models and studied the properties of fine-tuning for low ressources languages and draw the following conclusion:

- When the language has a small contribution on XLSR-53, training the more data the better.
- When the language has a bigger contribution on XLSR-53, we didn't observe the same correlation.
- We proved that pretraining is without a doubt useful but didn't manage to prove that XLSR-53 was more adapted to fine-tuning than a model only trained on English
- Finally, pre-training on close languages seems to have a positive effect on fine-tuning

Thank you for listening

March 25th, 2021

Julien Hauret & Nicolas Dufour