# Meeting: 05/05/2021

Cécile Macaire



Model: facebook/wav2vec2-large-xlsr-53

Corpus:

Corpus		Yongning Na	Japhug	
	Size	≈ 5 hours	≈ 30 hours	

#### **Evaluation metrics:**

Word Error Rate (WER) / Character Error Rate (CER) / Phoneme Error Rate (PER)

# Recap / Fine-tuning wav2vec2 model

Performances on the training data					
Language	Model	Train size	WER_huggingface	CER_huggingface	
Na	model_60epochs	150 sentences (15 minutes)	0.93921568627451	0.250009224071136	
Na	model_60epochs	300 sentences (30 minutes)	0.785490196078431	0.170423938309412	
Na	model_60epochs	450 sentences (45 minutes)	0.732549019607843	0.155407150499945	
. Na	model_60epochs	600 sentences (1h)	0.708627450980392	0.143046895177656	
Na	model_60epochs	900 sentences (1h20)	0.667058823529412	0.125816330295539	
Na	model_60epochs	1200 sentences (2h)	0.628627450980392	0.13125945606849	
Na	model_60epochs	1500 sentences (2h30)	0.61921568627451	0.115908336100963	
Na	model_60epochs	1786 sentences (3 h)	0.593333333333333	0.111221816303185	
Na	model_60epochs	2100 sentences (3h30)	0.5839215686274509	0.110742093804199	
Japhug	model_60epochs	3186 sentences (5h31)		0.115567978880348	

# Recap / Fine-tuning wav2vec2 model

#### Performances on the test data

Lang	Model	Tones / Clean ?	Train size	Test size	WER_huggingfa ce	CER_huggingfa ce	PER_difflib	CER_levenshtei n	WER_levenshtei n
Na	model_60epochs	Tones = 1	1786 sentences (3 h)	336 sentences (44 minutes)	0.490509059534	0.092234824878	0.079223381141	0.07008193383	0.319199182513
Na	model_60epochs	Tones = 0	1786 sentences (3 h)	336 sentences (44 minutes)	0.398619499568 594	0.082450008594 511		0.059381510518 605	0.332173909457 734
J	model_60epochs	Keep all test data	3186 sentences (5h31)	350 sentences (44 minutes)	0.239303843364 757	0.088794498381 877		0.059904970148 425	0.257242915932 687
J	model_60epochs	Remove data with unmatched timecodes	3187 sentences (5h31)	335 sentences (44 minutes)	0.218903591682 42	0.068786045857 364		0.046614529768 536	0.243567762735 33

### Recap / Fine-tuning wav2vec2 model - Na results

#### What are the main errors?

 $\rightarrow$  segmentation errors on word-level.

#### Examples:

	Ref	['hat', 'thytty1', 'myt', 'thitkhw1']	length = 4	
	Pred	['haldwltyl', 'mylthilkhw1']	length = 2	
	Ref	['połdzwimridzrjhij', 'dwisej', 'tałqai', 'talqai', 'talqai', 'pii']	length = 6	
	Pred	['podqzw1', 'mxdzxJhĩJqwJseJ', 'tædqædtædqædtadqadpid']	length = 3	
→ <b>tones</b> (uni tones = {"J", "7", "4"}, bi tones = {"4", "/4", "√4", "√4"})				

#### Example:

Ref	te <sub>u</sub> . Inique fii soy leqboquoq biqqqoq soqp <mark>sq qmqqquq w.diq</mark> biqtsmqwdq aqsoq
Pred	tehr-Ini-lle-lji] zo/l le-lpo1ho] pi-ldzo] zo-lbæ-l dm-ly-lum-lmz-lni] pi]tsm]my] a]Ro-l

### Recap / Fine-tuning wav2vec2 model - Japhug results

#### What are the main errors?

→ segmentati	on errors on word-level.	$\rightarrow$ chinese borro	$\rightarrow$ chinese borrowings.		
Example:		→ update transc	criptions?		
Ref Pred	['ki', <mark>'kwra</mark> ', 'ɲwkhama'] ['kw <mark>kwra</mark> ', 'nɲwkhama']	$\rightarrow$ wrong timecodes in the transcriptions.			
→ <b>the extern</b> a o).	l sandhi.	ightarrow vowel fusions	s (w-x > x, x-x > x, w-a > a, x-a > a, u-o >		
Example:		Example:			
Ref Pred	kγmbγom mγra ma azo aβlu tu kγmbγo mara ma zo aβlu tu	Ref Pred	աա- <mark>Ն</mark> աէշիօռ ատ- <mark>Ն</mark> աէշիօռ		

### **Current work / N-best hypothesis with beam search**

- "synchronous" beam search algorithm.
- generation of the 100-best hypothesis for all the test data.

d_sentence	Reference	Prediction
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhi-lnu.l le-lwo1 e-lmy-lki1	thin ledgælmy I myrdthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhi-lnu.l le-lwo1 e-lmy-lki1	thin ledgælmy I myrthad dzol thin hidmul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq?dzo.l thi/lhi-lnu.l le-lwo1 e-lmy-lki?	thin ledgælmy I myrthad dzol thin hidmul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhī-lnu.l le-lwo1 ə-lmy-lki1	thin ledgælmy I mydthad dzol thin hidmul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 e-lmy-lki1	thin ledge lunte to the transfer of the transf
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 e-lmy.lki1	thin ledge lund that logg to the tyme for the thing the
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhi-lnu.l le-lwo1 e-lmy-lki1	thin ledge lund that logg that the final ledge should be the same that
0	thi/lle-lgæ1/my.lmy.lthq1dzo.l/thi/lhī-lnu.lle-lwo1/e-lmy.lki1	thin ledgælmy I mydthad daol thin hidham Lowled high
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 a-lmy.lki1	thin ledge lund that logg that the family ledge should be the four files that the same of
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhj1-lnu.l le-lwo1 e-lmy.lki1	thin ledge lund to the thing long that the same thing ledge the same things the same that the same things the
0	thi/lle-lgæ1/my.lmy.lthq1dzo.l/thi/lhī-lnu.lle-lwo1/e-lmy.lki/l	thin ledge lund that logg that the standard from the thing ledge that
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhi-lnu.l le-lwo1 a-lmy-lki1	thin ledge lux that logg that the family ledge for the thing speed with
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 a-lmy.lki1	thin ledge lund fight for the thing for the first form of the thing for
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 a-lmy.lki1	thin ledge lund full logg that the family ledge from the following thin the things of the family states the family state
0	thi/lle-lgæ1lmy.lmy.lthq1dzo.l[thi/lhi-lnu.llle-lwo1la-lmy-lki1]	thin ledge lund that logg that the family ledwod almyddid
0	thi/lle-lgæ1lmy.lmy.lthq1dzo.l[thi/lhi-lnu.llle-lwo1la-lmy-lki1]	thin ledge Imy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1lmy.lmy.lthq1dzo.l[thi/lhi-lnu.llle-lwo1la-lmy.lki1]	thin ledge Imy I mydthan dzol thin hidgel ledwod edmydkil
0	thi/lle/lgæ1/my/my/tha/ldzo/lthi/lhi/nu/lle/wo1/e/my/ki/l	thin ledge lowled Luntin hid logb to the log ledge logic
0	thi/lle/lgæ1/my/my/tha/ldzo/lthi/lhi/nu/lle/wo1/e/my/ki/l	thin ledge lowed Luntin hid logb to the lower specific thin the
0	thi/ le-lgæ1 my.lmy.lthq1dzo.l thi/ hji-lnu.l le-lwo1 a-lmy-lki1	thin ledge lmy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 a-lmy-lki1	thin ledge lmy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 e-lmy-lki1	thin ledge Imy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 e-lmy-lki1	thin ledge Imy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l[thi/lhi-lnu.l le-lwo1 a-lmy-lki1	thin letgælmy) mytthat dzol thin hithuul letwot etmytkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhi-lnu.l le-lwo1 e-lmy-lki1	thin ledge lmy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/ le-lgae1 my.lmy:\lthq1dzo.l thi/ hji-lnu.l le-lwo1 e-lmy-lki1	thin ledge lmy I mydthad dzol thin hidgul ledwod edmydkil
0	thi/lle-lgæ1 my.lmy.lthq1dzo.l thi/lhji-lnu.l le-lwo1 e-lmy-lki1	thin ledge lmy I mydthad dzol thin hidnul ledwod edmydkil
0	thi/lle-lae-1/my/my/thaidzo/jthi/lhi-lnu/lle-lwo1/e-lmy-lkii/	this leage my mysthat dzol this highwal leaves earnyskil

### Current work / N-best hypothesis with beam search

#### Computation of the oracle scores :

Metrics	WER_HuggingFace	CER_HuggingFace	WER_lev	CER_lev
Oracle score	0.4027450980392157	0.06922764677663383	0.24350706955983692	0.05870925083702372
Gain	8,78 %	2,3 %	7,56 %	1,13 %

# **Objectives**

#### **Short-term**

- Generate the predictions with the use of a language model in the decoder.
- Use the n-best hypothesis to improve the predictions.
- Implement a word segmentation model.

### Long-term

- Use the knowledge of dictionaries to improve the performances.
- > Fine-tune the model on the all japhug corpus.

# **Open questions - Japhug**

Working on related languages (i.e. rGyalrong languages)?

Explore the transfer learning between speakers?