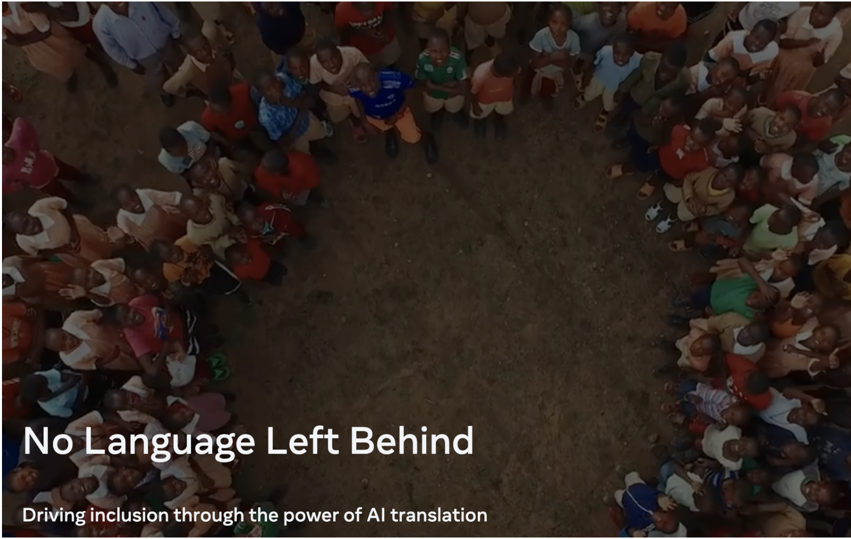
Massively Multilingual Text and Speech Mining

NLLB Team
presented by
Kevin Heffernan and Holger Schwenk

MT Marathon
September 5th 2022

Agenda for talk

- No Language Left Behind: translating 200+ languages
- LASER: Language Agnostic SEntence Representations
- Mining text
- Mining speech
- Conclusion



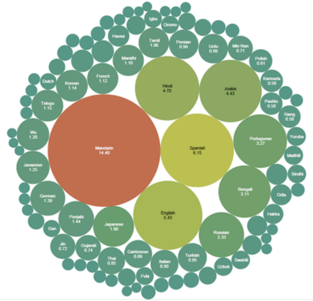
No Language Left Behind

Driving inclusion through the power of AI translation

Context and Motivation

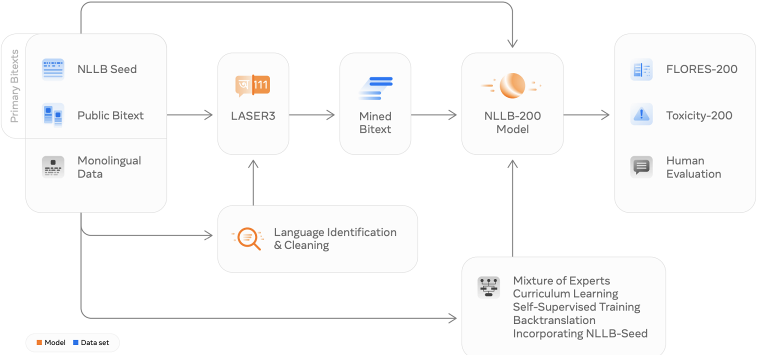
- 7 151 living languages
- 40% are endangered
- 23 languages account for half the population
- 200 languages ⇒ 88%
- ≈ 4 000 with developed writing system
- Multilingual approaches: ≈ 130 languages

Native speakers



⇒ How can we scale well beyond 100 languages?

Scaling to 200+



Primary Bitexts

NLLB Seed

Public Bitext

Monolingual Data

LASER3

Mined Bitext

NLLB-200 Model

FLORES-200

Toxicity-200

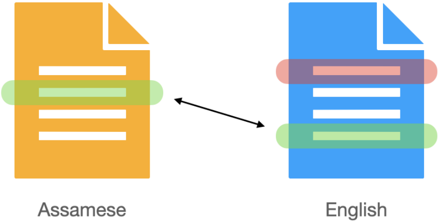
Human Evaluation

Language Identification & Cleaning

Mixture of Experts
Curriculum Learning
Self-Supervised Training
Backtranslation
Incorporating NLLB-Seed

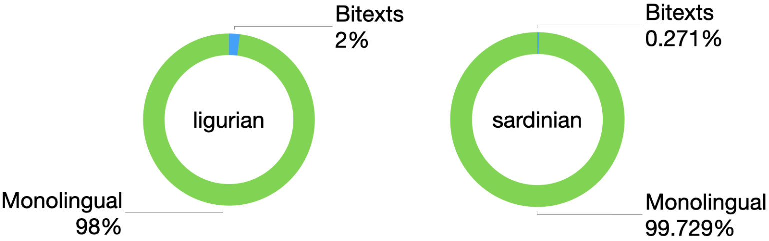
🟧 Model 🟦 Data set

Bitext mining



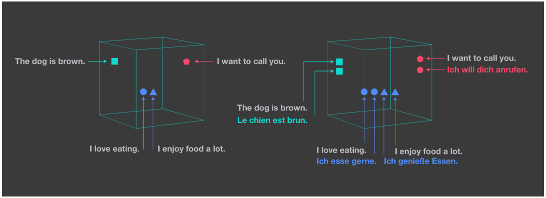
- Search for similar sentence meanings across different languages.
- Use proposed alignments to help supplement training data for NMT.

Bitext mining: data availability



- Only a small fraction of bitexts available in comparison to monolingual data.
- Leaves huge scope for bitext mining to help close this gap.

Multilingual Sentence Embeddings



- Sentences with similar meaning are close (paraphrases)
- Independently of the language they are written in

Popular approaches

- LASER, *Artexe and Schwenk, arXiv Dec'18, TACL'19*
- mBART, *Liu et al, arXiv'20*
- XLM-R, *Conneau et al, ACL'20*
- LaBSE, *Feng et al, arXiv'20*
- . . .

NLLB Team

NLLB
Motivation
Pipeline
Bitext mining

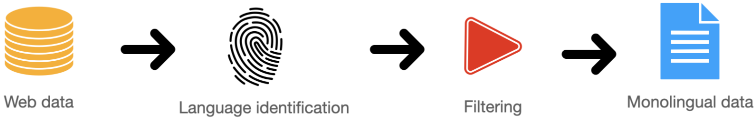
LASER3
Teacher-Student
xsim

Evaluation
Europe
Creole
Berber
African

Multimodality
Speech LASER
Mining

Conclusion

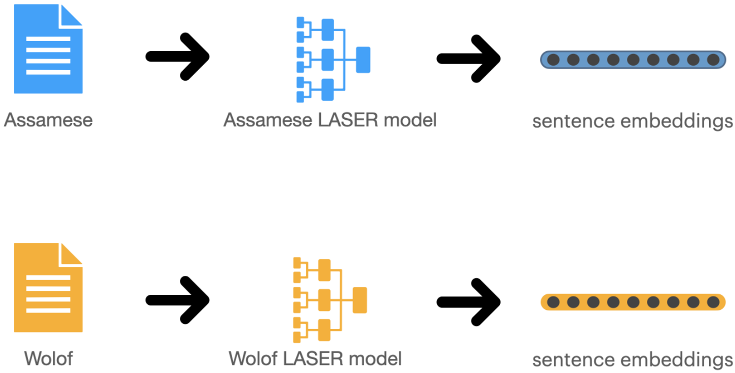
Mining bitexts: step 1



Web data → Language identification → Filtering → Monolingual data

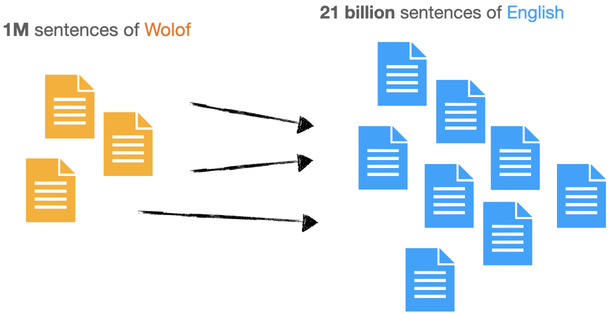
- Monolingual data (commoncrawl snapshots).
- Language identification model.
- Filtering such as sentence splitting, sentence deduplication, etc.
- Result is clean monolingual data, ready for mining!

Mining bitexts: step 2



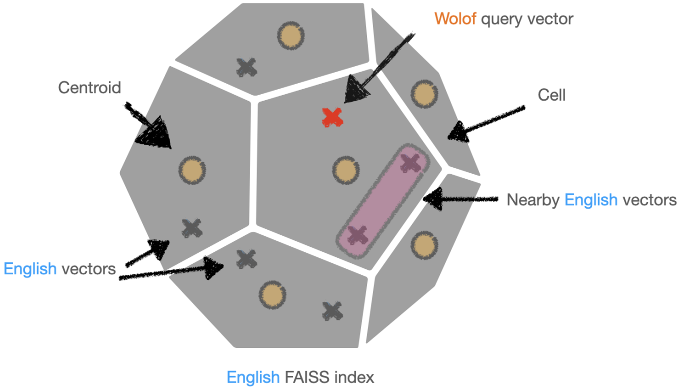
- Input (clean) monolingual data.
- Encode using specialised LASER model.

Mining complexity



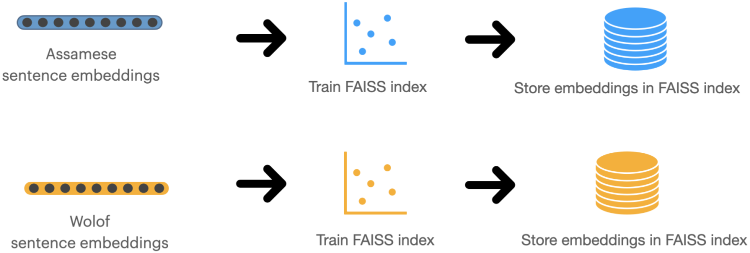
- How can we search efficiently among such large volumes of data?
- Even when one language is low-resource, we still have many billions of sentences to compare against?!

FAISS: Voronoi cells



- Efficient search amongst billions of sentences.
- Query lands in an initial cell, and then searches within that cell only (or neighboring cells if requested).

Mining bitexts: step 3

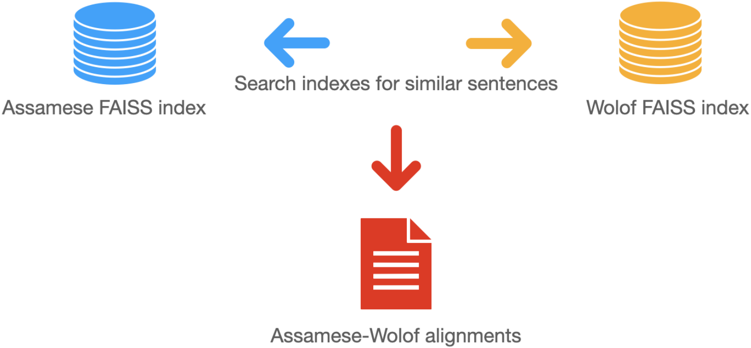


Assamese sentence embeddings → Train FAISS index → Store embeddings in FAISS index

Wolof sentence embeddings → Train FAISS index → Store embeddings in FAISS index

- Input sentence embeddings
- FAISS index learns clusters (Voronoi cells) using embedding data sample.
- Once clusters learned, then input all embeddings into index (i.e., assign all to various clusters/cells). This enables fast k-nn search!

Mining bitexts: step 4



Assamese FAISS index

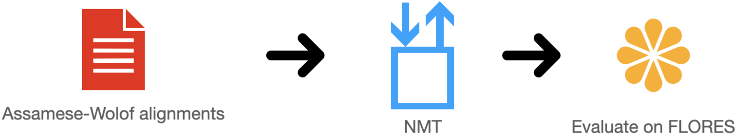
Search indexes for similar sentences

Wolof FAISS index

Assamese-Wolof alignments

- Search indexes for similar sentences.
- Output new alignments!

Mining bitexts: step 5



Assamese-Wolof alignments → NMT → Evaluate on FLORES

- Use new alignments to a train bilingual NMT system.
- Evaluate system on FLORES using metrics such as BLEU
- Such metrics can act as a proxy for “goodness” of the proposed alignments.

Stopes

NO LANGUAGE LEFT BEHIND

Driving inclusion through machine translation

stopes

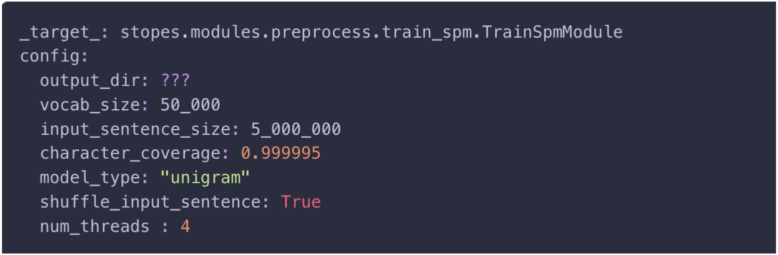
Large-Scale Translation Data Mining

https://facebookresearch.github.io/stopes/

Stopes

- End-to-end pipeline
- Processing of monolingual data
- Global mining
  - Text encoding using either LASER or any encoder available from HuggingFace
- Integrated caching (pick up where you left off).
- Job launching system, which can make use of either local GPUs or “submitit” jobs via SLURM.
- Bilingual NMT training of mined bitexts using `fairseq`.
- Configurable via Hydra so no need to edit code!

Stopes: Hydra integration



```
_target_: stopes.modules.preprocess.train_spm.TrainSpmModule
config:
  output_dir: ???
  vocab_size: 50_000
  input_sentence_size: 5_000_000
  character_coverage: 0.999995
  model_type: "unigram"
  shuffle_input_sentence: True
  num_threads : 4
```

- Configure modules without needing to edit code.
- Uses YAML files to store configuration.
- Allows for easy command-line overrides as well (i.e. no need to edit configuration file if you don't want to).

Massively Multilingual Models

One-for-all approach

- NMT, sentence representations, . . .
- Low-resource languages benefit from high-resource ones
  - e.g. Nepali/Hindi or Icelandic/German
- But accounting for the huge size difference is tricky
- Can new low-resource languages be efficiently learned

⇒ *Curse of multilinguality*

- Do we expect gains combining "unrelated languages"?
  - does Wolof benefit of Indonesian or Italian?
  - does Assamese benefit of Arabic or Albanian?
- Some low-resource languages are rather isolated (Quechua, Inuit, . . .)

Massively Multilingual Models

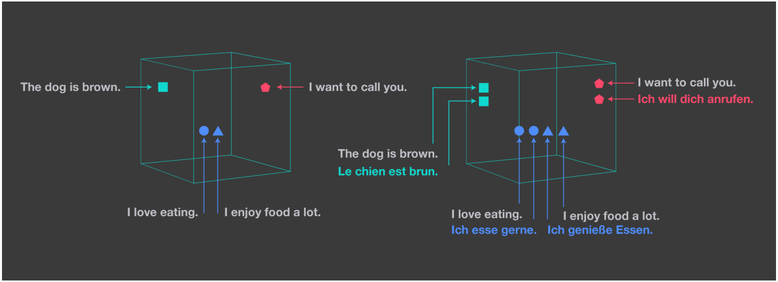
Switch to training multiple models

- Train models by groups of similar languages
- Ideally, each group contains a high-resource language

⇒ How can we make sure that these individual models are mutually compatible?

  - e.g. an African and Turkic language

Motivation: Extending the embedding space



- New model will learn a completely new space.
- Not compatible with existing models.
- Comparison will be apples to oranges.
- Bitext mining will quickly become intractable.

LASER3

- Substantially improved LASER sentence embeddings

NO LANGUAGE LEFT BEHIND

Driving inclusion through machine translation

LASER3

Language-Agnostic SEntence Representations

Meta AI

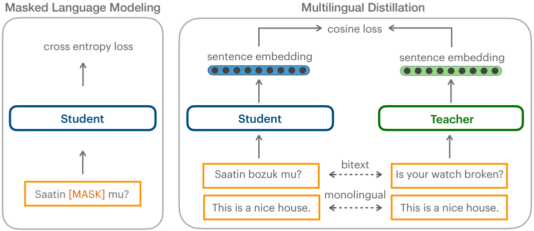
LASER3: No Language Left Behind (NLLB)

- Encoders to support more than 200 languages.
- `github.com/facebookresearch/LASER/tree/main/nllb`

LASER3 Teacher-Student Training

Idea

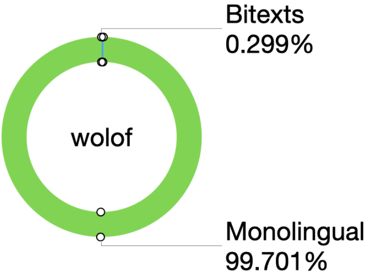
- Do not train new models from scratch (for new languages)
- Extend existing embedding space to more languages



Advantages

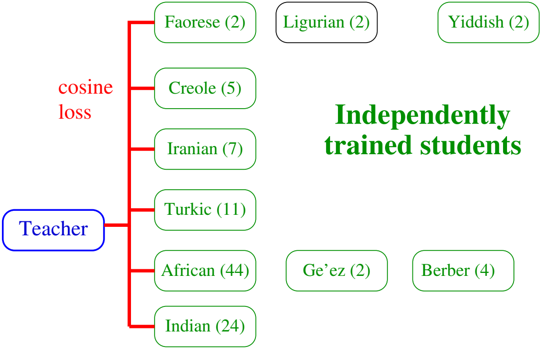
- Likely, less resources are needed
- Can be combined with masked LM training
- Fast turnaround (e.g. model for Ligurian trained in $<$ 1hr)

LASER3 Teacher-Student Training



- ∼6k bitexts available to train Wolof.
- ∼4 million sentences of monolingual data available.
- As monolingual data comes from commoncrawl (internet data), we found it needs to be high quality in order to work for masked language modelling: quality filtering very helpful.

Using Multiple Students



- Multiple students using the same teacher

⇒ The students are mutually compatible

- Each student can be separately optimized (architecture, capacity, vocabulary, convergence, . . .)

Script differences

Scripts

- **Amharic**: ge’ez script

  ቾው! ቾው! እኔ የባቡር ሾፌር ነኝ! ባቡሬ ወደ ሩቅ
  ከተሞች ይጓዛል። ዋውውው

- **Tamashek**: Tifinagh script

  ⵛⵓⵓ! ⵛⵓⵓ! ⴰⵈⴰⵍⵍⴰⵜ ⴰⵎⴰⵏⵓⵂⴰⵍ ⴰⵏ ⵜⴰⵔⴰⴽⵏⵉ. ⵜⴰⵔⵉⴽⵏⵉ
  ⵜⴰⵏⵉⵙⵙ ⵜⴰⵂⴰⵍ ⴽⵗⴰⵔⵎⴰⵜⵉⵏ ⵓⴽⵏ ⴰⵅⴰⵍⴰ ⵡⵓⵀⵓⵀⵓ

- Rare scripts likely to cause many [UNK] tokens in a shared vocabulary.

Evaluation of Multilinguality

Scaling multilingual models

- We may find training data in >1000 languages (e.g. bible)
- But high-quality evaluation data is more limited
  - Tatoeba is very noisy and unbalanced

Evaluation of Multilinguality

Scaling multilingual models

- We may find training data in >1000 languages (e.g. bible)
- But high-quality evaluation data is more limited
  - Tatoeba is very noisy and unbalanced

FLORES

- FLORES-101: ≈1000 sentences in 101 languages
- N-way parallel, sampled from Wikipedia
- NLLB: extension to 204 languages:
  - mostly low-resource languages
  - freely available
- Recently extended to speech (FLEURS-101)

Evaluation of Multilinguality

Bitext mining

- Final goal: improve MT performance
- Costly: train encoder, mine bitexts, train SMT → BLEU

Evaluation of Multilinguality

Bitext mining

- Final goal: improve MT performance
- Costly: train encoder, mine bitexts, train SMT → BLEU

Proxy: multilingual similarity search `xsim`

- Given a parallel test data (FLORES)
- Search translation with highest margin score

$$\text{score}(x,y) = \frac{cos(x,y)}{\sum_{z \in NN_k(x)} \frac{cos(x,z)}{2k} + \sum_{v \in NN_k(y)} \frac{cos(y,v)}{2k}}$$

- `xsim`: error rate of wrongly matched sentences in FLORES
- Easy to use open-source implementation

Evaluation of LASER3

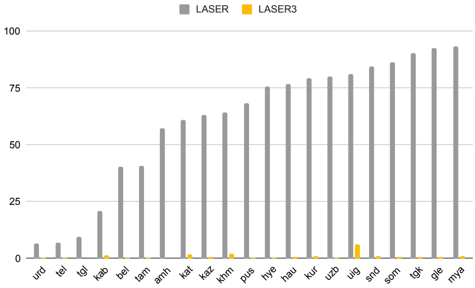
Methology

- Trained LASER3 models for 148 languages
- Transformers perform better than BiLSTM
- Select best model based on `xsim` on FLORES dev
- Mine bitexts against 21.5 billion English sentences
- Train NMT systems
- Compare BLEU on *“human”* versus *“human + mined”*

Evaluation of LASER3

Improving the original LASER

- Originial LASER performed badly on several languages



- Retrained models: avrg `xsim` 61→0.9%
  - Burmese: 93→0.9%, Irish 92→0.8%
  - on-pair with LaBSE

Malayo-Polynesian Languages

Lang.	bitexts	BLEU	xsim %	Monol.	Mined	BLEU
Acehnese	39.2k	0	2.4	2.2M	1.4M	10.3
Buginese	21.8k	0	1.6	0.7M	717k	4.2
Cebuano	1.1M	34.4	0.1	23.6M	8.1M	39.0
Indonesian	11M	-	0.1	-	-	-
Javanese	86k	11.1	0.1	27.2M	8.5M	31.2
Malay	2.3M	34.4	0.0	640M	40.5M	41.4
Pangasinan	327k	15.6	0.7	3.9M	1.9M	18.5
Sundanese	32.3k	1.5	0.6	8.2M	6.1M	28.5
Tagalog	1.3M	40.2	0.1	89M	33M	43.8
Warray	331k	26.5	0.2	26.9M	4.9M	36.5

Malayo-Polynesian Languages

Lang.	bitexts	BLEU	xsim %	Monol.	Mined	BLEU
Acehnese	39.2k	0	2.4	2.2M	1.4M	10.3
Buginese	21.8k	0	1.6	0.7M	717k	4.2
Cebuano	1.1M	34.4	0.1	23.6M	8.1M	39.0
Indonesian	11M	-	0.1	-	-	-
Javanese	86k	11.1	0.1	27.2M	8.5M	31.2
Malay	2.3M	34.4	0.0	640M	40.5M	41.4
Pangasinan	327k	15.6	0.7	3.9M	1.9M	18.5
Sundanese	32.3k	1.5	0.6	8.2M	6.1M	28.5
Tagalog	1.3M	40.2	0.1	89M	33M	43.8
Warray	331k	26.5	0.2	26.9M	4.9M	36.5

- Very low `xsim` error rates for most languages despite <100k bitexts for some languages

⇒ Training a language specific encoder seems to be beneficial

Malayo-Polynesian Languages

Lang.	bitexts	BLEU	xsim %	Monol.	Mined	BLEU
Acehnese	39.2k	0	2.4	2.2M	1.4M	10.3
Buginese	21.8k	0	1.6	0.7M	717k	4.2
Cebuano	1.1M	34.4	0.1	23.6M	8.1M	39.0
Indonesian	11M	-	0.1	-	-	-
Javanese	86k	11.1	0.1	27.2M	8.5M	31.2
Malay	2.3M	34.4	0.0	640M	40.5M	41.4
Pangasinan	327k	15.6	0.7	3.9M	1.9M	18.5
Sundanese	32.3k	1.5	0.6	8.2M	6.1M	28.5
Tagalog	1.3M	40.2	0.1	89M	33M	43.8
Warray	331k	26.5	0.2	26.9M	4.9M	36.5

- Large amounts of monolingual data

⇒ Optimal conditions for mining

Malayo-Polynesian Languages

Lang.	bitexts	BLEU	xsim %	Monol.	Mined	BLEU
Acehnese	39.2k	0	2.4	2.2M	1.4M	10.3
Buginese	21.8k	0	1.6	0.7M	717k	4.2
Cebuano	1.1M	34.4	0.1	23.6M	8.1M	39.0
Indonesian	11M	-	0.1	-	-	-
Javanese	86k	11.1	0.1	27.2M	8.5M	31.2
Malay	2.3M	34.4	0.0	640M	40.5M	41.4
Pangasinan	327k	15.6	0.7	3.9M	1.9M	18.5
Sundanese	32.3k	1.5	0.6	8.2M	6.1M	28.5
Tagalog	1.3M	40.2	0.1	89M	33M	43.8
Warray	331k	26.5	0.2	26.9M	4.9M	36.5

- BLEU gain >20: Javanese and Sundanese

Malayo-Polynesian Languages

Lang.	bitexts	BLEU	xsim %	Monol.	Mined	BLEU
Acehnese	39.2k	0	2.4	2.2M	1.4M	10.3
Buginese	21.8k	0	1.6	0.7M	717k	4.2
Cebuano	1.1M	34.4	0.1	23.6M	8.1M	39.0
Indonesian	11M	-	0.1	-	-	-
Javanese	86k	11.1	0.1	27.2M	8.5M	31.2
Malay	2.3M	34.4	0.0	640M	40.5M	41.4
Pangasinan	327k	15.6	0.7	3.9M	1.9M	18.5
Sundanese	32.3k	1.5	0.6	8.2M	6.1M	28.5
Tagalog	1.3M	40.2	0.1	89M	33M	43.8
Warray	331k	26.5	0.2	26.9M	4.9M	36.5

- BLEU gain >20: Javanese and Sundanese
- High resource languages also improve

European Minority Languages

Lang.	fao	fur	lij	lim	lmo	ltz	srd	szl	vec	ydd
Addtl. Lang	deu	ita	ita	nld	ita	deu	ita	pol	ita	deu
Bitexts [k]	6.6	6.3	2.2	5.4	1.3	9.8	1.4	6.4	1.2	6.2
BLEU	0	0	0	0	0	0	0	0	0	0
xsim **[%]**	2.57	0.1	0.2	16.1	1.09	0.59	0.1	0.69	2.77	0.1
Monolingual	1.2M	737k	106k	15M	61M	123M	515k	2.5M	12M	12M
Mined	1.6M	532k	631k	2.0M	4.1M	5.5M	723k	1.0M	2.5M	3.3M
BLEU	10.6	23.5	13.4	5.5	20.7	37.0	20.9	18.9	17.8	30.1

- Pairing low-resource with similar high-resource language is very effective
- BLEU $>$ 20: Faroese, Lombard and Sardinian
- BLEU $>$ 30: Luxemburgish and Yiddish

Creole Languages

Lang.	hat	kea	pap	sag	tpi
Addtl. Lang	fra	por	spa por	lin	eng
Bitexts	334	6	5	282	458
BLEU	20.2	0	0	4.8	14.7
`xsim` **[%]**	1.19	1.19	0.1	8.6	0.2
Monolingual	14M	227k	28M	645k	1.7M
Mined	8.0M	656k	7.3M	1.9M	1.2M
BLEU	29.2	4.9	40.9	5.3	16.1

- Papiemento: mono=28M → BLEU=40.9
- Tok Pisin: mono=1.7M → BLEU=16.1
- Kabuverdianu: mono<300k → BLEU=4.9

Creole Languages

Lang.	hat	kea	pap	sag	tpi
Addtl. Lang	fra	por	spa por	lin	eng
Bitexts	334	6	5	282	458
BLEU	20.2	0	0	4.8	14.7
`xsim` **[%]**	1.19	1.19	0.1	8.6	0.2
Monolingual	14M	227k	28M	645k	1.7M
Mined	8.0M	656k	7.3M	1.9M	1.2M
BLEU	29.2	4.9	40.9	5.3	16.1

- Papiemento: mono=28M → BLEU=40.9
- Tok Pisin: mono=1.7M → BLEU=16.1
- Kabuverdianu: mono<300k → BLEU=4.9

⇒ The amount of monolingual data is crucial

Berber Languages (14M speakers)

Lang. **Script**	Kabyle Latin	Tifinagh Latin	Tifinagh Tifinagh	Tamazight Tifinagh
bitexts	72k	10.2k	4k	6.2k
BLEU	1.2	0	0	0
xsim [%]	0.99	24.11	35.57	3.66
Monolingual	3.4M	23k	5k	59k
Mined	3.1M	240k	-	111k
BLEU	6.2	1.2	-	3.8

- Extremely limited resources, except Kabyle
- Kabyle: some mined bitexts and BLEU>6
- Tamazight: very modest BLEU score of ≈ 4
- Tifinagh: insufficient monolingual data

Berber Languages (14M speakers)

Lang.	Kabyle	Tifinagh	Tifinagh	Tamazight
Script	Latin	Latin	Tifinagh	Tifinagh
bitexts	72k	10.2k	4k	6.2k
BLEU	1.2	0	0	0
`xsim` [%]	0.99	24.11	35.57	3.66
Monolingual	3.4M	23k	5k	59k
Mined	3.1M	240k	-	111k
BLEU	6.2	1.2	-	3.8

- Extremely limited resources, except Kabyle
- Kabyle: some mined bitexts and BLEU>6
- Tamazight: very modest BLEU score of ≈ 4
- Tifinagh: insufficient monolingual data

⇒ Typical examples of very low-resource languages for which it is very hard to collect written material

African Languages

- 1.2 billion people, estimated 2000 languages
- Existing systems support only few African languages
  - LaBSE: 14 (+4)
  - Google translate: 22
- We trained encoders for 55 languages, 48 are low resource
- Specific encoder for languages with Ge'ez script: Amharic and Tigrinya
- Average over 44 languages: BLEU 11.0 → 14.8 with mined data

Challenges

- It seems very difficult to crawl textual resources for several languages

African Languages

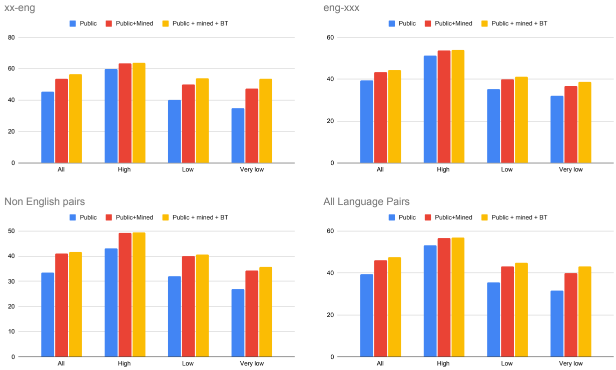
Languages with Ge’ez script

Training	SPM	#train	amh	tir
LASER2	50k joint	220M	34.9	92.9
Semitic	50k joint	9M	0.2	1.19
Ge’ez	8k specific	0.7M	0.1	0.89
LaBSE	501k joint	$\approx$ 6B	0	13.74

- Teacher-student model performs much better than LASER2
- Using an student specific SPM vocabulary yields further improvements
- The much bigger LaBSE model does not perform better

Massively Multilingual NMT

Impact of mined bitexts (chrF++)



- Substantial gains in chrF++ when adding mined data
  - very low-resource xx/eng: +12.5 chrF++
  - very low-resource eng/xx: +4.7 chrF++

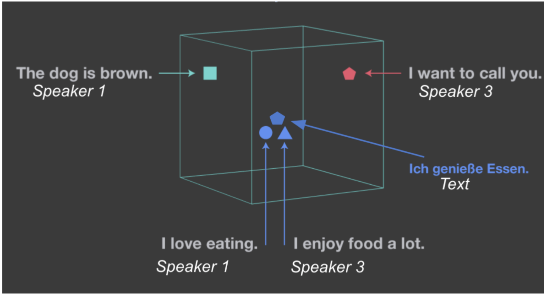
⇒ Mined data is crucial for very low-resource languages

Going Multimodal

What about other modalities?

- Many languages are rather spoken than written

⇒ Multilingual and multimodal fixed-size sentence representation



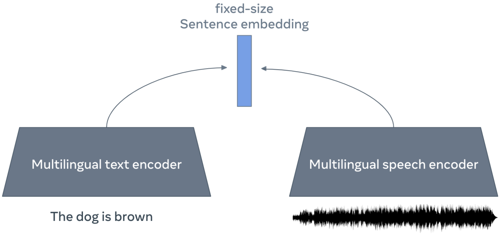
Going Multimodal

How to build a joint audio/text multilingual sentence embedding space?

- Challenges:

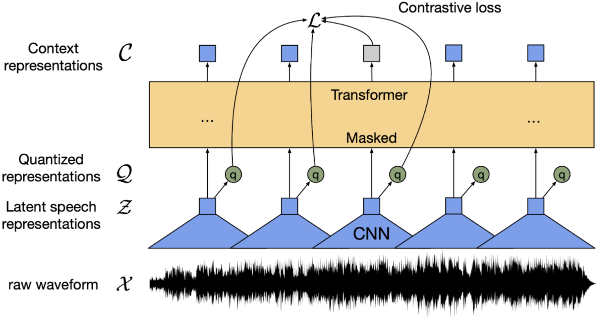
⇒ Semantic properties of the resulting embedding space

⇒ Encode a variable-length audio input into a single vector



Wav2vec 2.0 / XLSR

Leveraging self-supervised learning for multilingual speech



Going Multimodal

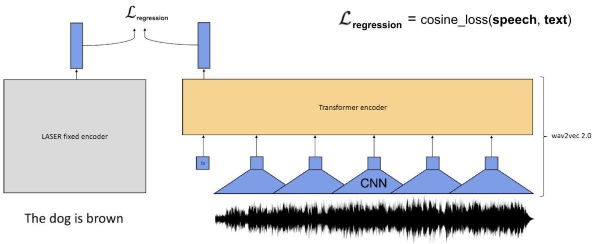
Speech LASER

- Apply teacher-student approach to speech

⇒ Fit fixed-size **speech** representation to LASER2

  - train with transcriptions, translations or both

- NeurIPS'21 paper:
  P.-A. Duquenne, H. Gong, H. Schwenk, *Multimodal and Multilingual Embeddings for Large-Scale Speech Mining*


$\mathcal{L}_{\text{regression}}$
$\mathcal{L}_{\text{regression}}$ = cosine_loss(**speech**, **text**)
LASER fixed encoder
Transformer encoder
wav2vec 2.0
1s
CNN
The dog is brown

Teacher/Student for Speech

Speech LASER

- Sentences are close in the embedding space if they have similar meanings independently of their language or their modality (either speech or text)

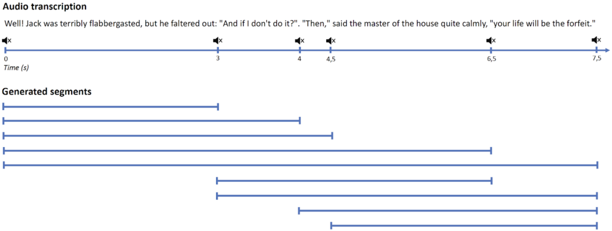
⇒ Align sentences across languages and modalities:
  - Speech-to-Text alignments in different languages
  - Speech-to-Speech alignments in different languages

- Generalize to unseen pairs: e.g.
  - Learn to align English audio with English text.
  - Then, align English audio with Turkish text.
- SpeechLASER compatible with LASER2 encoder

⇒ We can mine speech against all 200 NLLB languages !

Large-Scale Speech Mining

- Generate audio segments candidates based on Voice Activity Detection outputs



Audio transcription

Well! Jack was terribly flabbergasted, but he faltered out: "And if I don't do it?". "Then," said the master of the house quite calmly, "your life will be the forfeit."

- Audio segments matched with text sentences are kept
- Post-processing to get rid of overlapping audio

Large-Scale Speech Mining

Speech sources

- Librivox: a repository of open domain audio books in different languages
- We focus on English, German, French and Spanish audio

	De	Es	Fr	En
#audio books	633	257	343	13,292
#hours	3,529	1,535	1,770	73,511

- Mine these speech sources against texts from CommonCrawl

Speech-to-Text Mining

Mined S2T data

- Foreign audio against English texts

	de-en	fr-en	es-en
Mined [h]	1,074	543	668

- English audio against multiple languages

	en-fr	en-es	en-ru	en-ar	en-tr	en-vi
Mined [h]	6,289	6,544	3,330	1,549	1,656	1,390

- total approx. 20,000h of audio-text alignments

Speech-to-Text Mining

Train S2T translation systems, test on CoVoST2

- LNA approach builds on extensively pretrained pretrained models: wav2vec 2.0 and MBART
- Result summary:

Approach	Data	De-En	Es-En	Fr-En
Cascaded	human	23.2	31.1	29.1
LNA	human	24.4	29.2	30.7
LNA	human + mined	26.4	31.6	32.0

- Direct translation outperforms cascaded ASR + MT (except Es-En)
- Mined S2T data yields nice BLEU improvements

Speech-to-Speech Mining

Speech-to-speech mining

- Can we mine directly speech against speech?

Speech-to-Speech Mining

Speech-to-speech mining

- Can we mine directly speech against speech?
- Yes, directly in the embedding space!

Speech-to-Speech Mining

Speech-to-speech mining

- Can we mine directly speech against speech?
- Yes, directly in the embedding space!
- No need to transcribe or translate

Speech-to-Speech Mining

Speech-to-speech mining

- Can we mine directly speech against speech?
- Yes, directly in the embedding space!
- No need to transcribe or translate
- We run this on the Librivox speech data

Speech-to-Speech Mining

Speech-to-speech mining

- Can we mine directly speech against speech?
- Yes, directly in the embedding space!
- No need to transcribe or translate
- We run this on the Librivox speech data
- Challenges for S2S translation
  - previous S2S data was artificial
  - S2S didin't know how to use real data with many speakers

  ⇒ development of new speaker normalization algorithm
  - A. Lee et al., *Textless Speech-to-Speech Translation on Real Data*, NAACL'22

Speech-to-Speech Mining

Europarl test set

Lang	Train data	Train	BLEU	
			xx-en	en-xx
Es-En	human	522h	18.8	21.8
	+mined	+433h	21.2	24.1
Fr-En	human	515h	20.3	18.7
	+mined	+459h	22.1	20.3

- Mined data doubles the train data
- Improvement in BLEU of about 2 points

Speech-to-Speech Mining

CoVost test set

Es-En	9.2	16.3
Fr-En	9.6	16.7

- Huge improvement in BLEU 9.4 → 16.5
- Mined data seems to match very well domain

Conclusion

Scaling LASER

- Moved away from the popular one-for-all approach
  - train multiple mutually language specific models
  - alternative to adapters?
- Teacher-student approach with multiple mutually compatible encoders seems to be very efficient
- NLLB: mined more than 1 billion new bitexts
  (in addition to CCMatrix bitexts)
- Enabled scaling NMT to 200 languages and boosted performance
- First successful large-scale speech-to-speech mining

Conclusion

Challenges

- It is very hard to find textual resources for low-resource languages
- Does it make sense to scale translation to thousands of languages?

Conclusion

Challenges

- It is very hard to find textual resources for low-resource languages
- Does it make sense to scale translation to thousands of languages?
- Yes, but we should switch to the speech modality

Conclusion

Challenges

- It is very hard to find textual resources for low-resource languages
- Does it make sense to scale translation to thousands of languages?
- Yes, but we should switch to the speech modality
- Finding raw audio seems to be very tricky (legal problem)

Thank
You
Kiitos
Welalin
Terma Kasih
Salamat
Grazie
Asante
Dankon
Multumesc
Spasibo
Merci
Matondo
Nirringrazzjak
Ua Tsaug Rau Koj
Dank Je
Mochchakkeram
Obrigado
Arigato
Kia Ora
Raibh Maith Agat