Cross-lingual topic prediction for speech using translations

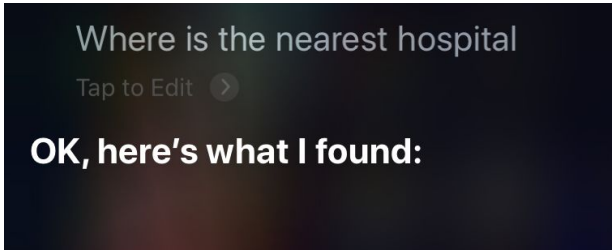
Sameer Bansal
Herman Kamper
Adam Lopez
Sharon Goldwater

THE UNIVERSITY *of* EDINBURGH
informatics

Automated speech-to-text

Translation

Information Retrieval



Current systems

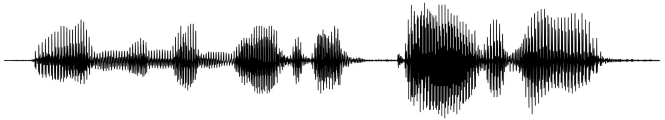
English audio:

?

downstream task:
translation, IR

Current systems

English audio:



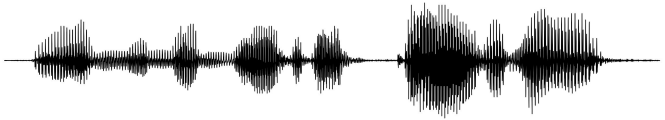
English text: *Where is the nearest hospital?*

Automatic Speech Recognition

downstream task: translation, IR

~100 languages supported by Google Translate ...

Unwritten languages

Mboshi Audio: 

~~Mboshi text:~~ ---

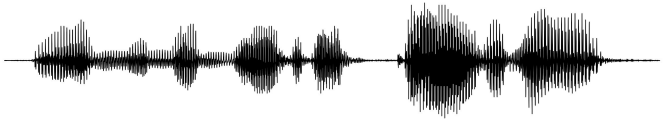
~~ASR~~

Aikuma: Bird et al. 2014,
LIG-Aikuma: Blachon et al. 2016
Godard et al. 2018

- ~3,000 languages with no writing system
- Traditional ASR based will not work!

Unwritten languages

Mboshi Audio:



~~Mboshi text:~~

French text

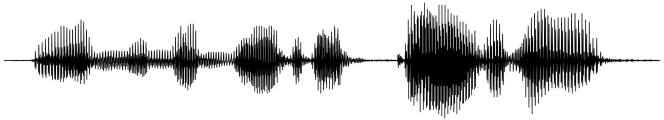
~~ASR~~

Aikuma: Bird et al. 2014,
LIG-Aikuma: Blachon et al. 2016
Godard et al. 2018

Efforts to collect speech and translations using mobile apps

Unwritten languages

Mboshi Audio:



~~Mboshi text:~~

French text

~~ASR~~

Aikuma: Bird et al. 2014,
LIG-Aikuma: Blachon et al. 2016
Godard et al. 2018

Build cross-lingual speech-to-text systems (ST)

Why speech input?

“For many Indians, searching by voice rather than text is their first choice.” Google

https://tnw.to/ieUbS

Radio content analysis in Uganda

55% households: radio main source of information

Quinn and Hidalgo-Sanchis, 2017

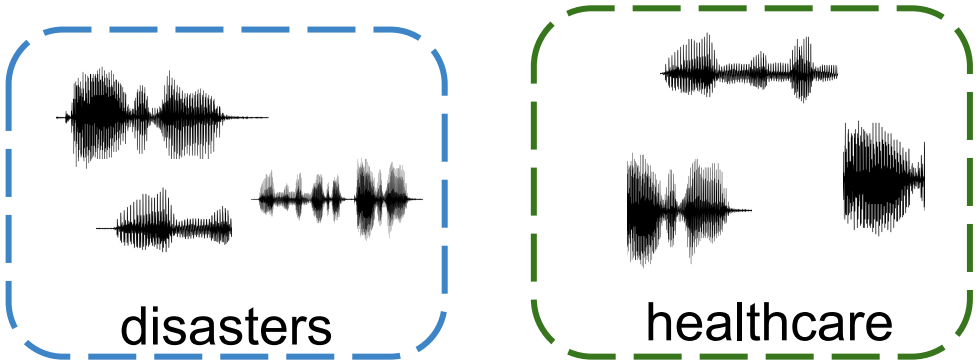
Radio content analysis in Uganda

Collect data from public radio conversations

Quinn and Hidalgo-Sanchis, 2017

Radio content analysis in Uganda

"Insights about the spread of infectious diseases, small-scale disasters, etc."



Quinn and Hidalgo-Sanchis, 2017

Radio content analysis in Uganda

Luganda audio

Topic?

Topic prediction task

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Luganda audio

Topic?

ASR

"Eddwaliro lyaffe temuli yadde …"

("… they have built health centers")

Speech to text system

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Keywords indicate topic information

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Availability of ASR!

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Luganda audio

healthcare

Topic prediction

~~ASR~~

~~**“Eddwaliro** lyaffe temuli yadde …”~~

(“… they have built **health centers**”)

Can we predict topics using ST?

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Luganda audio

healthcare

Topic prediction

~~ASR~~

~~**"Eddwaliro** lyaffe temuli yaddeko …"~~

("… they have built **health centers**")

Can we predict topics using ST?

https://radio.unglobalpulse.net/uganda

Radio content analysis in Uganda

Luganda audio

healthcare

Topic prediction

~~ASR~~

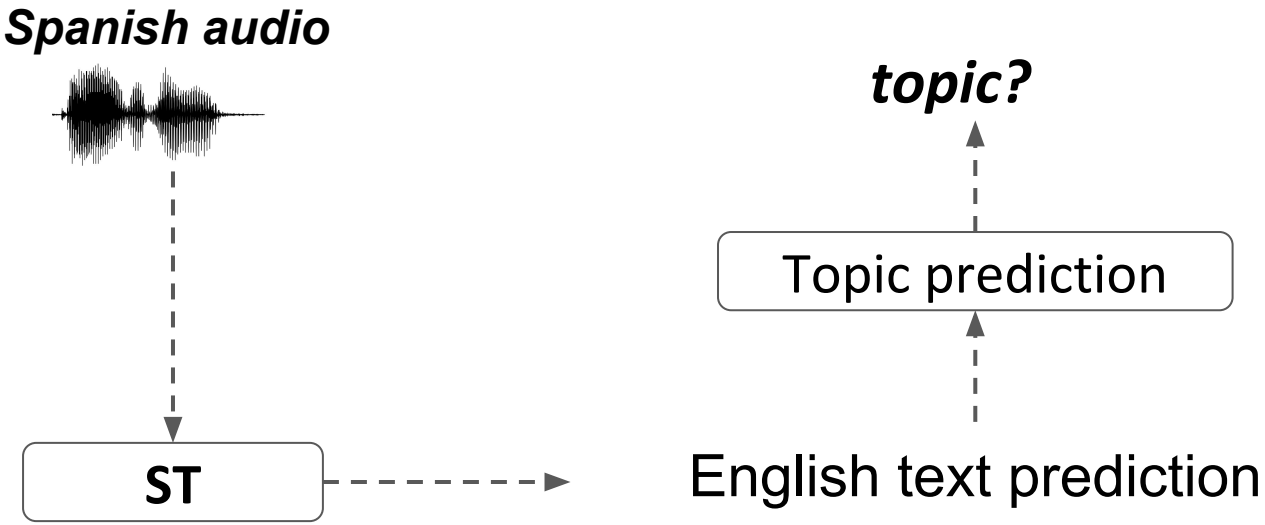
~~"**Eddwaliro** lyaffe temuli yadde …"~~

("… they have built **health centers**")

UN study dataset not available!

https://radio.unglobalpulse.net/uganda

Our work: topic prediction for Spanish speech



ST trained in simulated low-resource settings

ST performance in low-resource settings

Spanish-English	**BLEU**
160 hours - Weiss et al.	46

**for comparison text-to-text = 58*

Good performance if trained on 100+ hours

ST performance in low-resource settings

Spanish-English	**BLEU**
160 hours - Weiss et al.	46
20 hours - Bansal et al. 2019	**19**

**for comparison text-to-text = 58*

Mediocre performance in low-resource settings

ST performance in low-resource settings

Spanish-English	BLEU
160 hours - Weiss et al.	46
20 hours - Bansal et al. 2019	**19**

**for comparison text-to-text = 58*

"Good applications for crummy machine translation"
Church & Hovy, 1993

Sample translations

Spanish soy cat´olica pero no en realidad casi no voy a laiglesia

English i am catholic but actually i hardly go to church

Sample translations

Spanish soy cat´olica pero no en realidad casi no voy a laiglesia

English i am catholic but actually i hardly go to church

20h i’m catholics but reality i don’t go to the church

“Crummy” translation

Sample translations

Spanish soy cat´olica pero no en realidad casi no voy a laiglesia

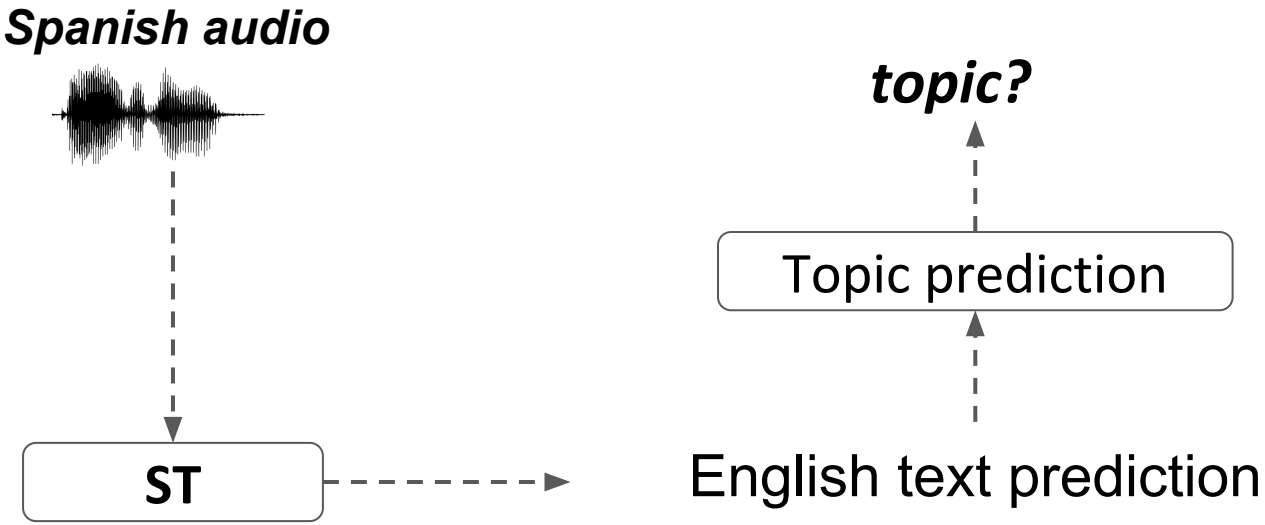
English i am catholic but actually i hardly go to church

20h i'm catholics but reality i don't go to the church

topic *religion*

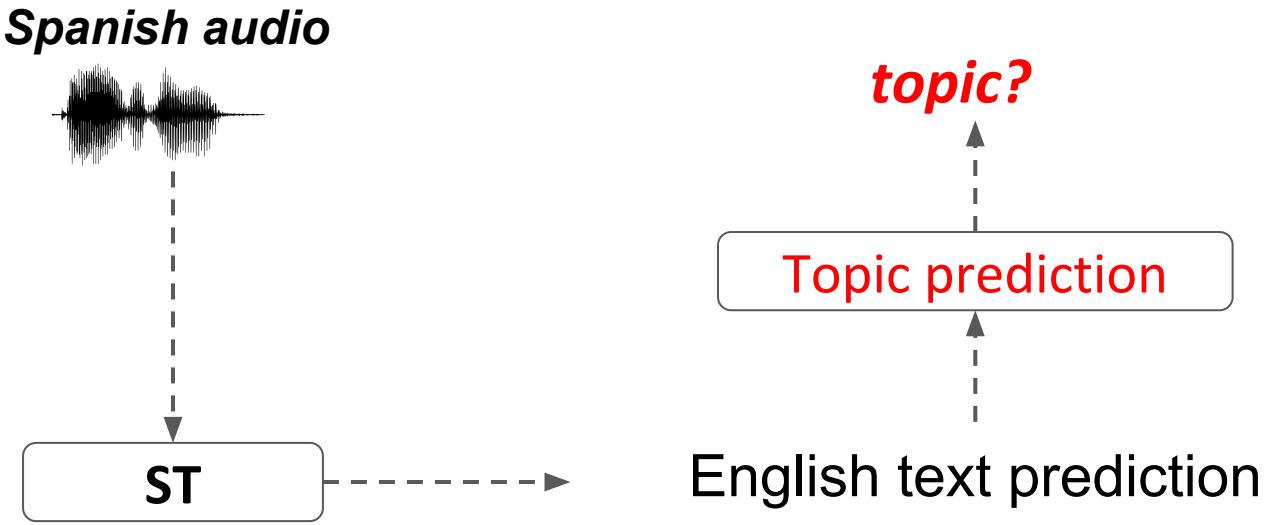
Keywords can be useful for topic prediction

Our work: topic prediction for Spanish speech



ST trained in simulated low-resource settings

Our work: topic prediction for Spanish speech



Gold topics labels not available!

Learning topic labels

Spanish audio

Gold topic label?

Learning topic labels

Spanish audio

Gold topic label?

I like to listen to jazz

Gold translation

Learning topic labels

Spanish audio

Gold topic label?

I like to listen to jazz

Gold translation

Use gold translations to infer topic labels

Learning topic labels

Spanish audio

Silver topic label

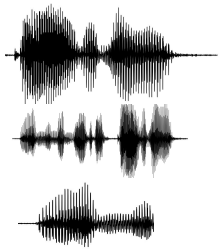
I like to listen to jazz

Gold translation

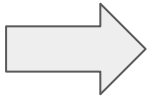
Use gold translations to infer topic labels

Learning topic labels

Spanish audio ***Gold human translation***



I listen to english music
I am catholic
hello how are you

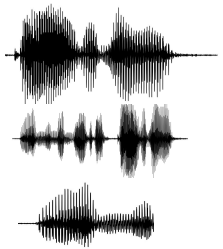


Topic model

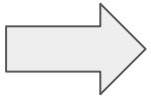
Training set

Learning topic labels

Spanish audio ***Gold human translation***



I listen to english music
I am catholic
hello how are you



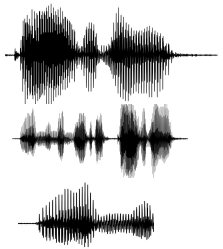
Topic model

Topic	Terms
small-talk	hello, fine, name
music	dance, listen, music
religion	god, bible, believe
...	...

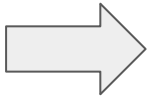
Training set

Learning topic labels

Spanish audio ***Gold human translation***



I listen to english music
I am catholic
hello how are you



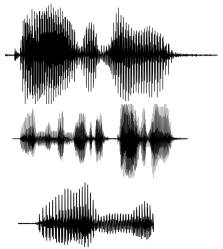
Topic model

Topic	Terms
small-talk	hello, fine, name
music	dance, listen, music
religion	god, bible, believe
...	...

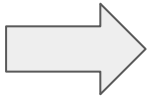
Number of topics set to 10

Learning topic labels

Spanish audio ***Gold human translation***



I listen to english music
I am catholic
hello how are you



Topic model

Topic	**Terms**
small-talk	hello, fine, name
music	dance, listen, music
religion	god, bible, believe
...	...

small-talk most frequent

Topic prediction and evaluation

Spanish audio

Topic model

Evaluation set

Topic prediction and evaluation

Gold translation

I like to listen to jazz

Silver

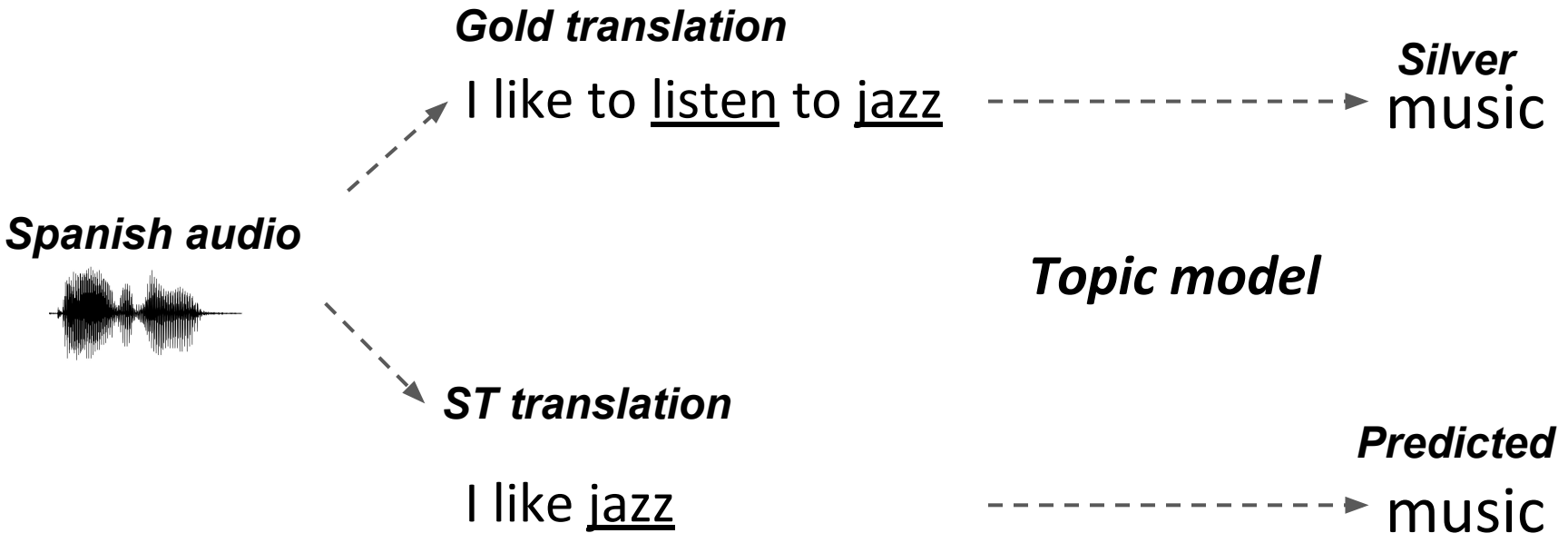
music

Spanish audio

Topic model

Evaluation set

Topic prediction and evaluation



Compare predicted and silver topic label

Topic prediction and evaluation

Gold translation

I like to listen to jazz

Silver

music

Spanish audio

Topic model

ST translation

I like jazz

Predicted

music

Good prediction

Topic prediction and evaluation

Gold translation

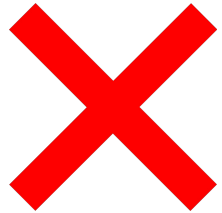
I like to listen to jazz

Silver

music

Spanish audio

Topic model



ST translation

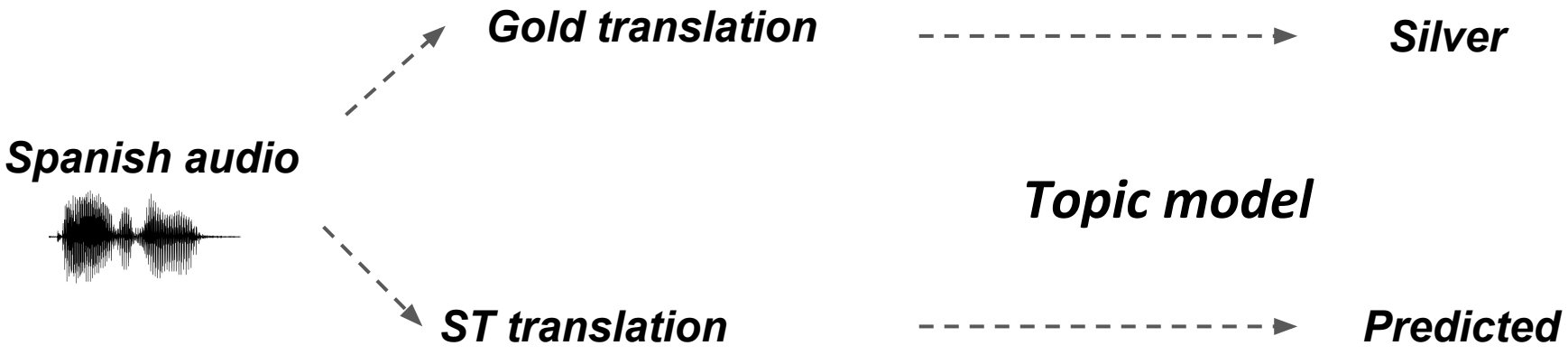
I like like

Predicted

small-talk

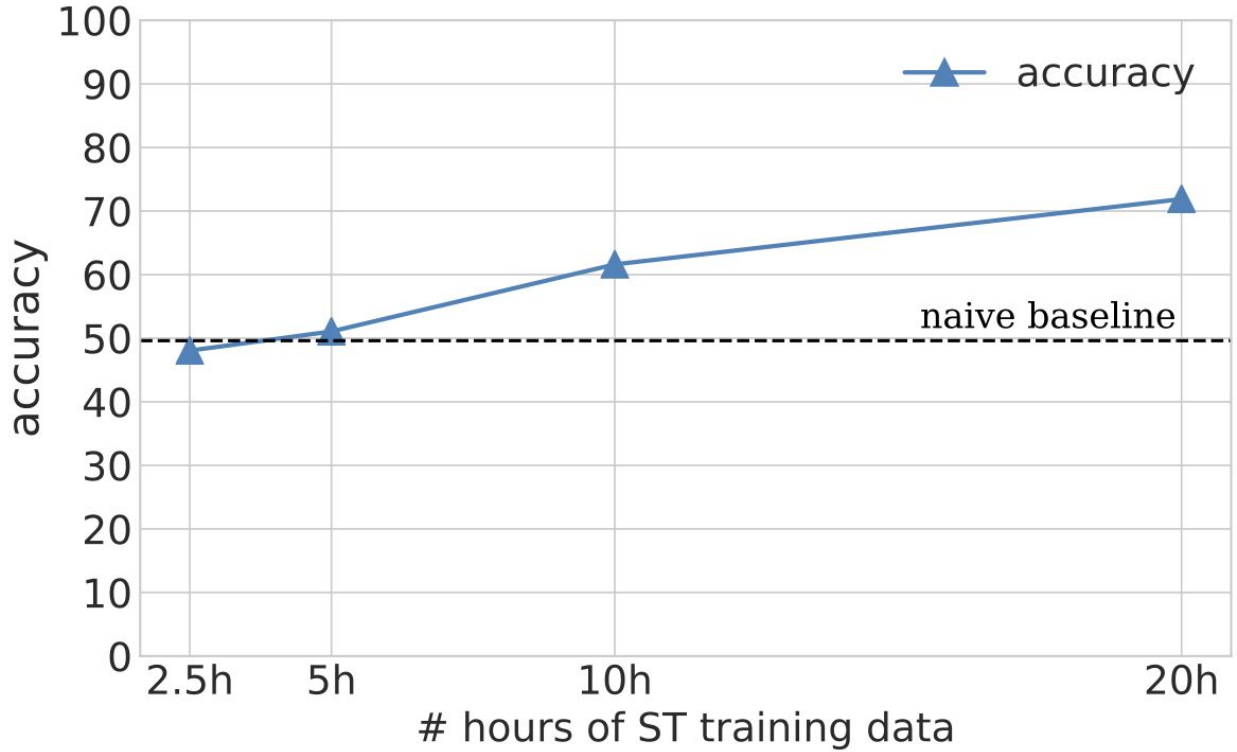
Poor prediction

Topic prediction and evaluation



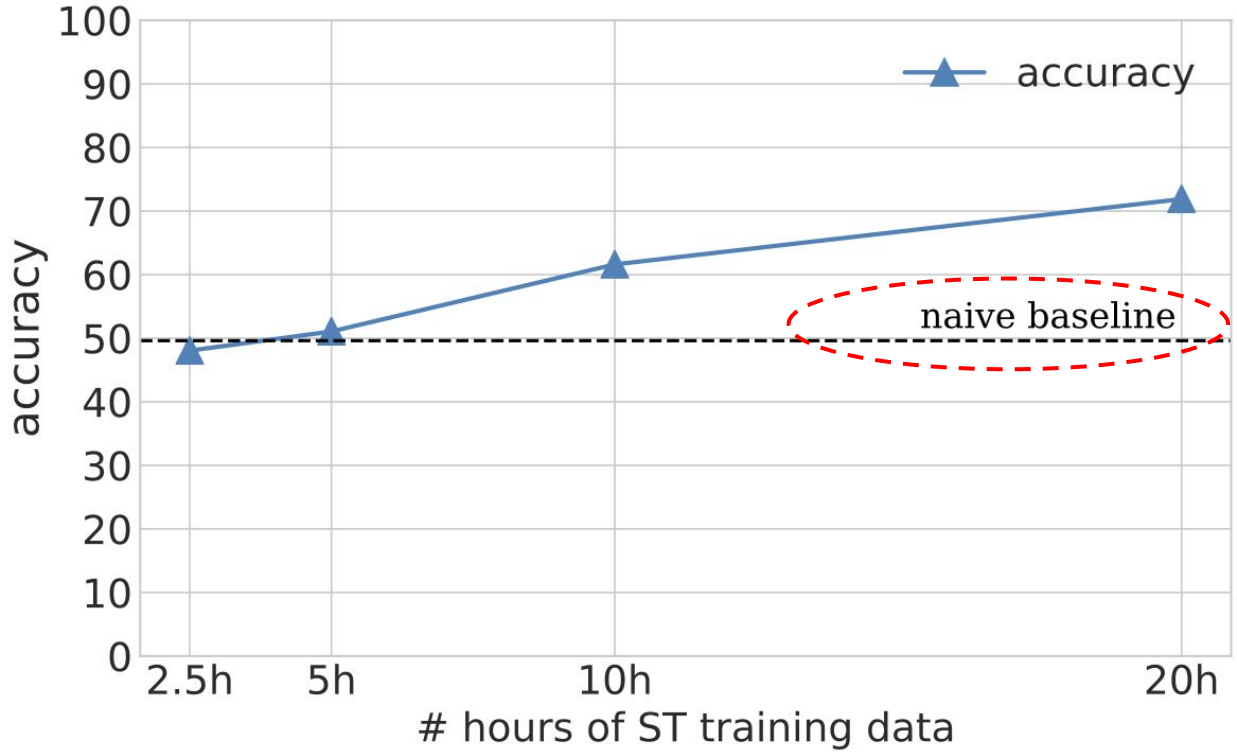
Evaluate over a 100 hour test set

Topic prediction accuracy



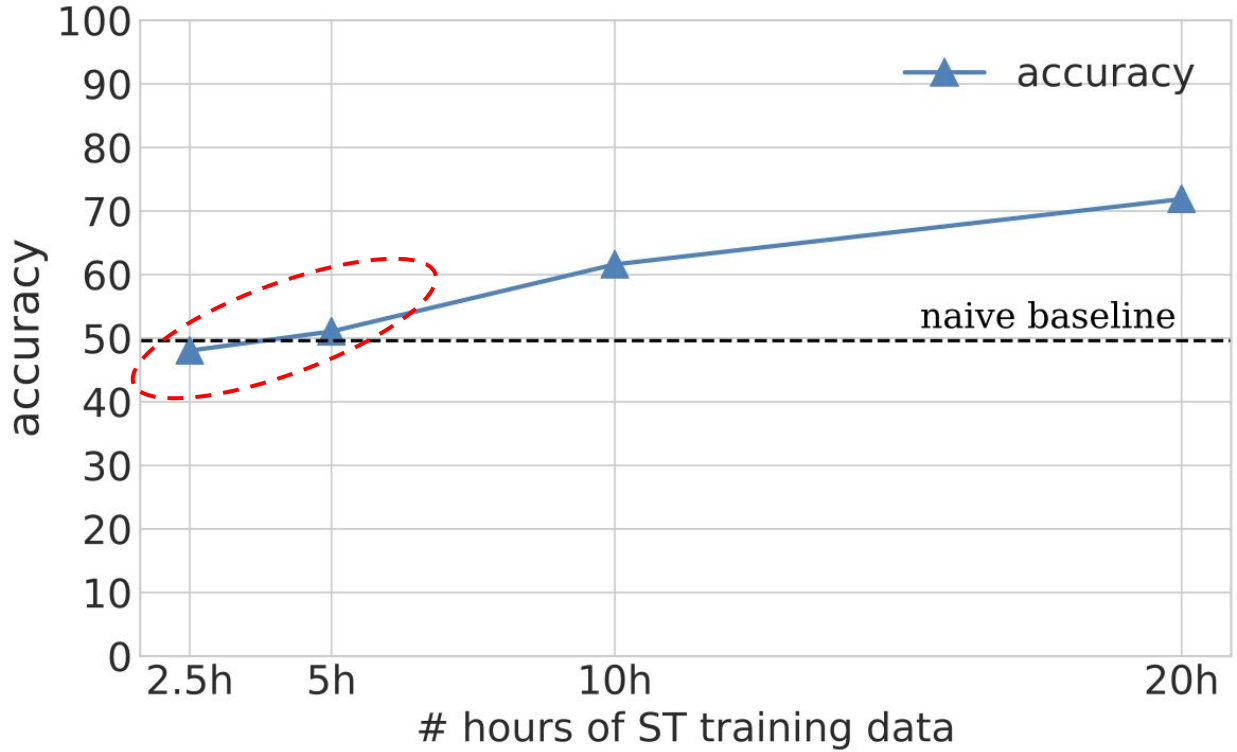
- ST trained on <= 20 hours of Spanish-English
- Pretrained on English ASR

Topic prediction accuracy



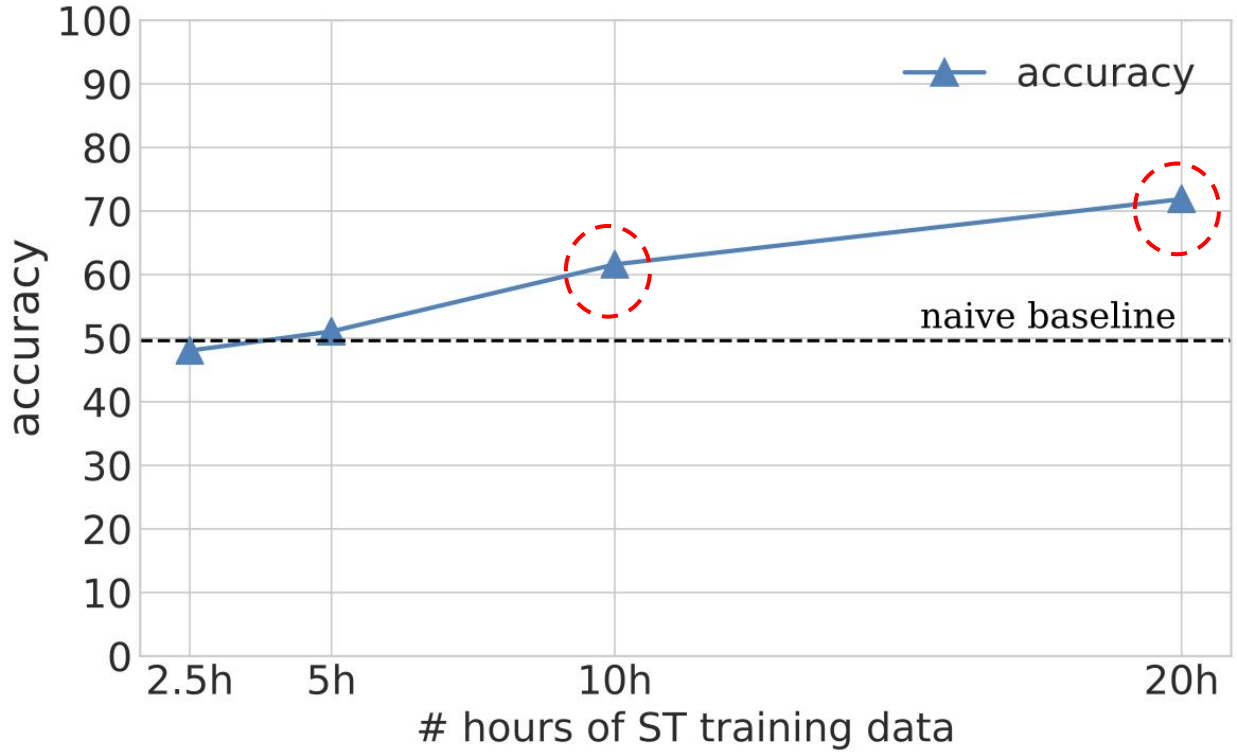
small-talk topic is the majority class baseline

Topic prediction accuracy



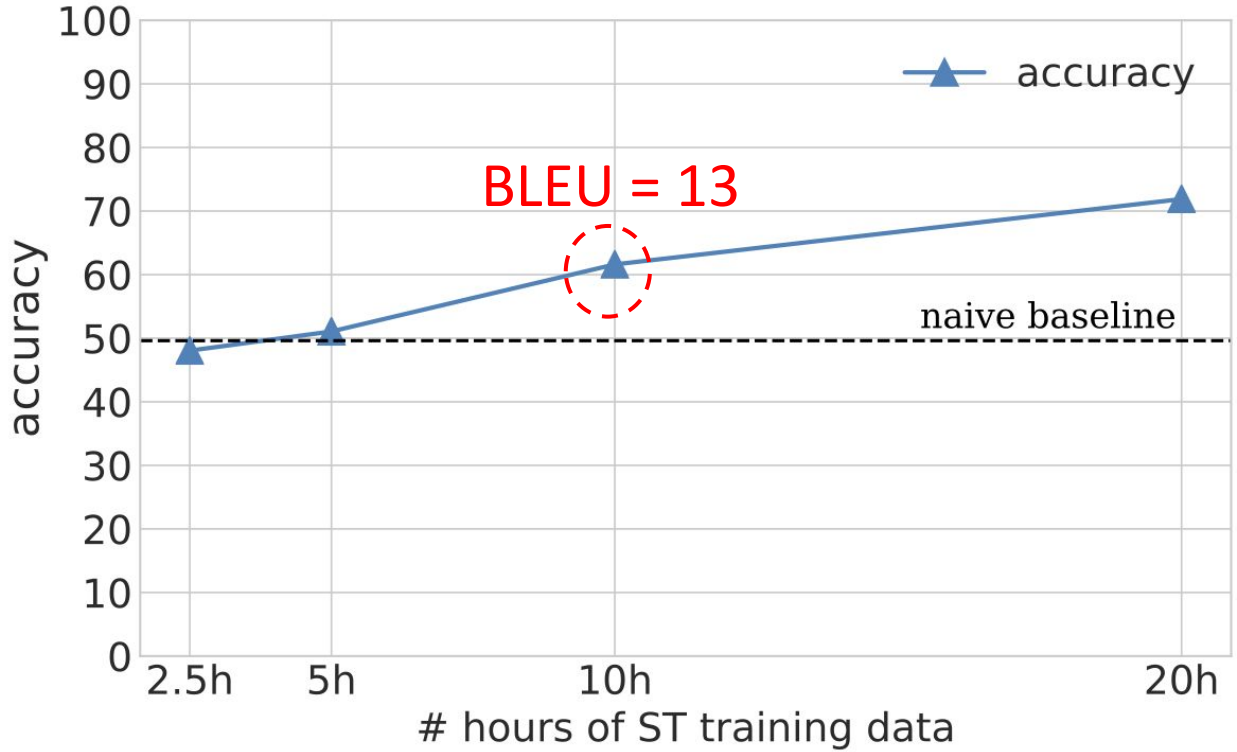
Poor performance <= 5 hours ST models

Topic prediction accuracy



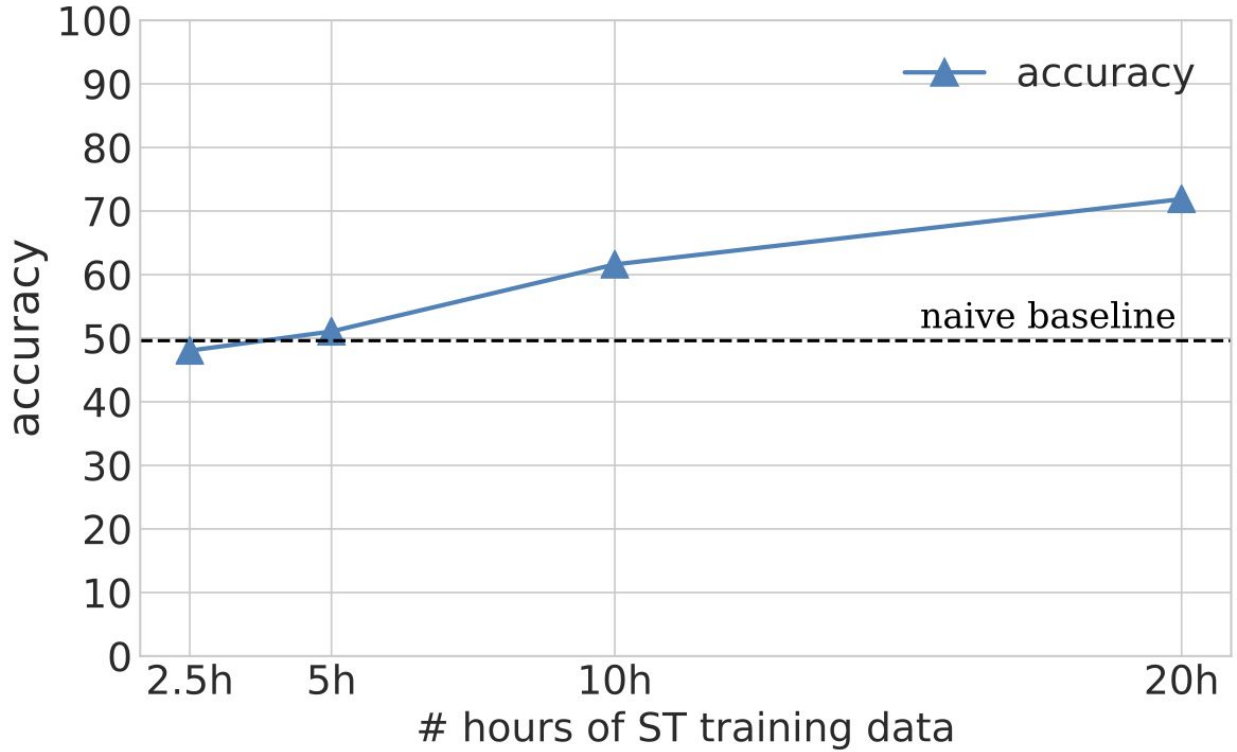
10-20h ST models outperform majority baseline

Topic prediction accuracy



10-20h ST models outperform majority baseline

Topic prediction accuracy



Takeaways

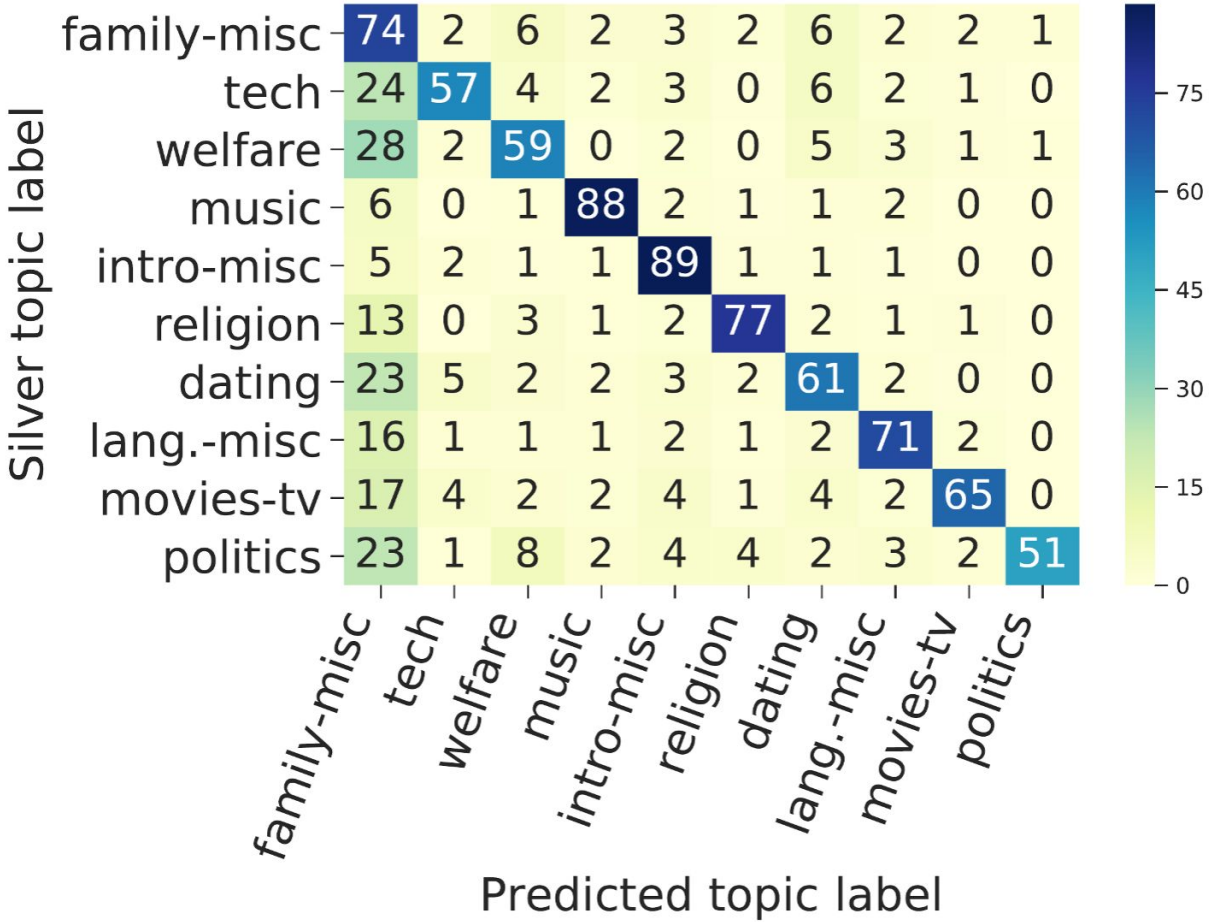
- Low-resource ST can still be useful for building downstream applications
- Silver evaluation for this preliminary study
  - Future: human evaluation
- Experiments on low-resource/unwritten languages
  - Datasets required
- Keyword spotting

Thanks!

- Check out: “Analyzing ASR pretraining for low-resource speech-to-text translation”, Stoian et al.

Backup

Topic prediction accuracy

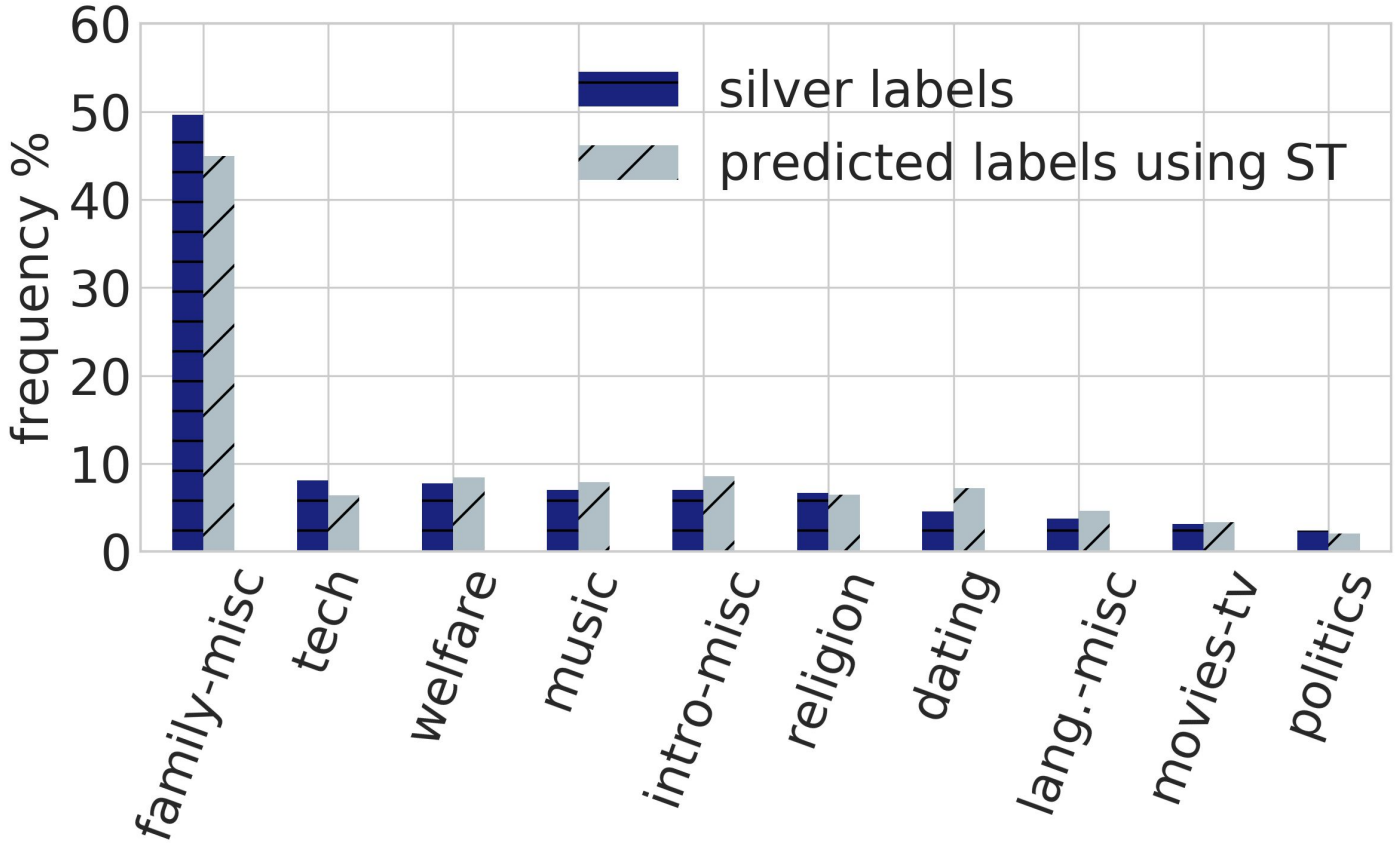


Silver labels

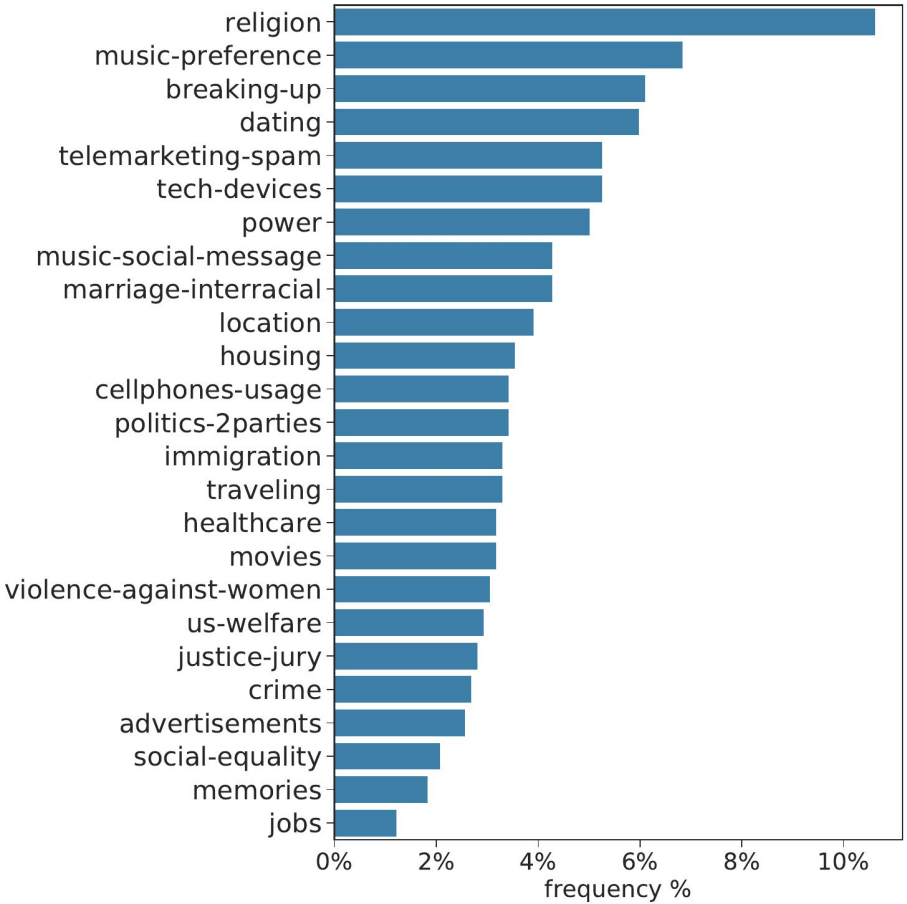
human translation	Assigned	Silver
hello good afternoon have you ever been in a jury in a trial	juries	intro-misc
i also receive many letters of life insurance from banks	spam	welfare
they tell us we have to talk about marriage	music	family-misc

Speakers were provided discussion prompts

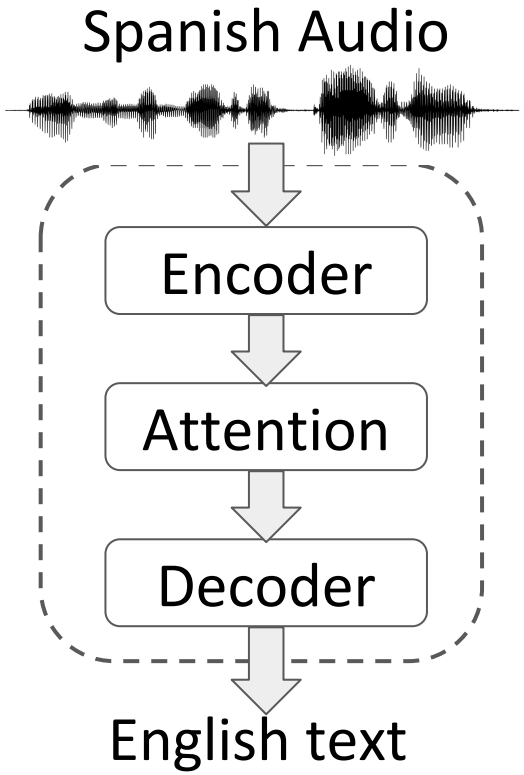
Topic labels



Spanish dataset discussion prompts



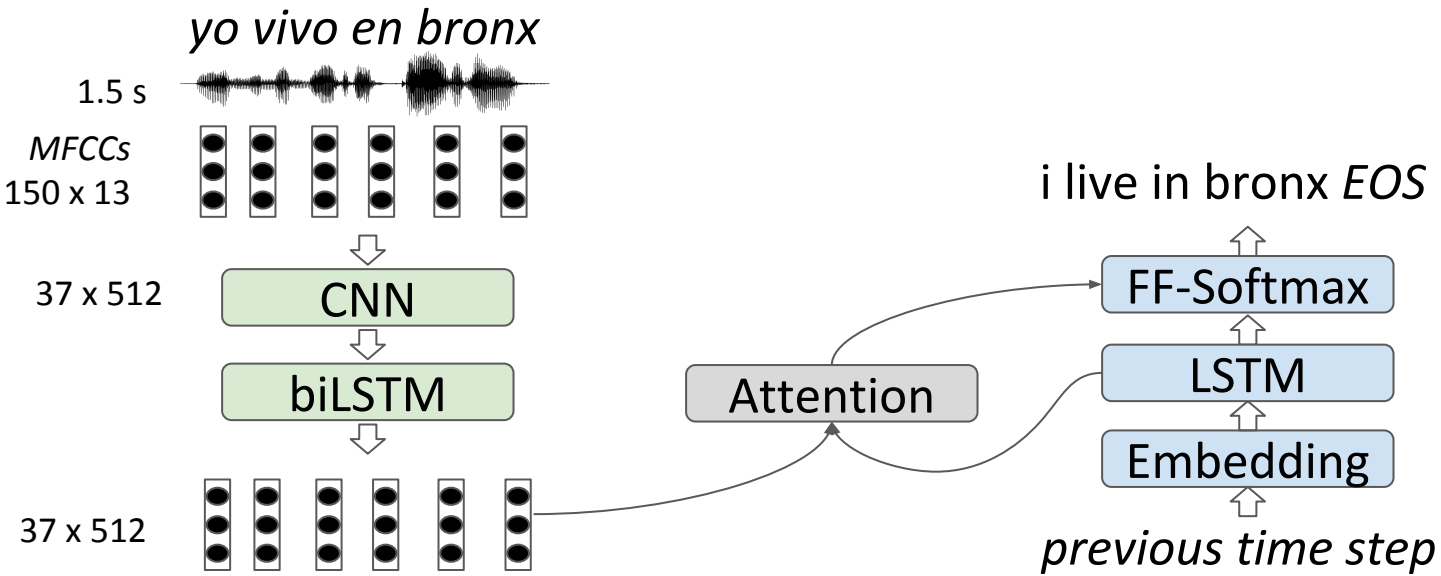
Spanish speech to English text



- Telephone speech (unscripted)
- Realistic noise conditions
- Multiple speakers and dialects
- Crowdsourced English text translations

Closer to real-world conditions

Neural ST model



Code available on Github

Cross-lingual applications for low-resource languages

- *Sheridan et al., 1997*
  - German speech retrieval system using French text queries.

- Projects *LORELEI, OpenCLIR* 


  - Query speech/text in a low-resource language using English (or similar high-resource).

- *Dredze et al. (2010) and Siu et al. (2014)*
  - Unsupervised clustering of speech into topics

- **Our work:** Speech paired with text translations