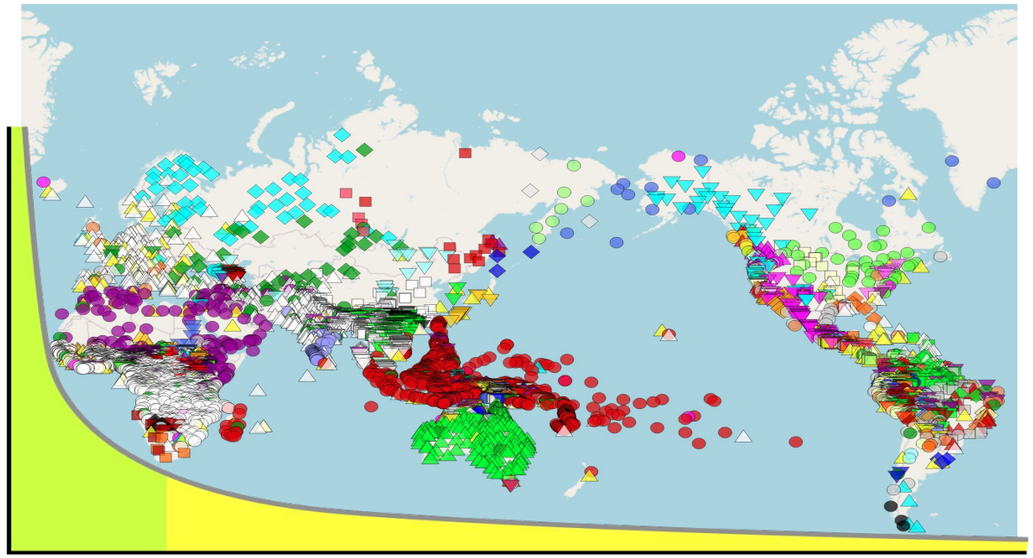
Sec 5.2

Multilingual ST

Multilingual ST

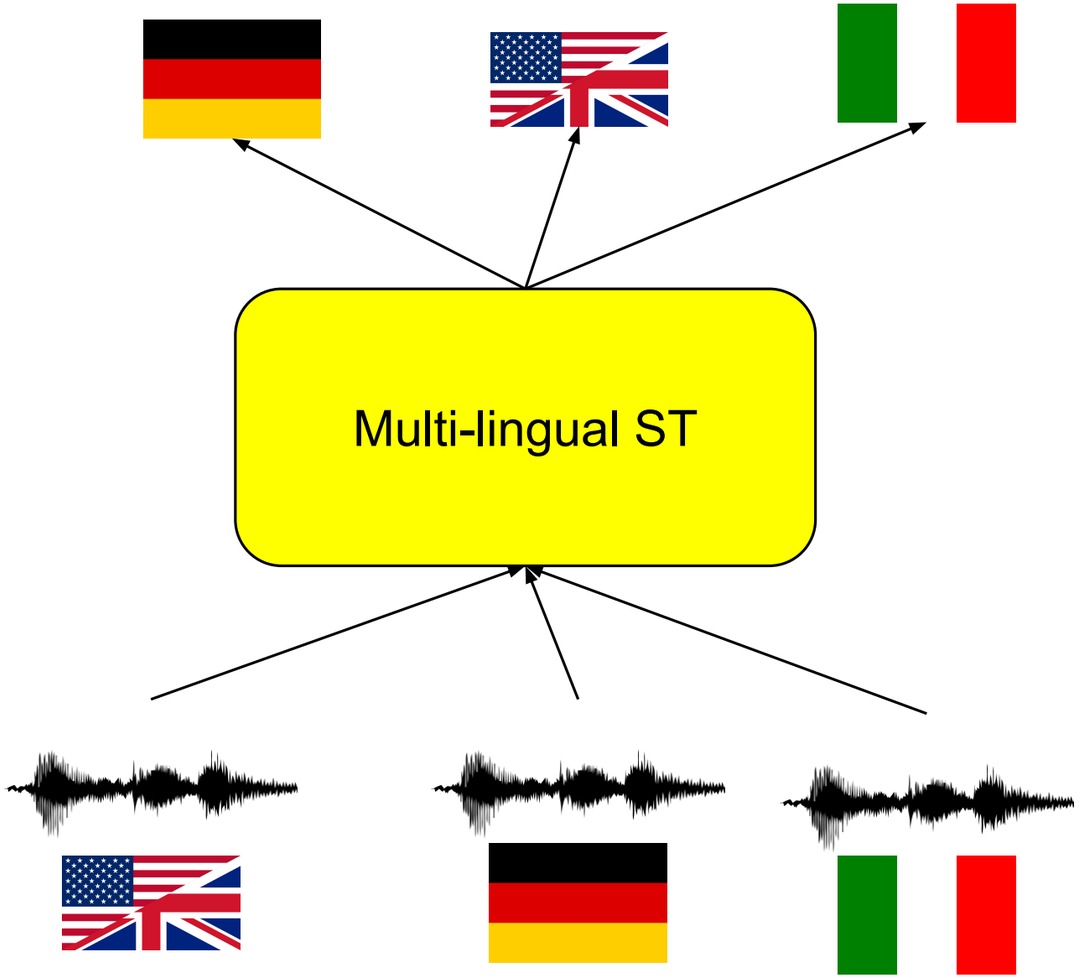
- Most research focuses on few languages
- More than *7,000 languages* in the world

- Challenges:
  - Scale to many languages
  - Limited resources



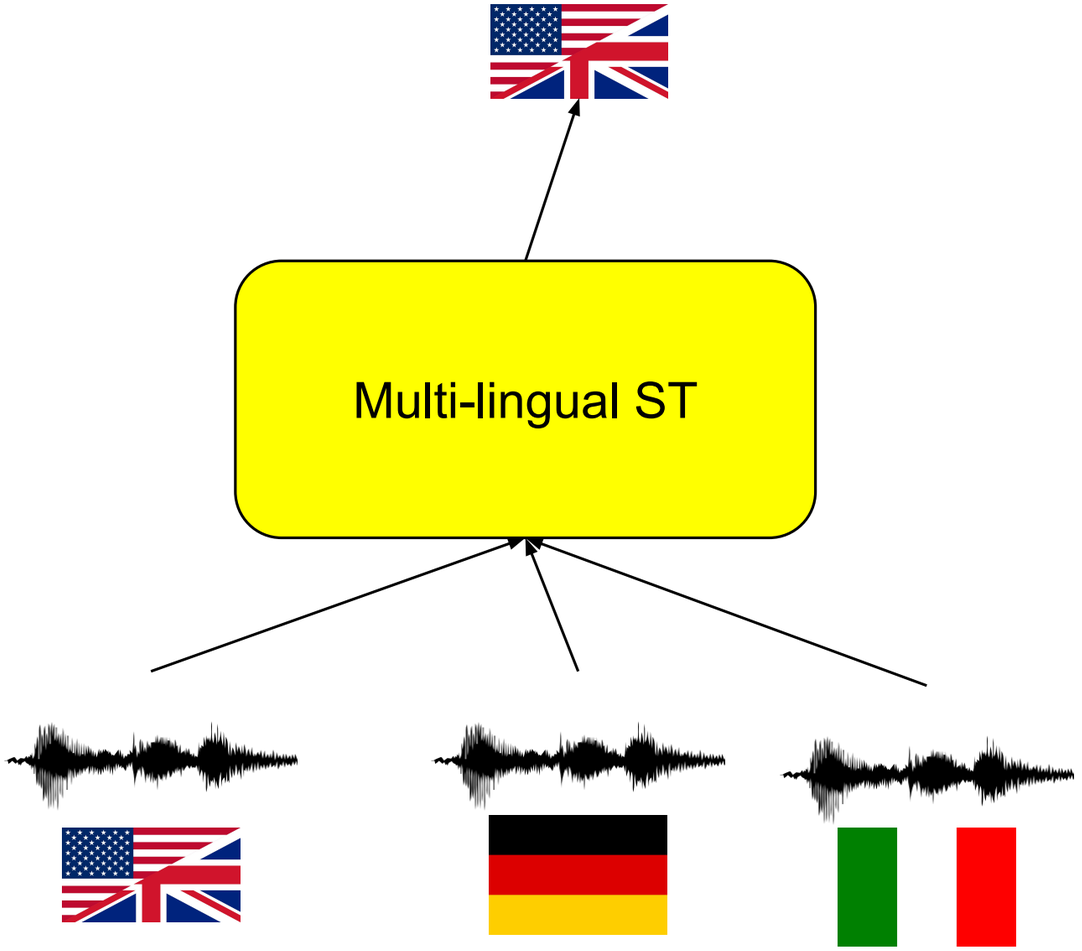
Multilingual ST

- Idea:
  - *Single model for many languages*
  - Motivated by text translation
- Advantages:
  - Less training data necessary
  - Handle several languages by single model
  - Zero-shot direction:
    - Translate between languages without training data



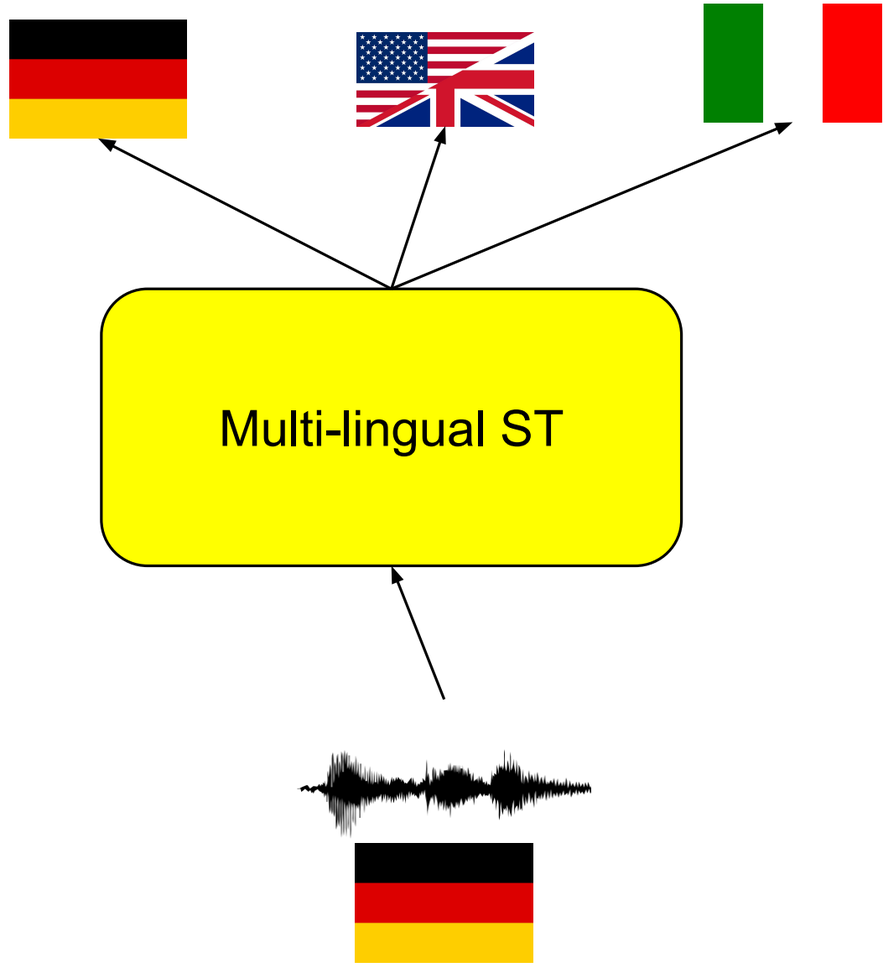
Multilingual ST

- Scenarios:
  - Many-to-One



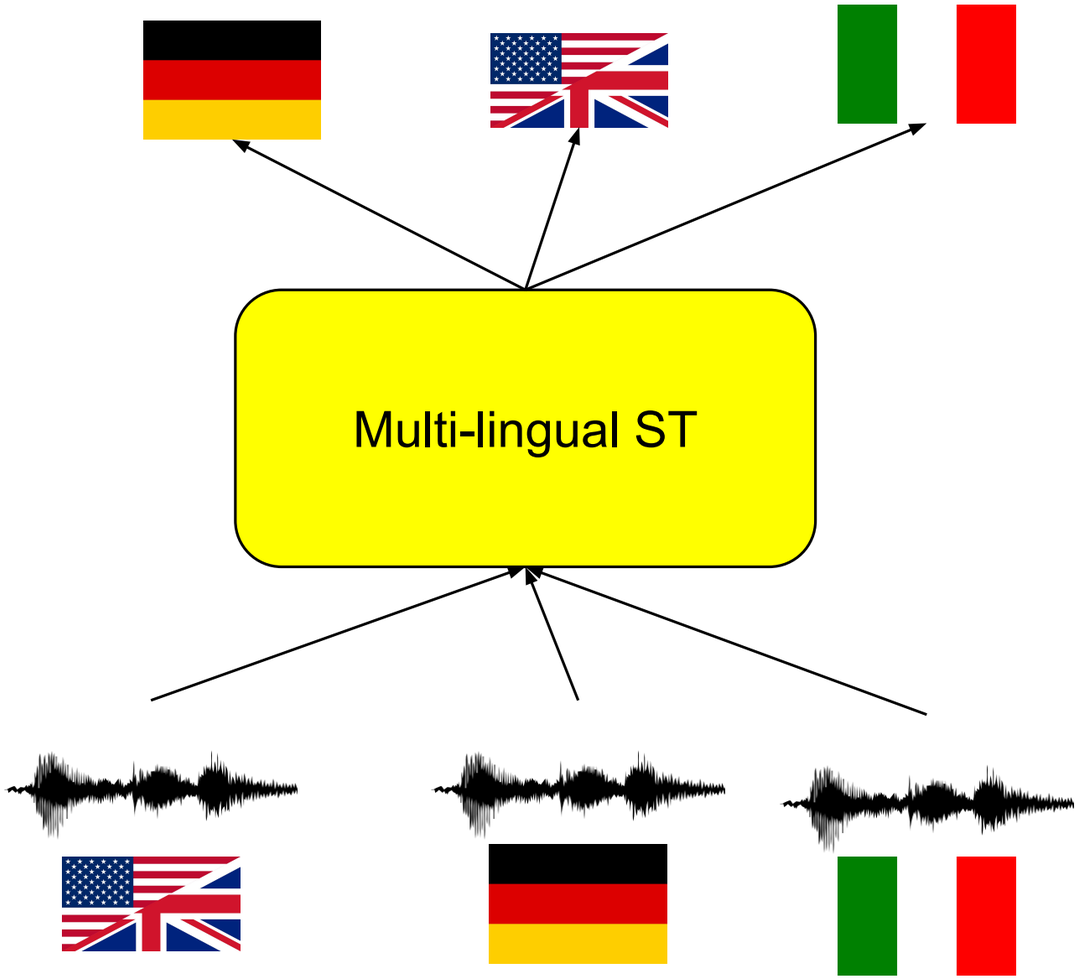
Multilingual ST

- Scenarios:
  - Many-to-One
  - One-to-Many



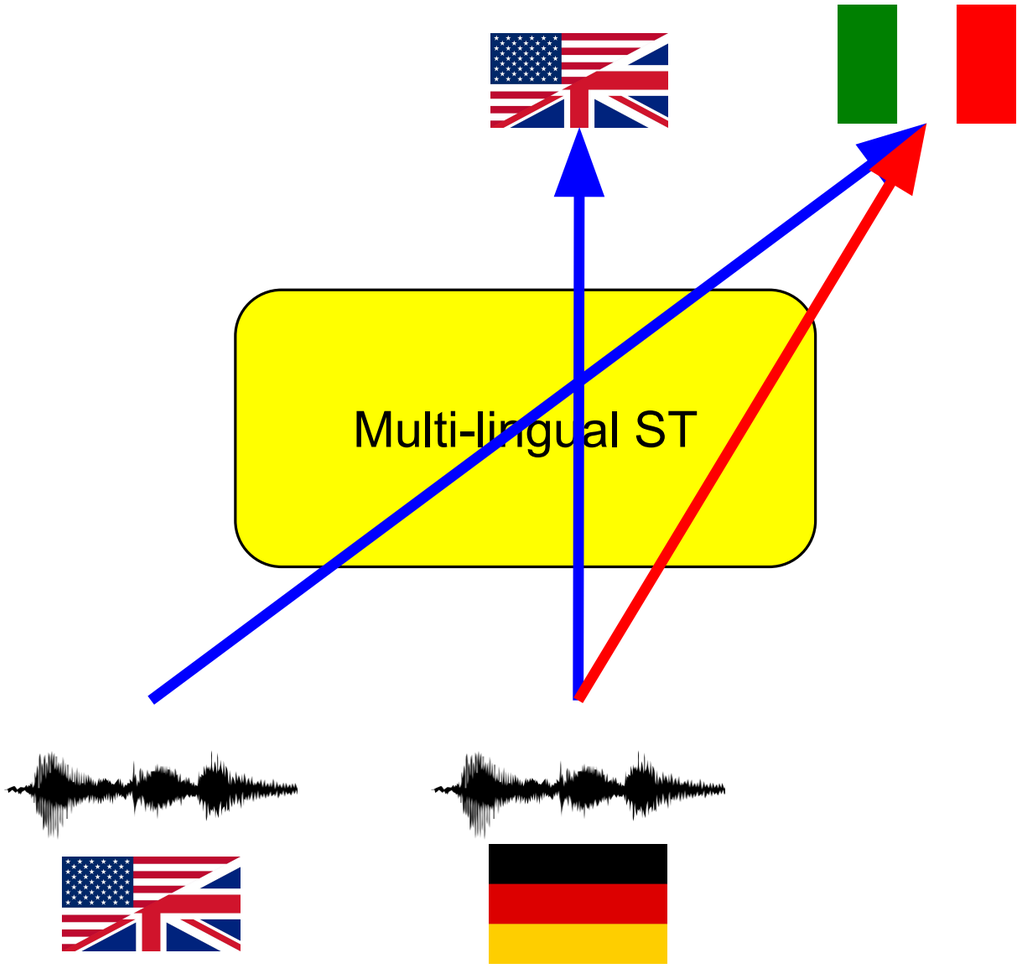
Multilingual ST

- Scenarios:
  - Many-to-One
  - One-to-Many
  - Many-to-Many



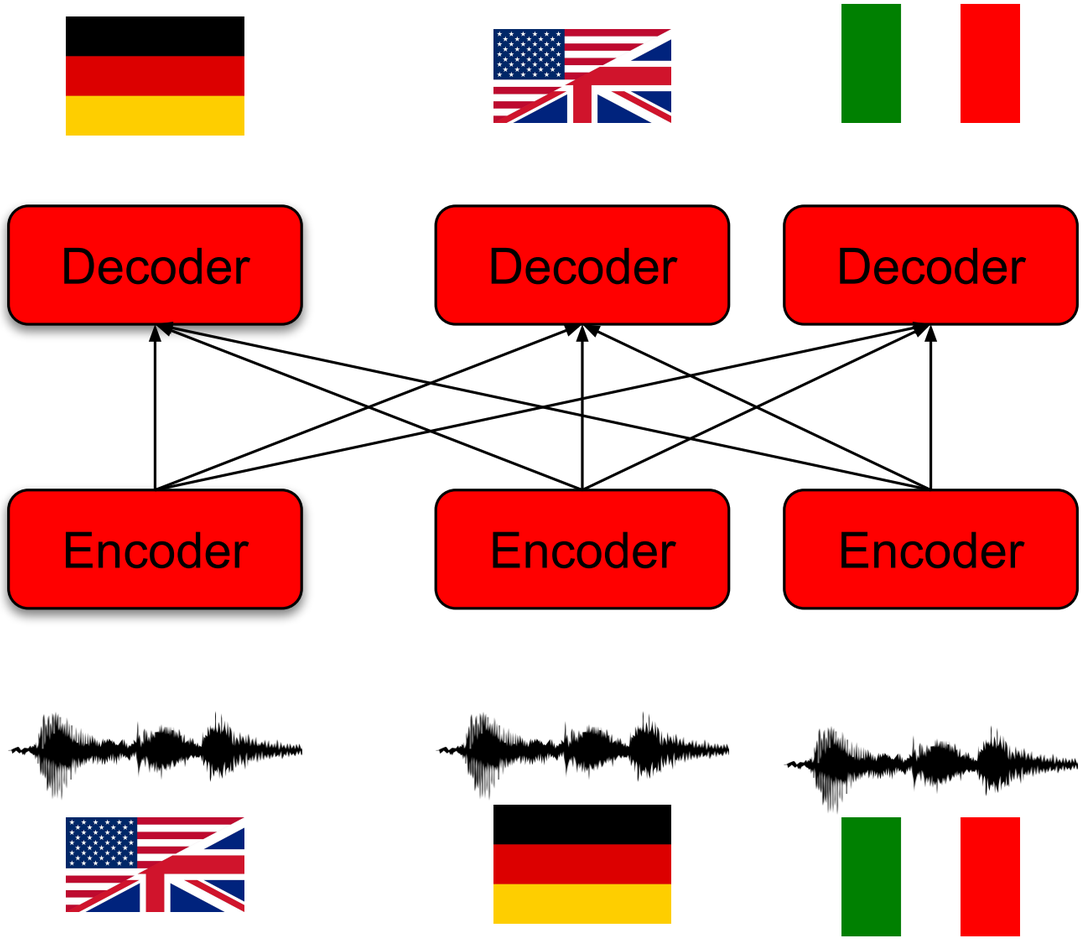
Multilingual ST

- Scenarios:
  - Many-to-One
  - One-to-Many
  - Many-to-Many
- Zero-shot:
  - No training data in test language pair



Training direction
Test direction

Multilingual ST - Architecture



Individual encoder and decoder for each language

(e.g. Escolano et al. 2020)

Multilingual ST - Architecture

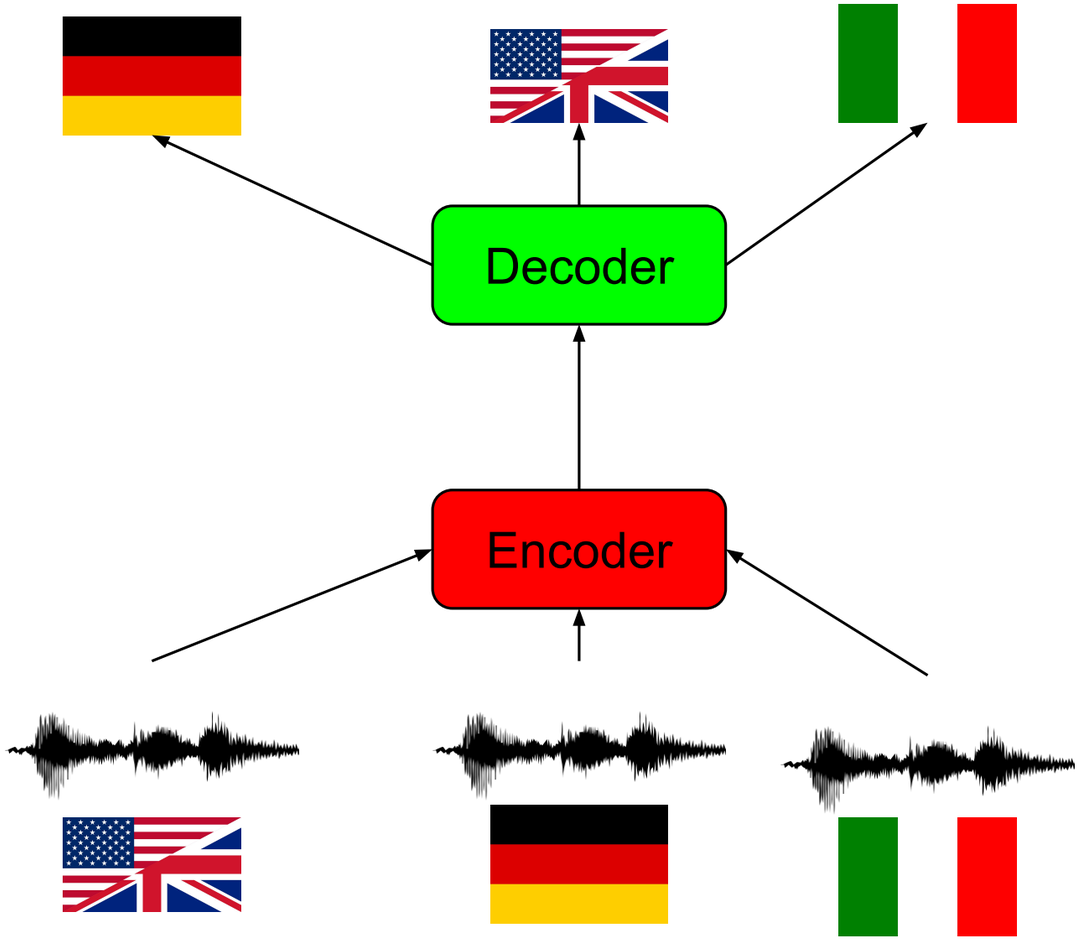
Joint encoder and decoder

Di Gangi et al., 2019

Inaguma et al., 2019

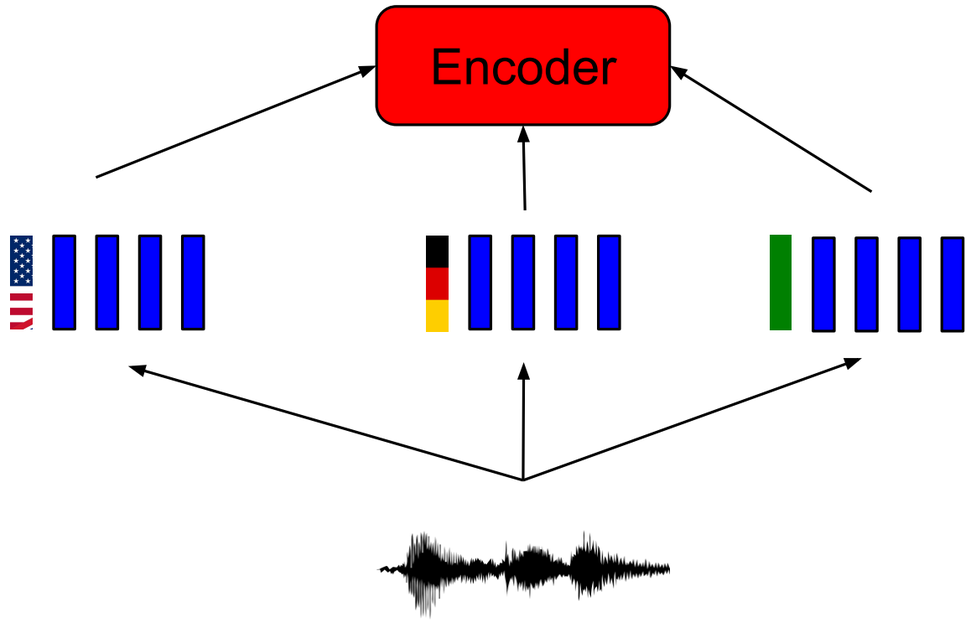
Challenge:

How to model different languages?



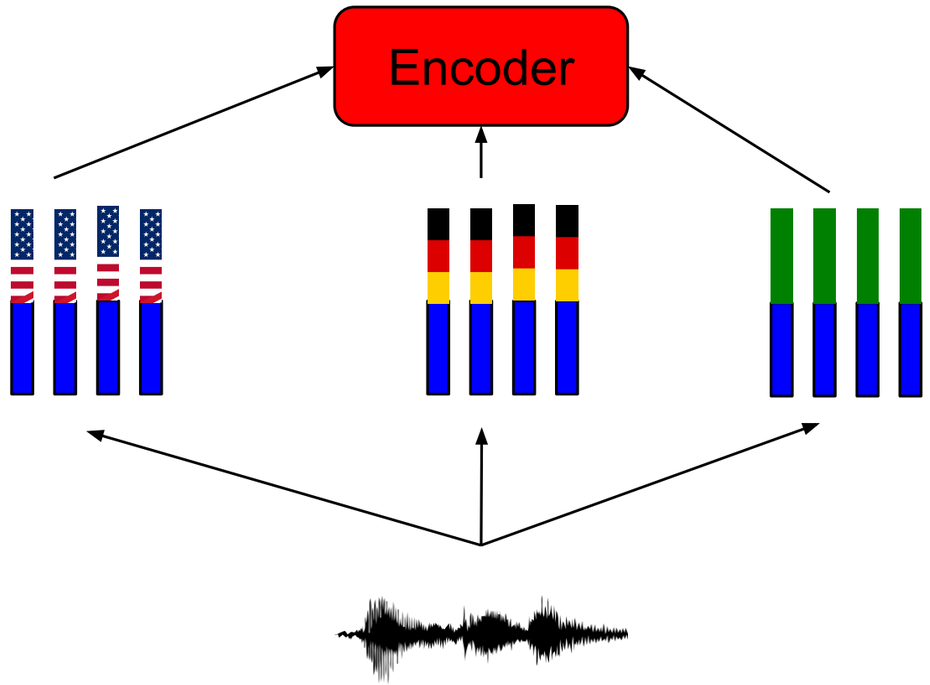
Multilingual ST - Language representation

- Encoder
  - Concat
    - Append learned language embedding for target language to audio features



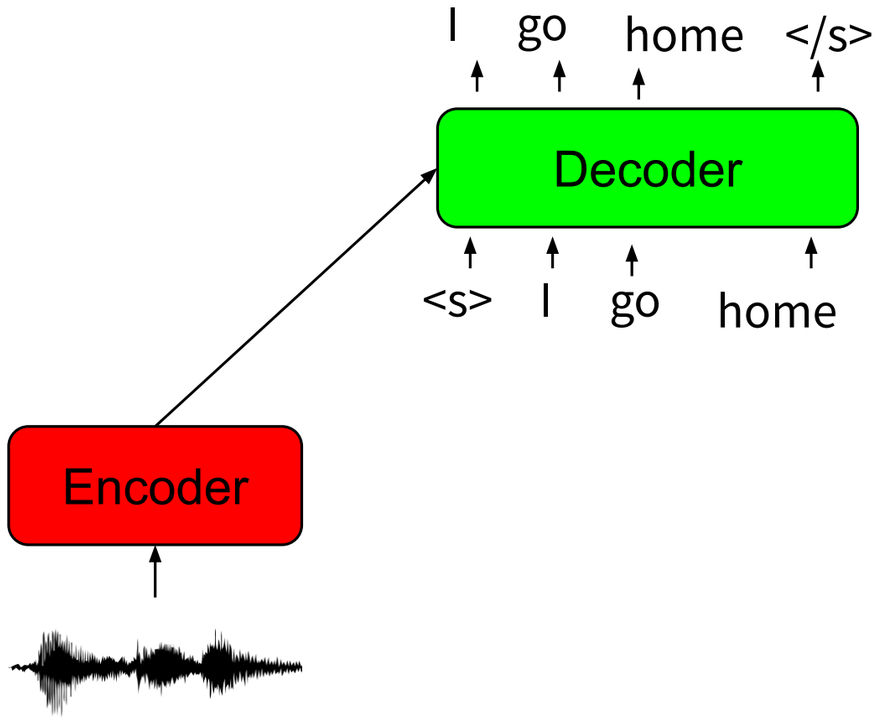
Multilingual ST - Language representation

- Encoder
  - Concat
    - Append learned language embedding for target language to audio features
  - Merge
    - Repeat language embedding for target language at each time step



Multilingual ST - Language representation

- Encoder
- Decoder



Multilingual ST - Language representation

- Encoder
- Decoder
  - Replace Begin of sentence by sentence embedding