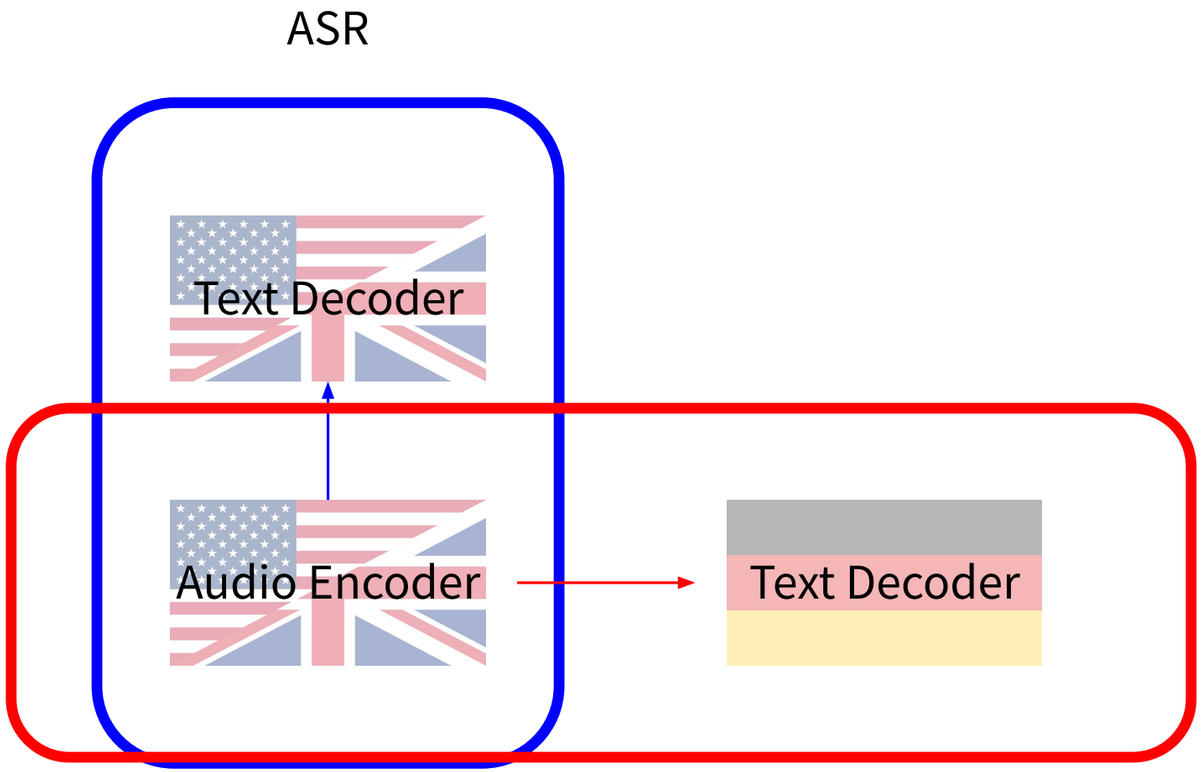
Sec 3.2.1

Multi-task Learning

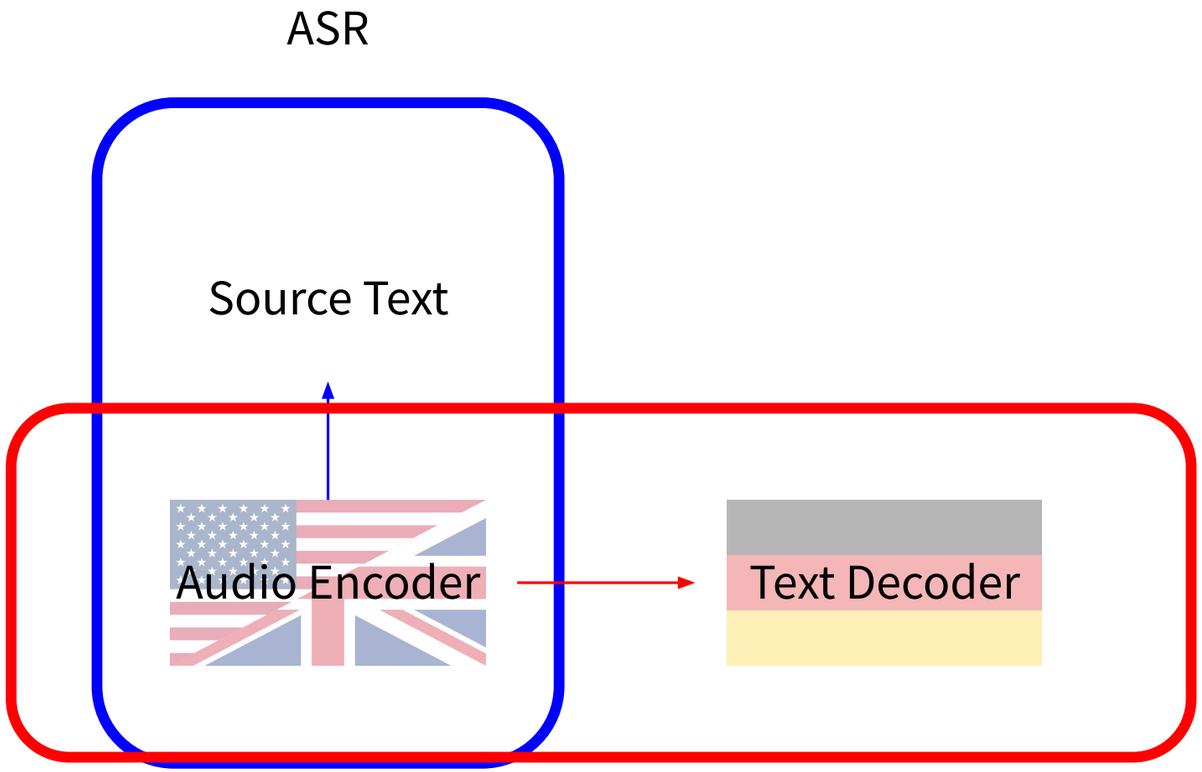
Multi-task learning

- Baseline
  - No changes to the architecture
- ST+ASR
  - One encoder
    - Source Language audio
  - Two decoder
    - Source Language text
    - Target language text
  - (Weis et al, 2017)



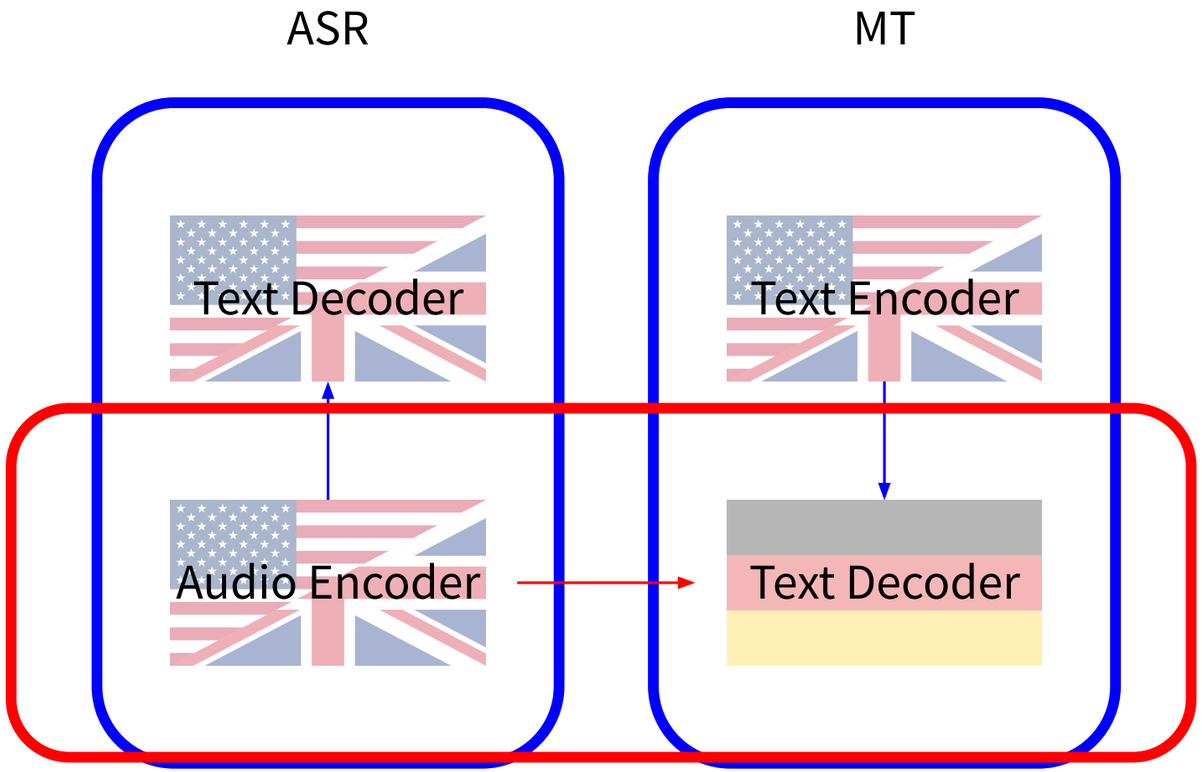
Multi-task learning

- Baseline
  - No changes to the architecture
- ST+ASR
  - One encoder
    - Source Language audio
  - Two decoder
    - Source Language text
    - Target language text
  - (Weis et al, 2017)
- ASR using CTC loss on encoder
  - (Hori et al, 2017)
  - (Bahra et al, 2019)



Multi-task learning

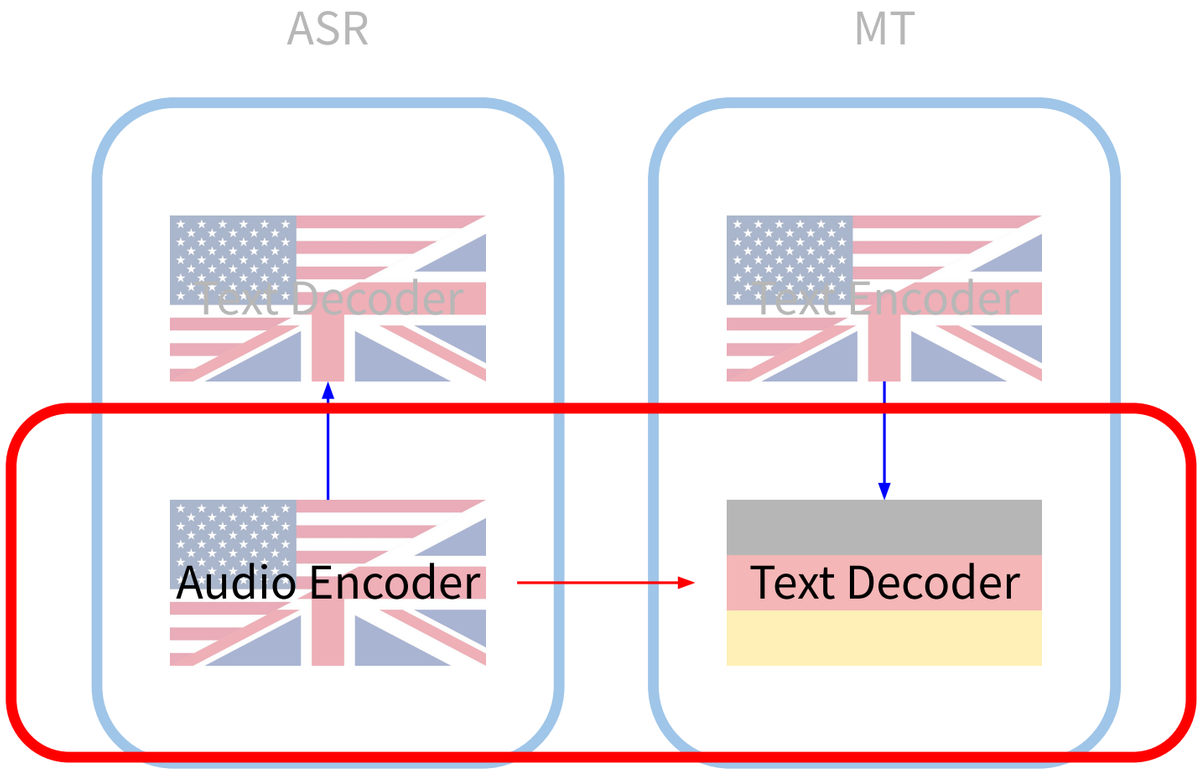
- Baseline
  - No changes to the architecture
- ST+ASR
- ST+ASR+MT
  - Two encoder
    - Source Language audio
    - Source Language text
  - Two decoder
    - Source Language text
    - Target language text
  - (Berard et al, 2018)



Multi-task learning

- Baseline
  - No changes to the architecture
- ST+ASR
- ST+ASR+MT
- Inference:
  - Direct translation
  - No use of additional parts

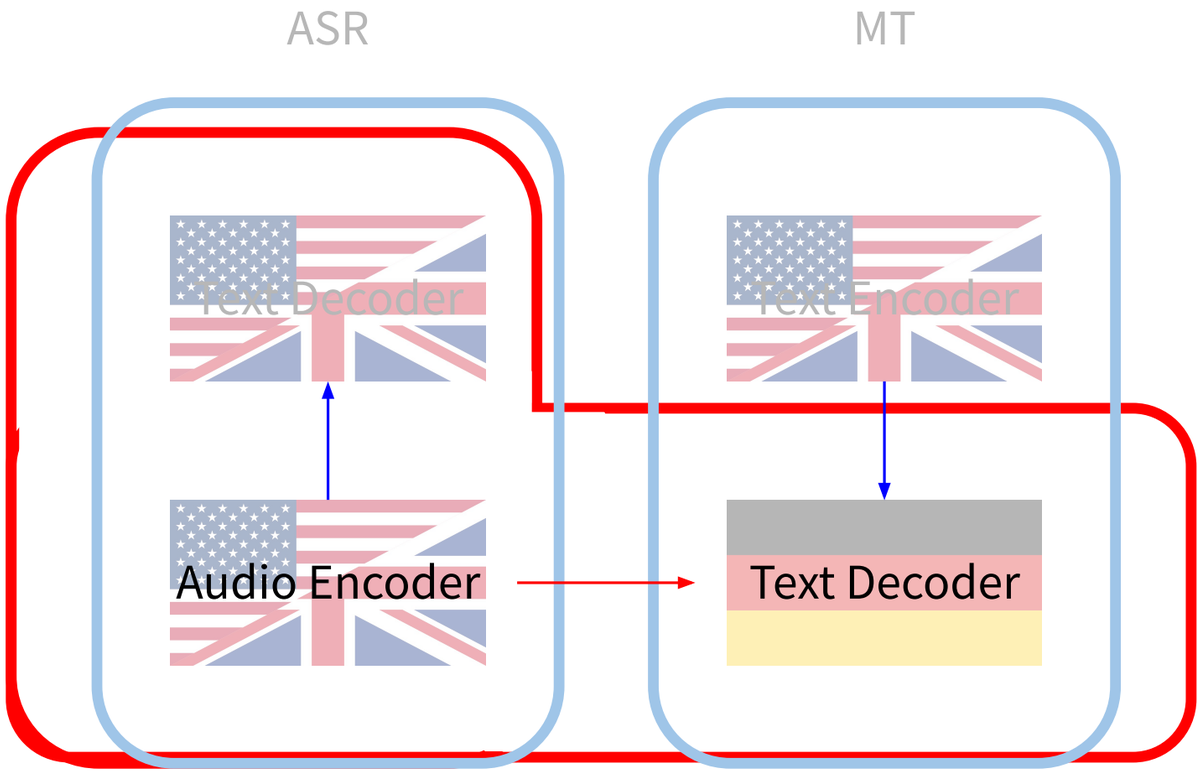
ST



2-stage models

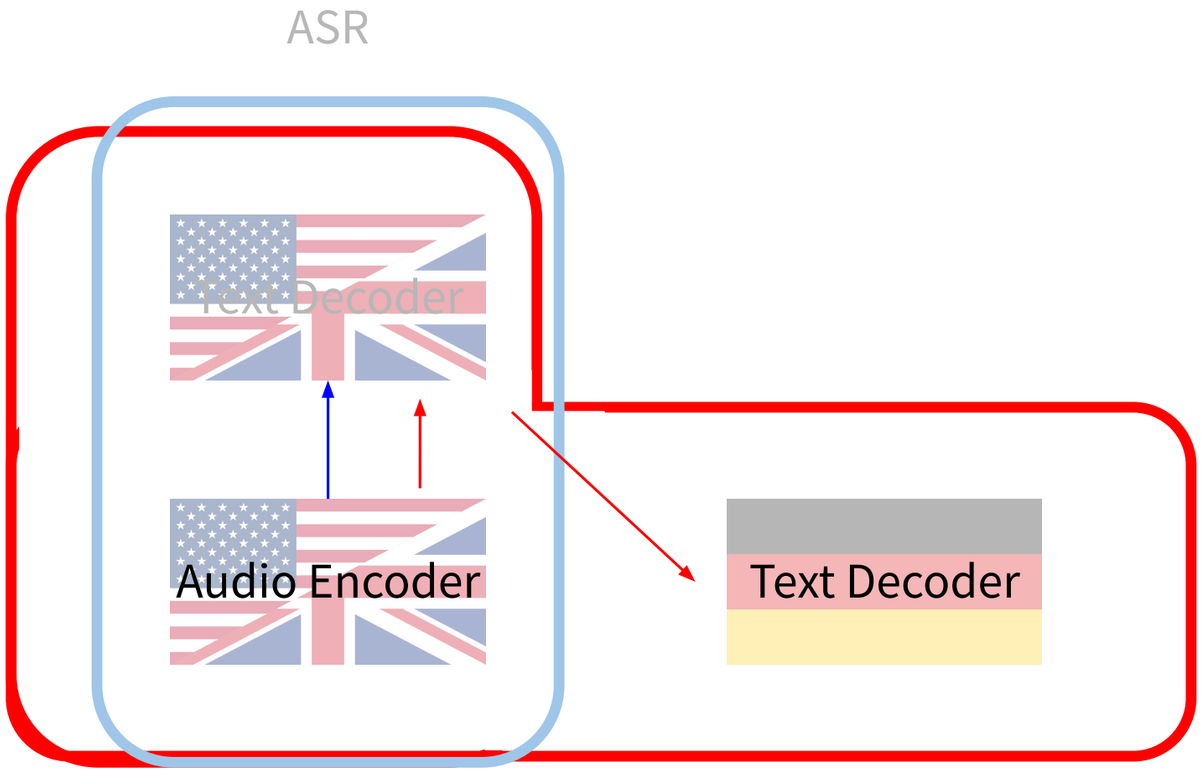
- Make use of additional model also during decoding
- *Simplify task*
  - using intermediate representation
- Comparison to cascade:
  - Full pipeline is trained
- Methods:
  - Adapt architecture
  - Preprocess data

ST



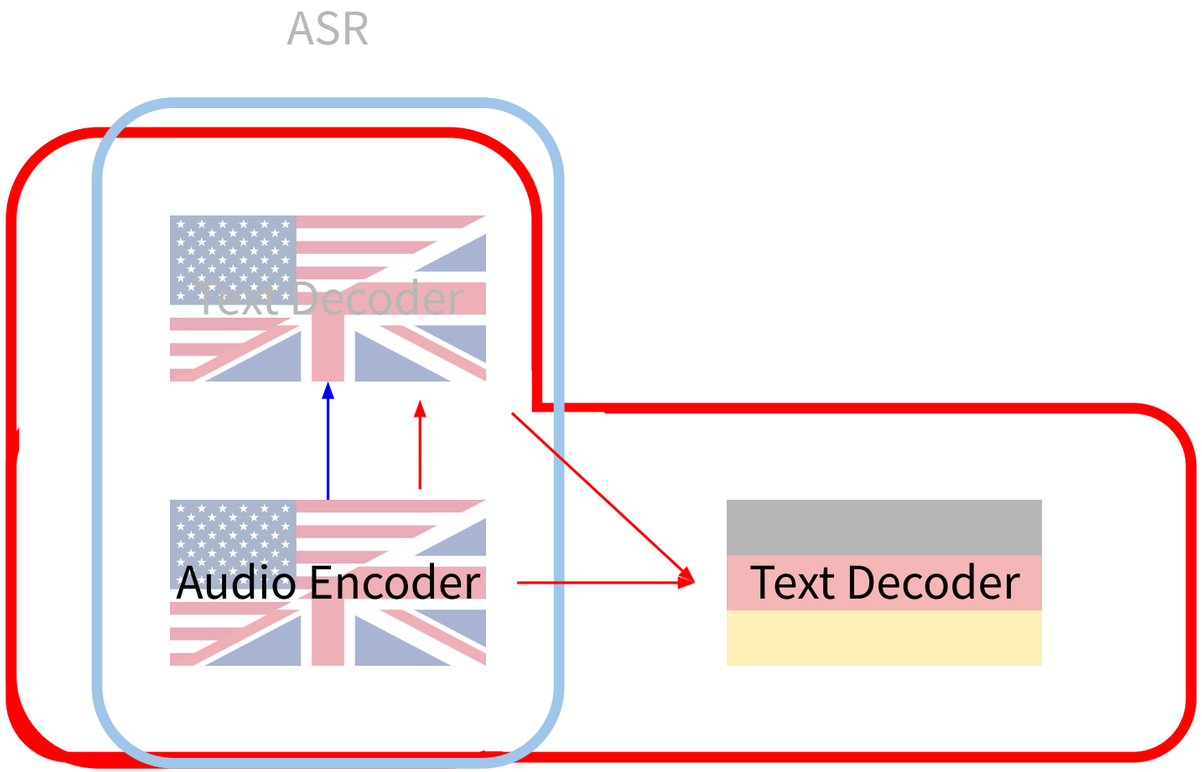
2-stage models

- Cascade:
  - Target language decoder attents to source text decoder
  - (Anastasopoulos Chiang, 2018)



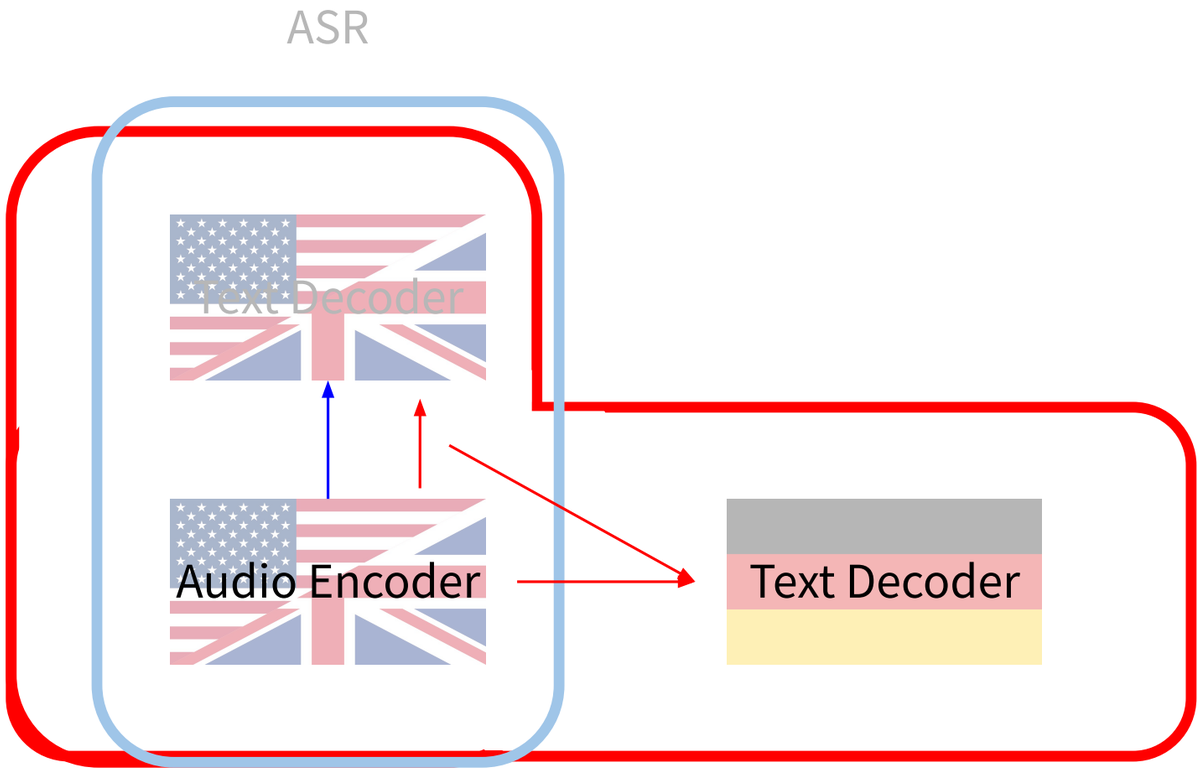
2-stage models

- Cascade:
- Triangle:
  - Target language decoder attents to source audio encoder and source text decoder
  - (Anastasopoulos Chiang, 2018)



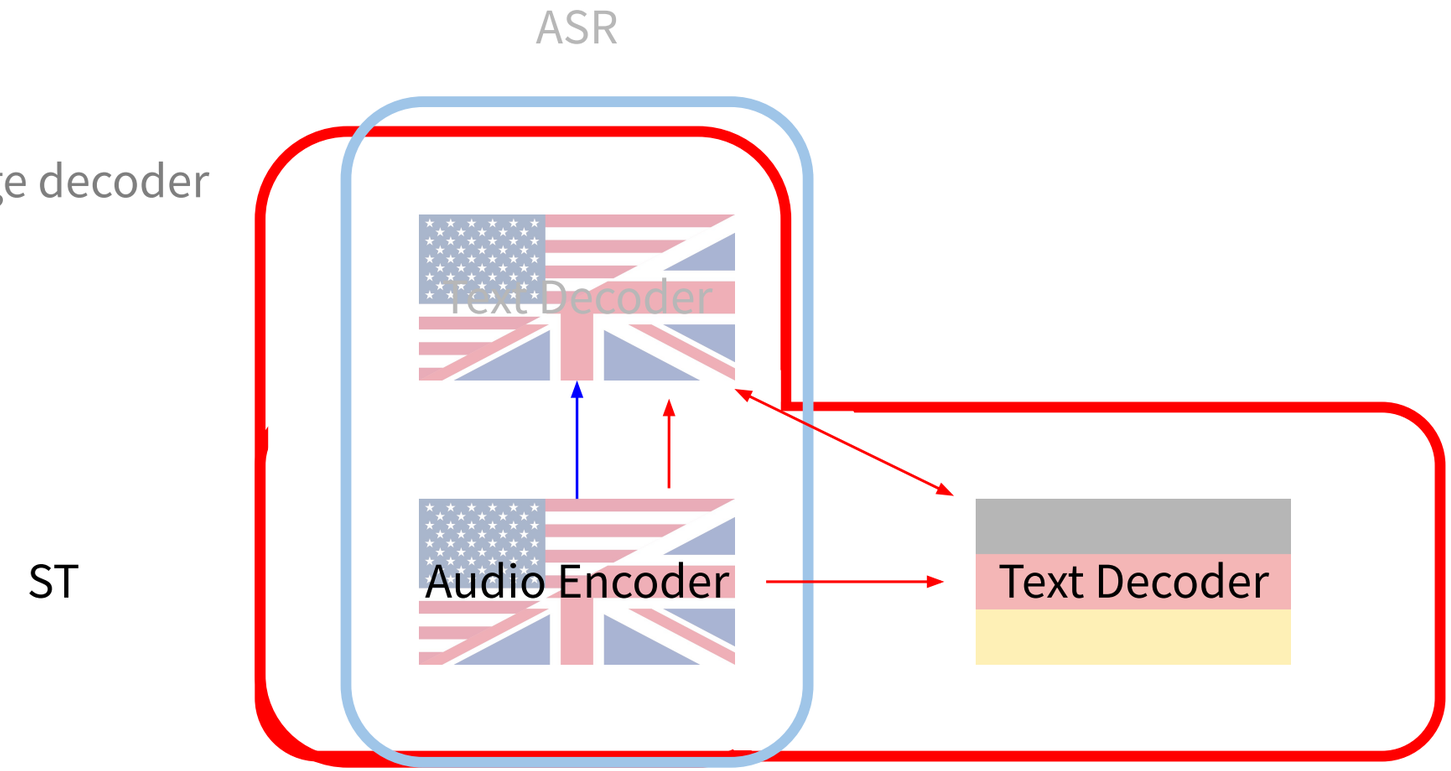
2-stage models

- Cascade:
- Triangle:
- Shared context vector
  - Target language decoder attents to source audio encoder and ASR context vectors
  - No direct influence of hard decisions of source text decoder
  - (Sperber et al, 2019)



2-stage models

- Cascade:
- Triangle:
- Shared context vector
- Dual Decoder
  - Source and target language decoder run in parallel
  - Attend to each other
  - (Le et al, 2020)



2-stage models

- Cascade:
- Triangle:
- Shared context vector
- Dual Decoder
- Concat
  - Single decoder generates source and target language
  - Output is concatenation
  - (Sperber et al, 2020)