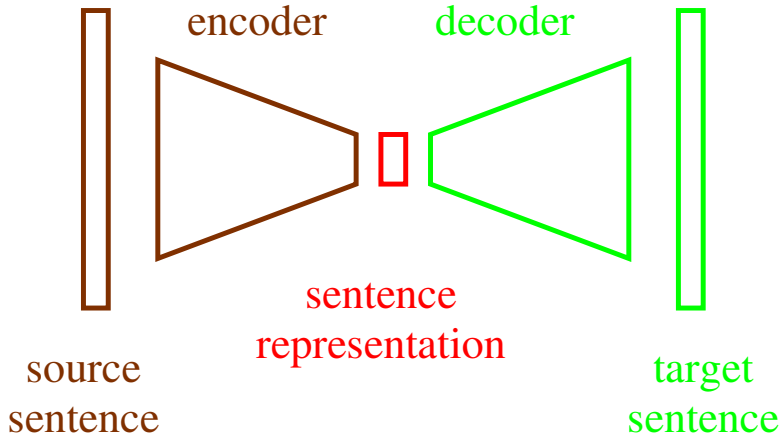
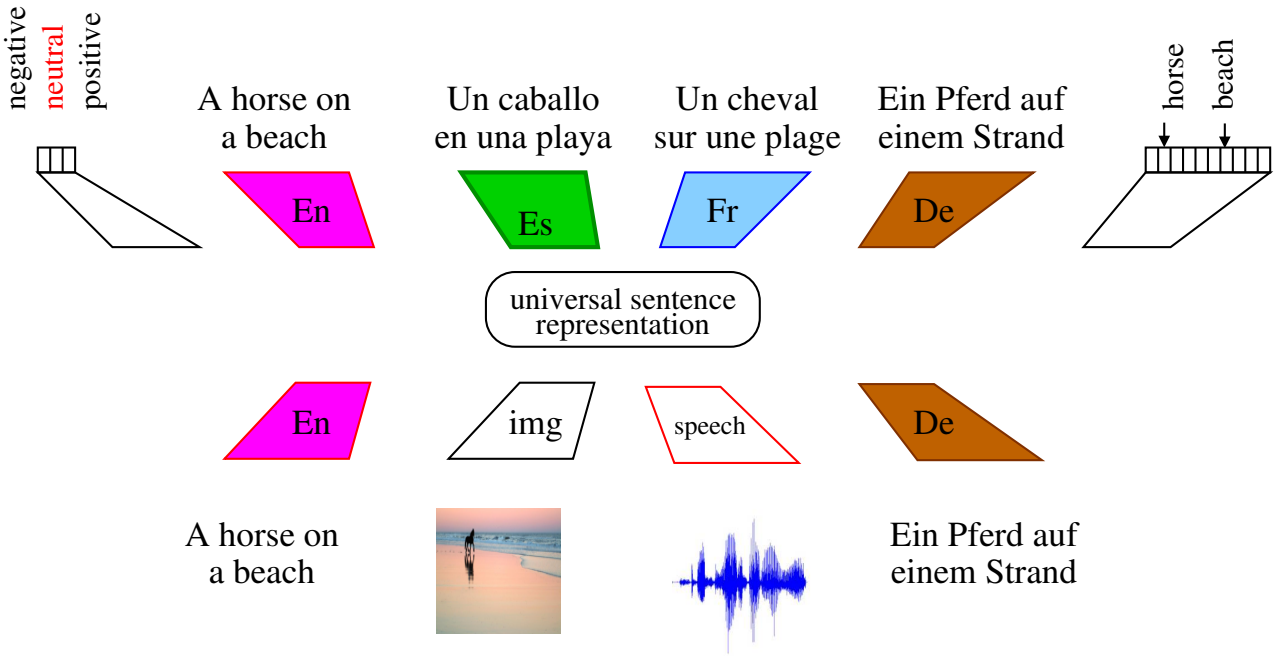
Sequence-to-Sequence Processing

General idea

- An encoder processes the source sentence and creates an compact representation
- This representation is the input to the decoder which generates a sequence in the target language
- Both encoder and decoder are RNNs

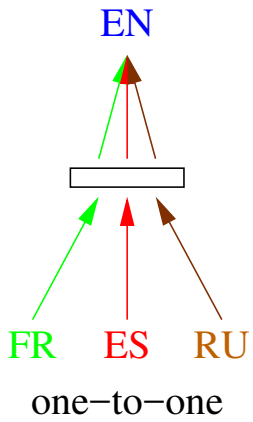


Multiple Encoder/Decoder Framework



- Use several encoders and decoders
  - different language pairs
  - other Seq2Seq tasks (speech)
  - sentence classification tasks (sequence-to-category)
  - image captioning (image-to-sequence)
- *"Force"* the representation to be identical for all encoders

Different Training Paths



One-to-one strategy

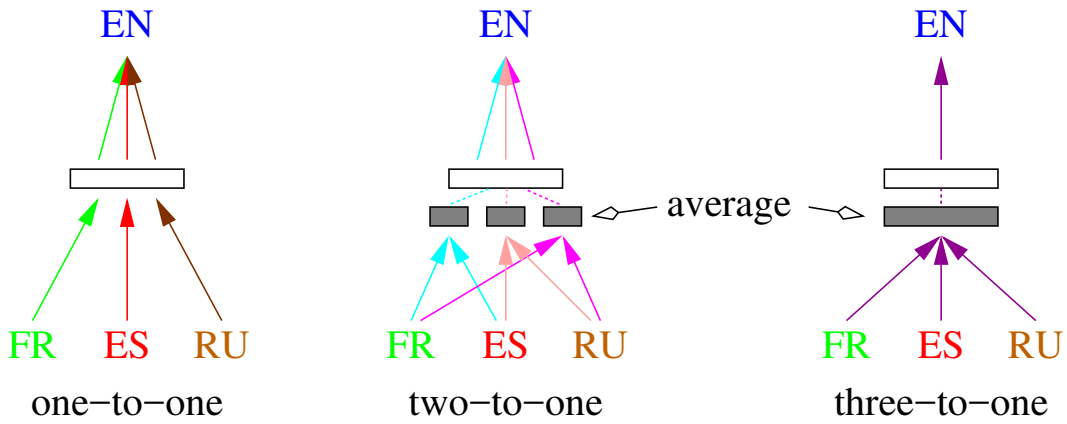
- Alternate between different language pairs with one common target

$\Rightarrow$ Encourages joint representation

$+$ Train with pairwise parallel data

$-$ No embedding for output language

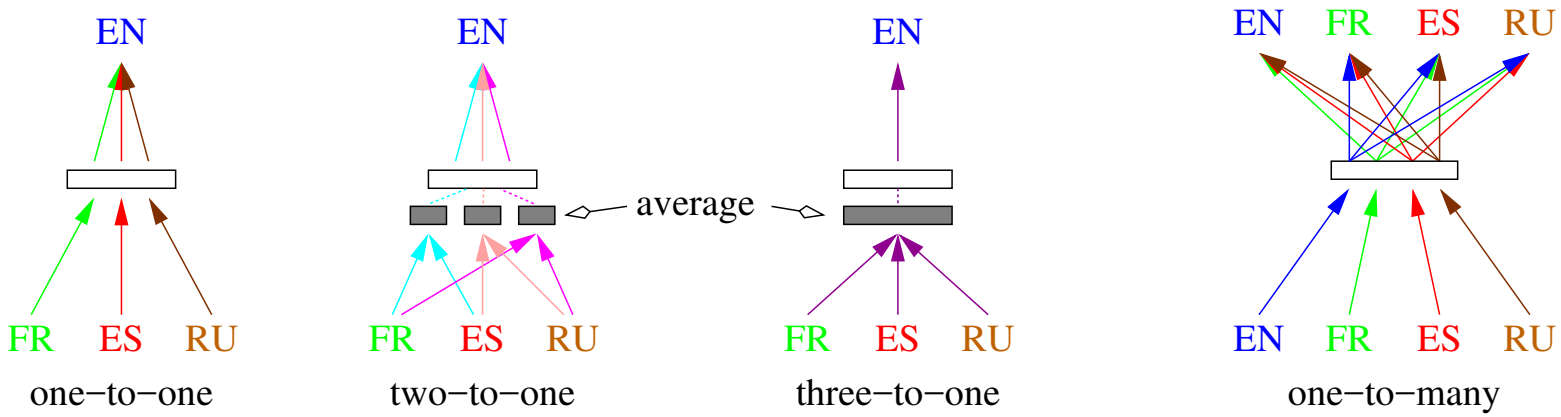
Different Training Paths



Many-to-one strategy

- Add a regularizer to explicitly force a joint representations, eg. average, correlation, ...
  - Still no embedding for output language

Different Training Paths



One-to-many strategy

- Translate from one to all other language, source excluded

⇒ Always at least one common target language

- Sentence embeddings for all languages

– Needs N-way parallel training corpora

- Extension to *"many-to-many strategy"* straightforward

Evaluation of Sentence Representations

Desired properties:

- **semantic closeness:** similar sentences should be close in the embeddings space
- **multilingual closeness:** identical sentences in different languages should be close
- **preservation of content:** task specific: NMT, classification, entailment, etc.
- **scalability to many languages:** limit the need of human labeling of data.