

Effective Strategies in Zero-Shot Neural Machine Translation

Thanh-Le Ha, Jan Niehues, Alexander Waibel

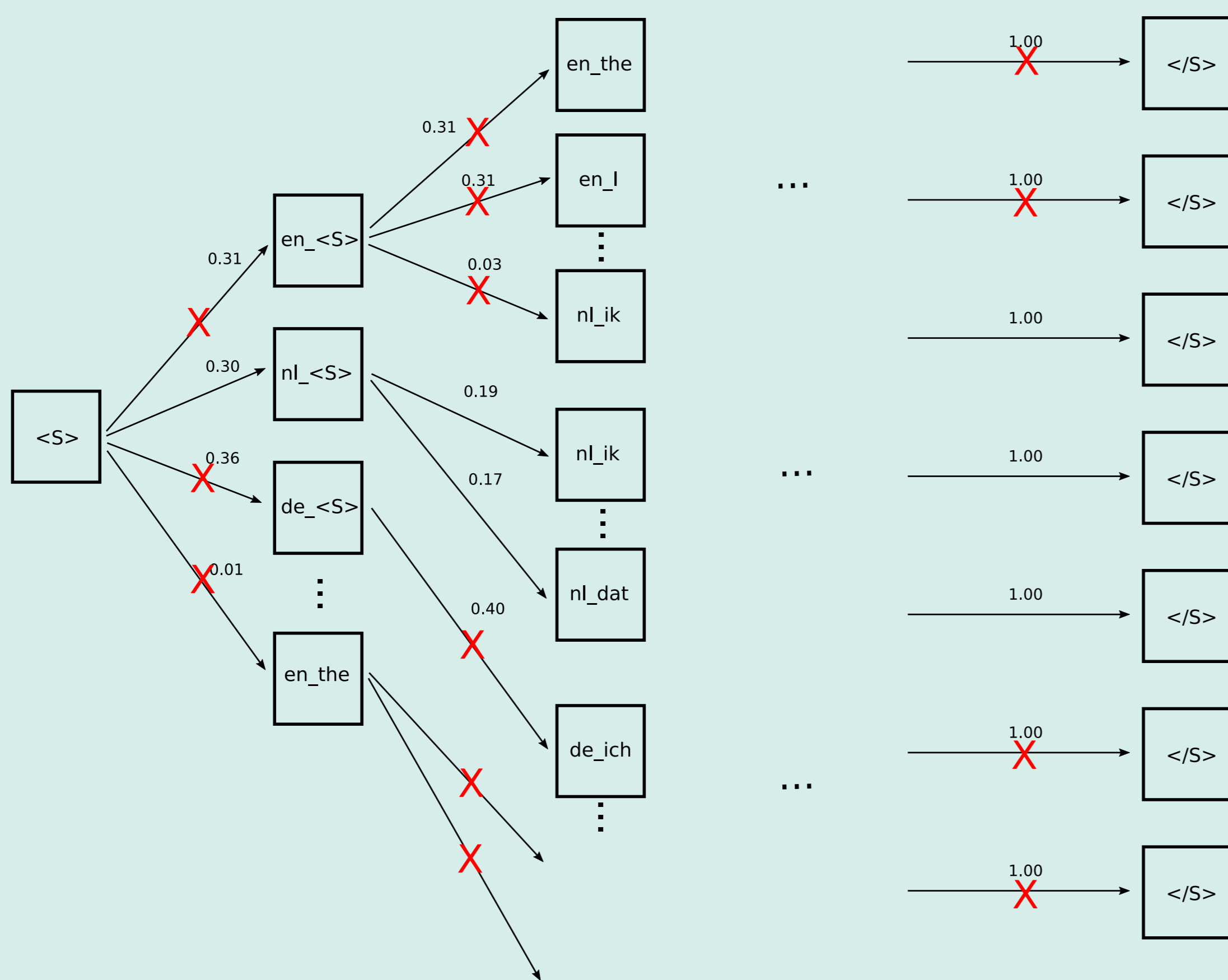
Zero-shot Translation

- Translate between two languages **without any parallel corpus** of that language pair
- E.g. IWSLT'17 Zero-shot tasks:
 - Given 16 pair-wise parallel corpora among {English, German, Dutch, Italian, Romanian},
 - Translate German↔Dutch and Italian↔Romanian

Multilingual Neural Machine Translation using Universal Encoder and Decoder

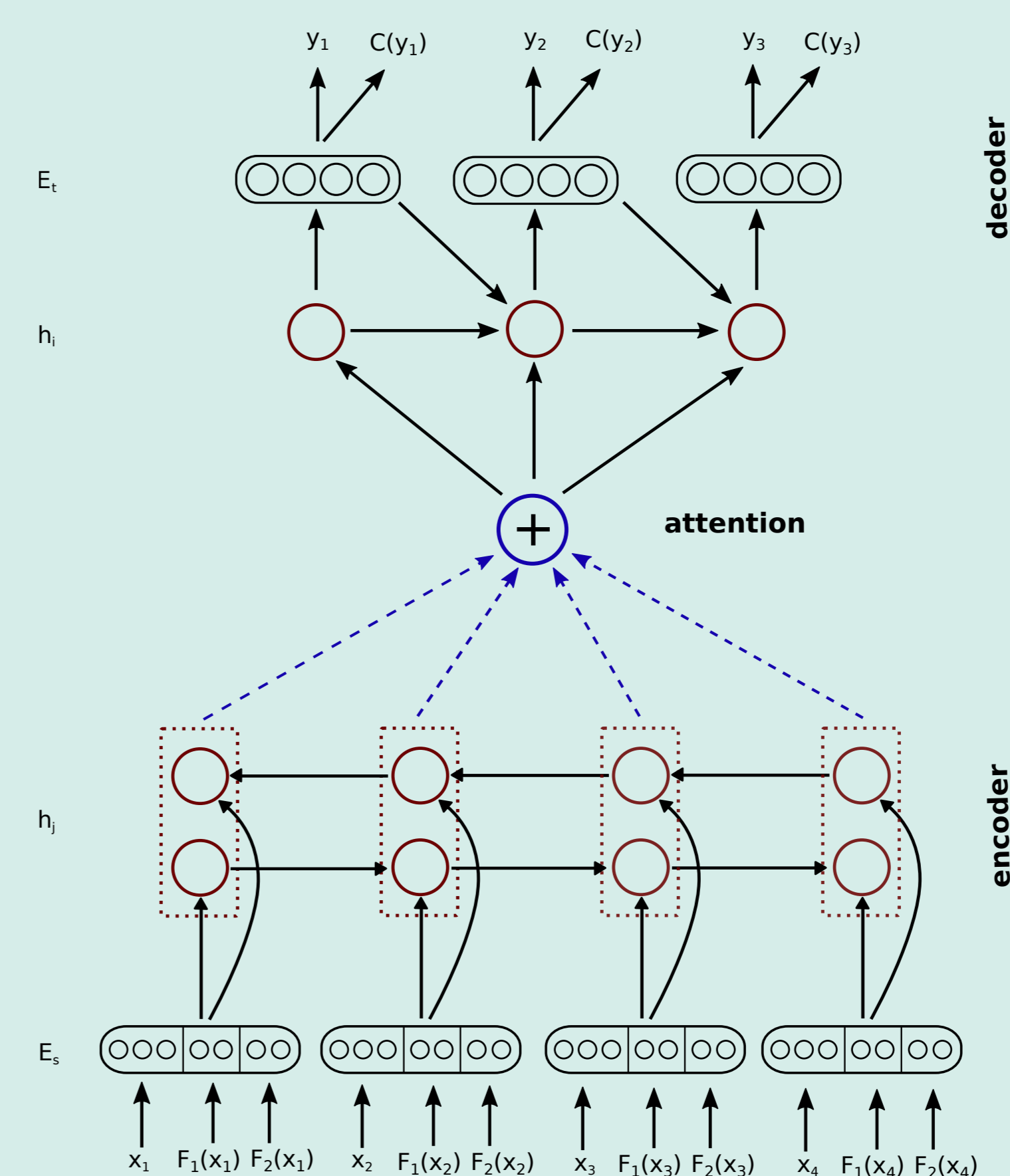
- Train a single model for many translation directions using a standard NMT architecture
- Shared encoder, attention and decoder for all directions
- Two-step preprocessing:
 - Language Coding**: attach language ID for each word
 - Target Forcing**: force the system to translate into a desired target language
- Example:
 - Source Sentence (German): **<en>** de_versetzen de_Sie de_sich de_mal de_in de_meine de_Lage **<en>**
 - Target Sentence (English): en__ en_put en_yourself en_in en_my en_position en_.
- Can apply in Zero-shot Translation.
- Disadvantage**:
 - Large Vocabulary Size (Merged from vocabularies of different languages)
 - In Zero-shot Translation: Language Bias and Ambiguities.

Target Vocabulary Filtering



- Filter out vocabulary entries of other languages prior to translation
- Reduce vocabulary size ⇒ Faster in decoding
- Reduce cross-language ambiguities.

Language as a Word Feature



- Modify the NMT architecture to take word features into account
- Use the source and target languages as word features
- Those language information help to reduce the translation ambiguities

Results

	System	Zero-shot?	German→Dutch		German→Romanian	
			dev2010	tst2010	dev2010	tst2010
(1)	Zero 2L	Yes	4.79	5.75	1.55	2.05
(2)	Zero 4L	Yes	6.31	7.93	3.15	3.73
(3)	Zero 6L	Yes	11.58	14.95	8.61	10.83
(3a)	Zero 6L Filtered Dict	Yes	12.50	16.02	9.10	11.00
(3b)	Zero 6L Lang Feature	Yes	13.95	17.15	9.88	11.37
(4)	Back-Trans	No	17.33	20.36	12.92	15.62
(4a)	Back-Trans Filtered Dict	No	17.13	20.22	13.10	15.67
(4b)	Back-Trans Lang Feature	No	17.48	20.24	13.43	15.70

German→Dutch			
System	#parameters (millions)	Vocab Size (thousands)	Training Time (hours/epoch)
Zero 6L	243	68	7.3
Lang Feature	130	28	1.5
German→Romanian			
System	#parameters (millions)	Vocab Size (thousands)	Training Time (hours/epoch)
Zero 6L	247	69	6.0
Lang Feature	122	31	1.3