NLP 南京大学 自然语言处理研究组 Natural Language Processing Group

南京大学 NANJING UNIVERSITY

机器翻译相关的研究方向

黄书剑

论文发表情况粗略统计

- ACL17
  - long 14/195; short 12/107
- EMNLP17
  - 26/322
- NAACL18
  - long 13/207; short 8/125
- ACL18
  - long 12/258; short 10/126
- Total:
  - 80+
- Other papers appear in IJCAI, AAAI, NIPS, ICLR or TACL, TASLP etc.

神经网络机器翻译的研究热点

- 网络结构变化
- 提升解码效率

→ 提高训练解码效率

- 无监督/低资源翻译
- 领域自适应
- 多模态翻译

→ 扩大适用范围

- 可解释性、可视化、分析
- 鲁棒性、对抗样本

→ 可理解性

- 引入句法结构信息
- 引入统计机器翻译的经验
- 记忆网络、注意力改进

→ 提高翻译质量

- ……

*下文列出了部分相关方向的论文作为参考

论文发表情况粗略统计

- ACL 22
  - long 44/604; short 7/98; srw 5/40
- EMNLP 22
  - 52/828
- NAACL 22
  - long 29/443

- 多语言机器翻译 Multilingual Machine Translation
- 语音翻译与同传 Speech/Simultaneous Translation
- 机器翻译与预训练 Pretraining and Translation
- 语言相关研究 Linguistics

多语言机器翻译

Sharing Encoders

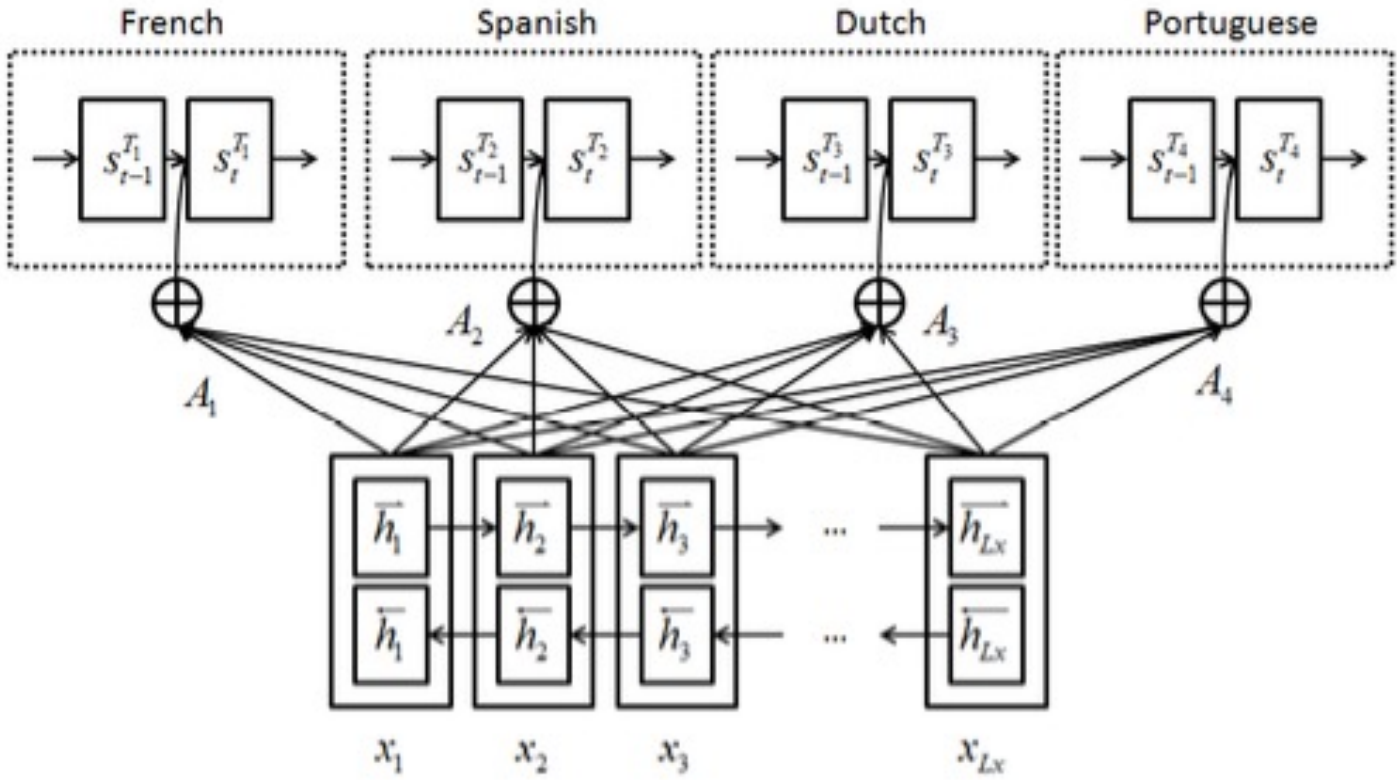
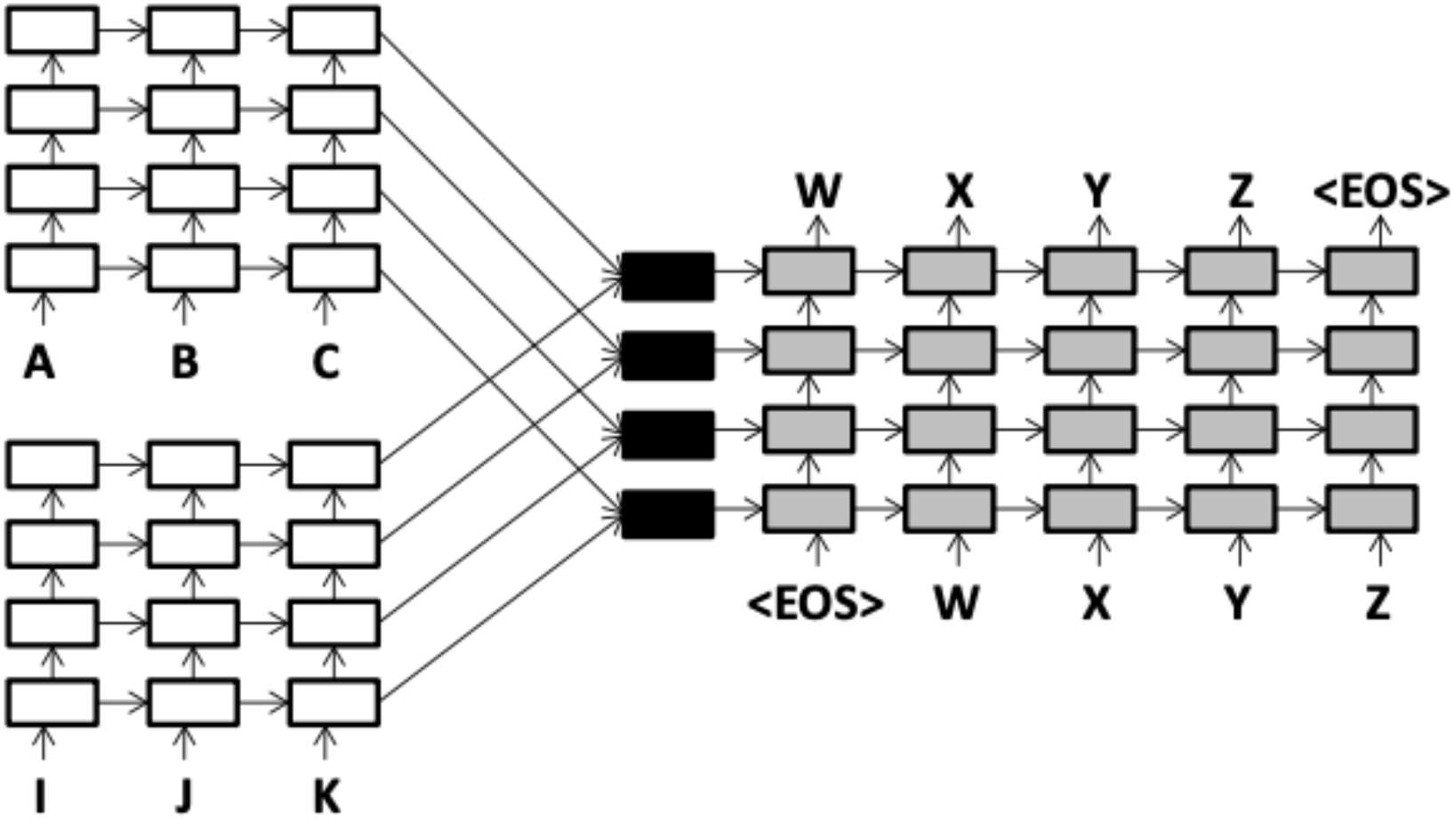


Figure 2: Multi-task learning framework for multiple-target language translation

Multi-Task Learning for Multiple Language Translation Dong et al. 2015

Sharing decoders



Multi-Source Neural Translation. Zoph and Knight 2016

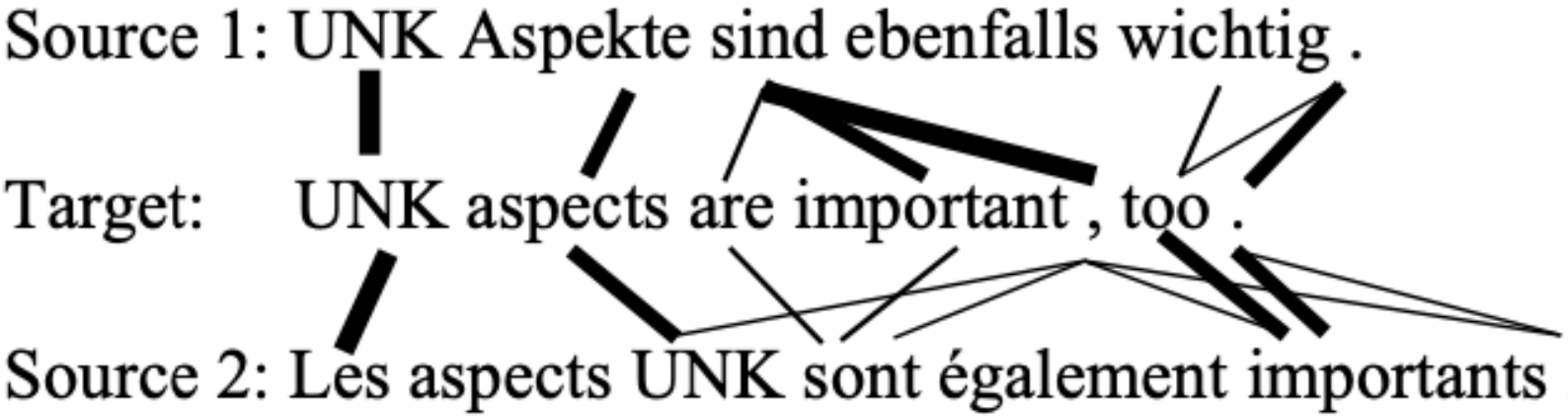
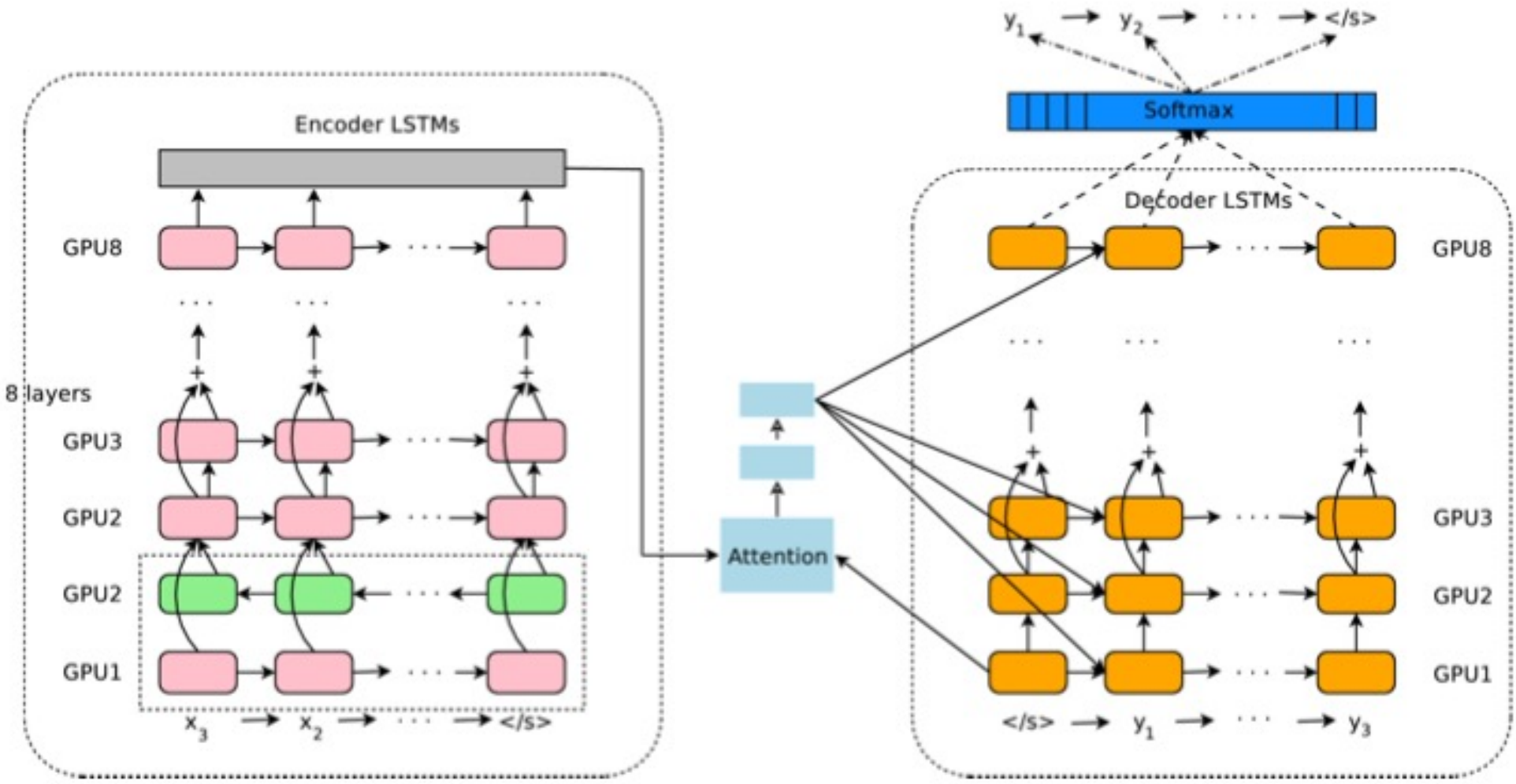


Figure 5: Action of the multi-attention model as the neural decoder generates target English from French/German sources (test set). Lines show strengths of $a_t(s)$.

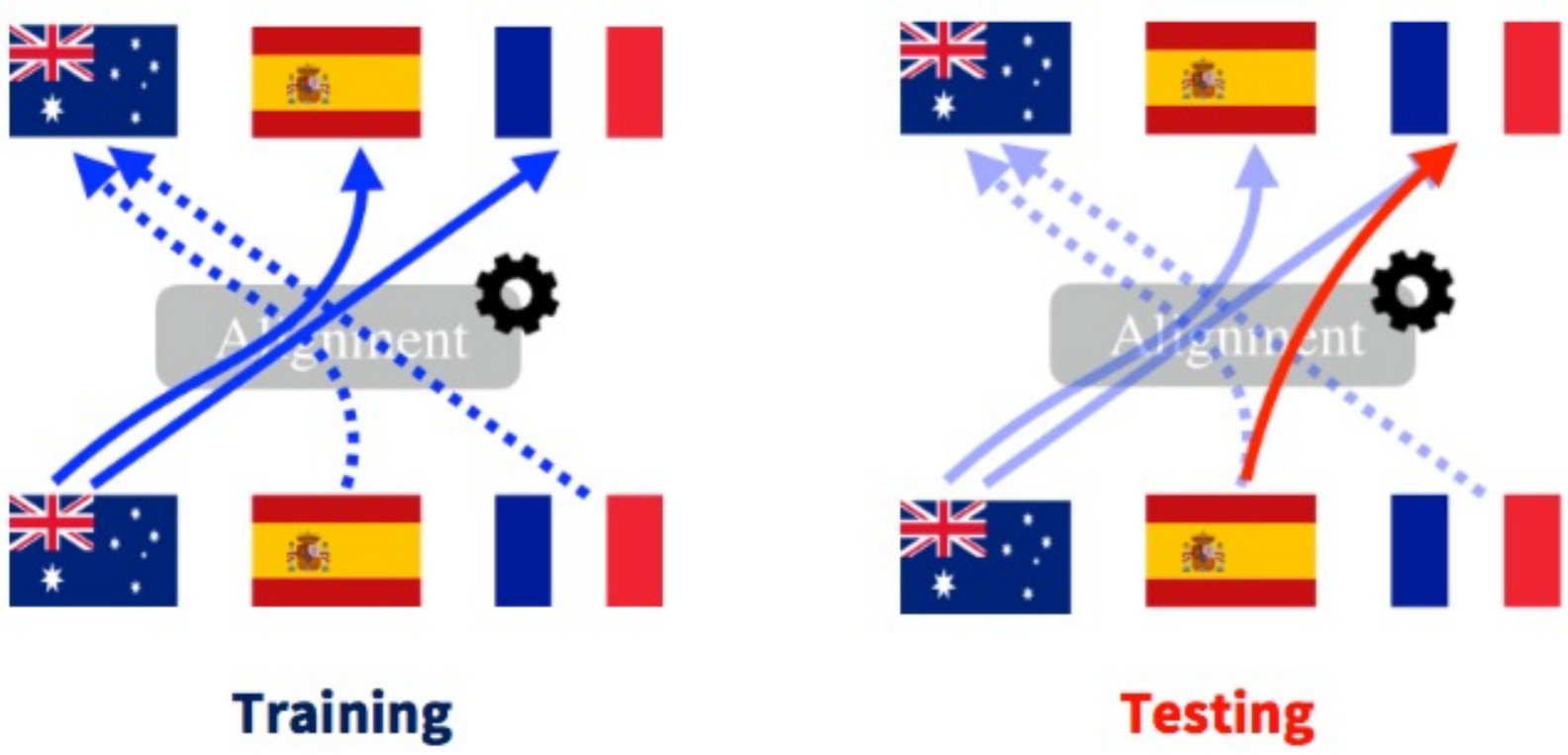
Multi-Source Neural Translation. Zoph and Knight 2016

Google NMT



Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. 2016

Sharing both encoder and decoder



Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation Johnson et al. 2017


a
b
c
ENGLISH
The stratosphere extends from about 10km to about 50km in altitude.
KOREAN
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.
JAPANESE
成層圏は、高度 10km から 50km の範囲にあります。

Sharing Attention

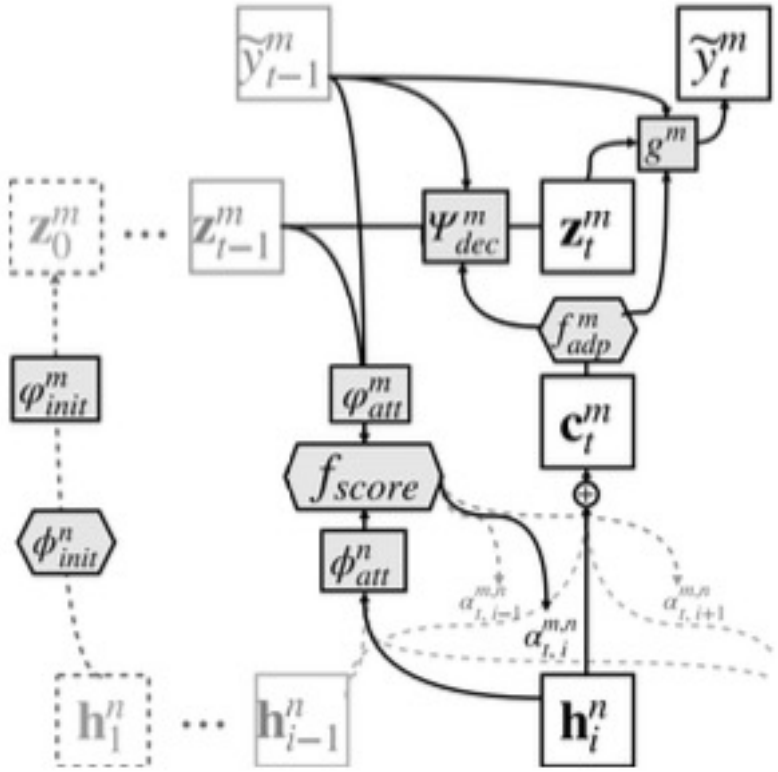


Figure 1: One step of the proposed multi-way. multilingual Neural Machine Translation model, for the n-th encoder and the m-th decoder at time step t. See Sec. 4 for details.

Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism
Firat et al. 2016

Universal Neural Machine Translation

- 共享不同语言之间的词表和翻译参数

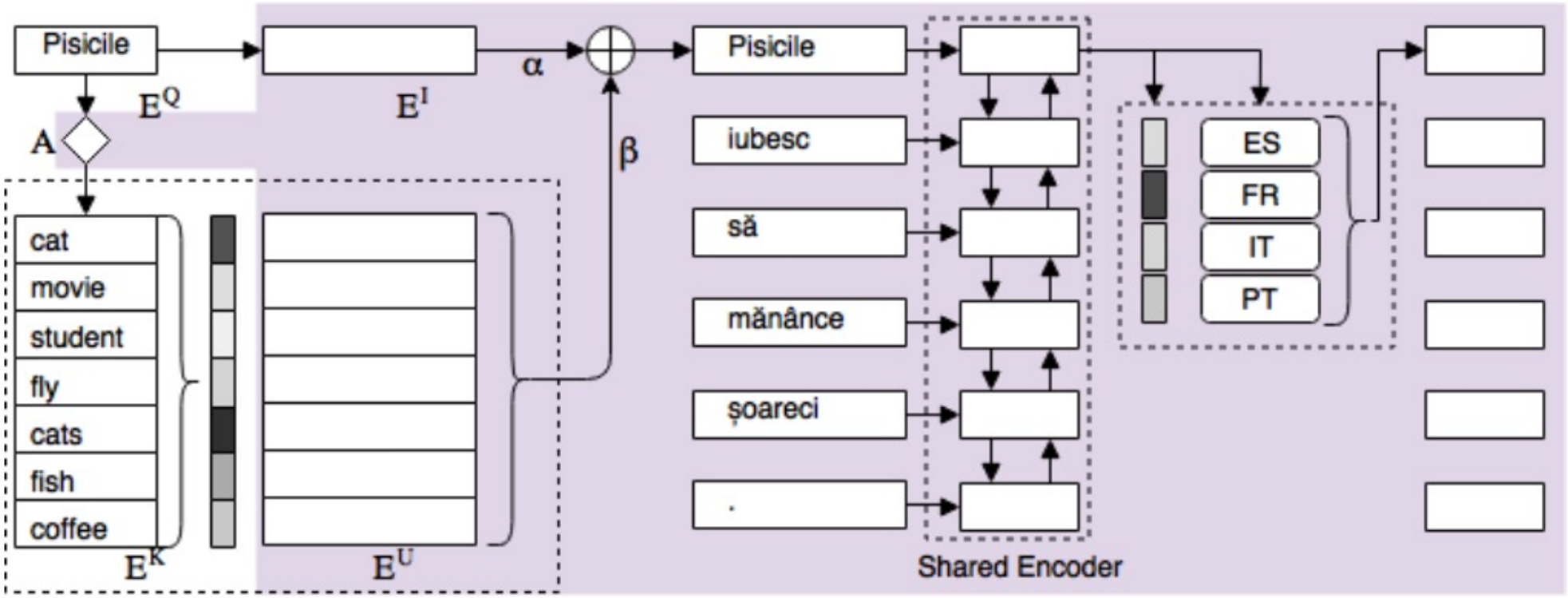


Figure 2: An illustration of the proposed architecture of the ULR and MoLE. Shaded parts are trained within NMT model while unshaded parts are not changed during training.

Universal Neural Machine Translation for Extremely Low Resource Languages. Gu et al. NAACL2018

- ACL 2022
  - Towards Making the Most of Cross-Lingual Transfer for Zero-Shot Neural Machine Translation
  - Alternative Input Signals Ease Transfer in Multilingual Machine Translation
  - EAG: Extract and Generate Multi-way Aligned Corpus for Complete Multilingual Neural Machine Translation
  - Multilingual Mix: Example Interpolation Improves Multilingual Neural Machine Translation

- Warren Weaver, “Translation”, 1949

Indeed, what seems to W.W. to be the most promising approach of all is one based on the ideas expressed in Section 2 above - that is to say, an approach that goes so deeply into the structure of languages as to come down to the level where they exhibit common traits.

—— Warren Weaver

Single Model for Multiple Tasks

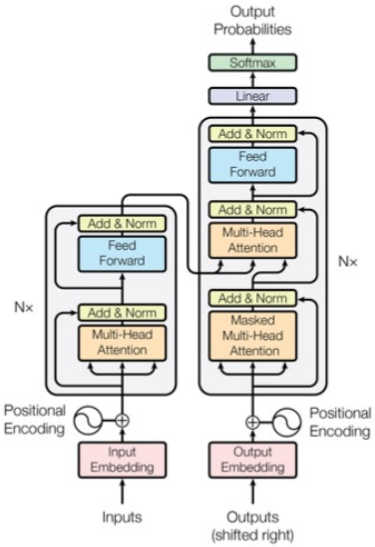


Figure 1: The Transformer - model architecture.

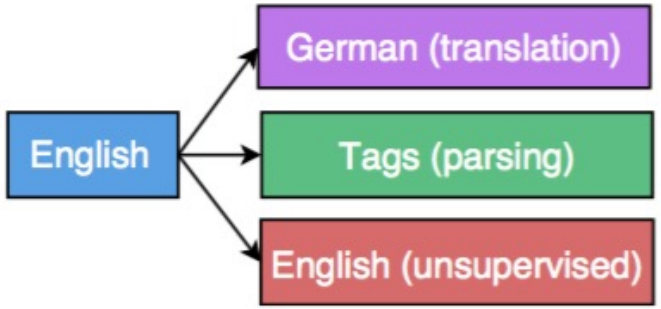
Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

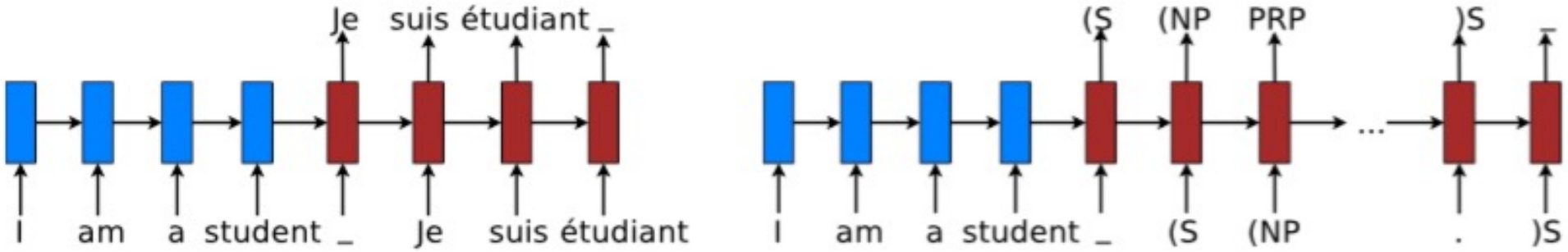
Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [17]	23.75			
Deep-Att + PosUnk [37]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [36]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [31]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [37]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [36]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	**41.29**	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	**$3.3 \cdot 10^{18}$**	
Transformer (big)	**28.4**	**41.0**	$2.3 \cdot 10^{19}$	

Table 4: The Transformer generalizes well to English constituency parsing (Results are on Section 23 of WSJ)

Parser	**Training**	**WSJ 23 F1**
Vinyals & Kaiser el al. (2014) [35]	WSJ only, discriminative	88.3
Petrov et al. (2006) [28]	WSJ only, discriminative	90.4
Zhu et al. (2013) [38]	WSJ only, discriminative	90.4
Dyer et al. (2016) [8]	WSJ only, discriminative	91.7
Transformer (4 layers)	WSJ only, discriminative	91.3
Zhu et al. (2013) [38]	semi-supervised	91.3
Huang & Harper (2009) [14]	semi-supervised	91.3
McClosky et al. (2006) [25]	semi-supervised	92.1
Vinyals & Kaiser el al. (2014) [35]	semi-supervised	92.1
Transformer (4 layers)	semi-supervised	92.7
Dyer et al. (2016) [8]	generative	93.3

Sharing between tasks





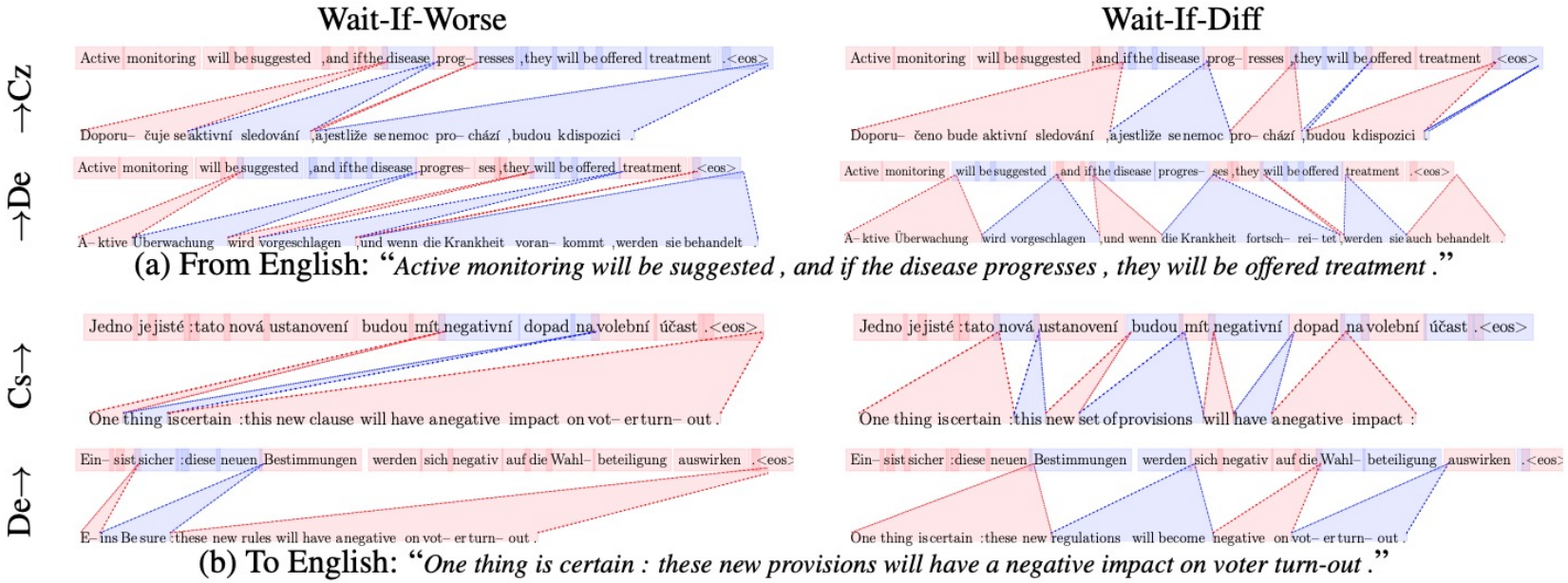
Learning to Parse and Translate Improves Neural Machine Translation. 2017

Simultaneous Translation , Speech-Text/Speech Translation

语音机器翻译

语音同传

- 在有限的时延约束下完成翻译
  - 何时可以进行翻译?
    - 设计解码算法，选择何时解码/等待



Can neural machine translation do simultaneous translation? Cho and Esipova 2016

语音同传

- 在有限的时延约束下完成翻译
  - 何时可以进行翻译?
    - 学习动作序列

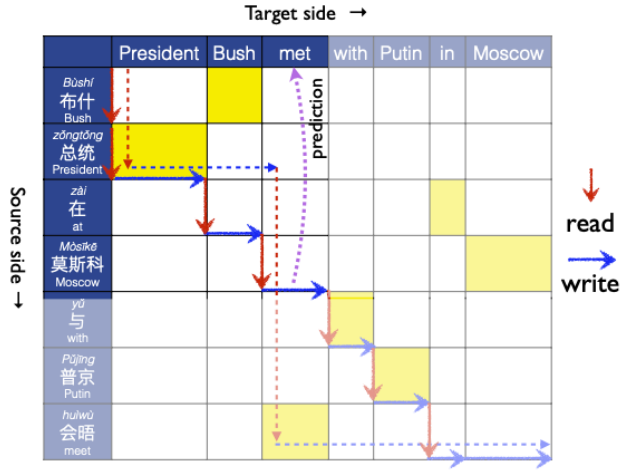
German	Ich	bin	mit	dem	Bus	nach	Ulm	gekommen
Gloss	I	am	with	the	bus	to	Ulm	come
Action	R W	R	R	R	W W W	R	R	W W W W
Translation	I				took the bus			to come to Ulm

Table 1: An example for READ/WRITE action sequence. R represents READ and W represents WRITE.

Simpler and Faster Learning of Adaptive Policies for Simultaneous Translation. Zheng et al., 2019

语音同传

- 在有限的时延约束下完成翻译
  - 学习低时延下的翻译



	Bùshí zǒngtǒng zài Mòsīkē *yǔ* *Pǔjīng* *huìwù* **布什 总统 在 莫斯科** 与 普京 会晤 **Bush president in Moscow** with/and Putin meet
(a) simultaneous: our wait-2	*...wait 2 words...* pres. bush **met** with putin in moscow
(b) non-simultaneous baseline	*..... wait whole sentence .....* pres. bush met with putin in moscow
(c) simultaneous: test-time wait-2	*...wait 2 words...* pres. bush in moscow and pol- ite meeting
(d) simultaneous: non-predictive	布什 总统 在 莫斯科 与 普京 会晤 *...wait 2 words...* pres. bush *..... wait 5 words .....* met with putin in moscow

STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework. Ma et al., 2019

端到端语音翻译

- 将语音、文本等模态信息联系起来

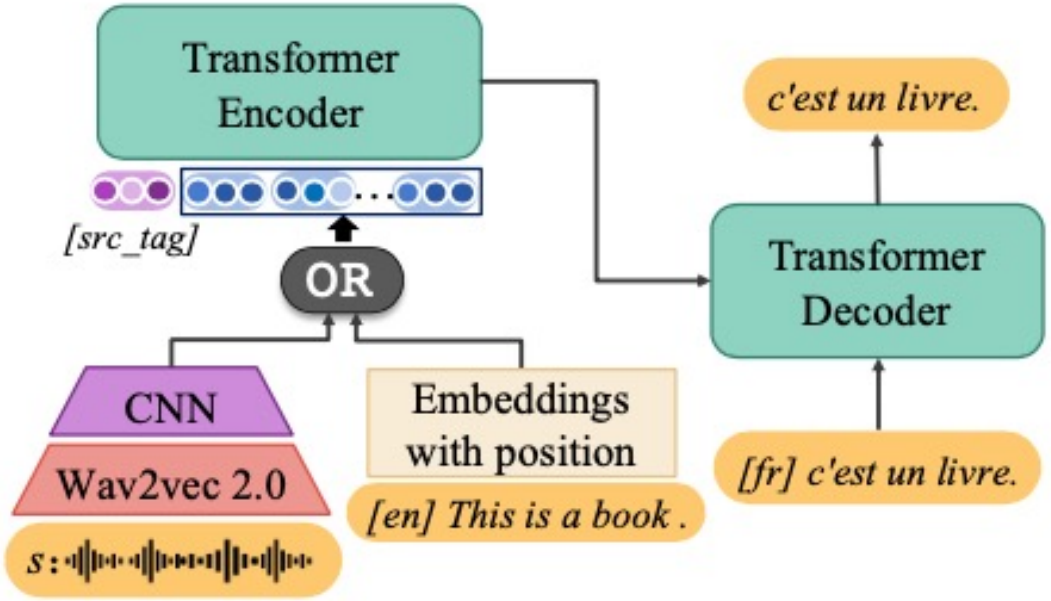


Figure 1: *The model structure of XSTNet. It accepts either audio or text input. For the audio input, we deploy wav2vec2.0 followed by two convolution layers to get the audio embedding. The Transformer encoder and decoder are shared for both modalities.*

End-to-end Speech Translation via Cross-modal Progressive Training. Ye et al, 2021

- ACL 22
  - Learning Adaptive Segmentation Policy for End-to-End Simultaneous Translation
  - From Simultaneous to Streaming Machine Translation by Leveraging Streaming History
  - Learning When to Translate for Streaming Speech
  - Modeling Dual Read/Write Paths for Simultaneous Machine Translation
  - Reducing Position Bias in Simultaneous Machine Translation with Length Aware Framework
  - Scheduled Multi-task Learning for Neural Chat Translation
  - Unified Speech-Text Pre-training for Speech Translation and Recognition

多模态翻译

- 利用图片中包含的信息帮助翻译结果生成
- ACL17
  - Doubly-Attentive Decoder for Multi-modal Neural Machine Translation
- EMNLP17
  - Incorporating Global Visual Features into Attention-based Neural Machine Translation
  - An empirical study of the effectiveness of images on Multi-modal Neural Machine Translation
- ACL18
  - Learning Translations via Images: A Large Multilingual Dataset and Comprehensive Study
- ……

src.	eine Gruppe junger Menschen trinkt Shots in einem Mexikanischen Setting .
ref.	a group of young people take shots in a Mexican setting .
NMT	a group of young people are **having fun** in an auditorium .
PBSMT	a group of young people drinking at a Shots Mexikanischen Setting .
IMG_{2W}	a group of young people having drinks in a **Mexican restaurant** .
IMG_{E}	a group of young people drinking apples in a **Mexican restaurant** .
IMG_{D}	a group of young people drinking food in a **Mexican restaurant** .
IMG_{2W+D}	a group of young people **having fun** in a Mexican room .
IMG_{E+D}	a group of young people drinking dishes in a **Mexican restaurant** .

die grauen mauern und grünen terrassen einer ruine auf einem berg , mit einem sehr markanten berg dahinter und einer bergkette im hintergrund .

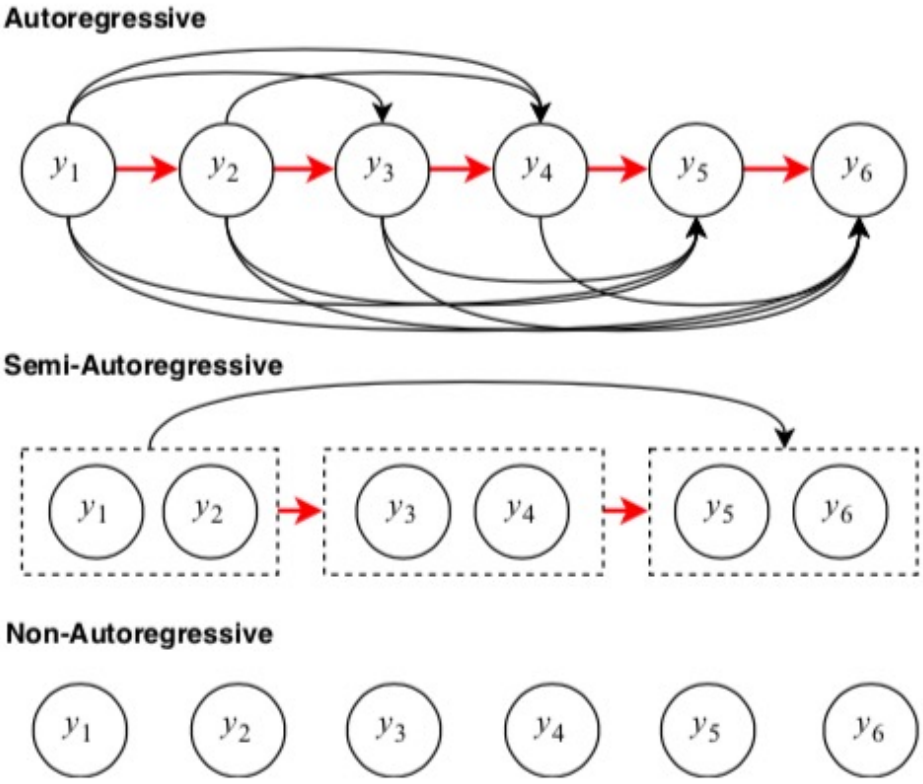
PRE.: a ruin with grey walls and green terraces in the foreground .
JOINT: the grey walls and green terraces of ruins on top of a mountain , with a very distinctive mountain behind them and a wooded mountain range in the background .

Incorporating Global Visual Features into Attention-Based Neural Machine Translation. Calixto and Liu 2017
Zero-Resource Neural Machine Translation with Multi-Agent Communication Game Chen et al. 2018

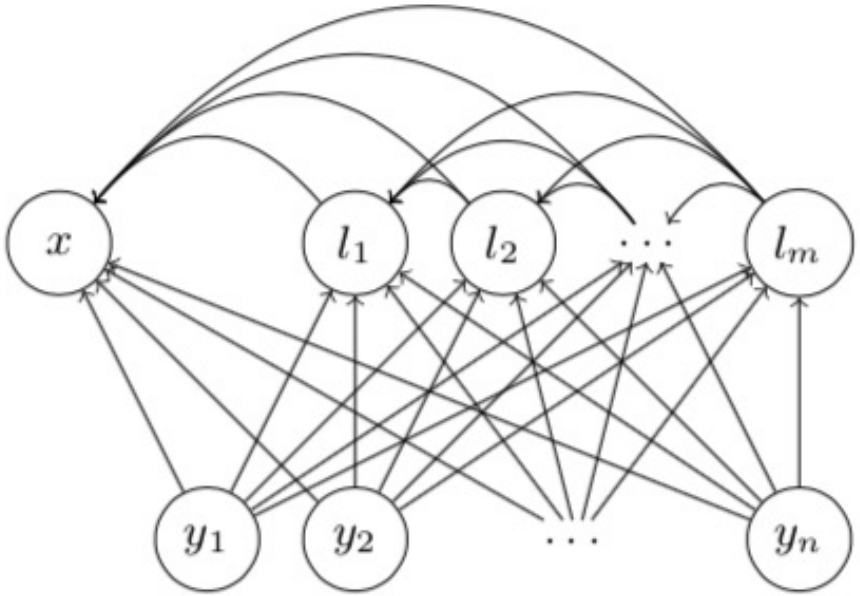
网络结构变化

- 序列处理v.s. 并行处理
- ACL17
  - A Convolutional Encoder Model for Neural Machine Translation (Convolutional Sequence to Sequence Learning)
- NIPS17
  - Attention Is All You Need.
- ICLR2018
  - Non-Autoregressive Neural Machine Translation
- ACL2018
  - The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation.
- EMNLP2018
  - Semi-Autoregressive Neural Machine Translation.

- 必须从左到右生成一个句子(单词生成依赖前序)
  - 并行化程度低



Semi-Autoregressive Neural Machine Translation. Wang et al. 2018



Fast Decoding in Sequence Models Using Discrete Latent Variables. Kaiser et al. 2018

鲁棒的翻译系统学习

- ACL2018
  - Towards Robust Neural Machine Translation
- AAAI2018
  - Neural Machine Translation with Gumbel-Greedy Decoding
- NAACL2018
  - Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets
- COLING2018
  - On Adversarial Examples for Character-Level Neural Machine Translation

机器翻译容易受到输入噪音的影响

Input	tamen *bupa* kunnan zuochu weiqi AI.
Output	They are not afraid of difficulties to make Go AI.

他们不怕困难做出围棋AI

Input	tamen *buwei* kunnan zuochu weiqi AI.
Output	They are not afraid to make Go AI.

他们不畏困难做出围棋AI

中国电子银行业务管理新规将于三月一日起实行

zhongguo dianzi yinhang yewu guanli xingui jiangyu sanyue yiri qi shixing
china's electronic bank rules to be implemented on march 1

中方电子银行业务管理新规将于三月一日起实行

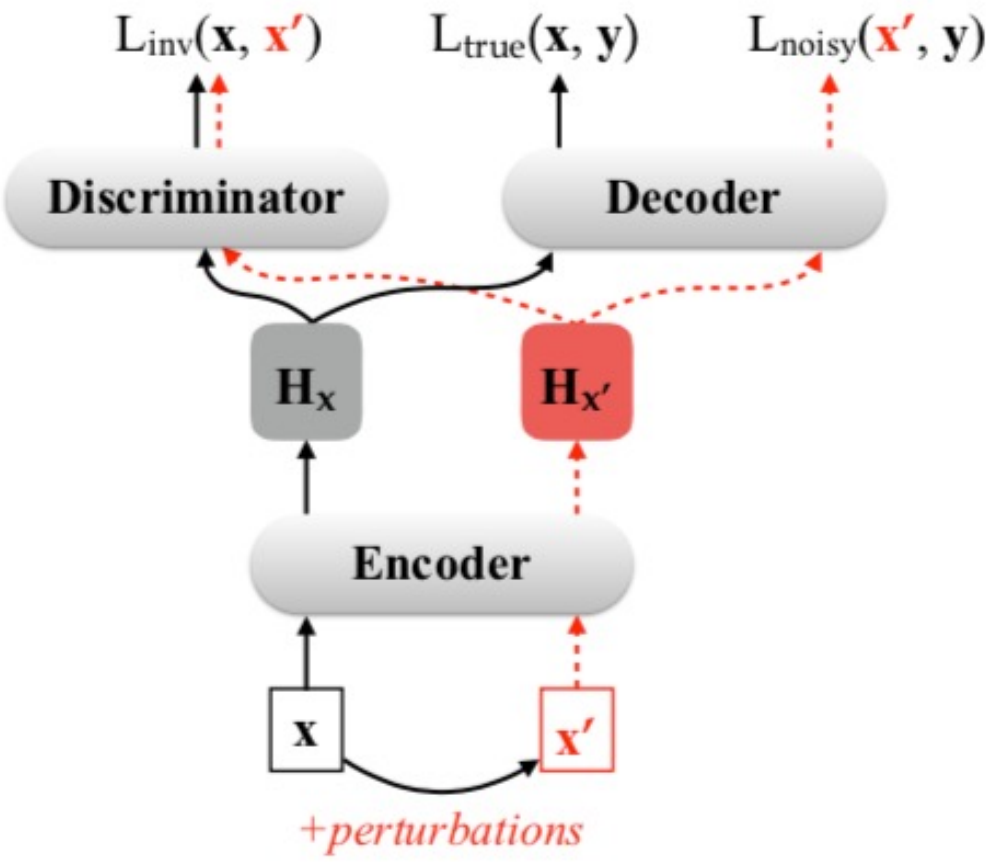
zhongfang dianzi yinhang yewu guanli xingui jiangyu sanyue yiri qi shixing
china to implement new regulations on business management

Towards Robust Neural Machine Translation. Cheng et al. 2018

1901 wurde eine Frau namens Auguste in eine medizinische Anstalt in Frankfurt gebracht.
1901 wurde eine Frau namens **Afuiguste** in eine medizinische Anstalt in Frankfurt gebracht.
In 1931, a woman named **Augustine** was brought into a medical institution in France.
In 1931, a woman named Rutgers was brought into a medical institution in France.
Das ist Dr. Bob Childs – er ist Geigenbauer und Psychotherapeut.
Das ist Dr. Bob Childs – er ist Geigenbauer und **Psy6hothearpeiut**.
This is Dr. Bob Childs – he's a wizard maker and a **therapist**'s **therapist**.
This is Dr. Bob Childs – he's a brick maker and a **psychopath**.

On Adversarial Examples for Character-Level Neural Machine Translation.
Ebrahimi et al. 2018

- 人工制造噪音，加强模型抗干扰能力



Towards Robust Neural Machine Translation. Cheng et al. 2018

神经翻译系统的训练技术（ACL22）

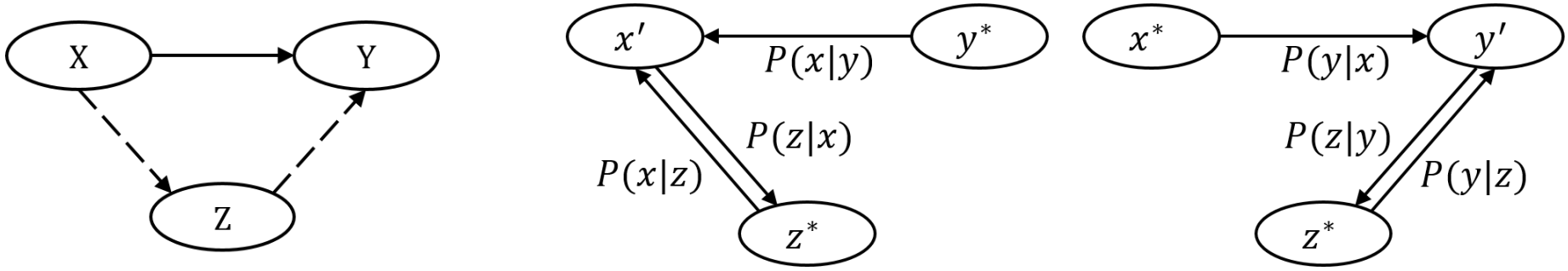
- Data Augmentation
  - CipherDAug: Ciphertext based Data Augmentation for Neural Machine Translation
  - Learning to Generalize to More: Continuous Semantic Augmentation for Neural Machine Translation
  - Prediction Difference Regularization against Perturbation for Neural Machine Translation
- Confidence
  - Confidence Based Bidirectional Global Context Aware Training Framework for Neural Machine Translation
  - Learning Confidence for Transformer-based Neural Machine Translation
  - Overcoming Catastrophic Forgetting beyond Continual Learning: Balanced Training for Neural Machine Translation

无监督/低资源翻译

- 尝试在少量甚至无平行数据的情况下学习机器翻译
- ACL17
  - Data Augmentation for Low-Resource Neural Machine Translation (short paper)
- ICLR18
  - Word Translation Without Parallel Data
  - Unsupervised Machine Translation Using Monolingual Corpora Only
  - Unsupervised Neural Machine Translation
- ACL18
  - Unsupervised Neural Machine Translation with Weight Sharing
  - Adaptive Knowledge Sharing in Multi-Task Learning: Improving Low-Resource Neural Machine Translation (short paper)
- ……

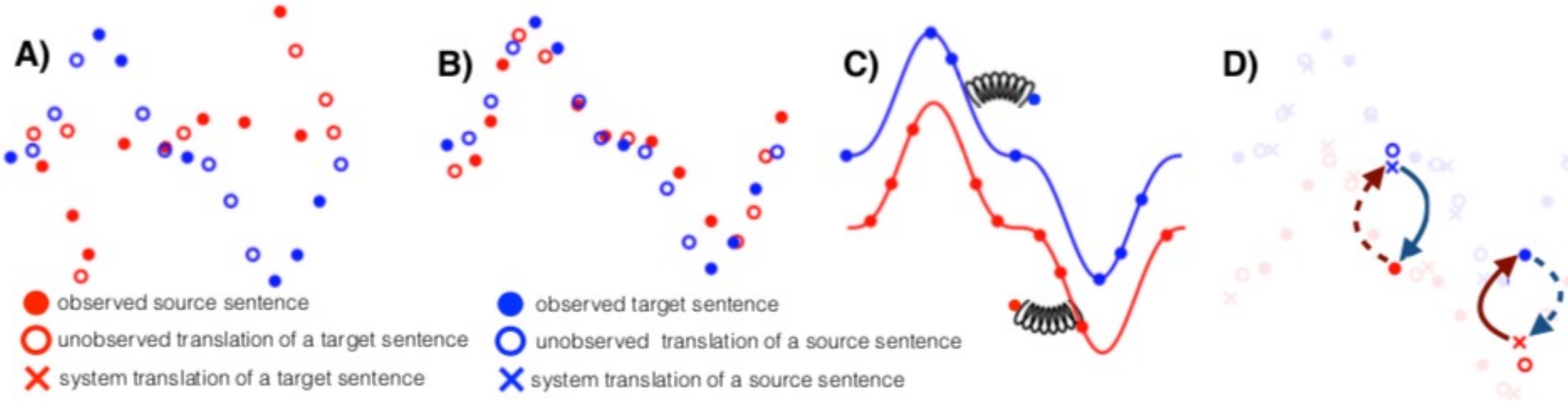
低资源翻译

- 利用中间语言



Triangular Architecture for Rare Language Translation. Ren et al. 2018

无监督翻译学习过程



- Monolingual
- Initialization
- Language Modeling
- Back Translation

- ACL22:
  - Bridging the Data Gap between Training and Inference for Unsupervised Neural Machine Translation
  - Flow-Adapter Architecture for Unsupervised Machine Translation

- ChatGPT related thoughts
  - Is ChatGPT A Good Translator? A Preliminary Study Jiao et al. 2023
  - Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis. Zhu et al. 2023

语言现象研究

- Compositionality
  - Can Transformer be Too Compositional? Analysing Idiom Processing in Neural Machine Translation
  - The Paradox of the Compositionality of Natural Language: A Neural Machine Translation Case Study

- Word Translation
  - DEEP: DEnoising Entity Pre-training for Neural Machine Translation
  - Improving Word Translation via Two-Stage Contrastive Learning
  - DiBiMT: A Novel Benchmark for Measuring Word Sense Disambiguation Biases in Machine Translation

- Bias:
  - Investigating Failures of Automatic Translation in the Case of Unambiguous Gender
  - Measuring and Mitigating Name Biases in Neural Machine Translation