中国中文信息学会《前沿技术讲习班》第二十五期（CIPS ATT25）

机器同传

何中军

2021-10-8

提纲

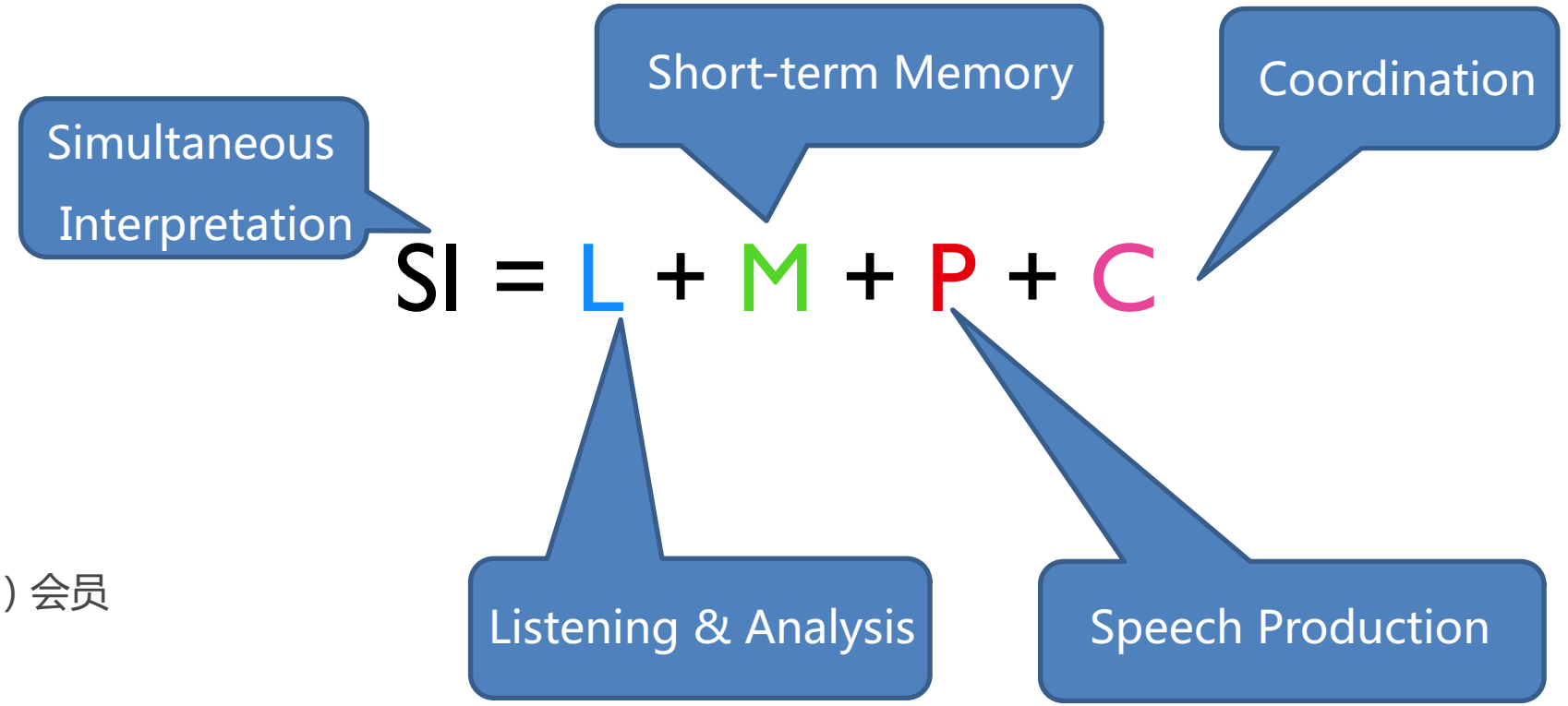
- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 数据、鲁棒性模型及评价方法
- 产品及应用
- 总结及展望

同声传译

- 极具挑战性的翻译方式

吉尔精力分配模式（Gile’s Effort Model)

欧洲翻译研究学会前主席
国际会议口译员协会（AIIC）会员



GILE, D. “Basic Theoretical Components in Interpreter and Translator Training.” DOLLERUP, C. and LODDEGAARD, A. (eds). *Teaching Translation and Interpreting: Training, Talent and Experience*. Amsterdam: John Benjamins, 1992.

同声传译

准确度及完整性：连续工作30分钟后，每过5分钟，准确率下降10%

工作方式：一般2人一组，轮流工作和休息

AIIC会员：全球约3000人

Vidal M (1997) New study on fatigue confirms need for working in teams. Proteus Newsl NAJIT 6.1

同传常用技巧

skipping, approximation, filtering, comprehension omission, and substitution

Al-Khanji R, El-Shiyab S, Hussein R (2000) On the use of compensatory strategies in simultaneous interpretation.

同传常用技巧

skipping, approximation, filtering, comprehension omission, and substitution

然后 这个 粉色 的 方框 是 算法 自动 预测 出来 的 结果 。

Translation： While the pink boxes were predicted automatically by an algorithm.

SI： And the pink one is through algorithms.

Al-Khanji R, El-Shiyab S, Hussein R (2000) On the use of compensatory strategies in simultaneous interpretation.

同传常用技巧

skipping, approximation, filtering, comprehension omission, and substitution

我们 在 2015 年 就 已经 取得 了 世界 第一 的 水平 ，错误 率 只有 0.23%。

Translation： We ranked No.1 in the world in 2015, with an error rate of only 0.23%.

SI： We rank the first in the world. And the error rate is really low

Al-Khanji R, El-Shiyab S, Hussein R (2000) On the use of compensatory strategies in simultaneous interpretation.

同传常用技巧

顺句驱动、适当重复

We are most pleased that we share identical views on a wide range of issues.

Translation：我们在很多问题上观点一致，非常高兴

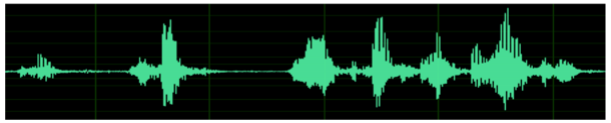
SI：我们非常高兴，我们的观点一致，在很多问题上观点一致

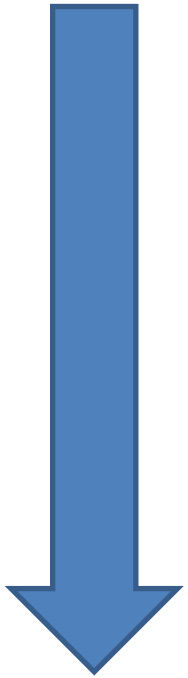
提纲

- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 数据、鲁棒性模型及评价方法
- 产品及应用
- 总结及展望

机器同传

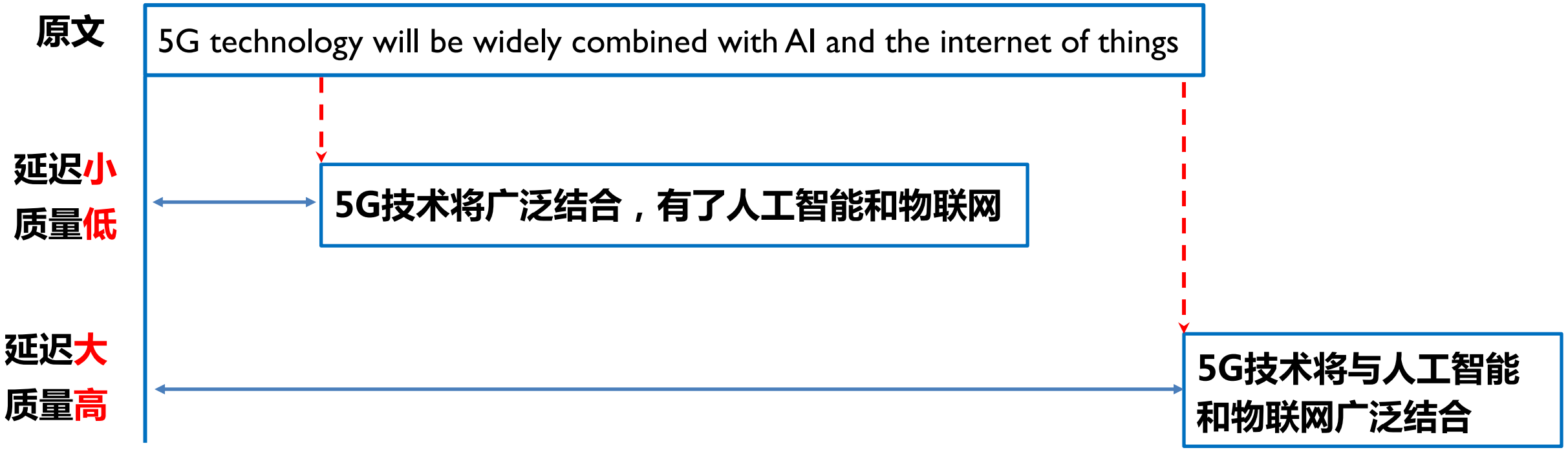
源语言



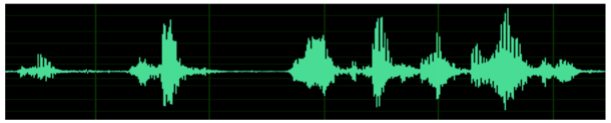


目标语言

挑战一：翻译质量与时延难以兼顾



挑战二：无断句标点

源语言  

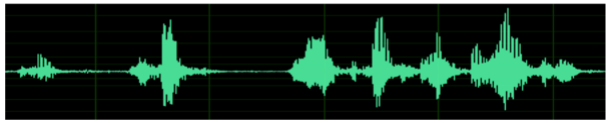
那么大家知道这个重庄家最怕的是出现病虫害一旦就是一个病虫害出现防治不及时的话会造成大量的这个减产

从哪里开始？到哪里结束？

目标语言  

挑战二：语音识别错误

源语言

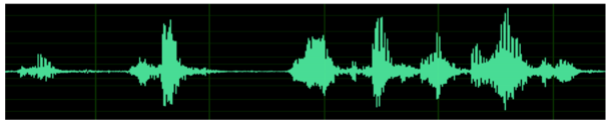


那么大家知道这个重庄家最怕的是出现病虫害一旦就是一个病虫害出现防治不及时的话会造成大量的这个减产

错误传递和放大：差之毫厘，谬以千里

目标语言

挑战三：口语化成分（冗余、停顿等）

源语言  

那么大家知道这个重庄家最怕的是出现病虫害一旦就是一个病虫害出现防治不及时的话会造成大量的这个减产

影响质量，增加时延

目标语言  

挑战四：数据稀缺

MT

Bilingual

Source Text — Target Text

我们齐心搭建世界联通新桥梁

Together, we have built new bridges for global connectivity.

ASR

Monolingual

Audio Signal — Transcription

那么我们今天呢就希望，从一个二十年的AI工作者来说，如何从专业的角度去解读一下 …

SimulTrans

Bilingual

Source Audio — Target Text /Audio

So today, as one who has been working on AI for twenty years, I wish I could give you a professional interpretation …

挑战五：无统一评价标准

然后 这个 粉色 的 方框 是 算法 自动 预测 出来 的 结果 。

Translation： While the pink boxes were predicted automatically by an algorithm.

SI： And the pink one is through algorithms.

评价文本翻译的指标（关注翻译的准确、语法等），

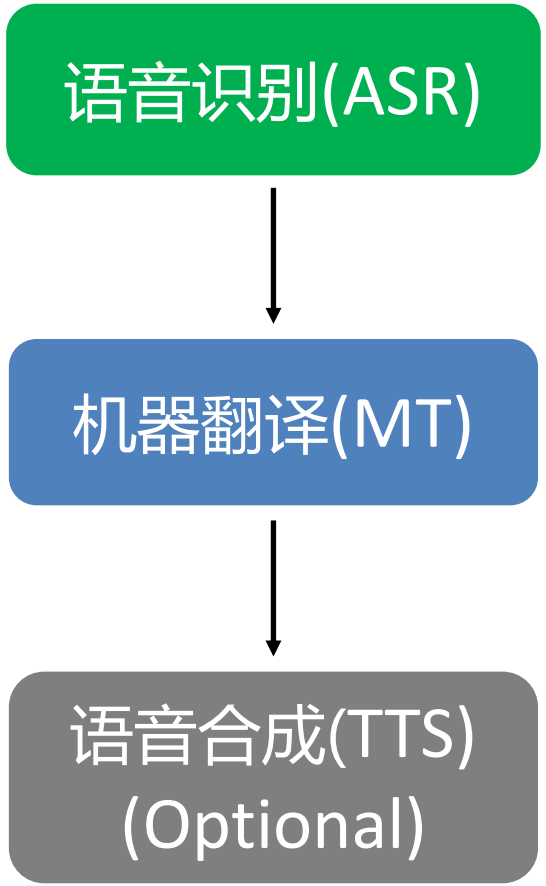
不适合评价同传（关注信息传递的质量和效率）

提纲

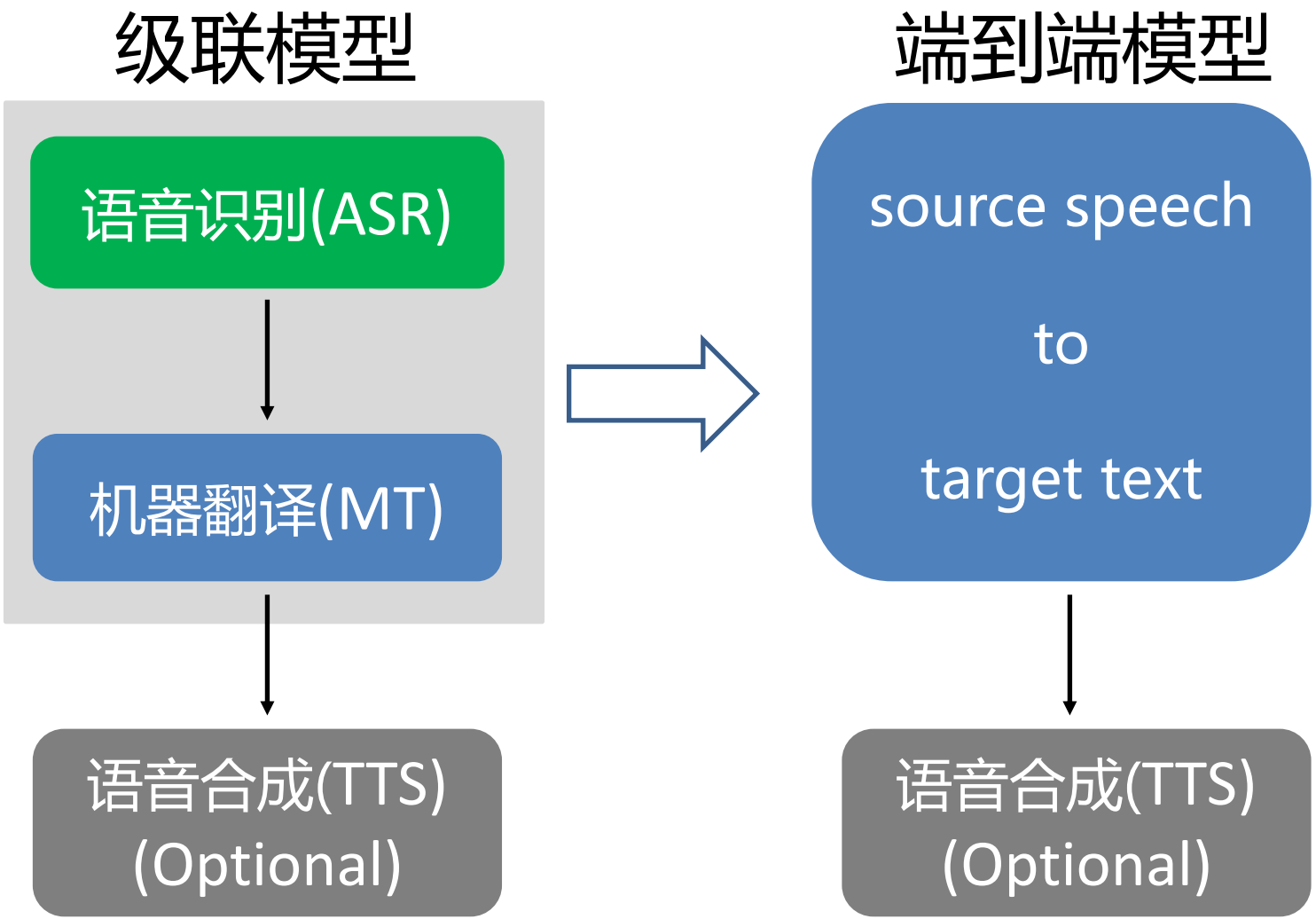
- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 数据、鲁棒性模型及评价方法
- 产品及应用
- 总结及展望

主流方法

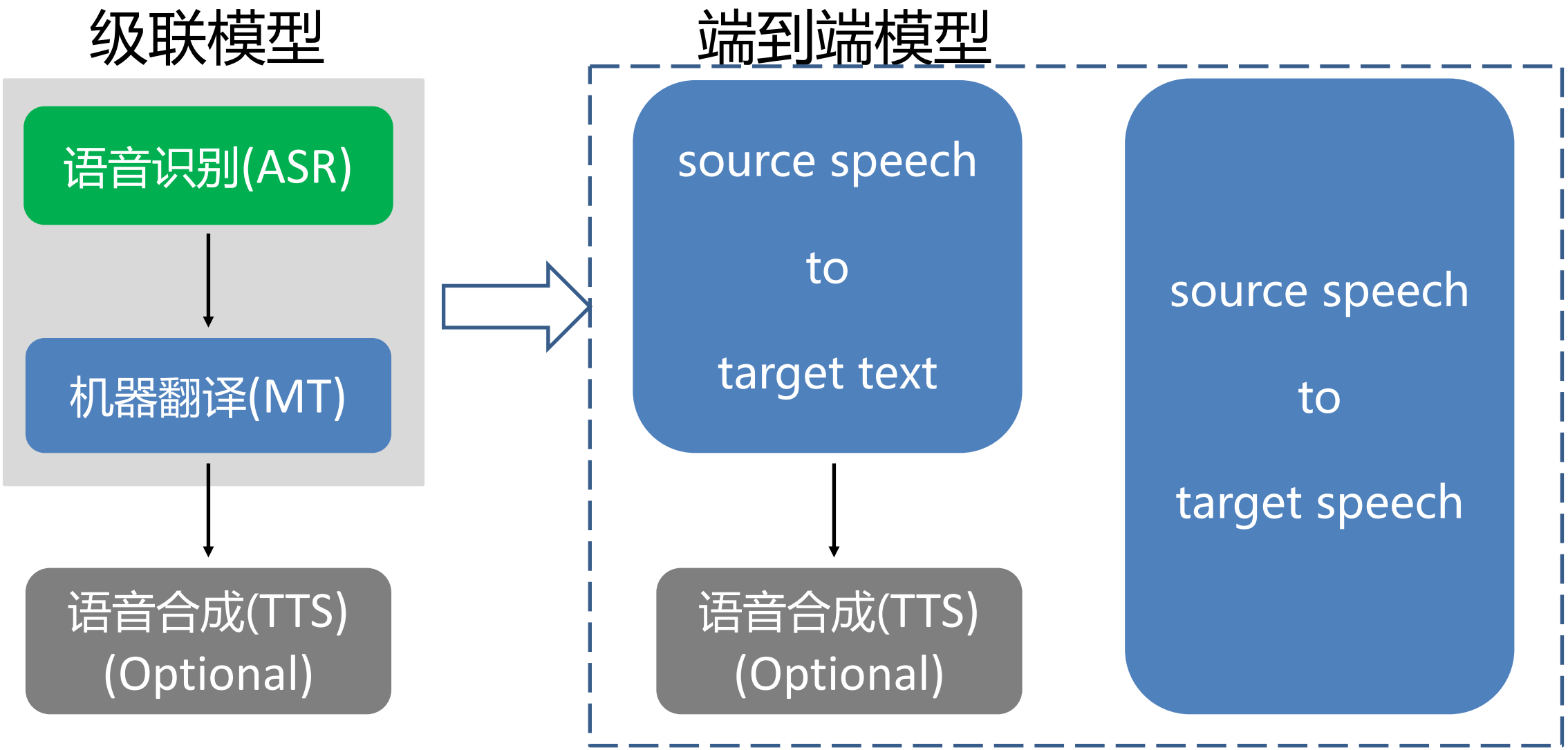
级联模型



主流方法


级联模型
语音识别(ASR)
机器翻译(MT)
语音合成(TTS)
(Optional)
端到端模型
source speech
to
target text
语音合成(TTS)
(Optional)

主流方法


级联模型
语音识别(ASR)
机器翻译(MT)
语音合成(TTS)
(Optional)
端到端模型
source speech
to
target text
语音合成(TTS)
(Optional)
source speech
to
target speech

级联模型

ASR

MT

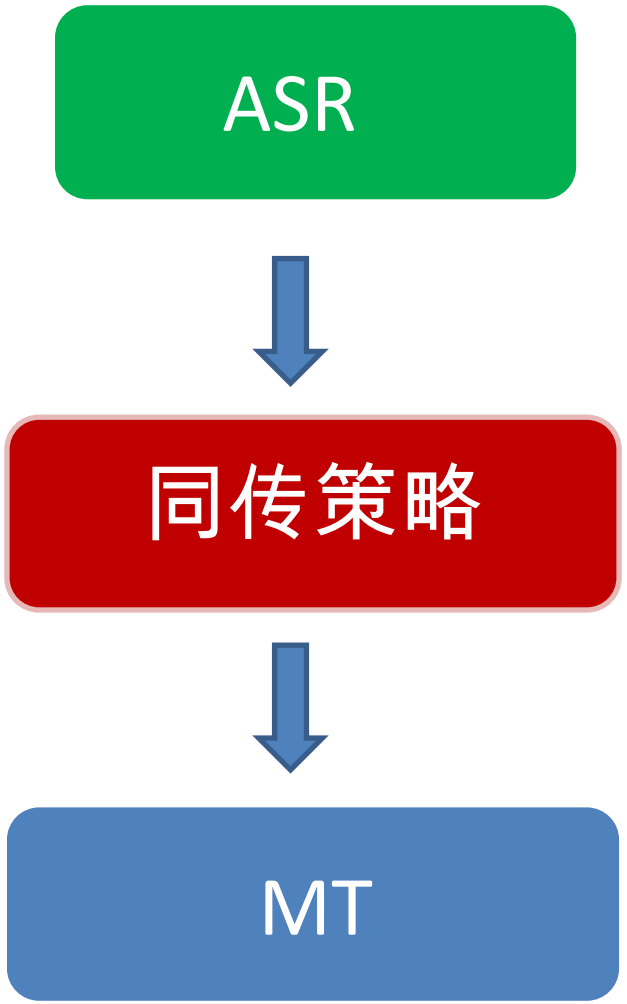
那么大家知道这个种庄稼最怕的是出现病虫害一旦就是一个病虫害出现防治不及时的话会造成大量的这个减产

ASR: 无断句标点

MT: 以句子为单位进行翻译

As we all know, The biggest fear of planting crops is the emergence of pests and diseases. Once there is a disease and pest, if the prevention and control is not timely, it will cause a lot of production reduction.

级联模型



那么大家知道这个种庄稼最怕的是出现病虫害一旦就是一个病虫害出现防治不及时的话会造成大量的这个减产

读入（READ）：读入流文本

翻译（WRITE）：翻译已经读入的文本

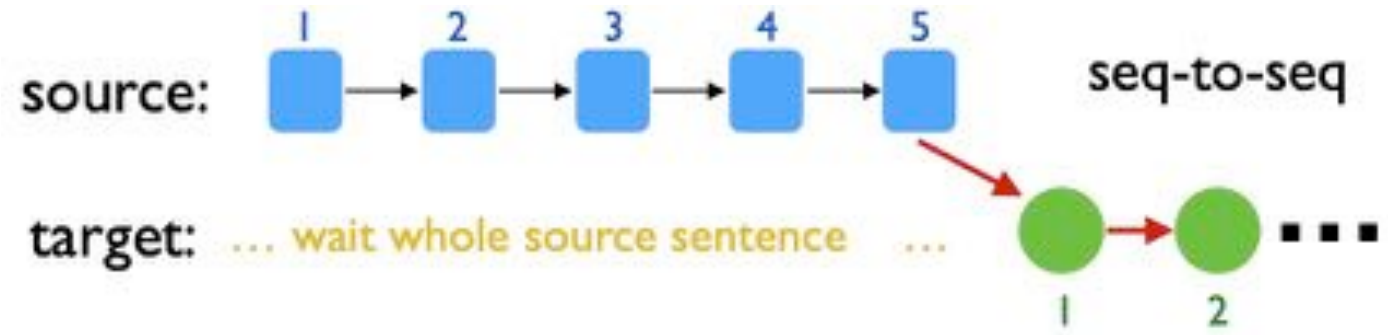
As we all know, The biggest fear of planting crops is the emergence of pests and diseases. Once there is a disease and pest, if the prevention and control is not timely, it will cause a lot of production reduction.

级联模型 – 同传策略

- 固定策略（Fixed Policy）：读入长度（翻译单元）固定
  - Static read-write (Dalvi et al., 2018)
  - STACL (Ma et al., 2018), etc.

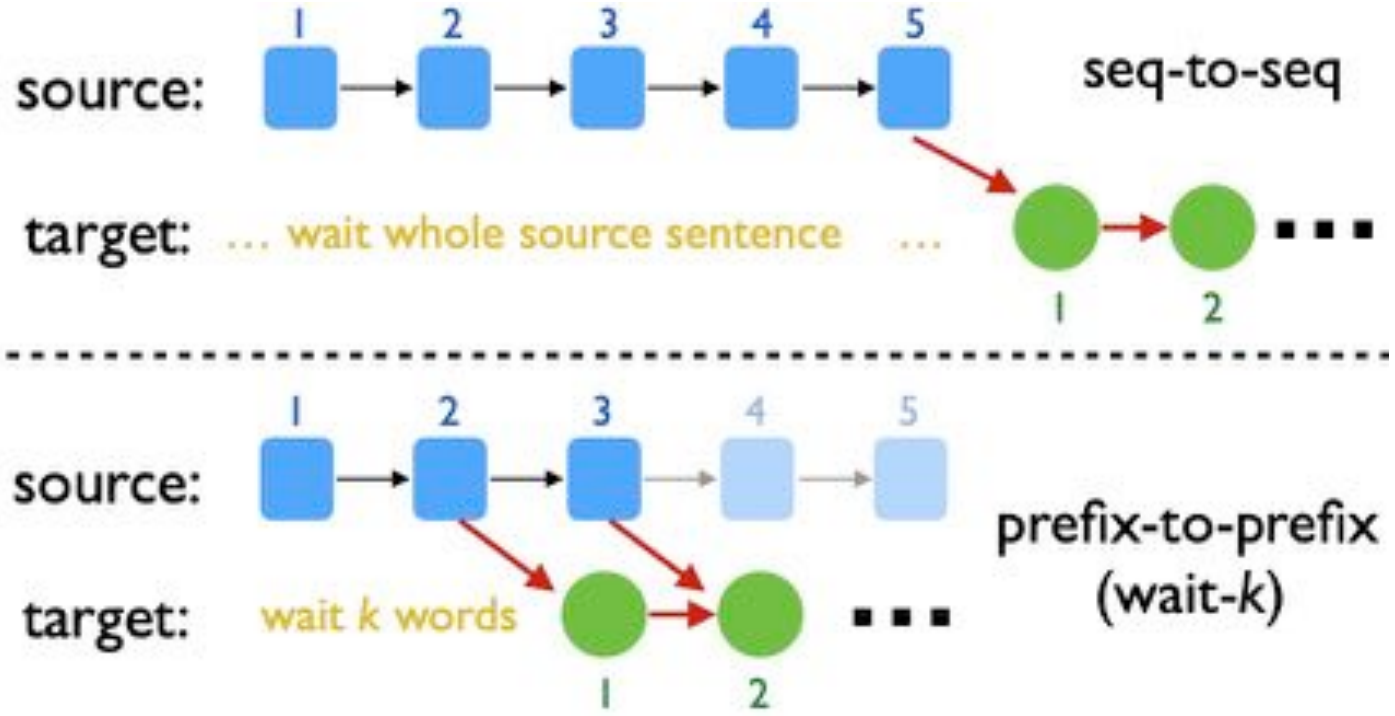
- 自适应策略（Adaptive Policy）：动态调整翻译单元长度
  - Rule-based （Cho et al., 2016)
  - RL-based (Gu et al., 2017）
  - Supervised policy (Zheng et al., 2019)
  - Meaningful unit (Zhang et al., 2020)
  - MILk (Arivazhagan et al., 2019)
  - Multihead monotonic attention (Ma et al., 2020), etc.

wait-*k*策略


source:
1
2
3
4
5
seq-to-seq
target:
… wait whole source sentence …
1
2

wait-*k*策略

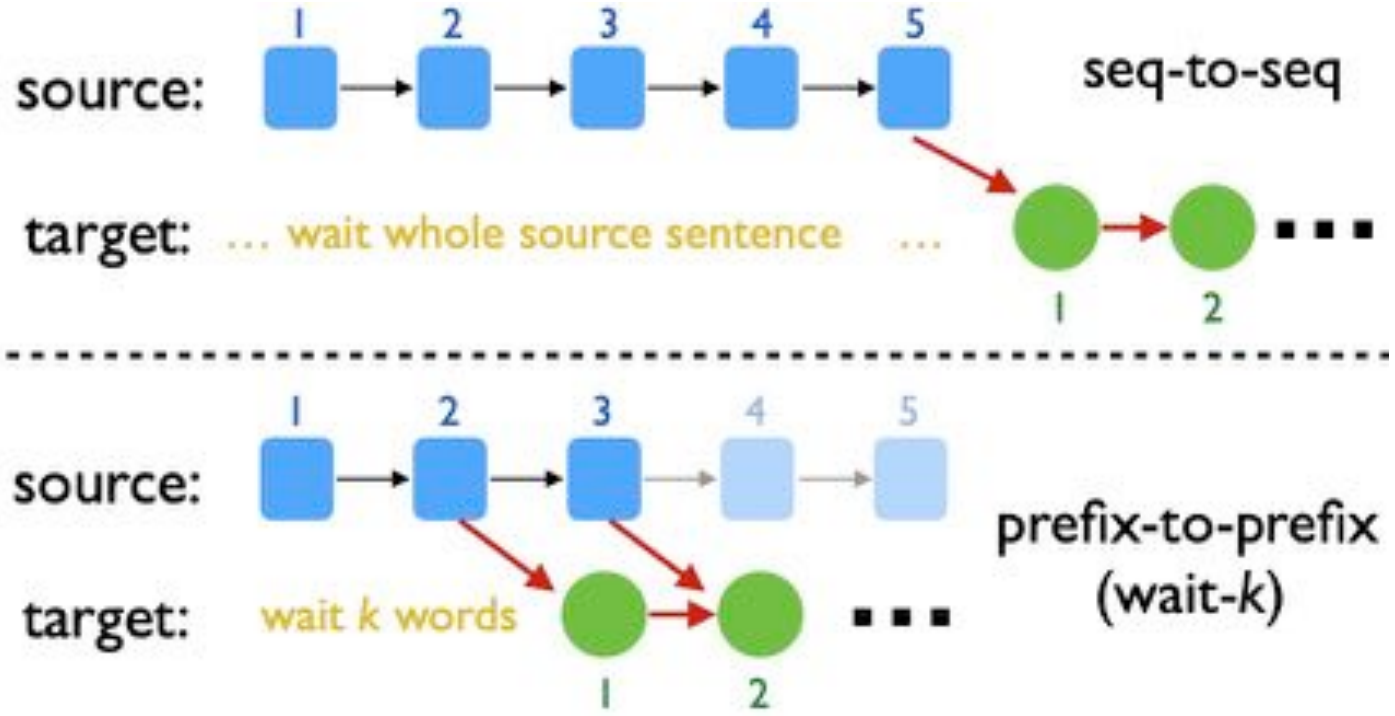
先读入*k*个词，然后开始翻译



Mingbo Ma, Liang Huang, Hao Xiong, Kaibo Liu, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, and Haifeng Wang. 2019. STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pages 3025—3036. July, 2019, Florence, Italy

wait-*k*策略

先读入*k*个词，然后开始翻译



$p(y_t|\{x_1,\dots,\boldsymbol{x_n}\},\{y_1,\dots,y_{t-1}\})$

$p(y_t|\{x_1,\dots,\boldsymbol{x_{t+k-1}}\},\{y_1,\dots,y_{t-1}\})$

$g(t)$:*number of source words used to predict* y_t

$p(y_t|\{x_1,\dots,x_{g(t)}\},\{y_1,\dots,y_{t-1}\})$

wait-*k*策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October

wait-*k*策略

两国　领导人　将　在　　十月　举行　会谈

wait-2

The leaders of　the　two countries will hold talks in October

两国	R										
领导人	R										
将											
在											
十月											
举行											
会谈											

wait-k策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October

	The										
两国	R										
领导人	R	W									
将											
在											
十月											
举行											
会谈											

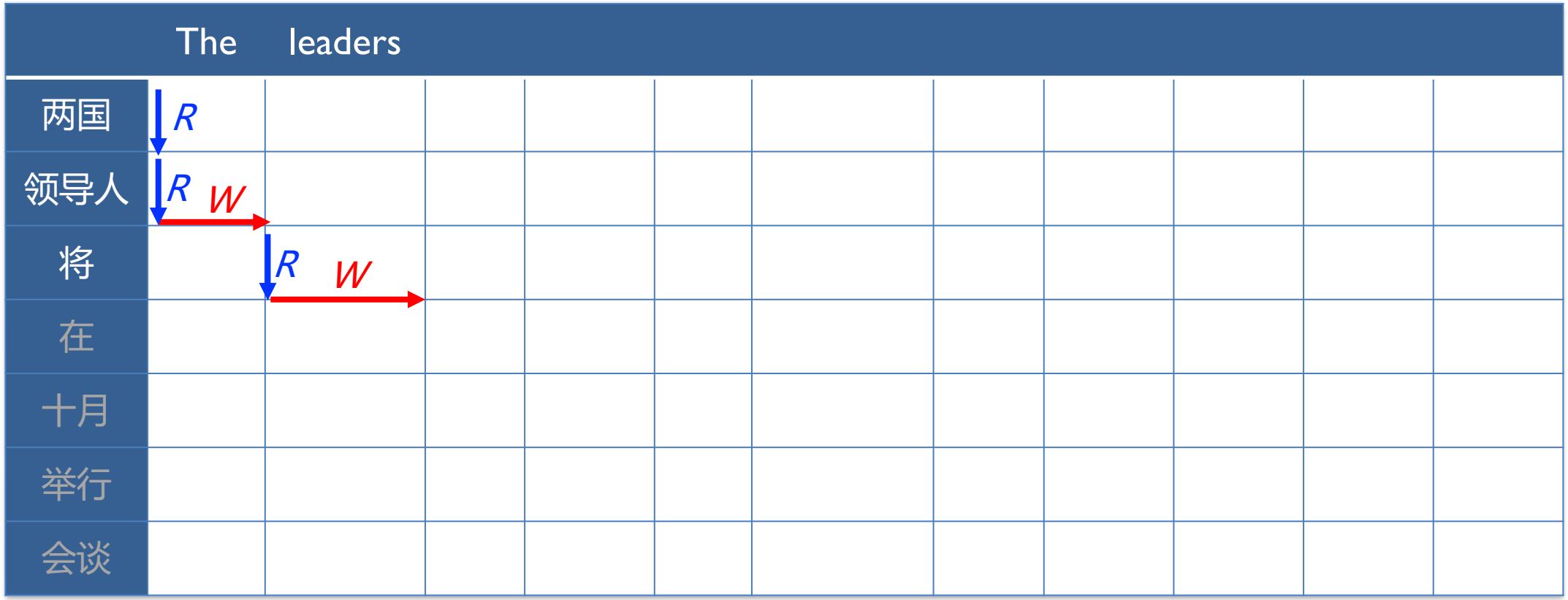
$g(1) = 2$

wait-k策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October

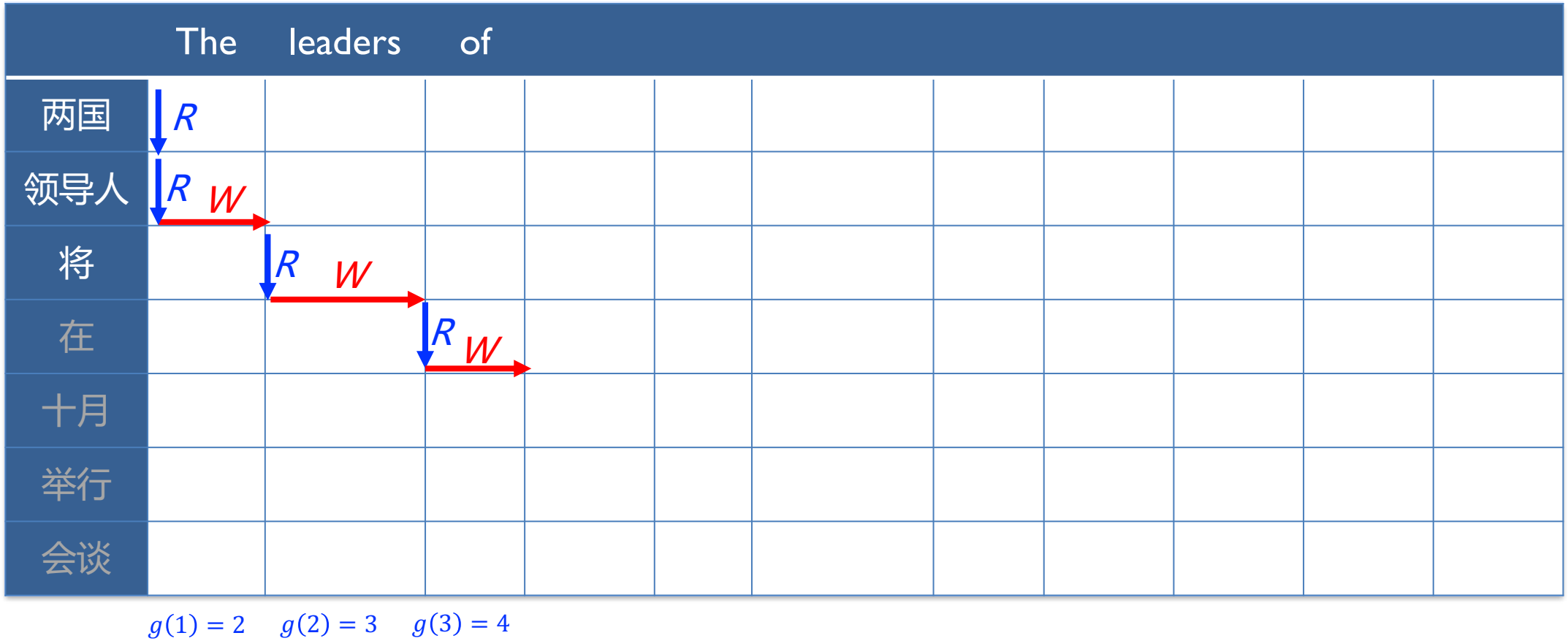


wait-k策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October

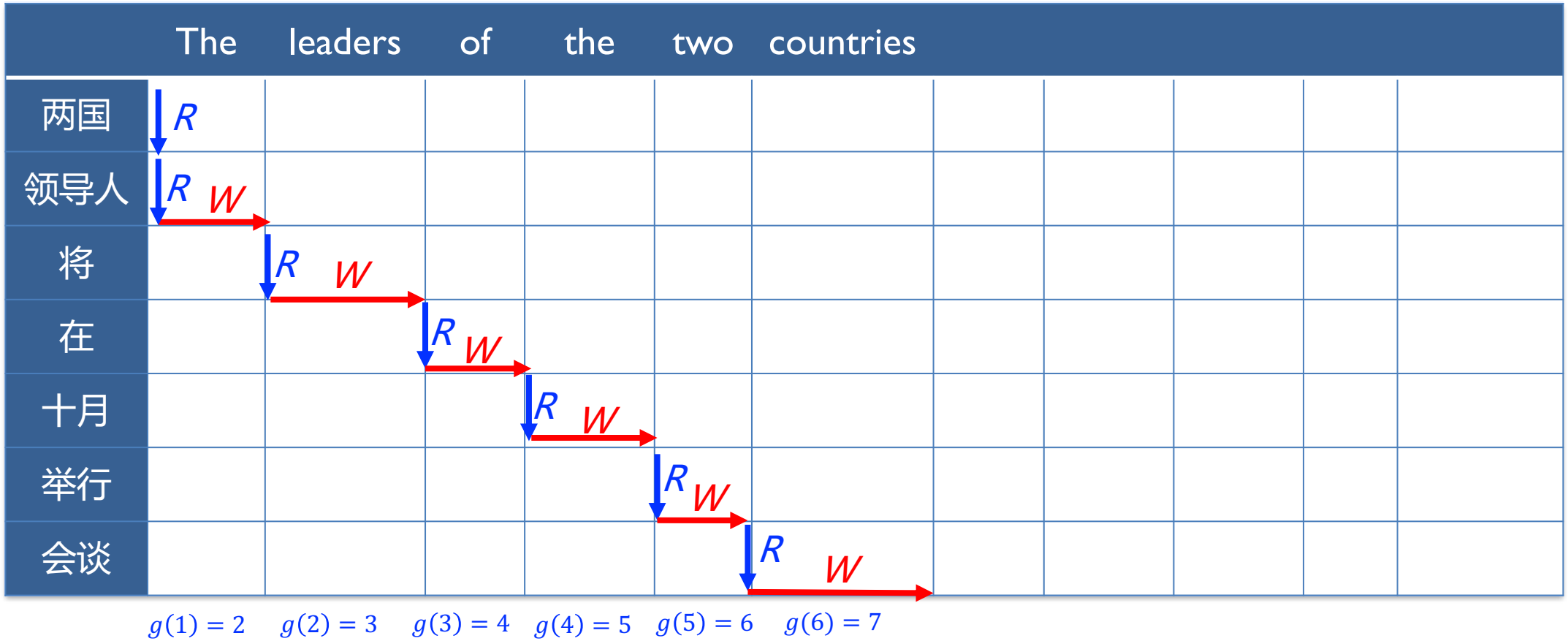


wait-k策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October

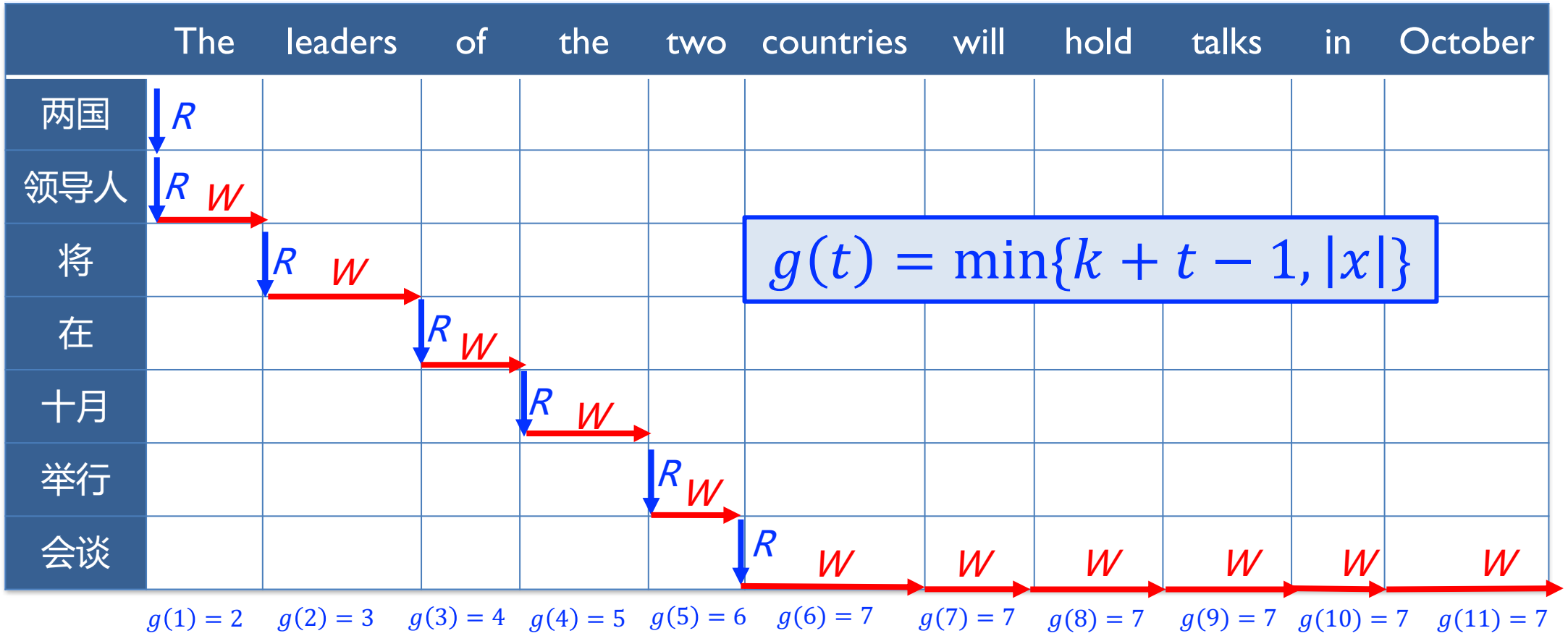


wait-k策略

两国 领导人 将 在 十月 举行 会谈

wait-2

The leaders of the two countries will hold talks in October



wait-k存在的问题

目标句子尾部信息量大，增加听众负担

两国　领导人　将　在　　十月　举行　会谈

wait-2　The　leaders　of　the　two　countries will hold talks in October

wait-k存在的问题

信息不充分导致译文预测错误

大家 对 他 取得 的 成绩 表示 怀疑

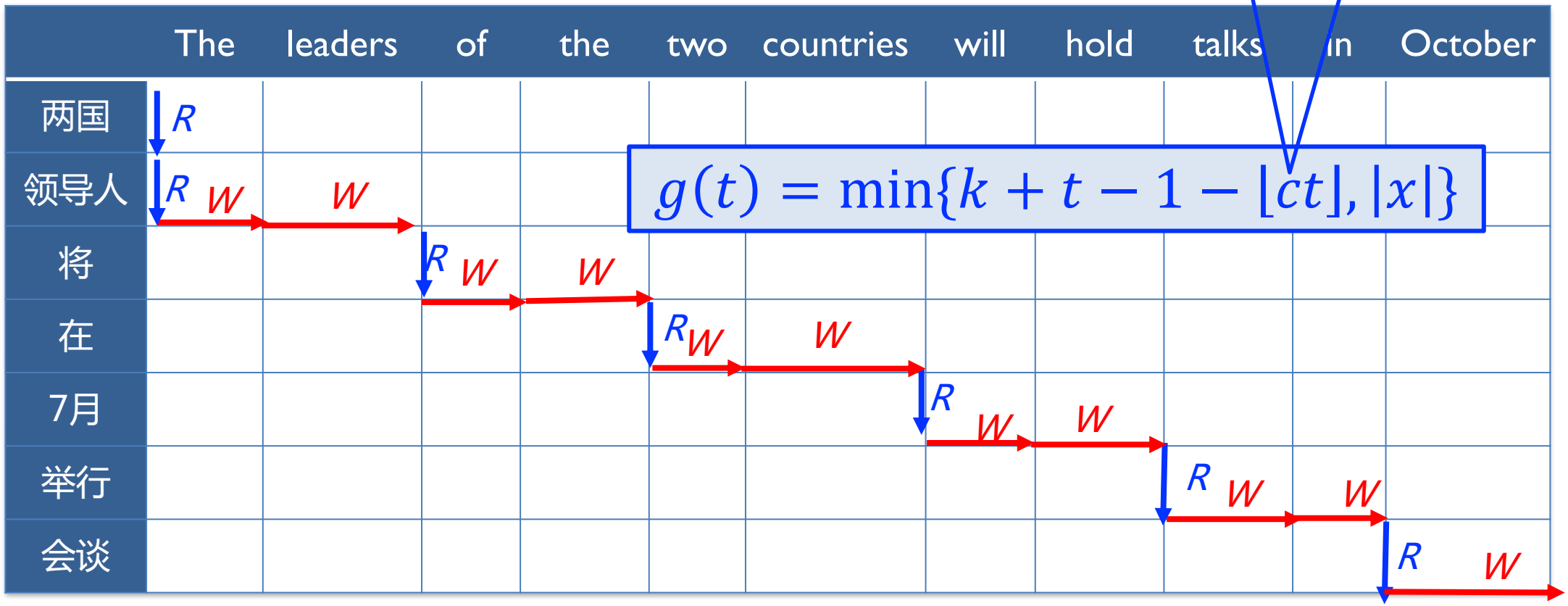
wait-2 Everyone is satisfied with his achievement

wait-5 Everyone is doubt with his achievement

wait-*k* catchup

-根据目标语言和源语言长度比，调整翻译速度

$$c = \frac{|y^*|}{|x|} - 1$$

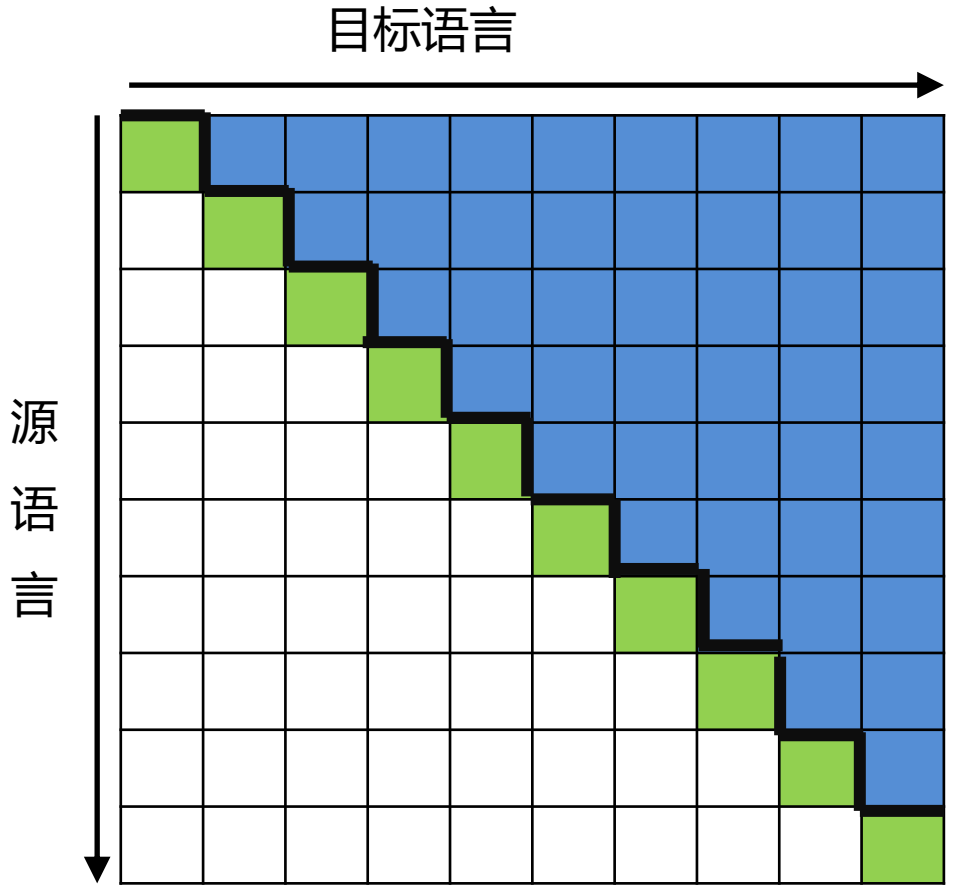


wait-*k*策略 —— Demo

https://github.com/PaddlePaddle/PaddleNLP/blob/develop/examples/simultaneous_translation/stacl/

同传时延的评价

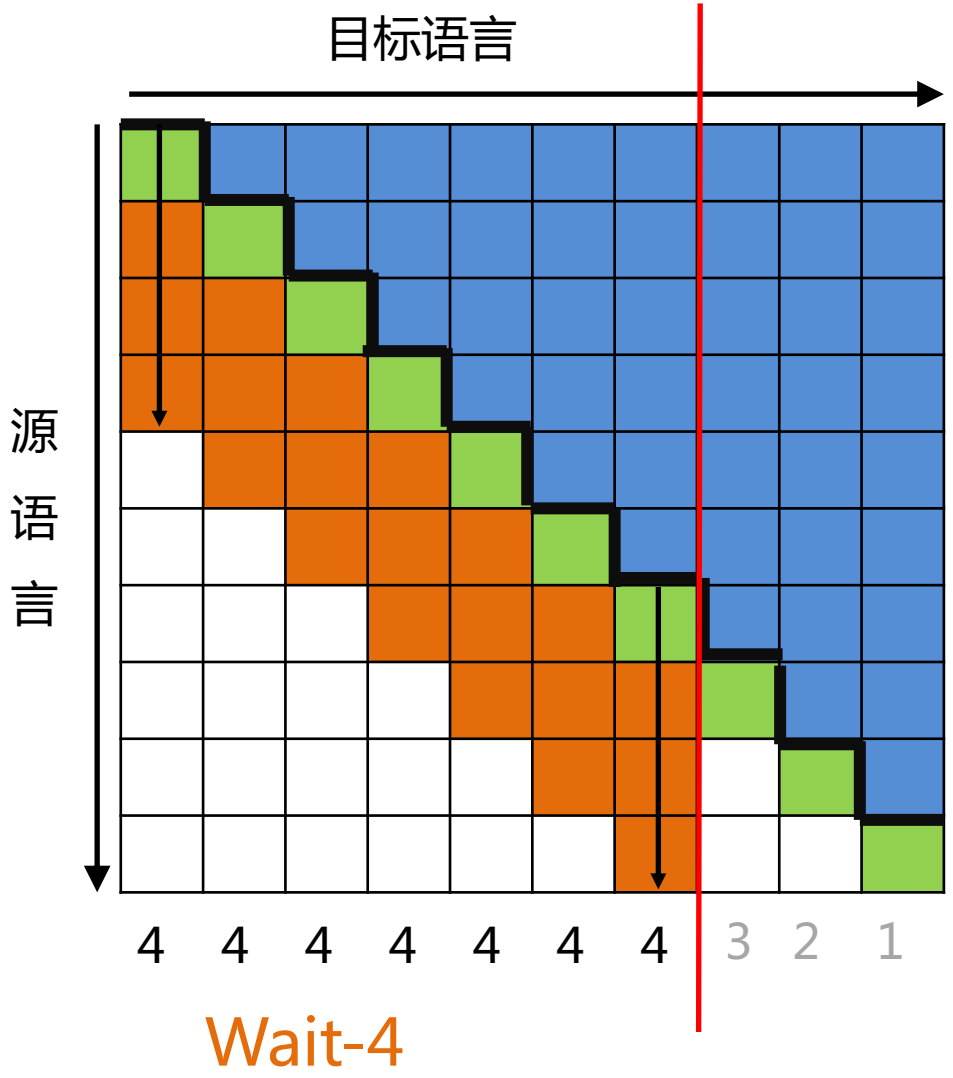
时延评价方法 – Average Lagging



理想情况（oracle）：AL = 0

先写1个词，然后再读1个词，依次进行

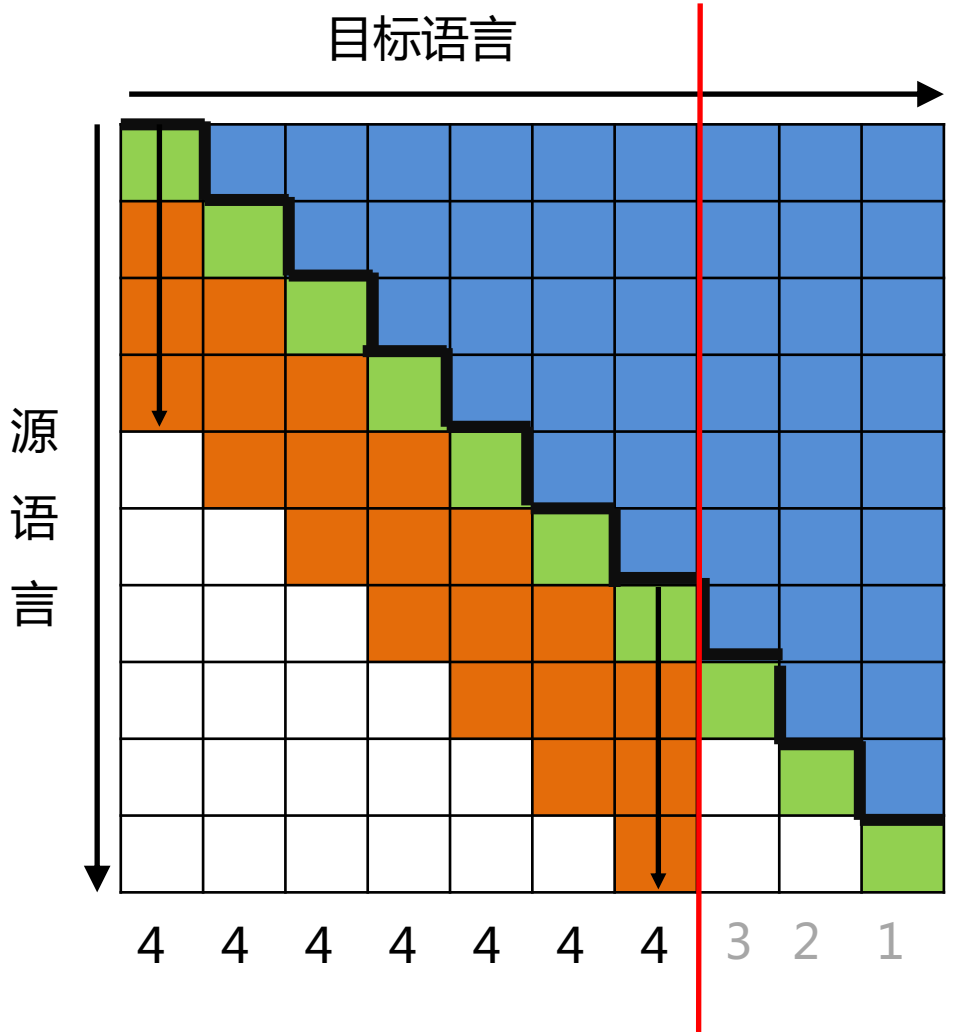
时延评价方法 – Average Lagging



AL = 34/10 = 3.4 ?

需要在源语言句子读完时，停止计算

时延评价方法 – Average Lagging



$$\mathrm{AL} = \frac{1}{\tau}\sum_{i=1}^{\tau} g_i - \frac{i-1}{\gamma}$$

理想情况

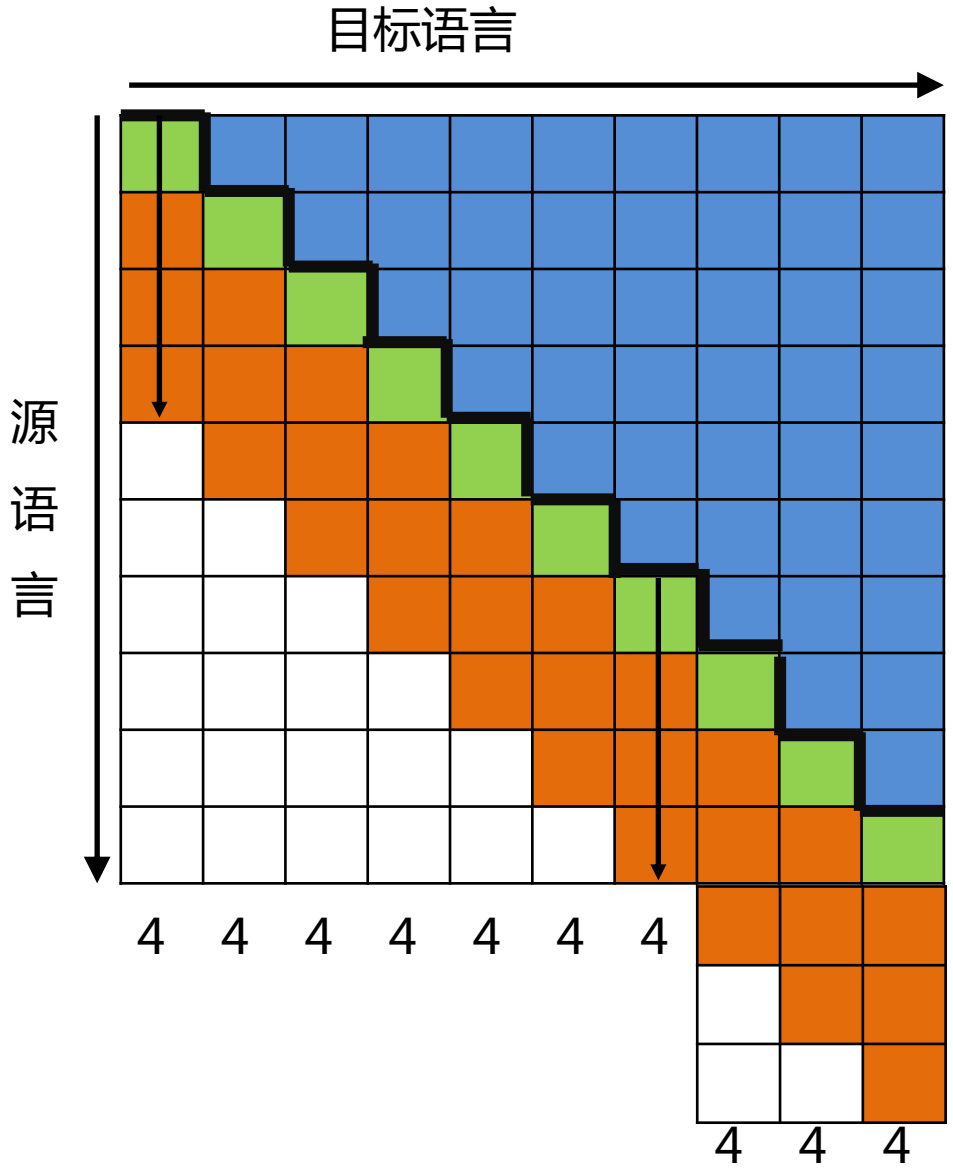
$$\tau = \mathrm{argmin}_i\, g_i = |\mathbf{x}|$$

源语言句子截止位置

$$\gamma = \frac{|y|}{|x|}$$

源语言与目标语言句子长度比

时延评价方法 – Differentiable Average Lagging (DAL)

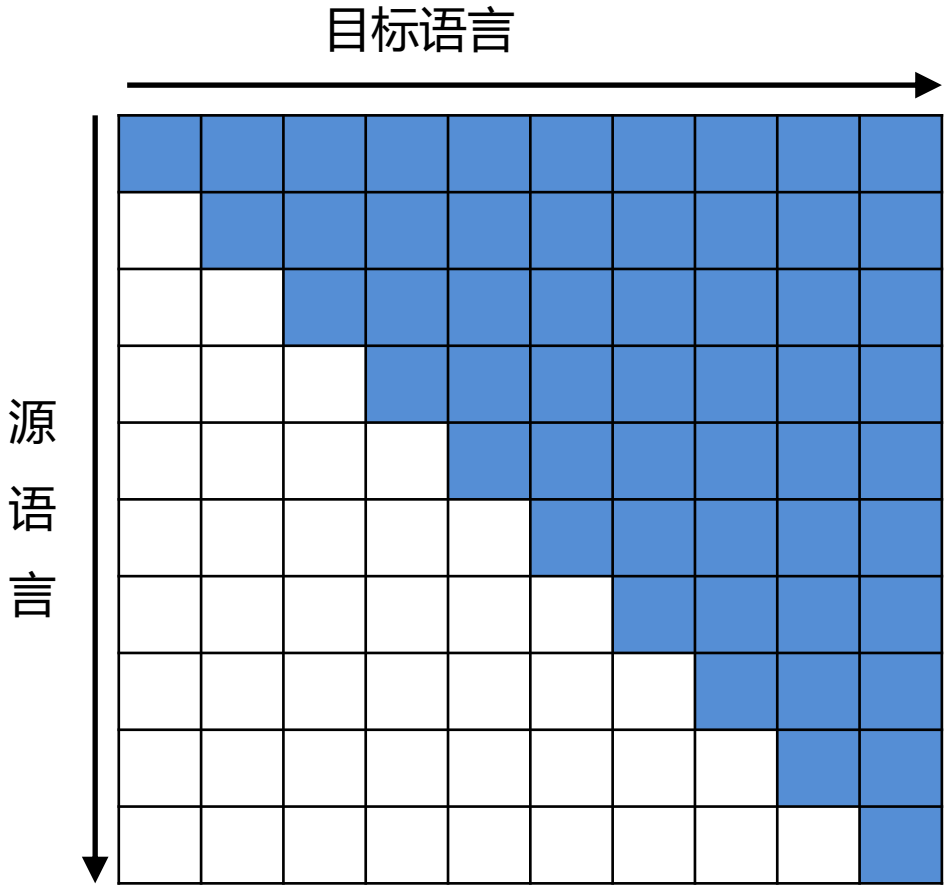


$$\mathrm{DAL} = \frac{1}{|Y|}\sum_{t=1}^{|Y|} g'_i - \frac{i-1}{\gamma}$$

$$g'_i = \max(g_i,\ g'_{i-1} + \frac{1}{\gamma})$$

可集成到损失函数，用于模型训练

时延评价方法 – Average Proportion (AP)



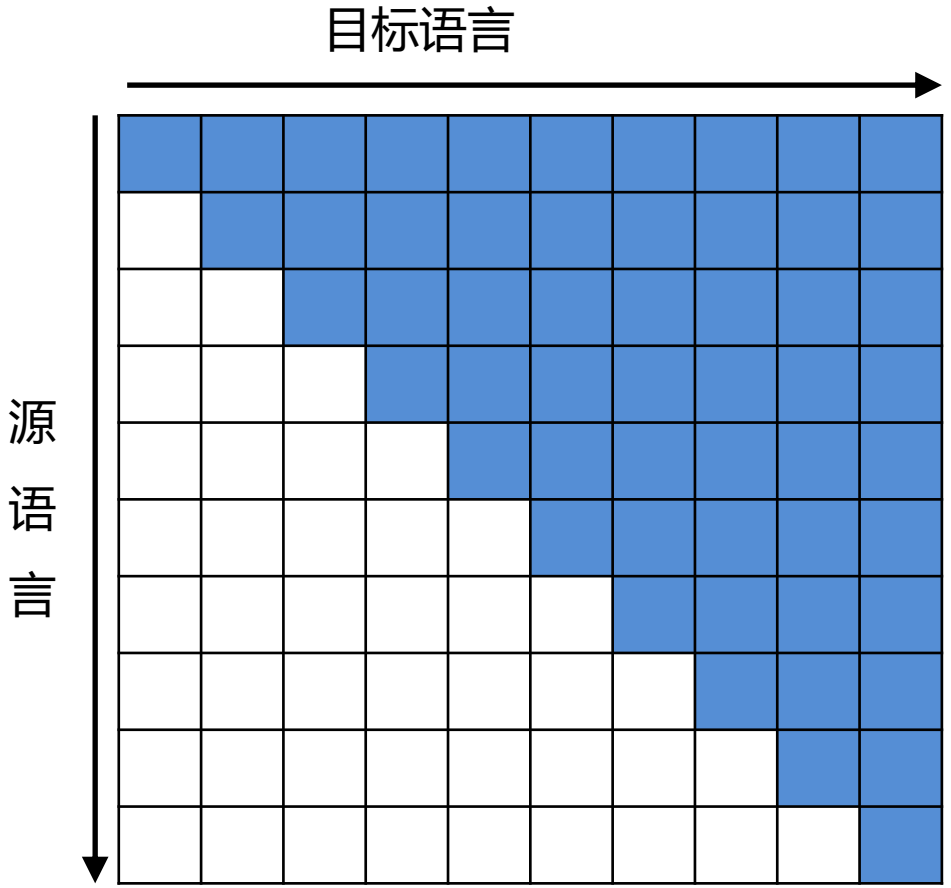
$$\mathrm{AP} = \frac{1 + 2 + \cdots + 10}{10 \times 10} = 0.55$$

the number of source words waited when decoding y_t

$$\mathrm{AP} = \frac{1}{|X||Y|} \sum_{t=1}^{|Y|} g(t)$$

Cho K, Esipova M. Can neural machine translation do simultaneous translation?[J]. arXiv preprintarXiv: 1606.02012,2016.

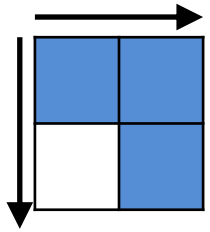
时延评价方法 – Average Proportion (AP)



$$\mathrm{AP} = \frac{1 + 2 + \cdots + 10}{10 \times 10} = 0.55$$

the number of source words waited when decoding y_t

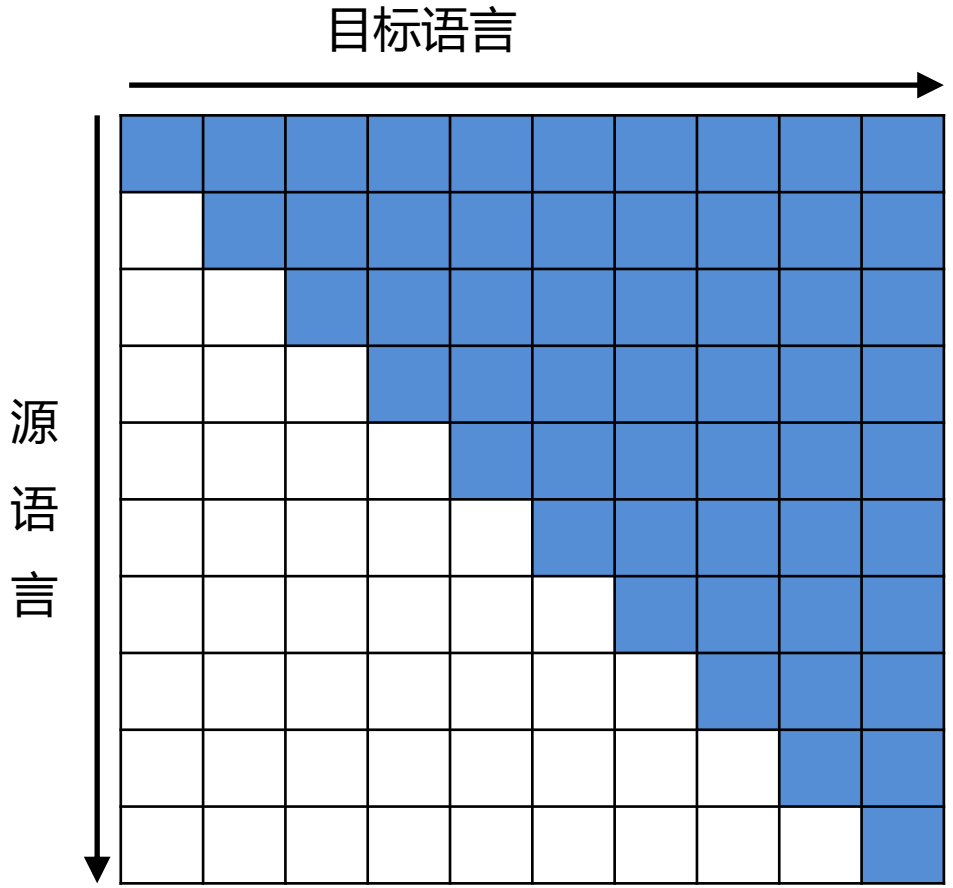
$$\mathrm{AP} = \frac{1}{|X||Y|} \sum_{t=1}^{|Y|} g(t)$$



$$\mathrm{AP} = \frac{1 + 2}{2 \times 2} = 0.75$$

对句子长度敏感

时延评价方法 – Consecutive Wait (CW)



$$\mathrm{CW} = 1$$

$$c_t = \begin{cases} c_{t-1} + 1 & a_t = \mathrm{READ} \\ 0 & a_t = \mathrm{WRITE} \end{cases}$$

$$\mathrm{CW} = \frac{\sum_{t=1}^{|Y|} c_t}{\sum_{t=1}^{|Y|} 1_{c_t>0}}$$

实际衡量的是源语言句子平均切分长度

Gu JT, Neubig G, Cho K, et al. Learning to translate in real-time with neural machine translation[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017:1053–1062.

级联模型 – 同传策略

- 固定策略（Fixed Policy）：读入长度（翻译单元）固定
  - Static read-write (Dalvi et al., 2018)
  - STACL (Ma et al., 2018), etc.

- 自适应策略（Adaptive Policy）：动态调整翻译单元长度
  - Rule-based （Cho et al., 2016)
  - RL-based (Gu et al., 2017）
  - Supervised policy (Zheng et al., 2019)
  - Meaningful unit (Zhang et al., 2020)
  - MILk (Arivazhagan et al., 2019)
  - Multihead monotonic attention (Ma et al., 2020), etc.

预先定义规则

- Wait-If-Worse (score comparison)：读入更多原文，预测概率下降

$$\Lambda(C, C \cup C'): \log p(\hat{y}|\hat{y}_{<t}, C) > \log p(\hat{y}|\hat{y}_{<t}, C \cup C')$$

$$\hat{y} = \text{argmax}_y \log p(y|\hat{y}_{<t}, C)$$

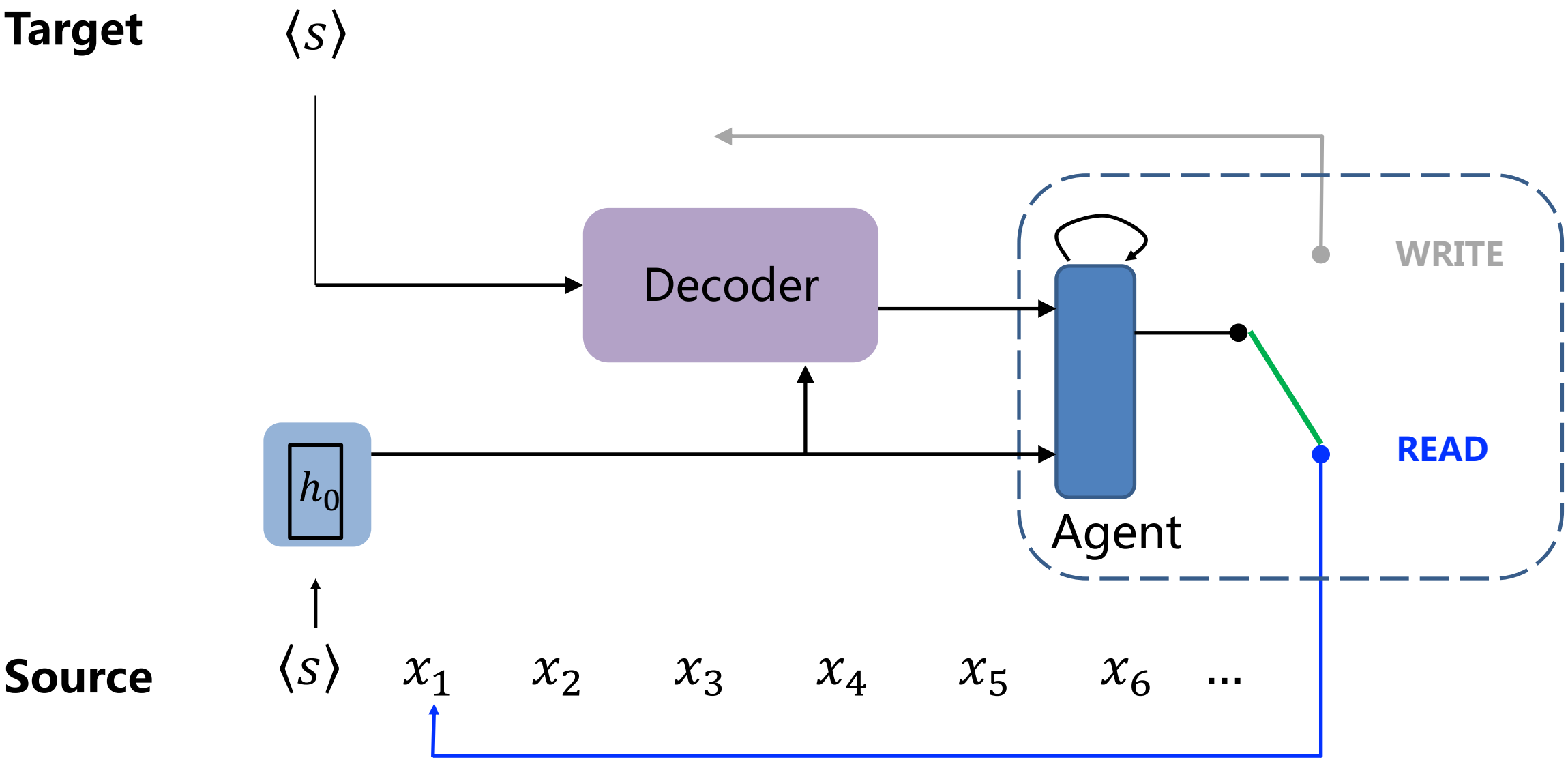
- Wait-If-Diff (Hypothesis comparison)：读入更多原文，预测单词不一致

$$\Lambda(C, C \cup C'): \hat{y} \neq \hat{y}'$$

$$\hat{y}' = \text{argmax}_y \log p(y|\hat{y}_{<t}, C \cup C')$$

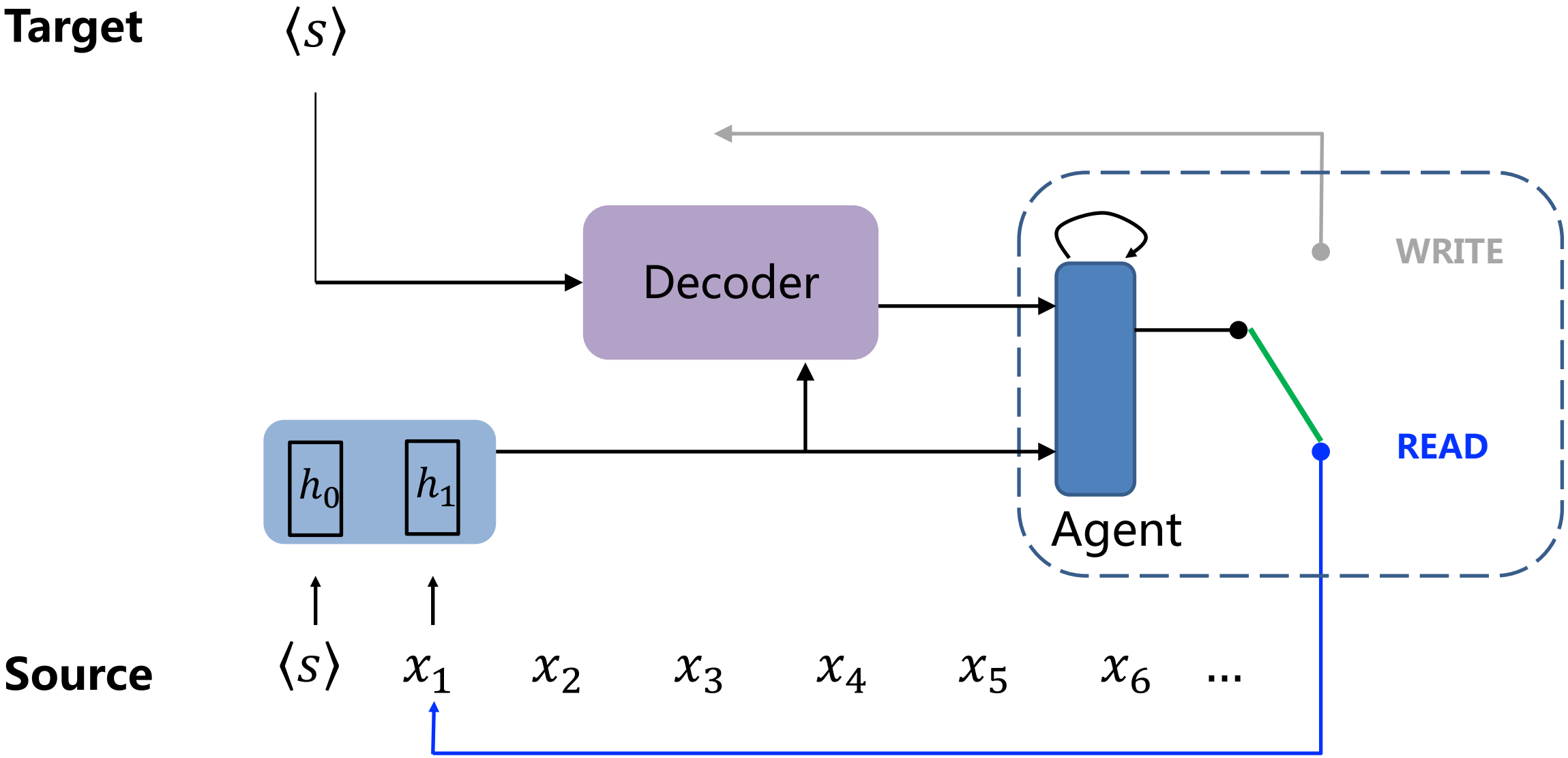
Cho K, Esipova M. Can neural machine translation do simultaneous translation?[J]. arXiv preprintarXiv: 1606.02012,2016.

基于强化学习的策略：READ/WRITE

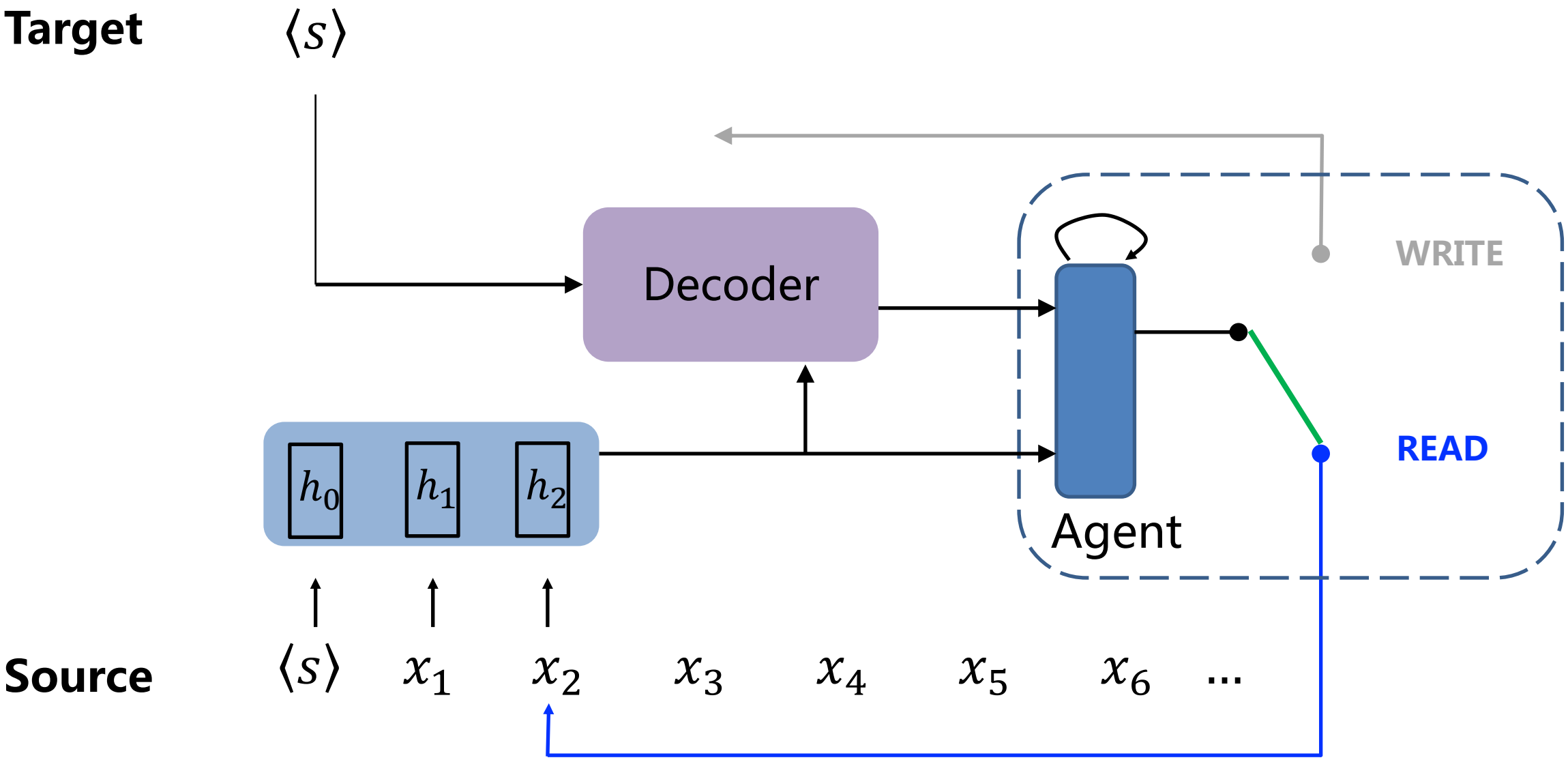


Gu JT, Neubig G, Cho K, et al. Learning to translate in real-time with neural machine translation[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017:1053–1062.

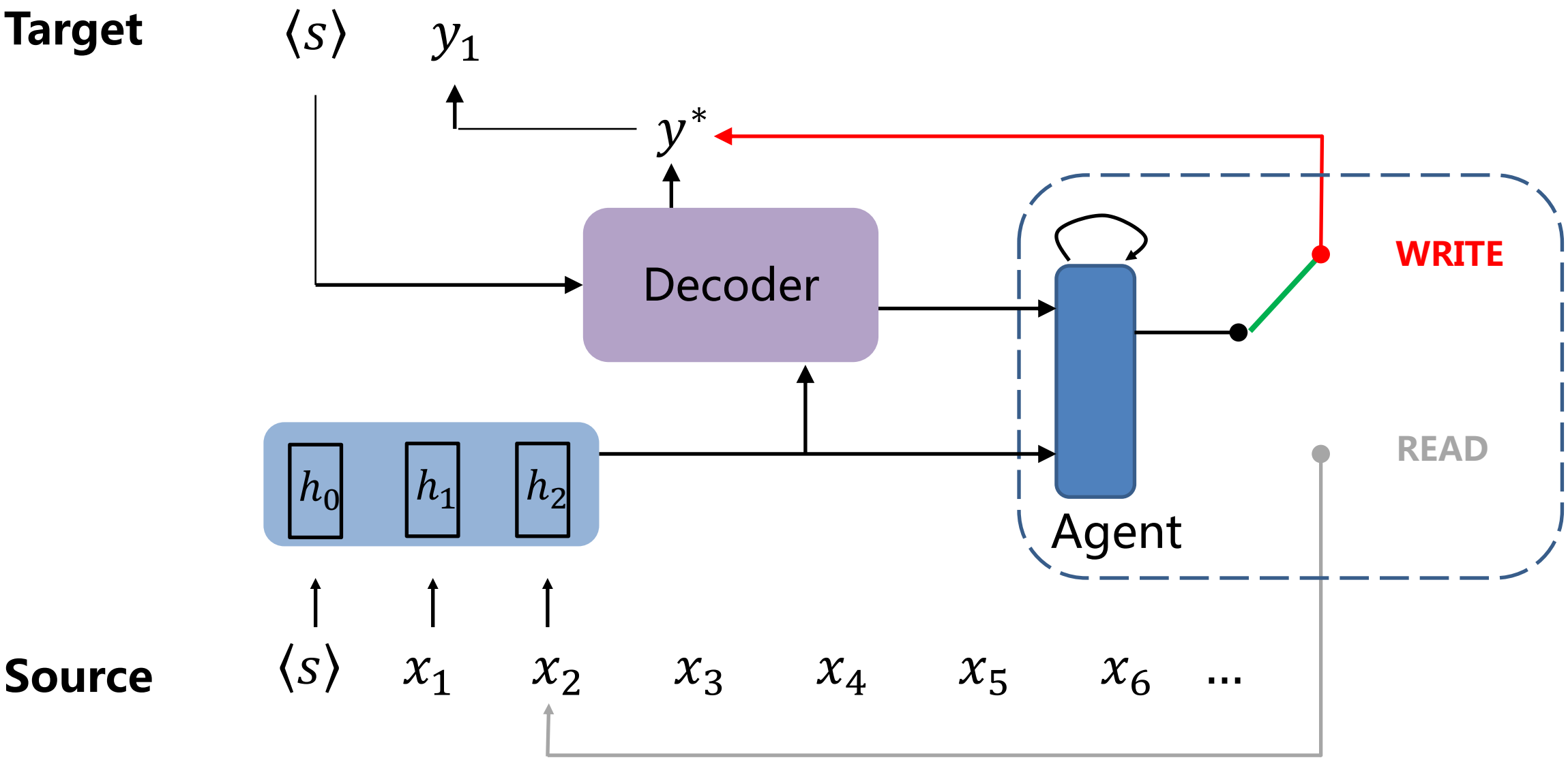
基于强化学习的策略：READ/WRITE



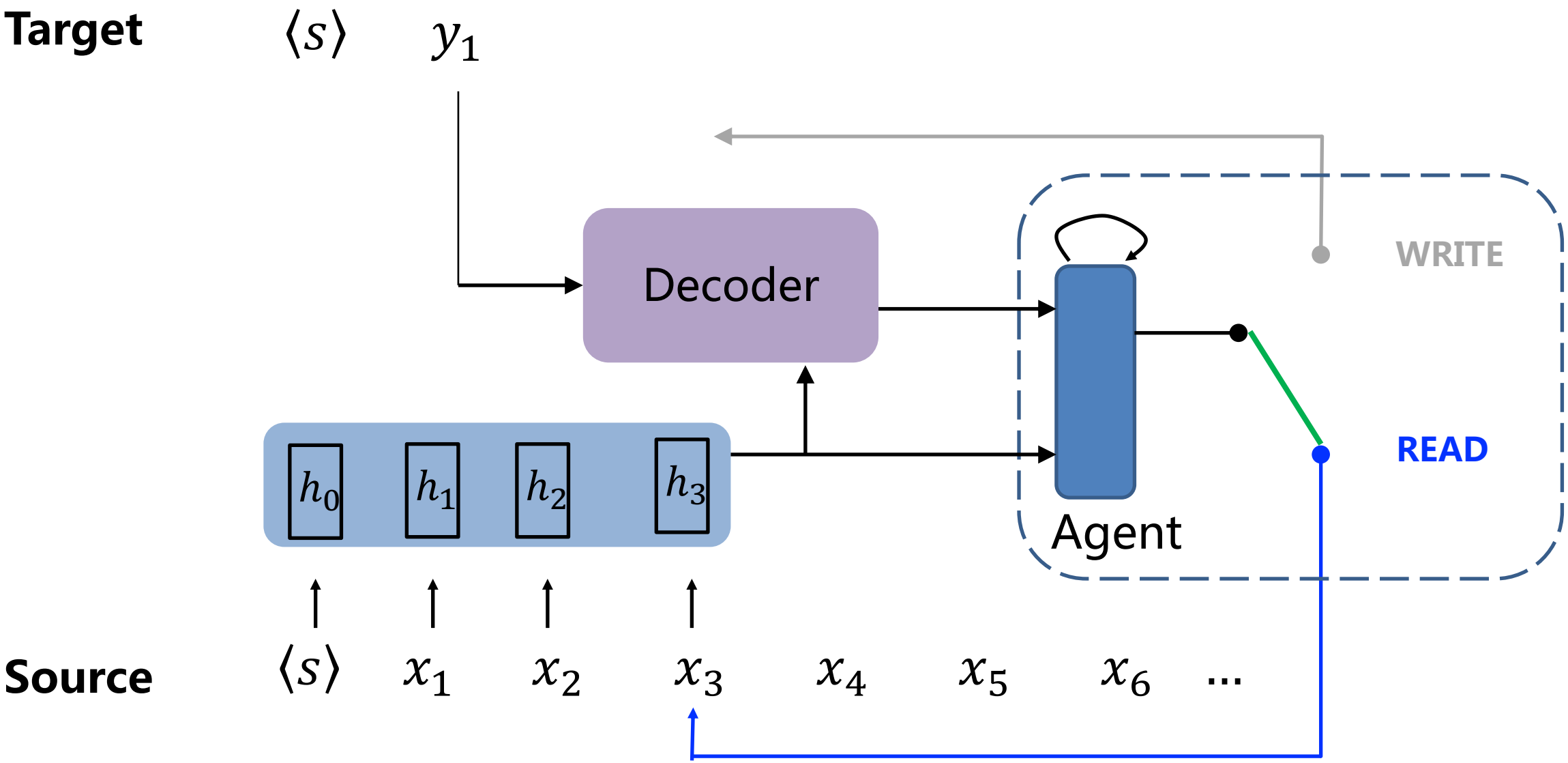
基于强化学习的策略：READ/WRITE



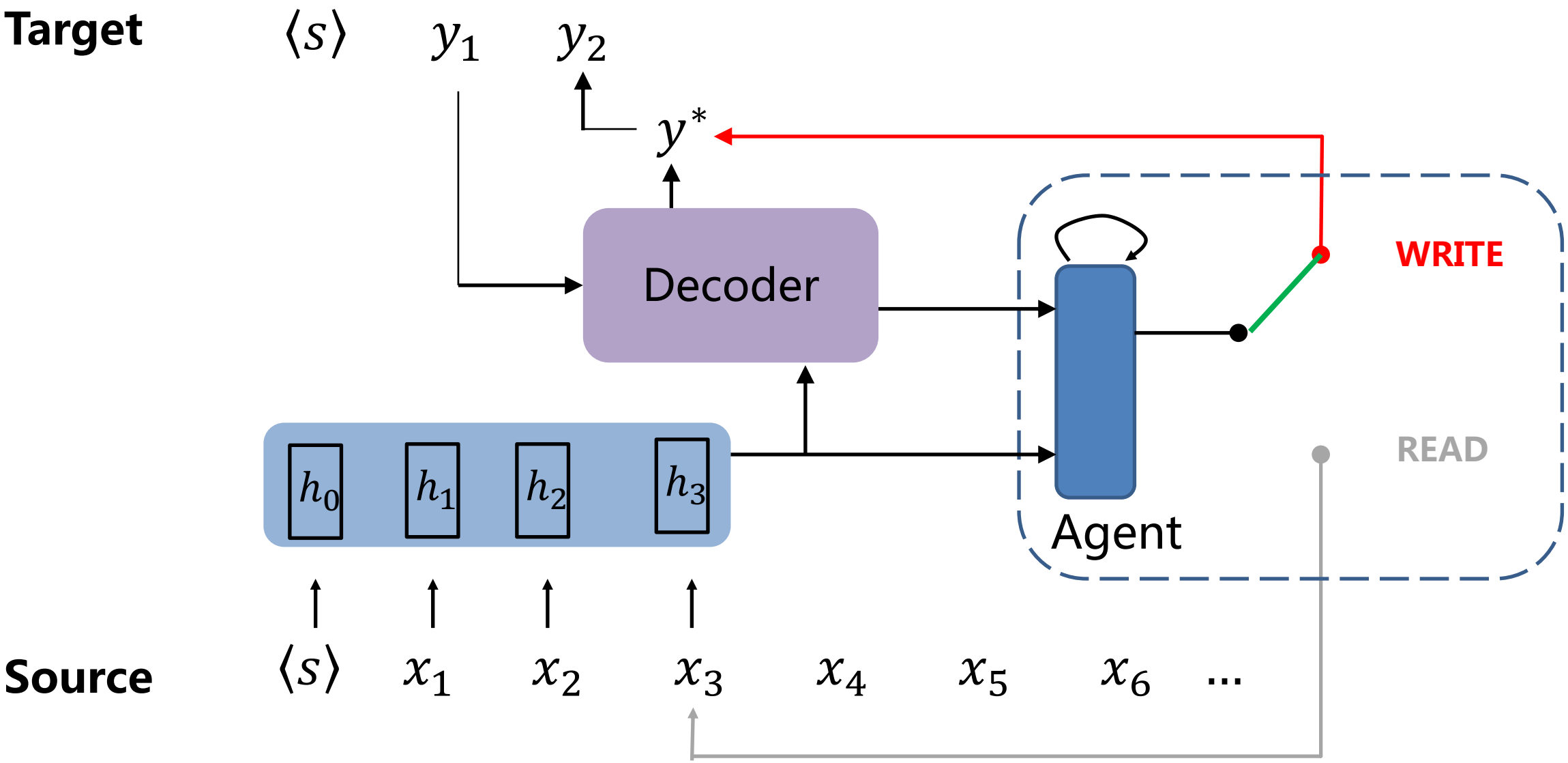
基于强化学习的策略：READ/WRITE



基于强化学习的策略：READ/WRITE



基于强化学习的策略：READ/WRITE



奖励函数（Reward Function）

翻译质量

$$r_t^Q = \begin{cases} \Delta BLEU^0(Y, Y^*, t) & t < T \\ BLEU(Y, Y^*) & t = T \end{cases}$$

$$r_t = r_t^Q + r_t^D$$

时间延迟

$$r_t^D = \underbrace{\alpha \cdot [\mathrm{sgn}(c_t - c^*) + 1]}_{CW} + \underbrace{\beta \cdot \lfloor d_t - d^* \rfloor_+}_{AP}$$

级联模型 – 同传策略

- 固定策略（Fixed Policy）：读入长度（翻译单元）固定
  - Static read-write (Dalvi et al., 2018)
  - STACL (Ma et al., 2018), etc.

- 自适应策略（Adaptive Policy）：动态调整翻译单元长度
  - Rule-based （Cho et al., 2016)
  - RL-based (Gu et al., 2017）
  - Supervised policy (Zheng et al., 2019)
  - Meaningful unit (Zhang et al., 2020)
  - MILk (Arivazhagan et al., 2019)
  - Multihead monotonic attention (Ma et al., 2020), etc.

Supervised Learning

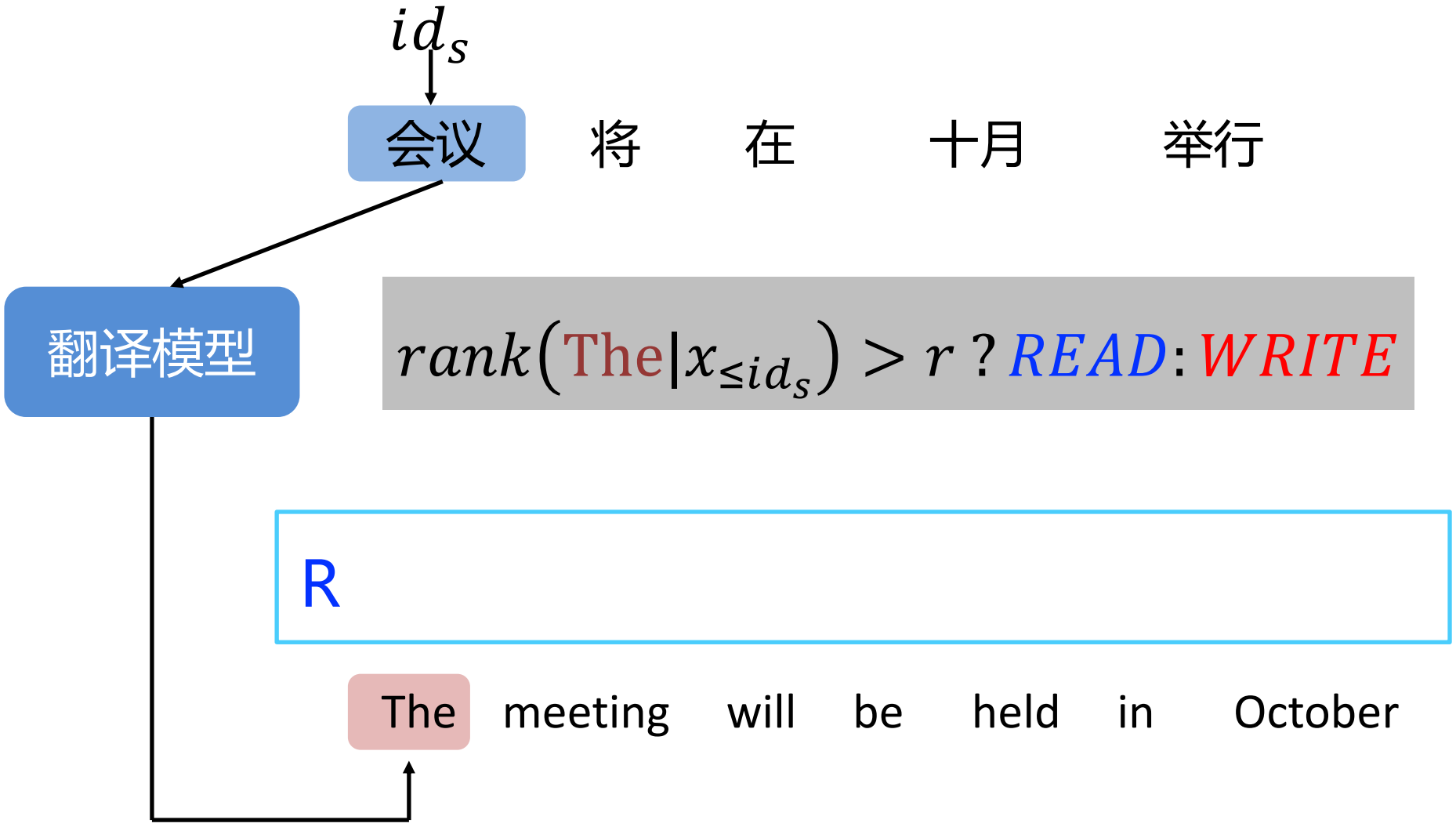
目标：预测动作序列

问题：没有标注数据，也没有标注标准

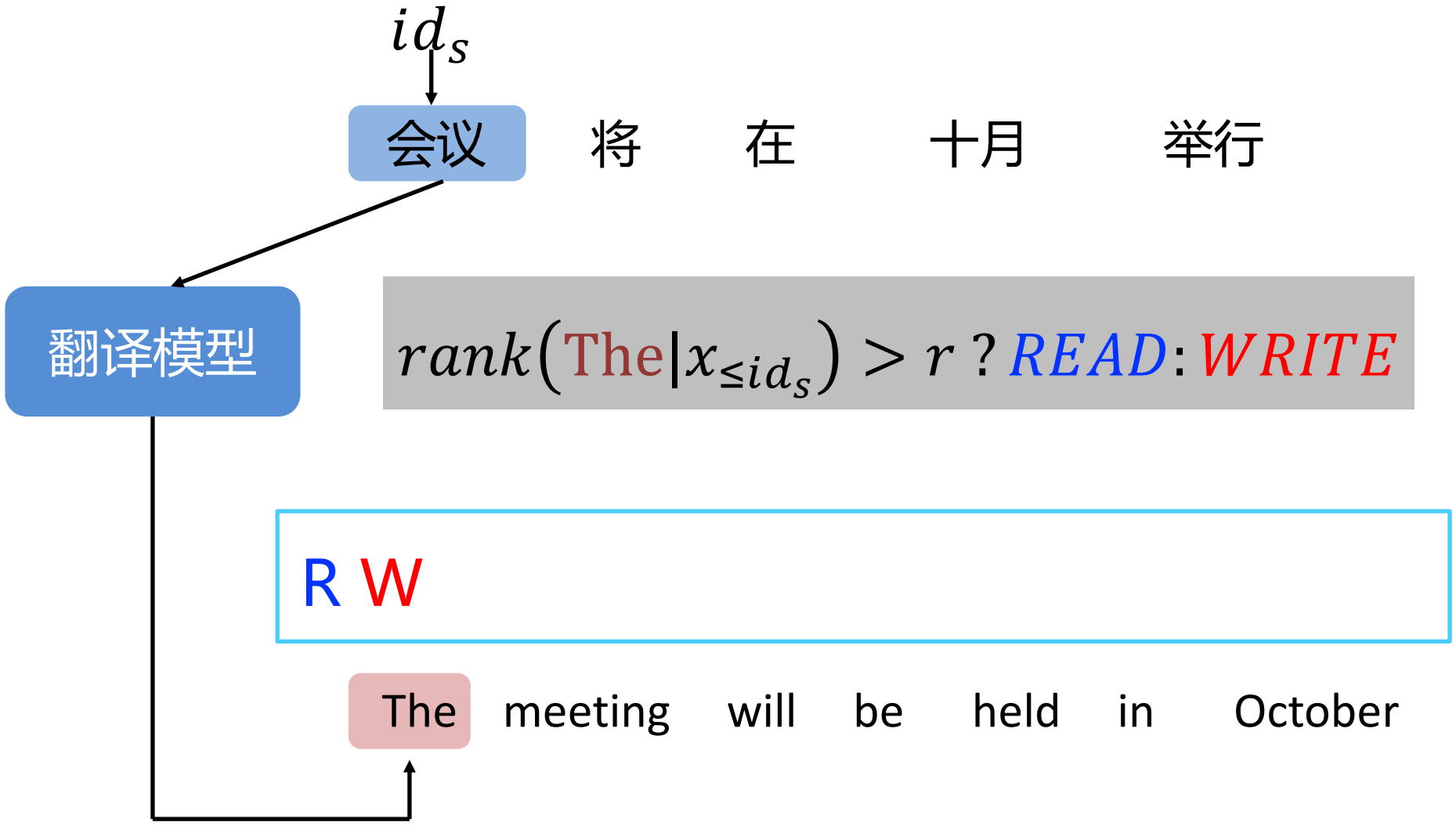
	会议			将				在	十月	举行			
Action	R	W	W	R	W	W	R	R	R	W	W	W	
Translation	The	meeting			will	be				held	in	October	

Baigong Zheng, Renjie Zheng, Mingbo Ma, and Liang Huang. 2019. Simpler and faster learning of adap-tive policies for simultaneous translation. In Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP).

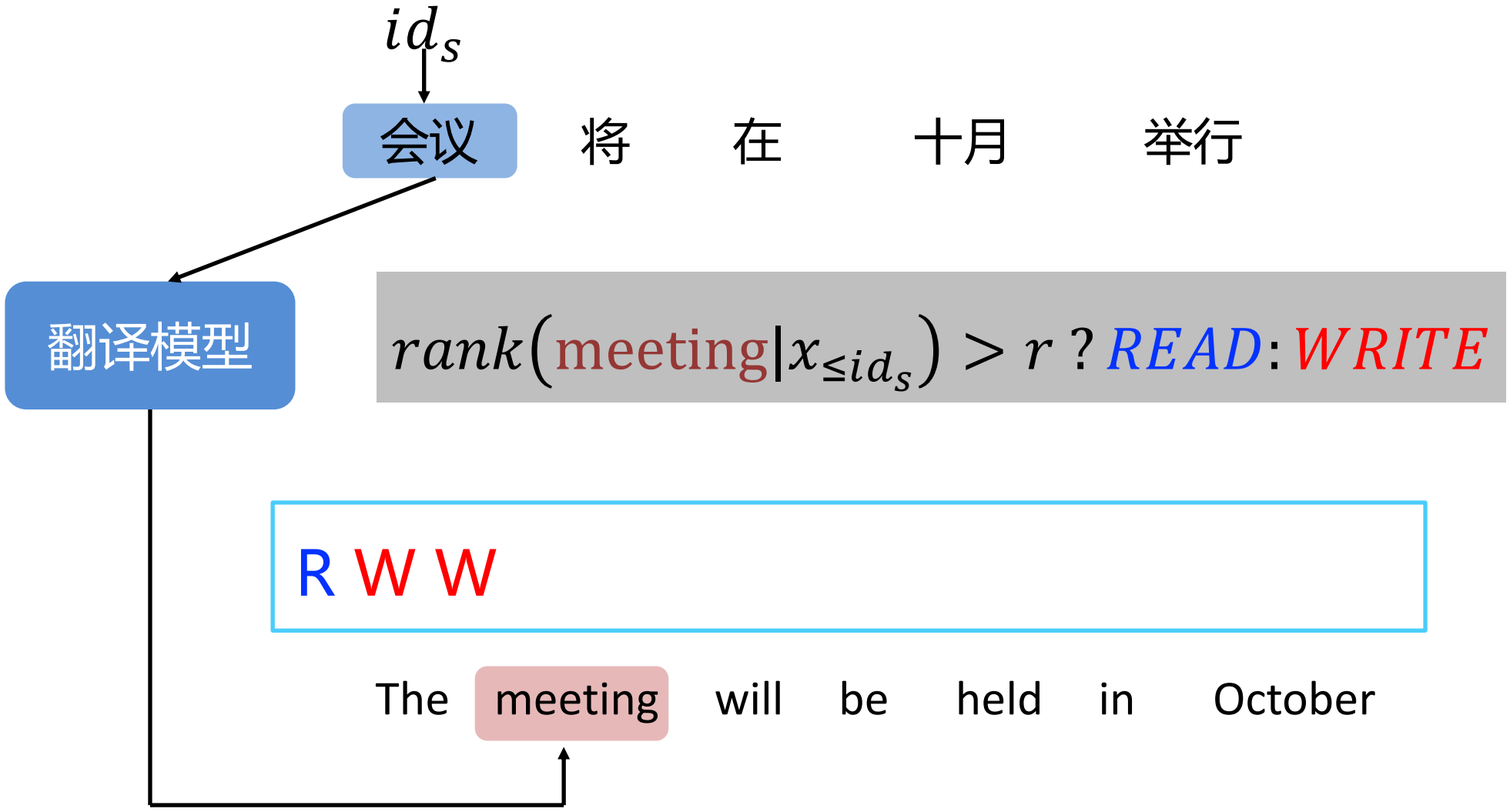
动作序列标注


id_s
会议
将
在
十月
举行
翻译模型
$rank(\text{The}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R
The
meeting
will
be
held
in
October

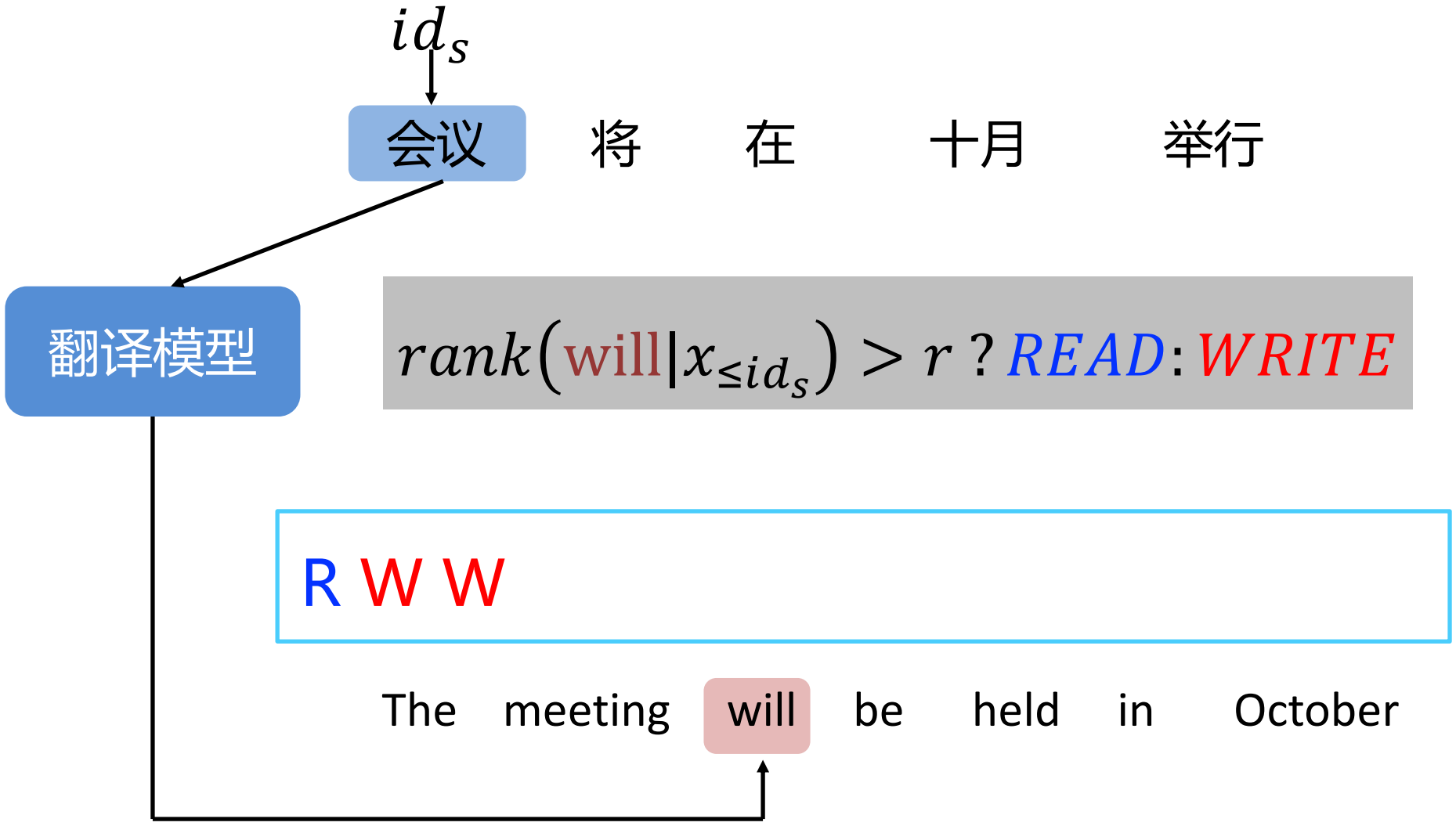
动作序列标注


id_s
会议
将
在
十月
举行
翻译模型
$rank(\text{The}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W
The
meeting
will
be
held
in
October

动作序列标注


id_s
会议
将
在
十月
举行
翻译模型
$rank(\text{meeting}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W W
The
meeting
will
be
held
in
October

动作序列标注


id_s
会议
将
在
十月
举行
翻译模型
$rank(\text{will}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W W
The meeting will be held in October

动作序列标注


id_s
会议 将 在 十月 举行
翻译模型
$rank(\text{will}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W W R
The meeting will be held in October

动作序列标注


id_s
会议 将 在 十月 举行
翻译模型
$rank(\text{will}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W W R W
The meeting will be held in October

动作序列标注


id_s
会议 将 在 十月 举行
翻译模型
$rank(\text{October}|x_{\leq id_s}) > r\ ?\ READ : WRITE$
R W W R W W R R R W W W
The meeting will be held in October

训练与解码

$$p(Action) = \text{argmax} \sum p(a_t | o_{\leq t}, a_{<t})$$

$$p(a_t | o_{\leq t}, a_{<t}) > \rho ? \text{READ} : \text{WRITE}$$

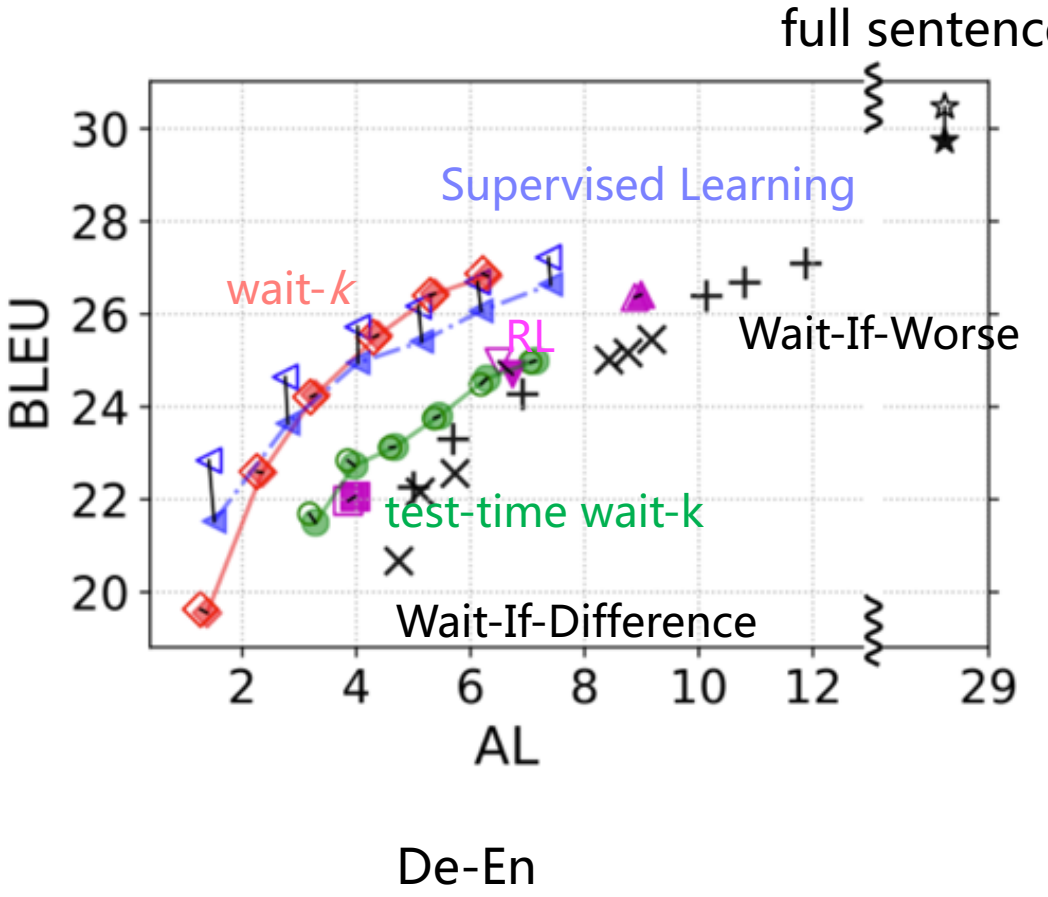
trade off between quality and latency

ρ 越大，倾向于WRITE操作，时延低

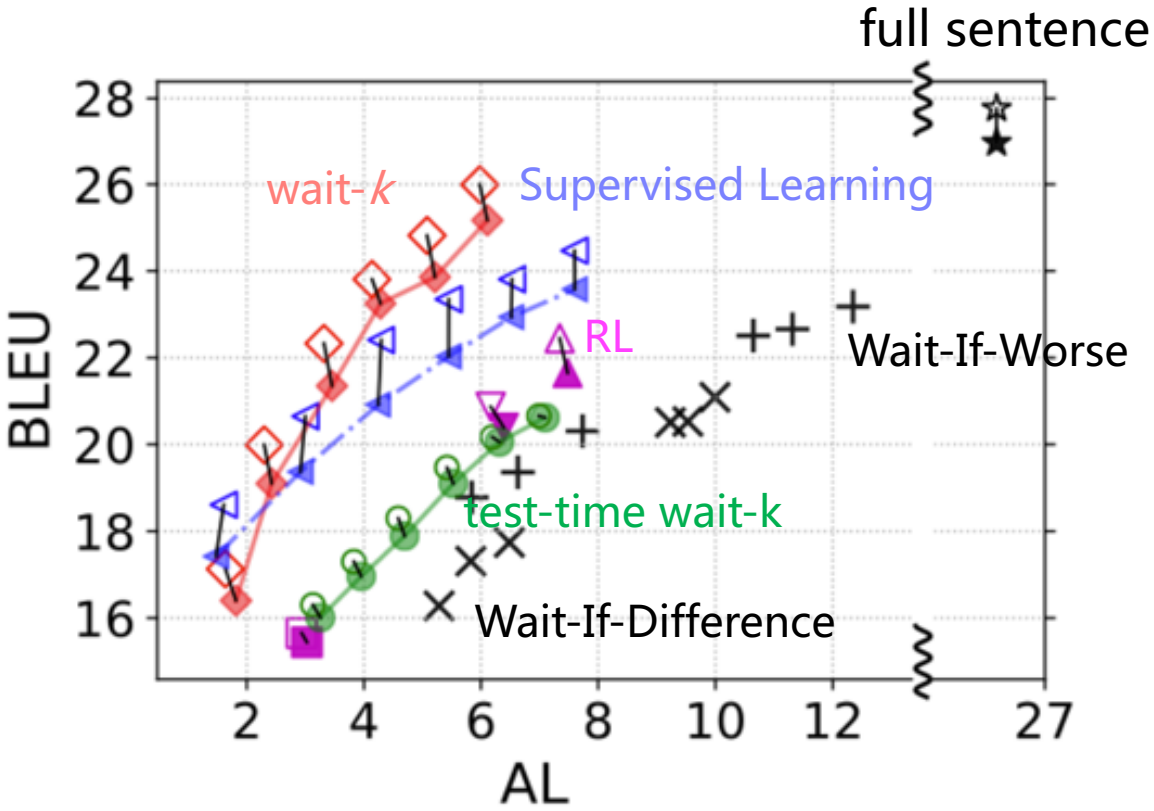
ρ 越小，倾向于READ操作，时延高

实验（EN-DE）

Trained on 4.5M sentence pairs (WMT 15)



De-En



En-De

基于语义单元（Meaningful Unit）的自适应策略

人类同传策略：顺句驱动、语义单元

Zhang RQ, Zhang CQ, He ZJ, et al. Learning adaptive segmentation policy for simultaneous translation[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2020: 2280-2289

如何定义语义单元？

- 尽量短，以缩短时延
- 含有较充分的信息（不依赖于后文），保证翻译质量

Meaningful Unit ?

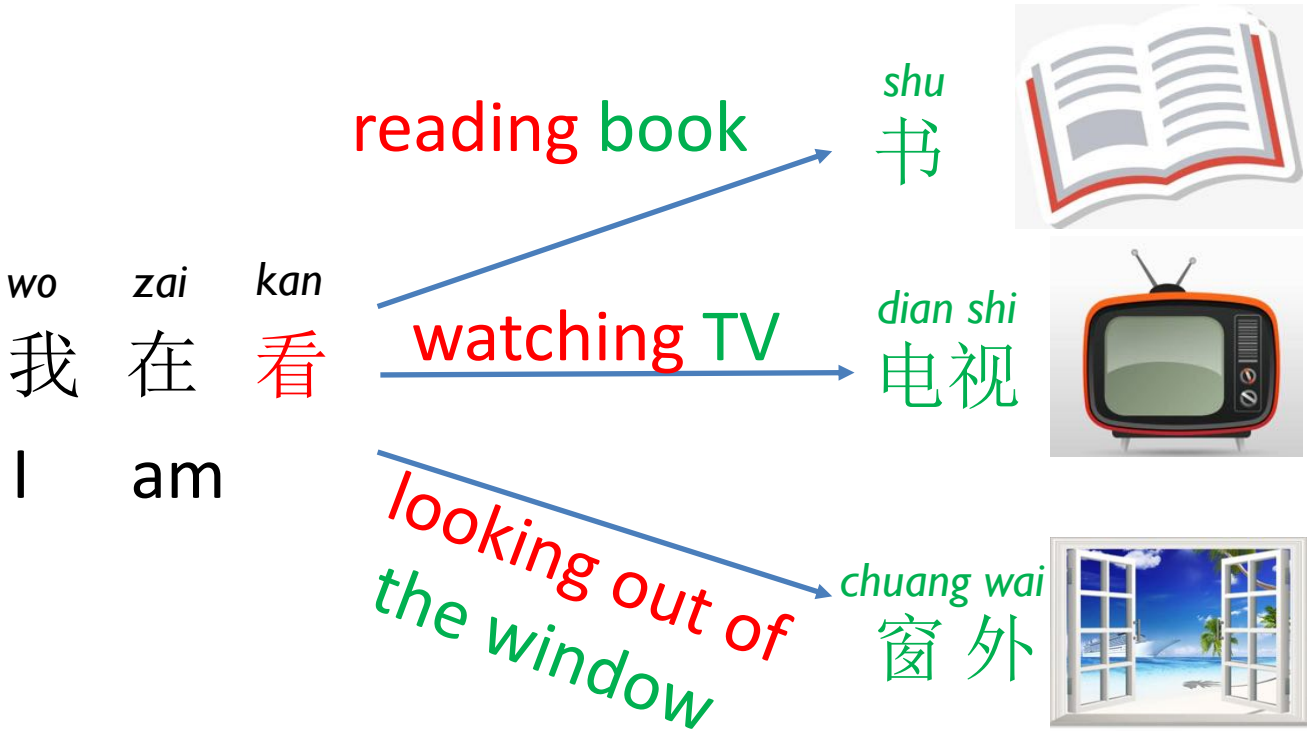
wo *zai* *kan*

我 在 看

I am

如何定义语义单元？

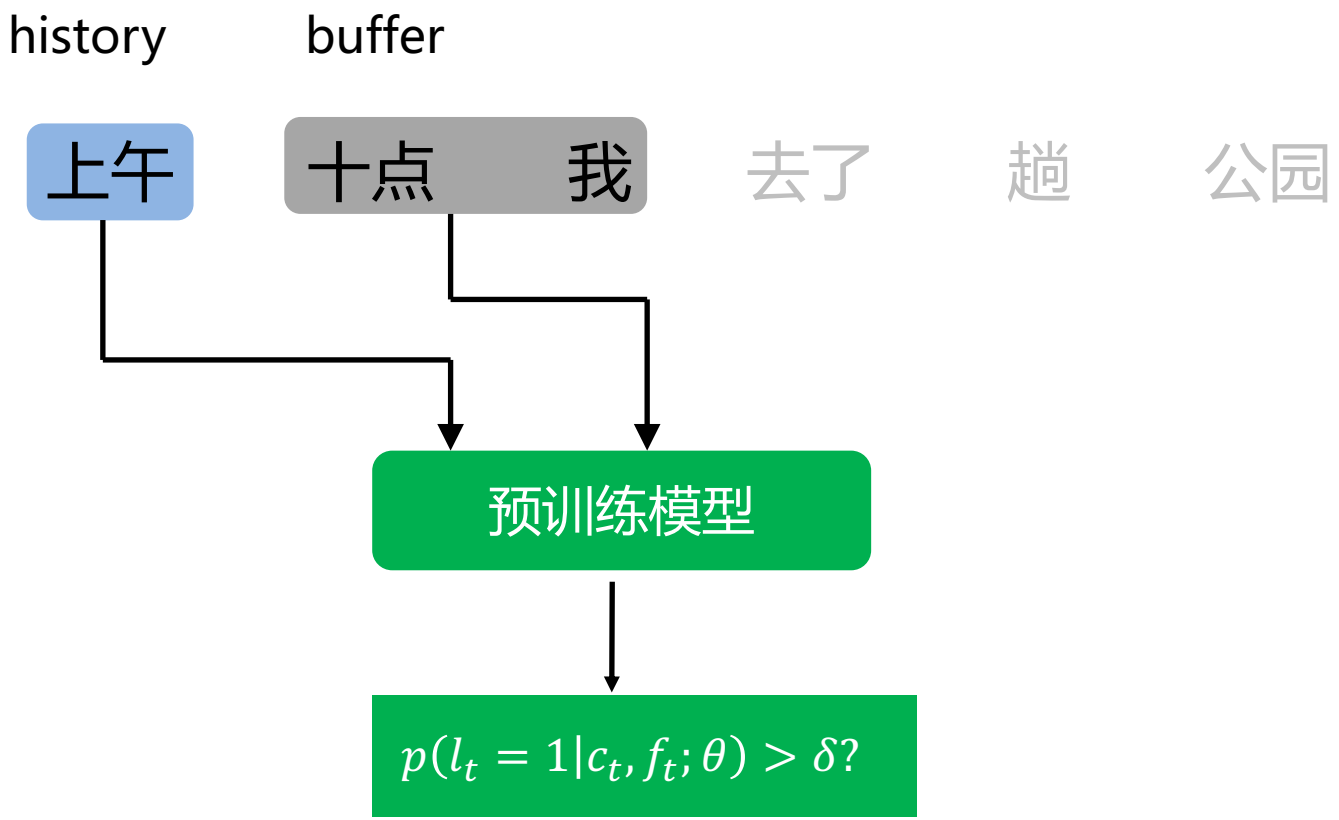
- 尽量短，以缩短时延
- 含有较充分的信息（不依赖于后文），保证翻译质量



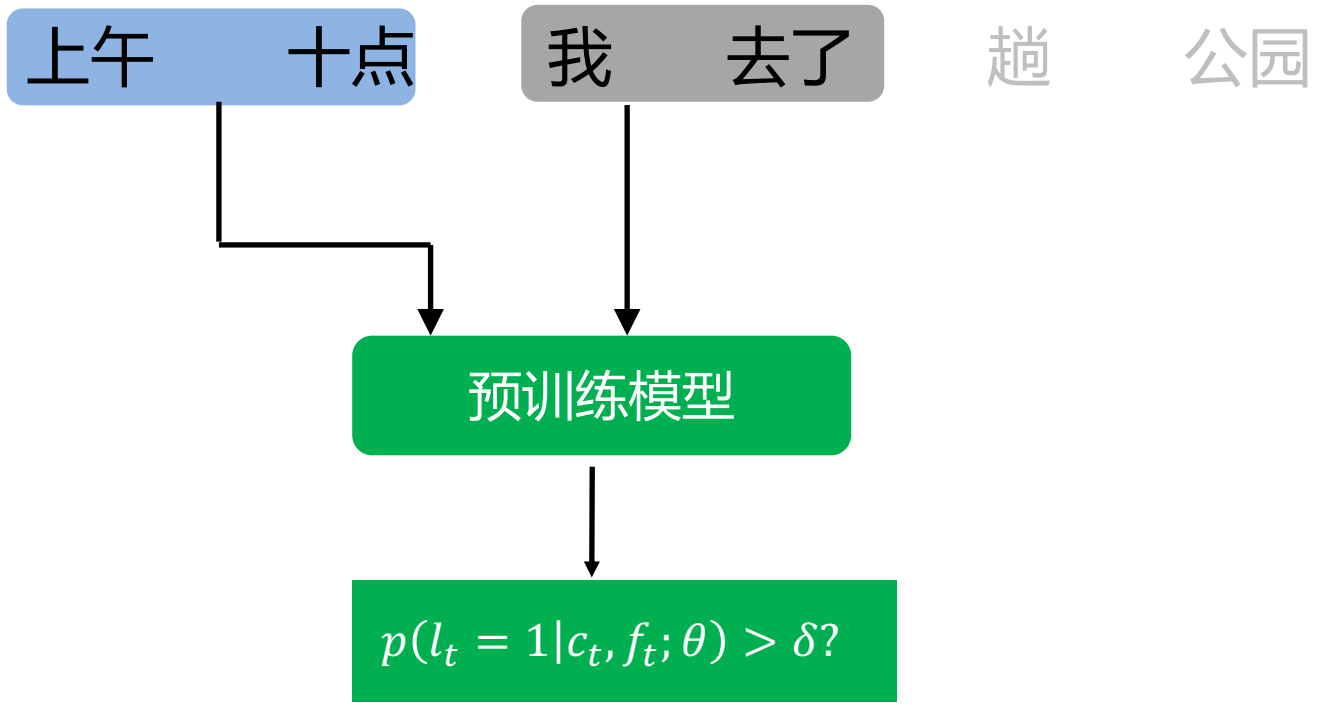
语义单元识别

- 分类问题，单词是否是语义单元边界？
- 基于预训练模型的边界识别 Pre-training & Fine-tuning

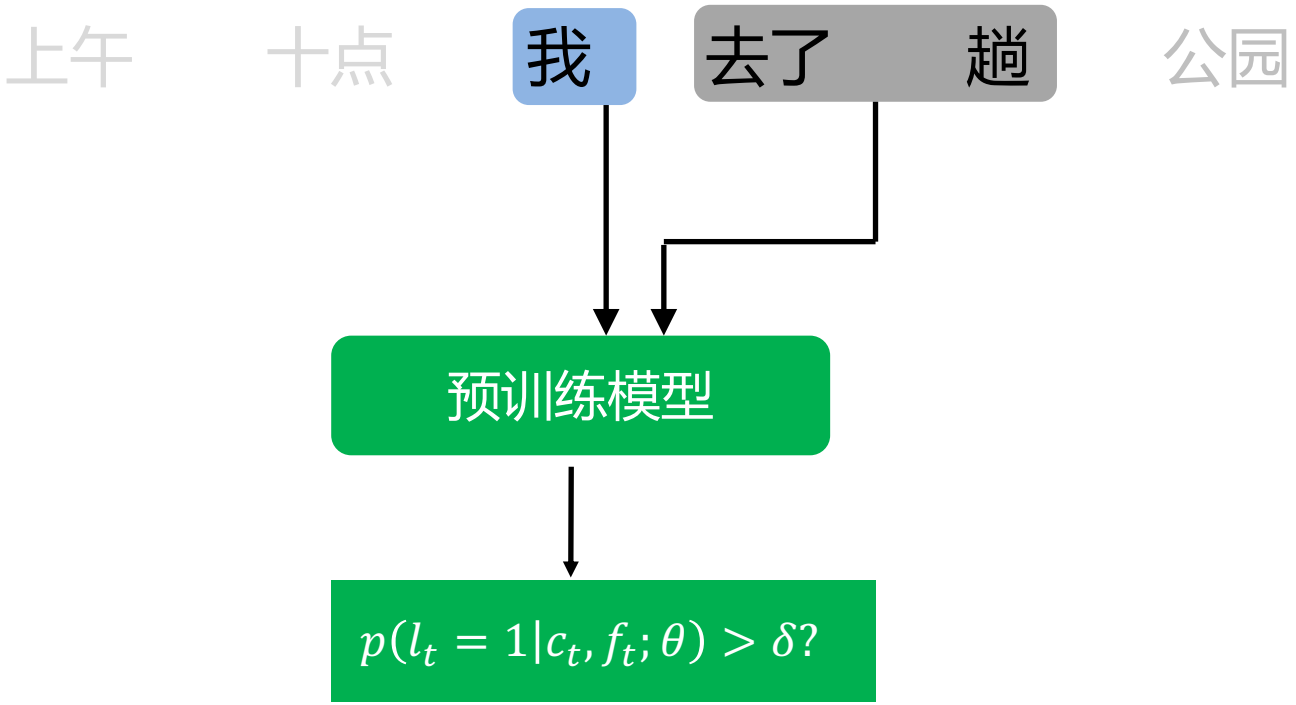
语义单元识别



语义单元识别



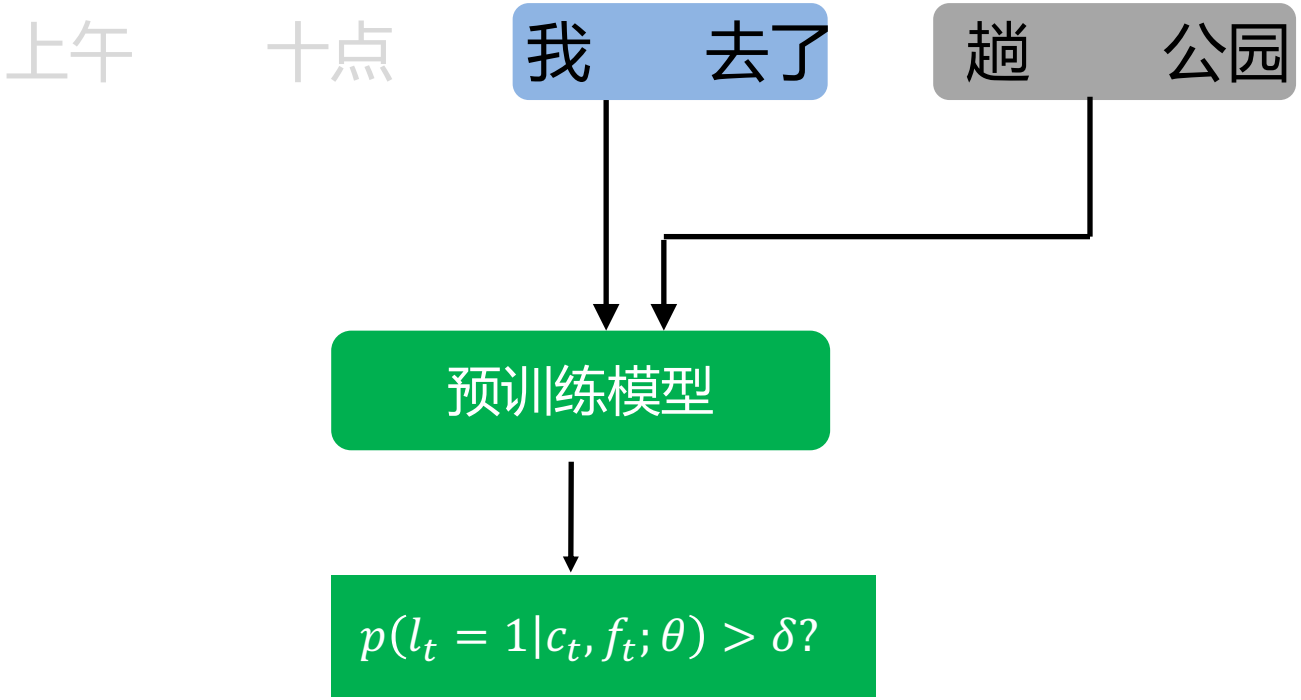
语义单元识别



上午　十点

At 10:00 a.m.

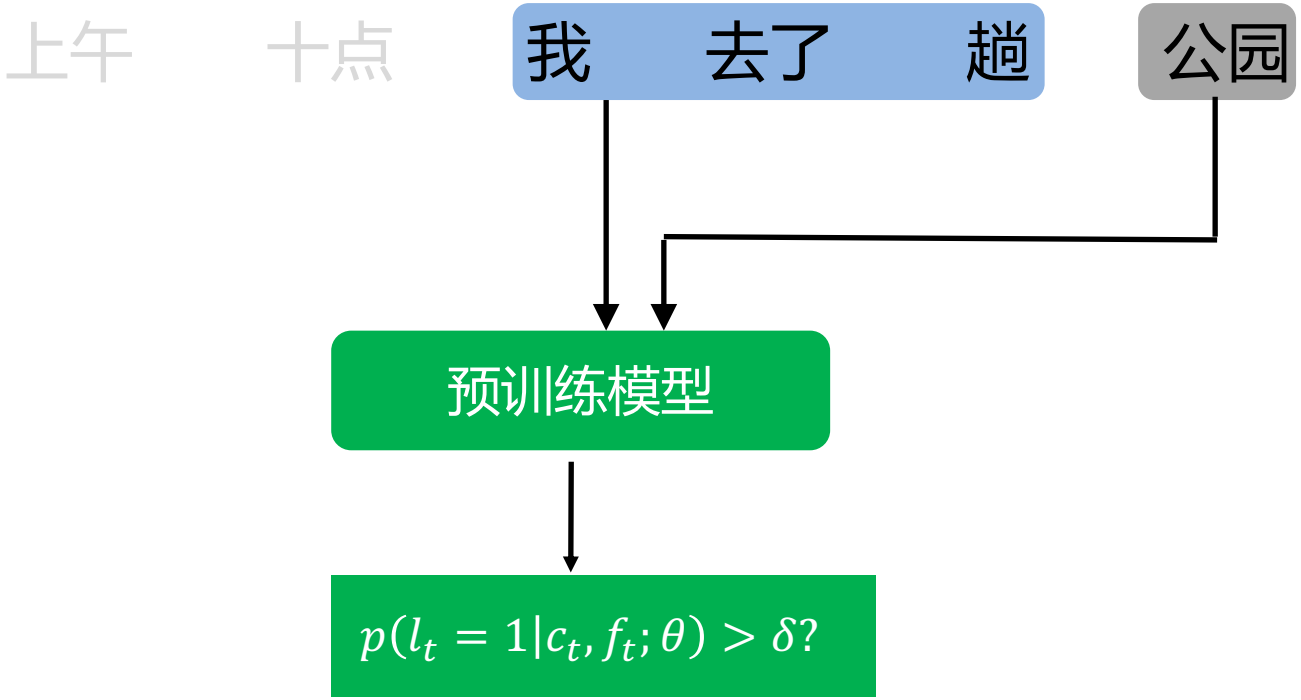
语义单元识别



上午 十点

At 10:00 a.m.

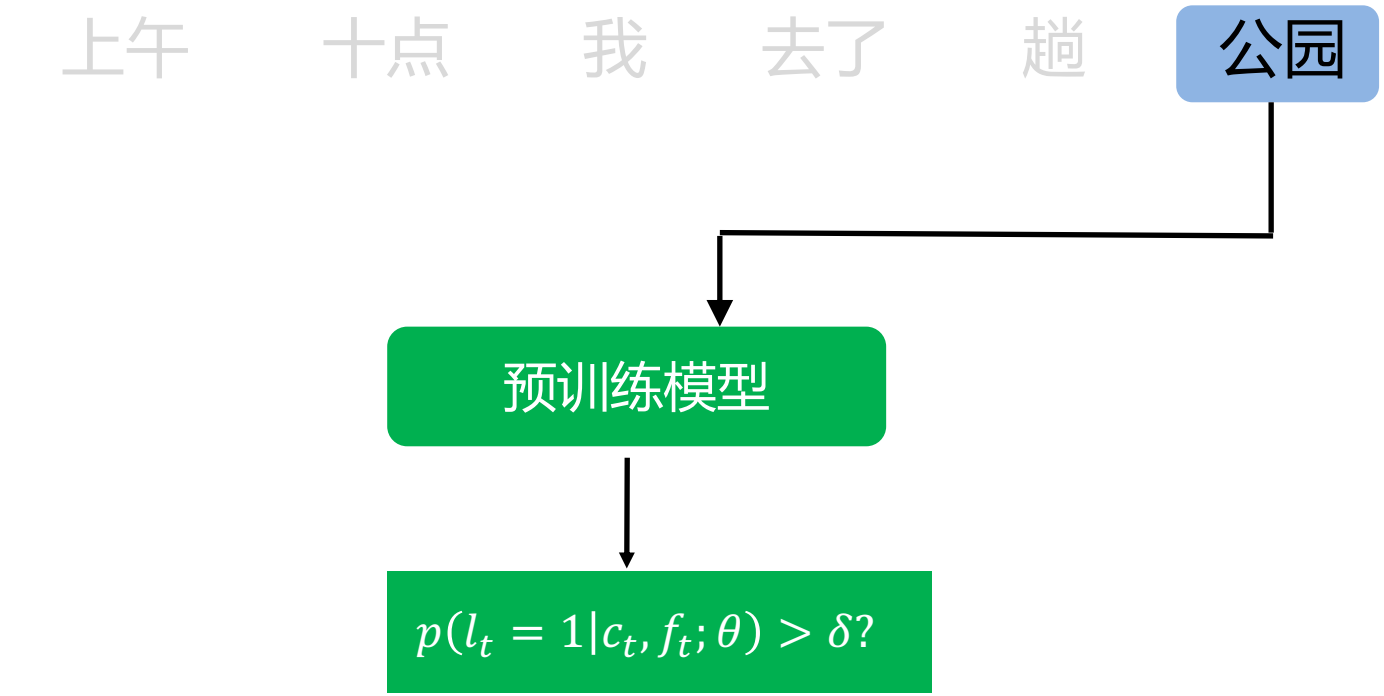
语义单元识别

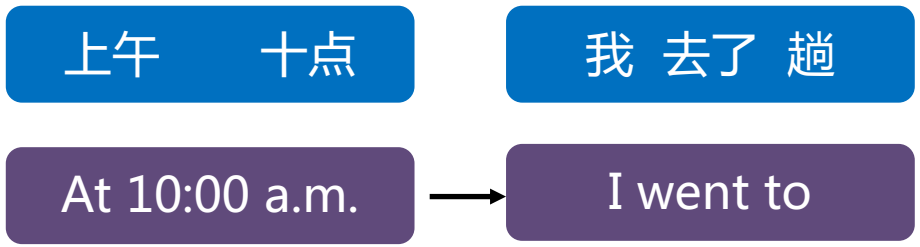


上午　十点

At 10:00 a.m.

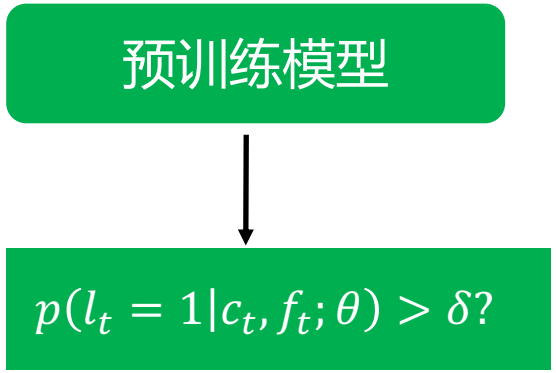
语义单元识别

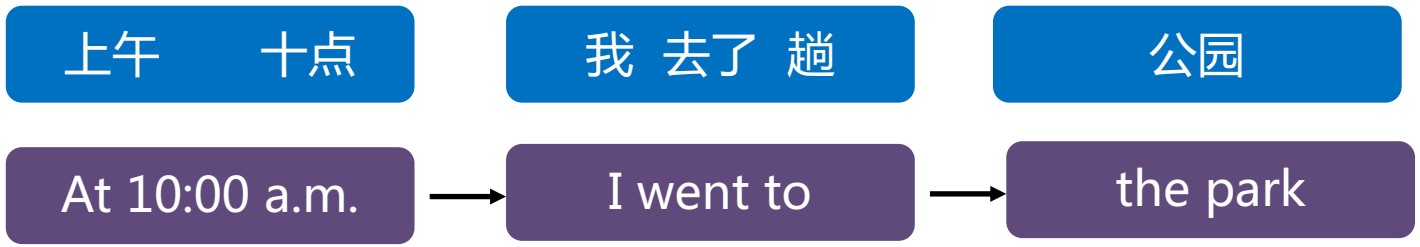

上午
十点
我
去了
趟
公园
预训练模型
$p(l_t = 1|c_t, f_t; \theta) > \delta?$


上午 十点
我 去了 趟
At 10:00 a.m.
I went to

语义单元识别

上午　十点　我　去了　趟　公园





构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.

构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
$M'_{nmt}(x_{\leq 1})$	Morning						

构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
$M'_{nmt}(x_{\leq 1})$	Morning						
$M'_{nmt}(x_{\leq 2})$	Morning 10						

构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
$M'_{nmt}(x_{\leq 1})$	Morning						
$M'_{nmt}(x_{\leq 2})$	Morning 10						
$M'_{nmt}(x_{\leq 3})$	At 10 a.m.						

构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
$M'_{nmt}(x_{\leq 1})$	Morning						
$M'_{nmt}(x_{\leq 2})$	Morning 10						
$M'_{nmt}(x_{\leq 3})$	At 10 a.m.						
$M'_{nmt}(x_{\leq 4})$	At 10 a.m.			me			
$M'_{nmt}(x_{\leq 5})$	At 10 a.m.			I went there			
$M'_{nmt}(x_{\leq 6})$	At 10 a.m.			I went to			

构造训练数据

- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
$M'_{nmt}(x_{\leq 1})$	Morning						
$M'_{nmt}(x_{\leq 2})$	Morning 10						
$M'_{nmt}(x_{\leq 3})$	At 10 a.m.						
$M'_{nmt}(x_{\leq 4})$	At 10 a.m.			me			
$M'_{nmt}(x_{\leq 5})$	At 10 a.m.			I went there			
$M'_{nmt}(x_{\leq 6})$	At 10 a.m.			I went to			
$M'_{nmt}(x_{\leq 7})$	At 10 a.m.			I went to			the park
Extracted MUs	shàngwǔ 10 diǎn			wǒ qùle tàng			gōngyuán

构造训练数据

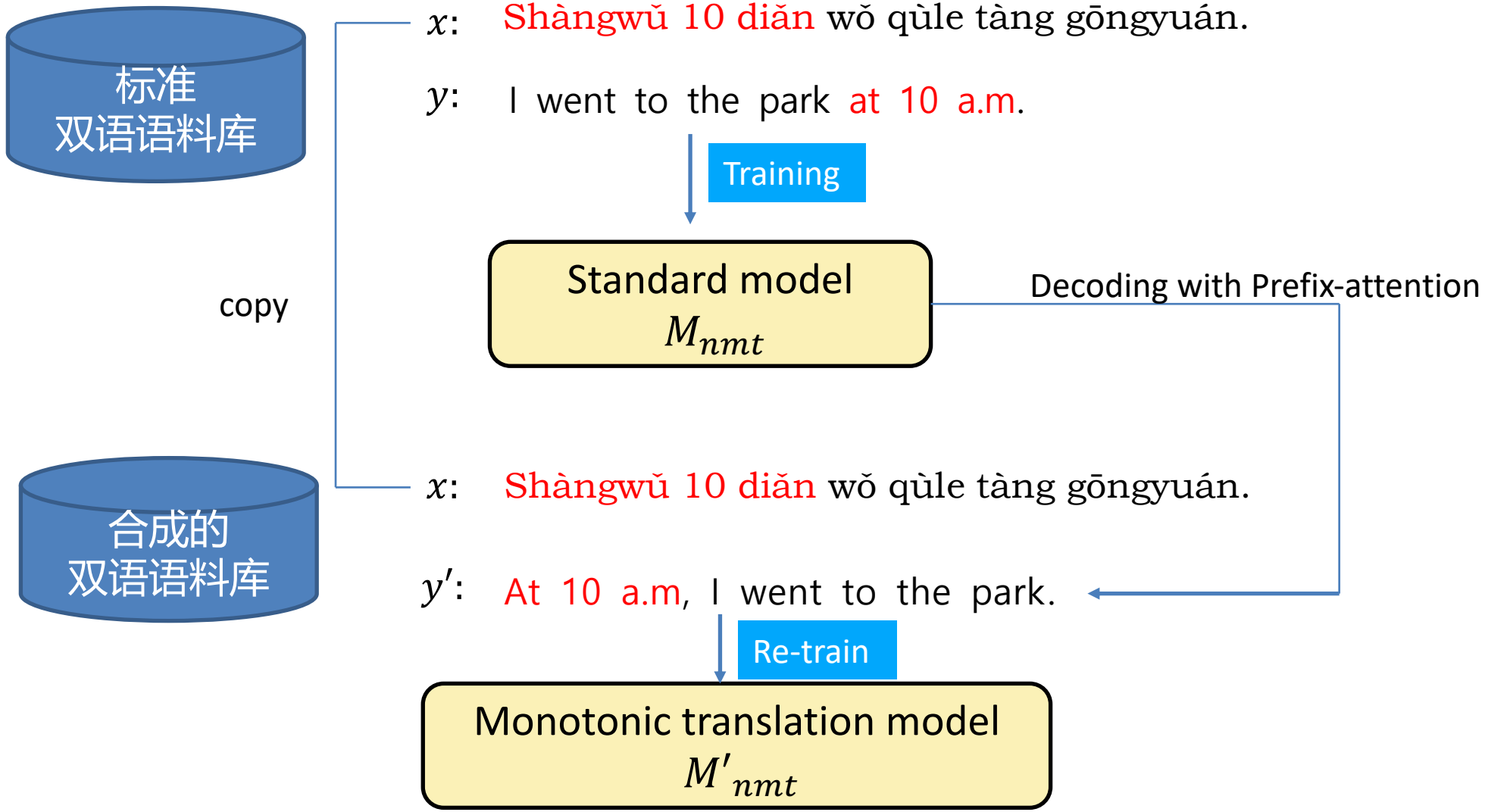
- source prefix whose translation is also a prefix of the full-sentence translation

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation	**At 10 a.m.**,			I went to			the park.
Extracted MUs	shàngwǔ 10 diǎn			wǒ qùle tàng			gōngyuán

- Long distance reorderings in full sentence translation generate long MUs

Source	**shàngwǔ**	**10**	**diǎn**	wǒ	qùle	tàng	gōngyuán
	上午	**10**	**点**	我	去了	趟	公园
full sentence translation with long reorderings	I went to			the park		at 10 a.m.	
Extracted MUs	shàngwǔ	10	diǎn	wǒ	qùle	tàng	gōngyuán

训练数据构建

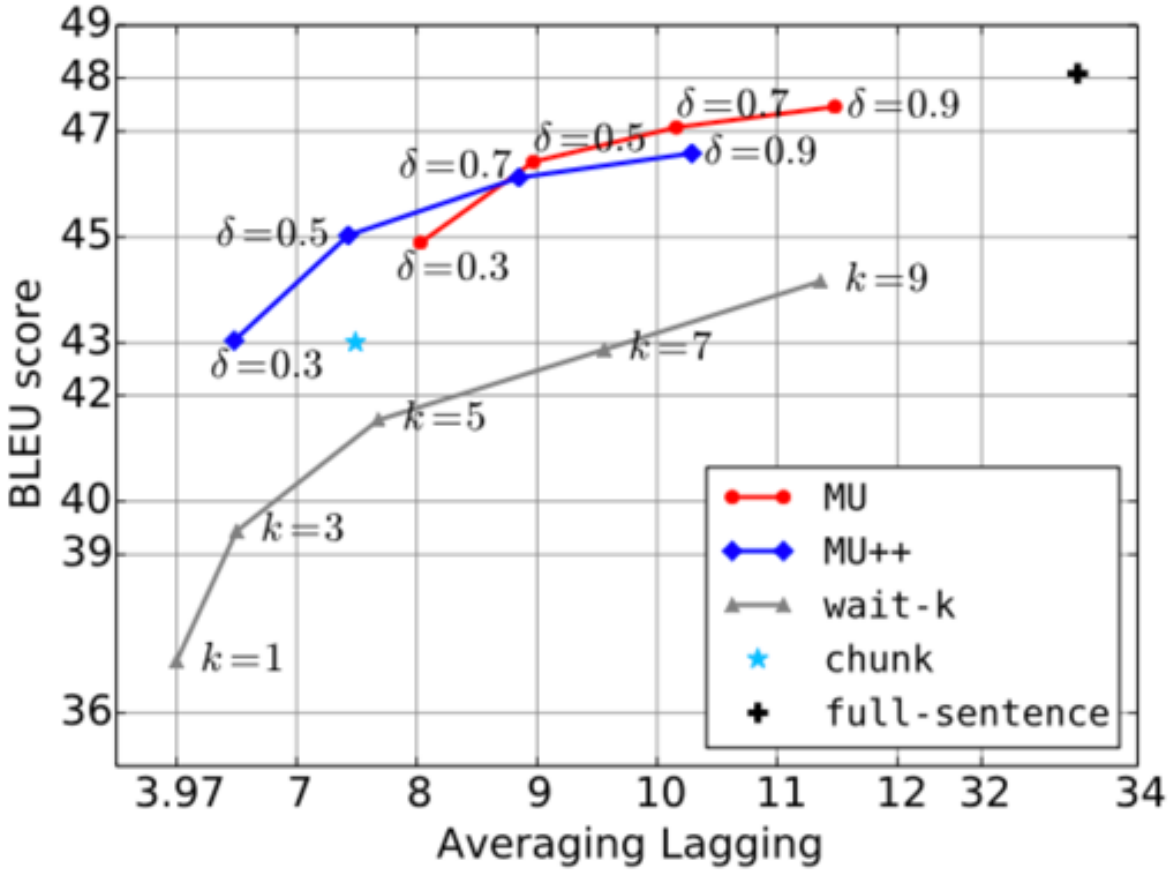

标准
双语语料库
x: Shàngwǔ 10 diǎn wǒ qùle tàng gōngyuán.
y: I went to the park at 10 a.m.
Training
Standard model
M_{nmt}
Decoding with Prefix-attention
copy
合成的
双语语料库
x: Shàngwǔ 10 diǎn wǒ qùle tàng gōngyuán.
y': At 10 a.m, I went to the park.
Re-train
Monotonic translation model
M'_{nmt}

训练数据构建

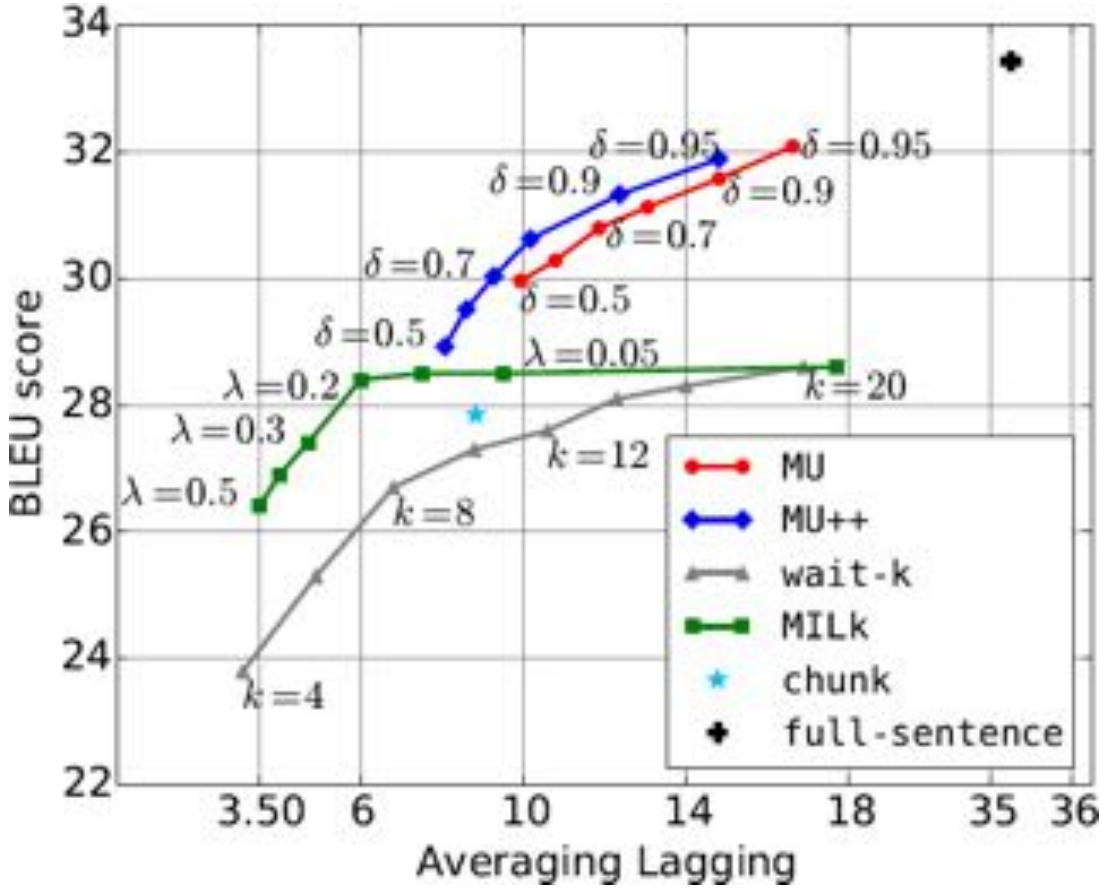
shàngwǔ 10 diǎn	wǒ qùle tàng	gōngyuán

t	history	future words	MU label
1	shàngwǔ	10 diǎn	0
2	shàngwǔ 10	diǎn wǒ	0
3	shàngwǔ 10 diǎn	wǒ qùle	1
4	shàngwǔ 10 diǎn wǒ	qùle tàng	0
5	shàngwǔ 10 diǎn wǒ qùle	tàng gōngyuán	0
6	shàngwǔ 10 diǎn wǒ qùle tàng	gōngyuán	1
7			
	...	...	...

实验结果



中英（NIST测试集）



德英（WMT测试集）

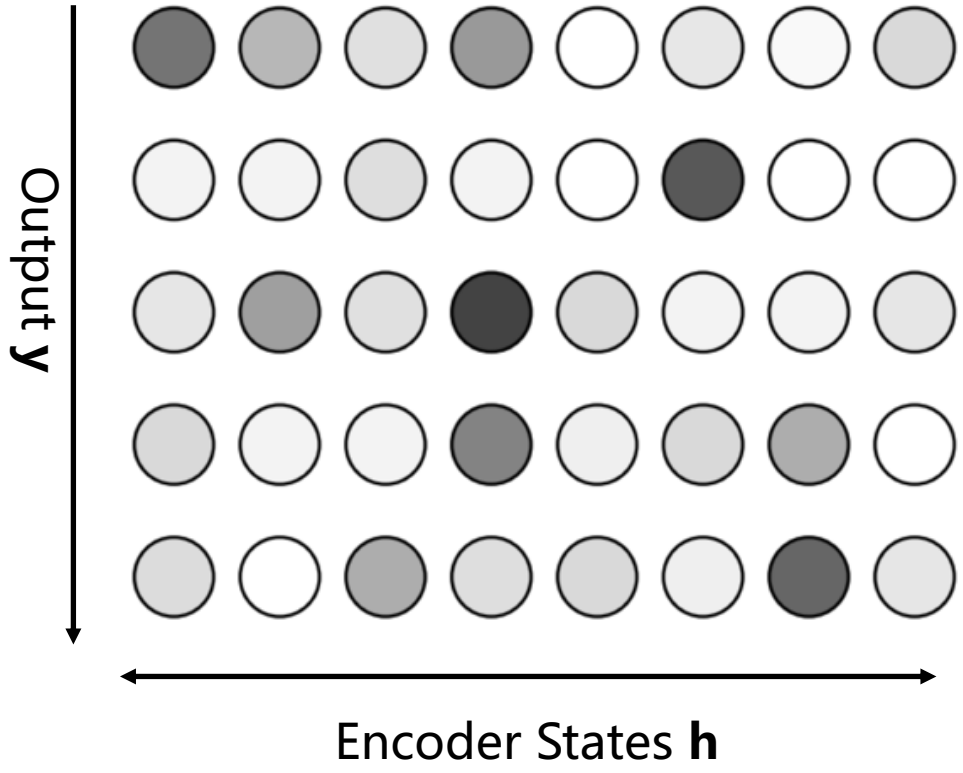
级联模型 – 同传策略

- 固定策略（Fixed Policy）：读入长度（翻译单元）固定
  - Static read-write (Dalvi et al., 2018)
  - STACL (Ma et al., 2018), etc.

- 自适应策略（Adaptive Policy）：动态调整翻译单元长度
  - Rule-based （Cho et al., 2016)
  - RL-based (Gu et al., 2017）
  - Supervised policy (Zheng et al., 2019)
  - Meaningful unit (Zhang et al., 2020)
  - MILk (Arivazhagan et al., 2019)
  - Multihead monotonic attention (Ma et al., 2020), etc.

Softmax Attention V.S. Monotonic Attention (单调注意力)

Softmax Attention



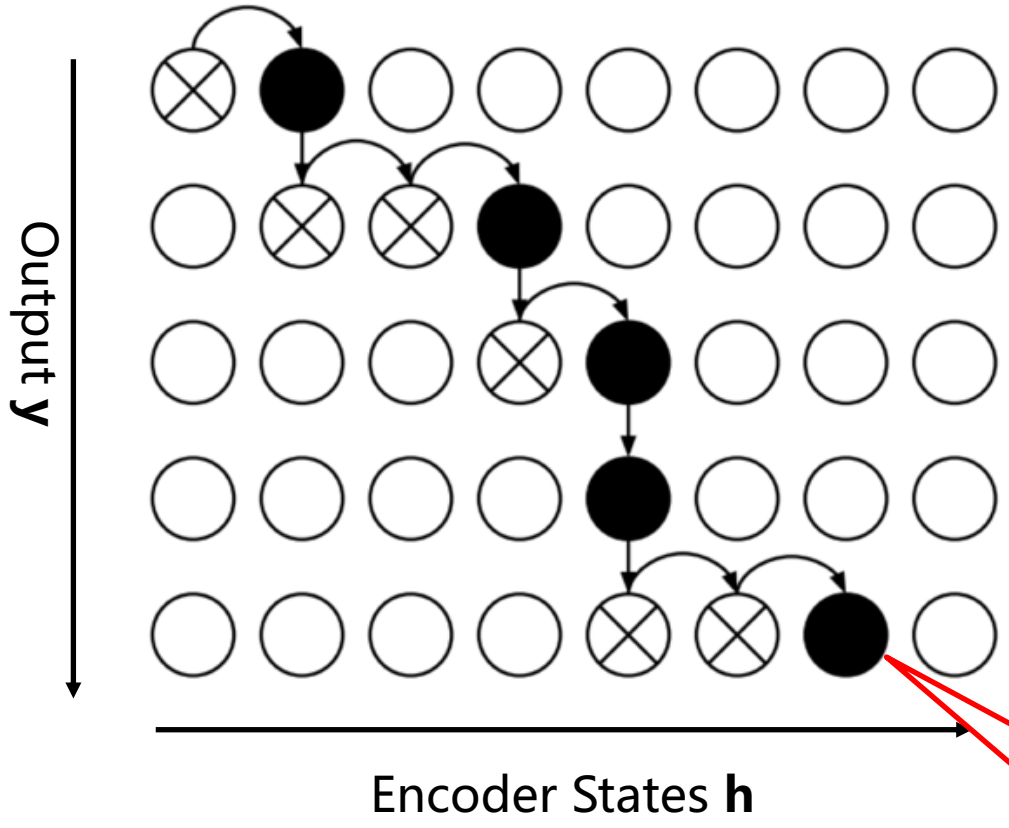
$$e_{i,j} = FeedForward(s_{i-1}, h_j)$$

$$\alpha_{i,j} = \frac{\exp(e_{i,j})}{\sum_{k=1}^{T} \exp(e_{i,k})}$$

$$c_i = \sum_{j=1}^{|\mathbf{x}|} \alpha_{i,j} h_j$$

Softmax Attention V.S. Monotonic Attention (单调注意力)

Monotonic Attention



$$e_{i,j} = a(s_{i-1}, h_j)$$

$$p_{i,j} = \sigma(e_{i,j})$$

$$z_{i,j} \sim \text{Bernoulli}(p_{i,j})$$

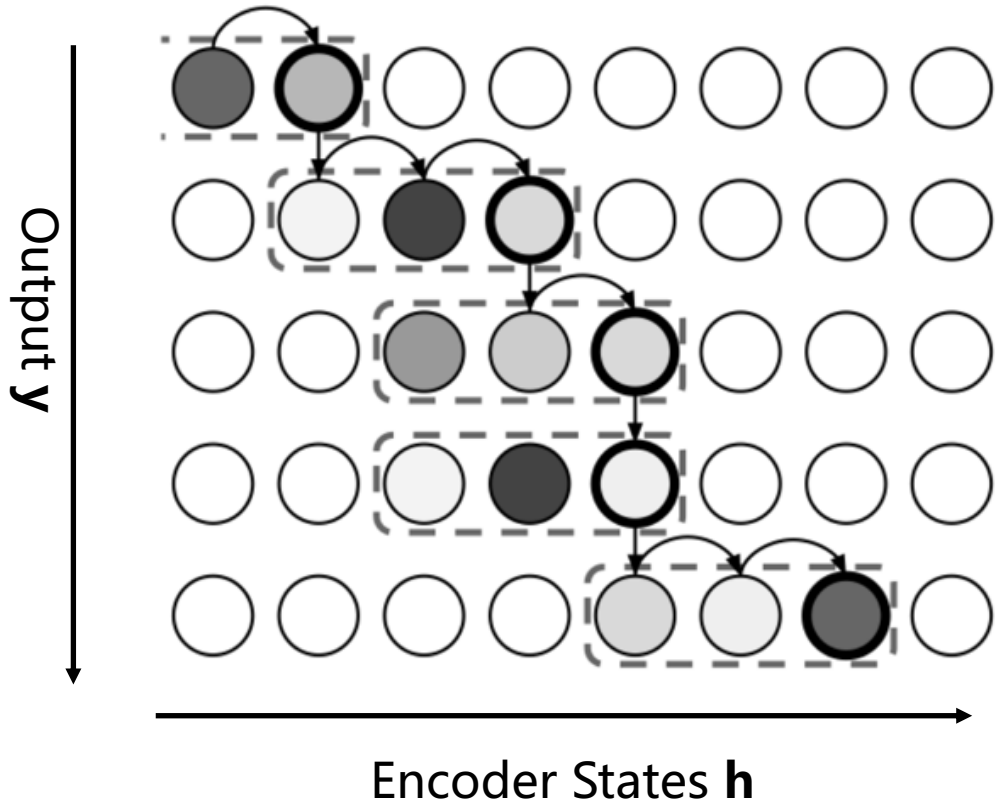
$$z_{i,j} = \begin{cases} 0 & \otimes \quad \text{Read} \\ 1 & \bullet \quad \text{Write} \end{cases}$$

$$c_i = h_j$$

Hard Attention

Raffel C, Luong MT, Liu PJ, et al. Online and linear-time attention by enforcing monotonic alignments[C]//Proceedings of the 34th International Conference on Machine Learning,2017：2837–2846.

Monotonic Chunkwise Attention



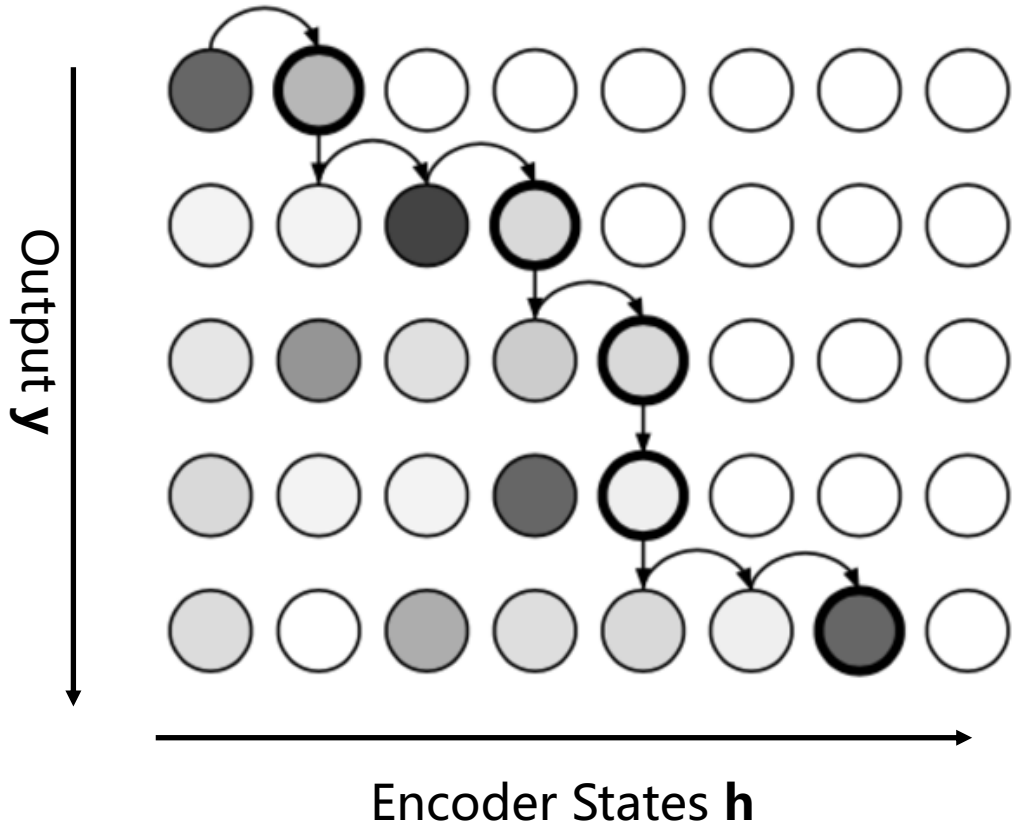
$$v = t_i - w + 1$$

$$u_{i,k} = \text{ChunkEnergy}(s_{i-1}, h_k), k \in \{v, v+1, \ldots, t_i\}$$

$$c_i = \sum_{k=v}^{t_i} \frac{\exp(u_{i,k})}{\sum_{l=v}^{t_i} \exp(u_{i,l})} h_k$$

Chiu CC， Raffel C. Monotonic chunkwise attention[C]//International Conference on Learning Representations. 2018.

Monotonic Infinite Lookback Attention



$$u_{i,k} = \mathrm{SoftmaxEnergy}(h_k, s_{i-1})$$

$$c_i = \sum_{j=1}^{t_i} \frac{\exp(u_{i,j})}{\sum_{l=1}^{t_i} \exp(u_{i,l})} h_j$$

Arivazhagan N, Cherry C, Macherey W, et al. Monotonic infinite lookback attention for simultaneous machine translation[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019:1313-1323.

提纲

- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 开放数据
- 产品及应用
- 总结及展望

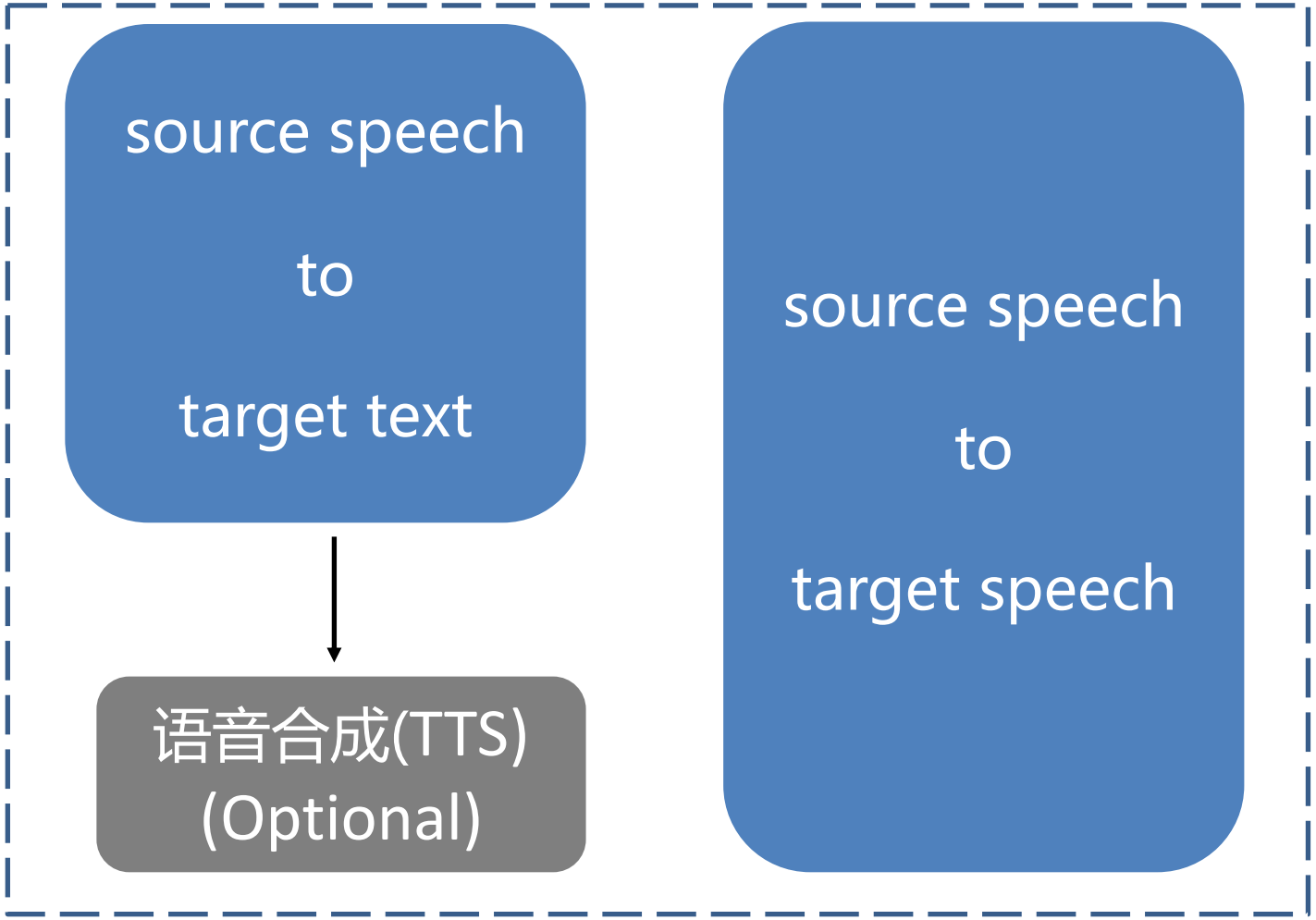
端到端模型

优点

- 缓解错误传递
- 降低时间延迟

挑战

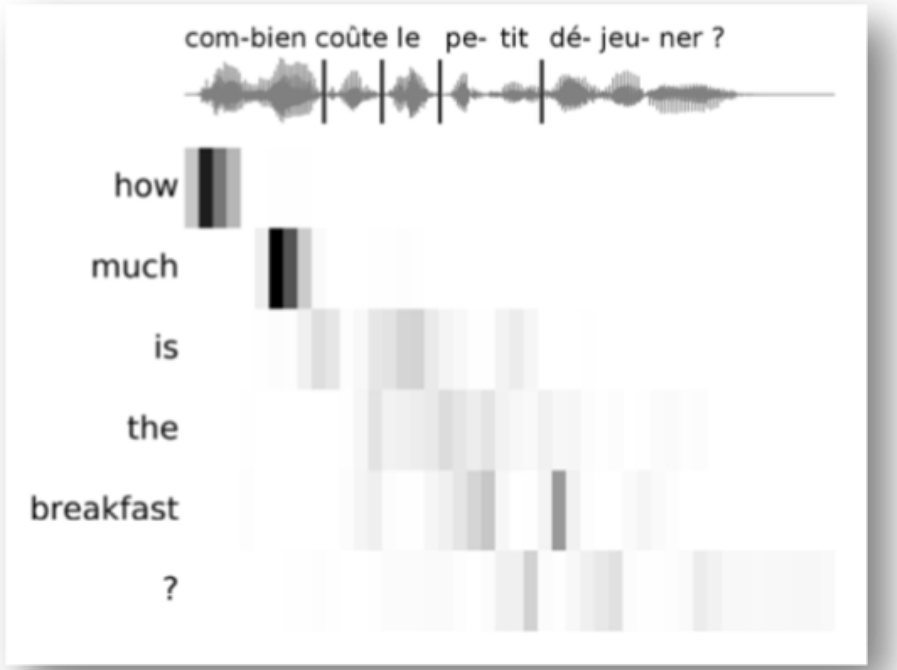
- 缺乏训练数据
- 异构知识共享

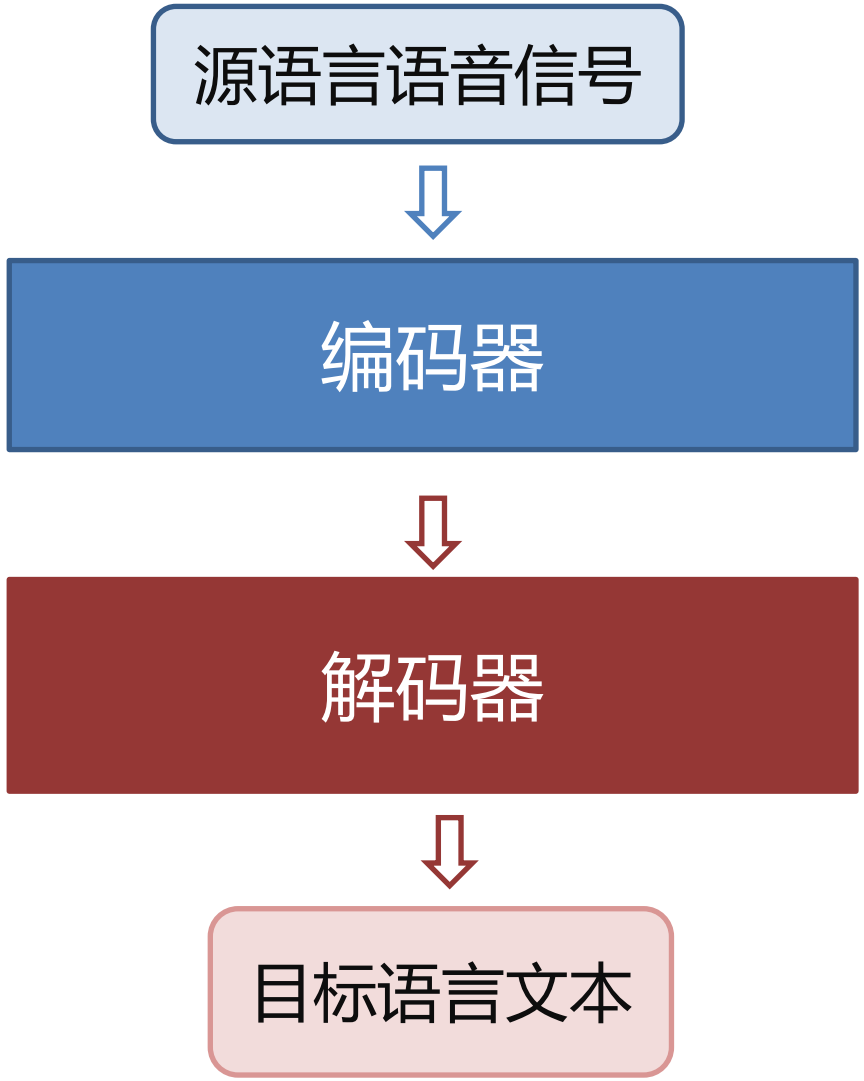


语音到文本端到端模型

French→English BTEC corpus

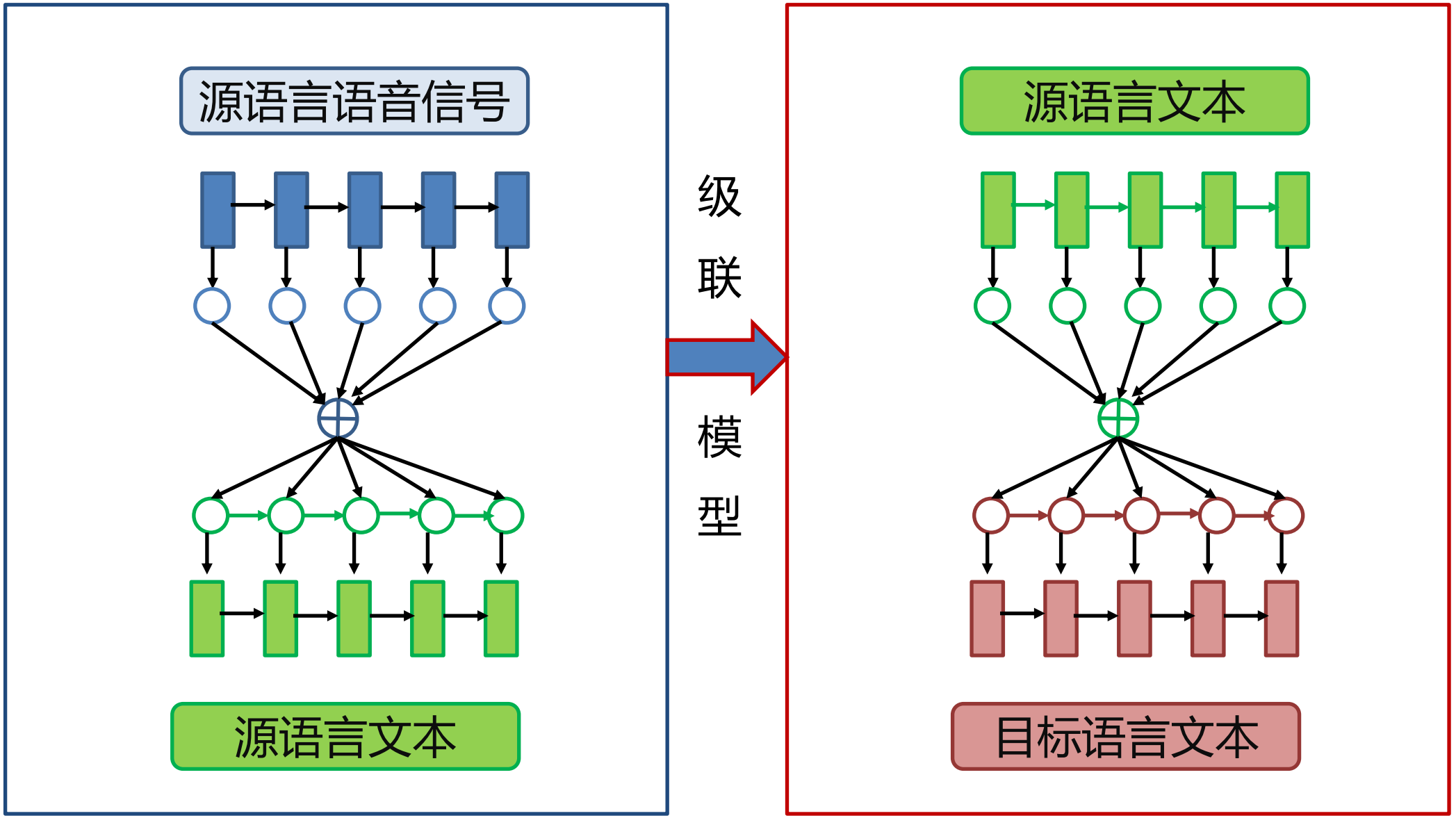
训练数据仅2万句对



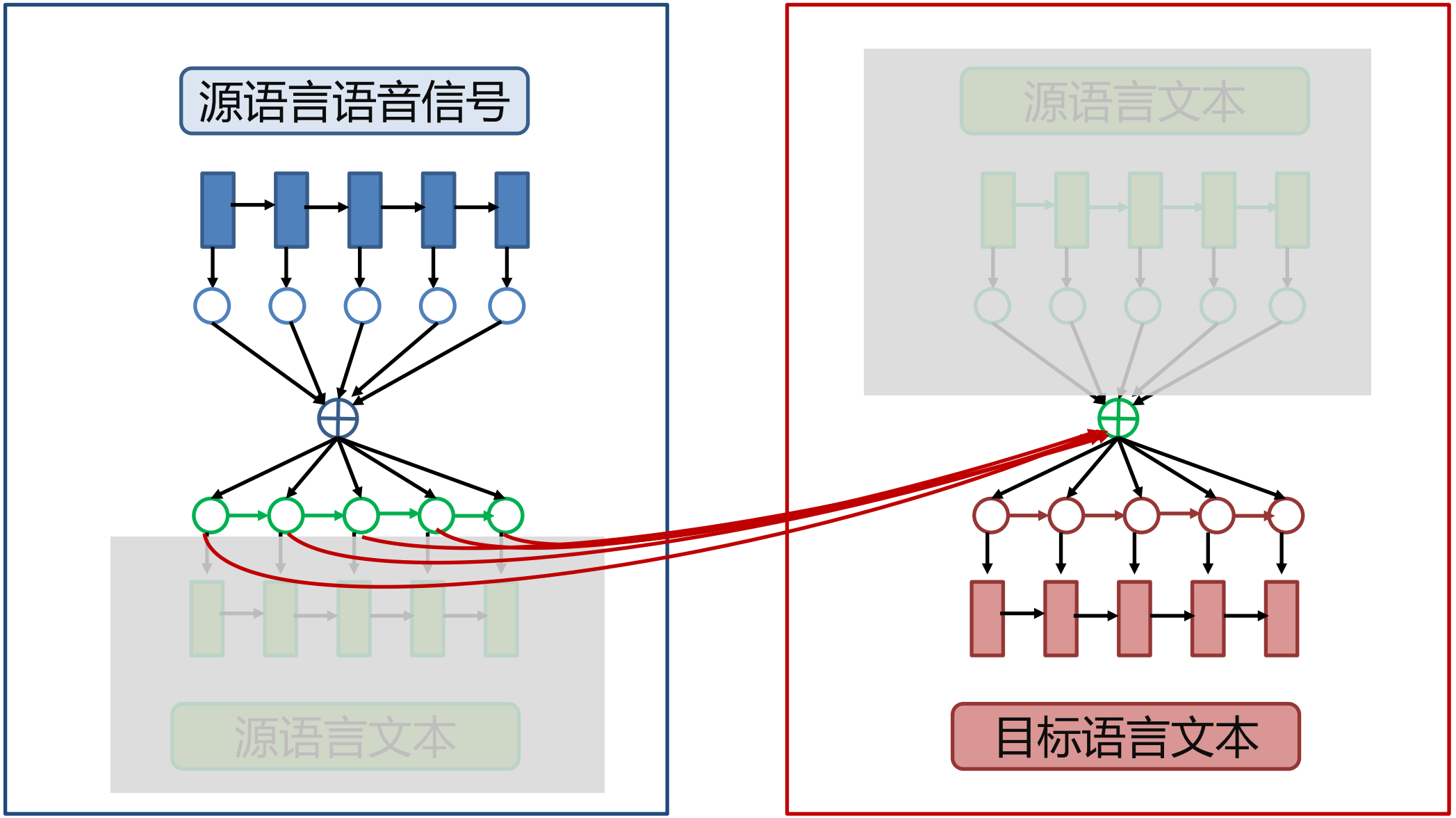


Berard A, Pietquin O, Servan C, et al. Listen and translate: a proof of concept for end-to-end speech-to-text translation[C]//Proceedings of NeurIPS Workshop on End-to-end Learning for Speech and Audio Processing, 2016.

两阶段（two-stage）模型

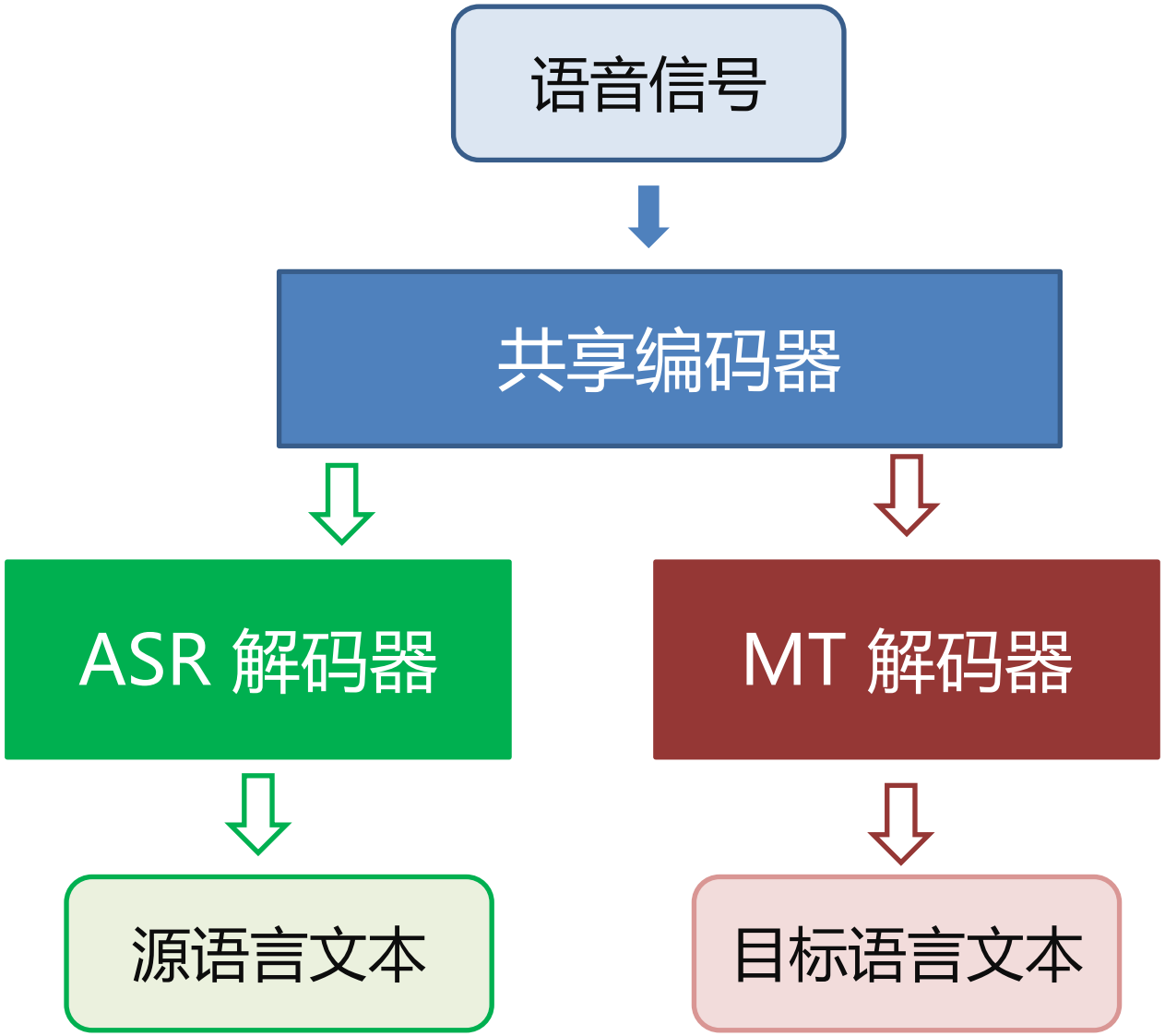


两阶段（two-stage）模型



Sperber M, Neubig G, Niehues J, et al. Attention-passing models for robust and data-efficient end-to-end speech translation[J]. Transactions of ACL, 2019, (7):313–325.

多任务学习

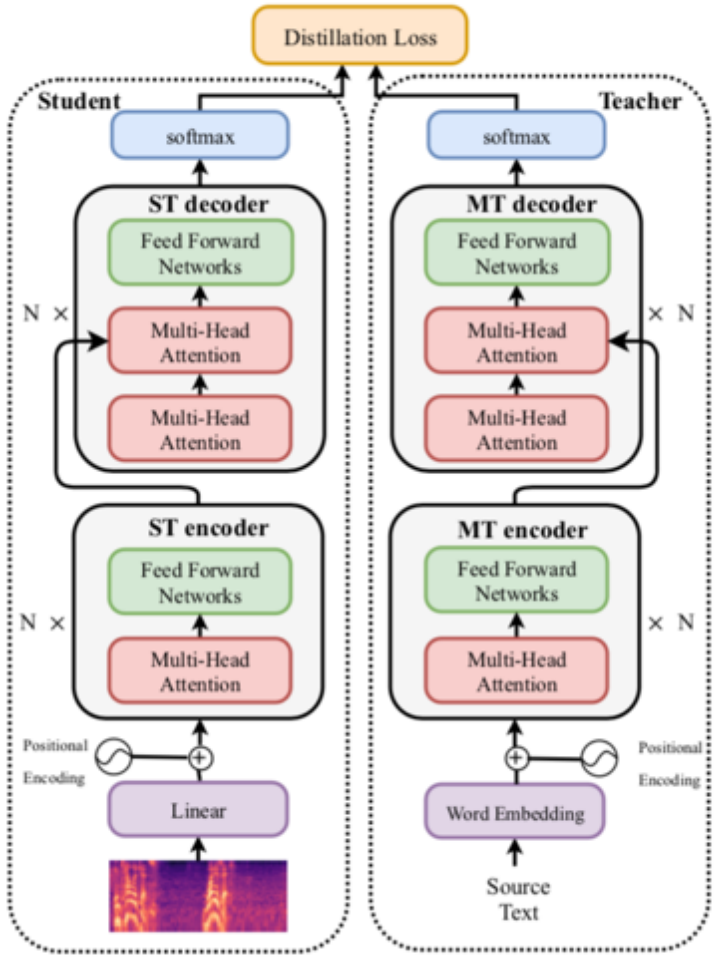


Anastasopoulos A, Chiang D. Tied multitask learning for neural speech translation[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018:82-91.

基于知识蒸馏的模型

$$L_{\mathrm{ALL}}(D;\theta;\theta_T) = (1-\lambda)L_{\mathrm{ST}}(D;\theta) + \lambda L_{\mathrm{KD}}(D;\theta,\theta_T)$$

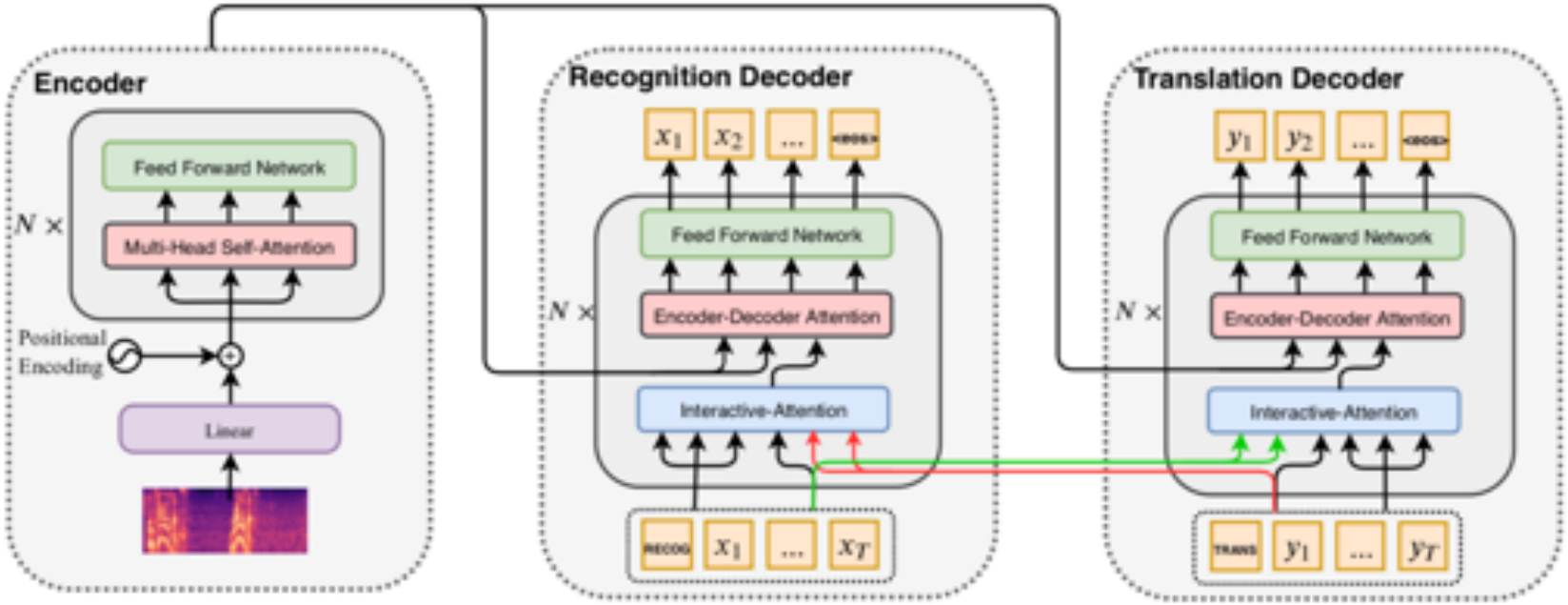
Student
Speech-Text
Translation Model



Teacher
Text-Text Translation
Model

Liu YC, Xiong H, He ZJ, et al. End-to-End Speech Translation with Knowledge Distillation[C]// Proceedings of the 20th Annual Conference of the International Speech Communication Association, 2019:1128-1132.

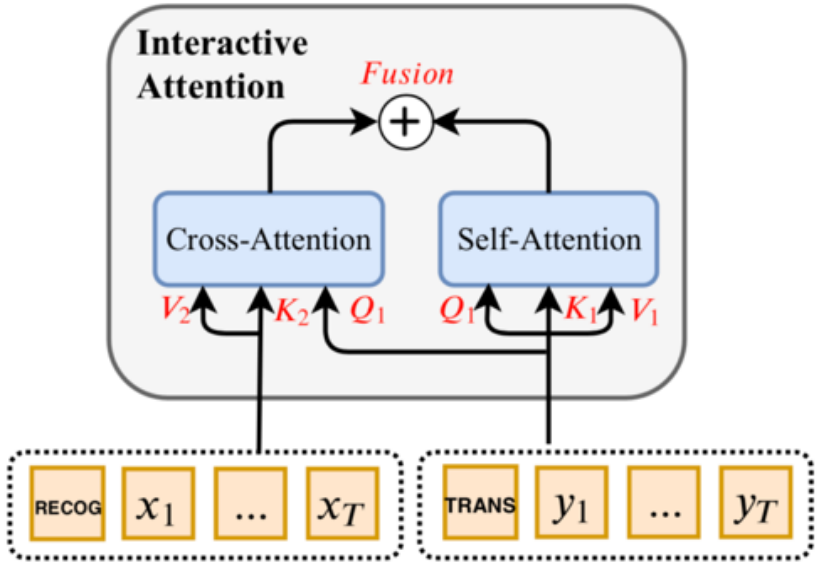
语音识别和翻译同步交互解码



$$\mathbf{H_{final}} = \mathbf{H_{self}} + \lambda * \mathbf{H_{cross}}$$

$$\mathbf{H_{self}} = \text{Attention}(\mathbf{Q_1}, \mathbf{K_1}, \mathbf{V_1})$$

$$\mathbf{H_{cross}} = \text{Attention}(\mathbf{Q_1}, \mathbf{K_2}, \mathbf{V_2})$$



实验

- Training set: TED corpus
  - Audio: English Speech
  - Translation pairs: English to German (De), French (Fr), Chinese (Zh), and Japanese (Ja)
- Dev and Test sets
  - IWSLT tst2014 / IWSLT tst2015

Model	En-De		En-Fr		En-Zh		En-Ja	
	WER(↓)	BLEU(↑)	WER(↓)	BLEU(↑)	WER(↓)	BLEU(↑)	WER(↓)	BLEU(↑)
Text MT	/	22.19	/	30.68	/	25.01	/	22.93
Pipeline	16.19	19.50	14.20	26.62	14.20	21.52	14.21	**20.87**
E2E	16.19	16.07	14.20	27.63	14.20	19.15	14.21	16.59
Multi-task	15.20	18.08	13.04	28.71	13.43	20.60	14.01	18.73
Two-stage	15.18	19.08	13.34	**30.08**	13.55	20.99	14.12	19.32
Interactive	**14.76**	**19.82**	**12.58**	29.79	**13.38**	**21.68**	**13.91**	20.06

提纲

- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 数据、鲁棒性模型及评价方法
- 产品及应用
- 总结及展望

开放数据集

同传数据集

Speech Translation	Languages	Hours
F-C (2013)	Es→En	38
KIT-Disfluency (2014)	De→En	13
BTEC (2016)	En→Fr	17
MSLT V1.0 (2016)	En↔Fr/De	23
	En→Zh/Jp	6
MSLT V1.1 (2017)	Zh→En	5
	Jp →En	9
Travel (2017)	Am→En	8
Aug-LibriSpeech (2018)	En→Fr	236
MuST-C (2019)	En→8 Euro langs	3617
Europarl-ST (2020)	9 Euro langs	1642
Covost (2020a; 2020b)	En↔21 langs	2880
Simultaneous Translation	**Languages**	**Hours**
CIAIR (2004)	En↔Jp	182
EPPS (2009)	En↔Es	217
Simul-Trans (2014)	En↔Jp	22
BSTC (ours)	Zh→En	68

Ruiqing Zhang, Xiyang Wang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Zhi Li, Haifeng Wang, Ying Chen, Qinfei Li. BSTC: A Large-Scale Chinese-English Speech Translation Dataset. In 2nd workshop on AutoSimulTrans

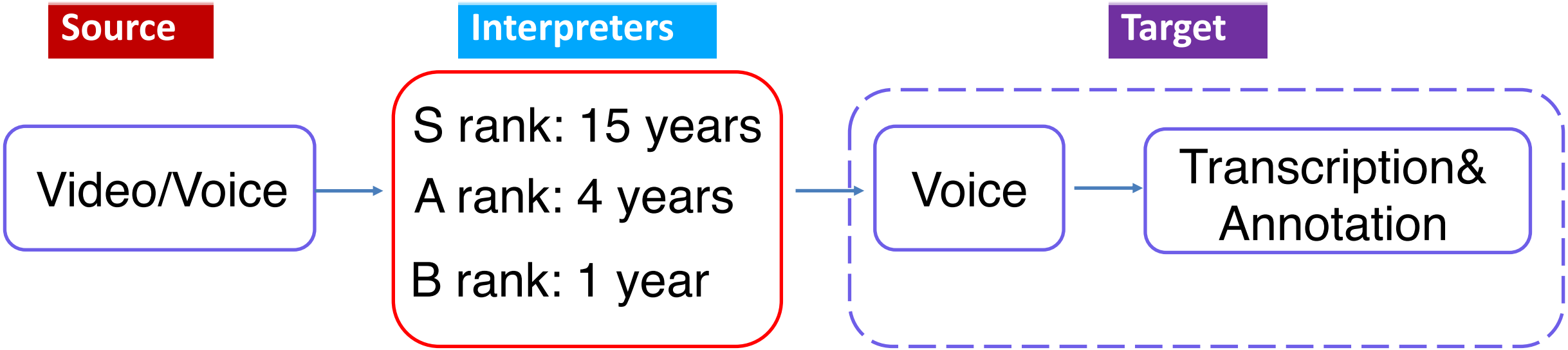
NAIST Corpus

Language	English-Japanese / Japanese-English
Domain	Academic Lectures,, News, General
Source Lang. Material	TED, CNN, CSJ(corpus of Spontaneous Japanese), NHK
Total Words	22 hours (387K words of transcribed data)
Link:	https://ahcweb01.naist.jp/resource/stc/

NAIST: Nara Institute of Science and Technology [Shimizu et al., 2014]

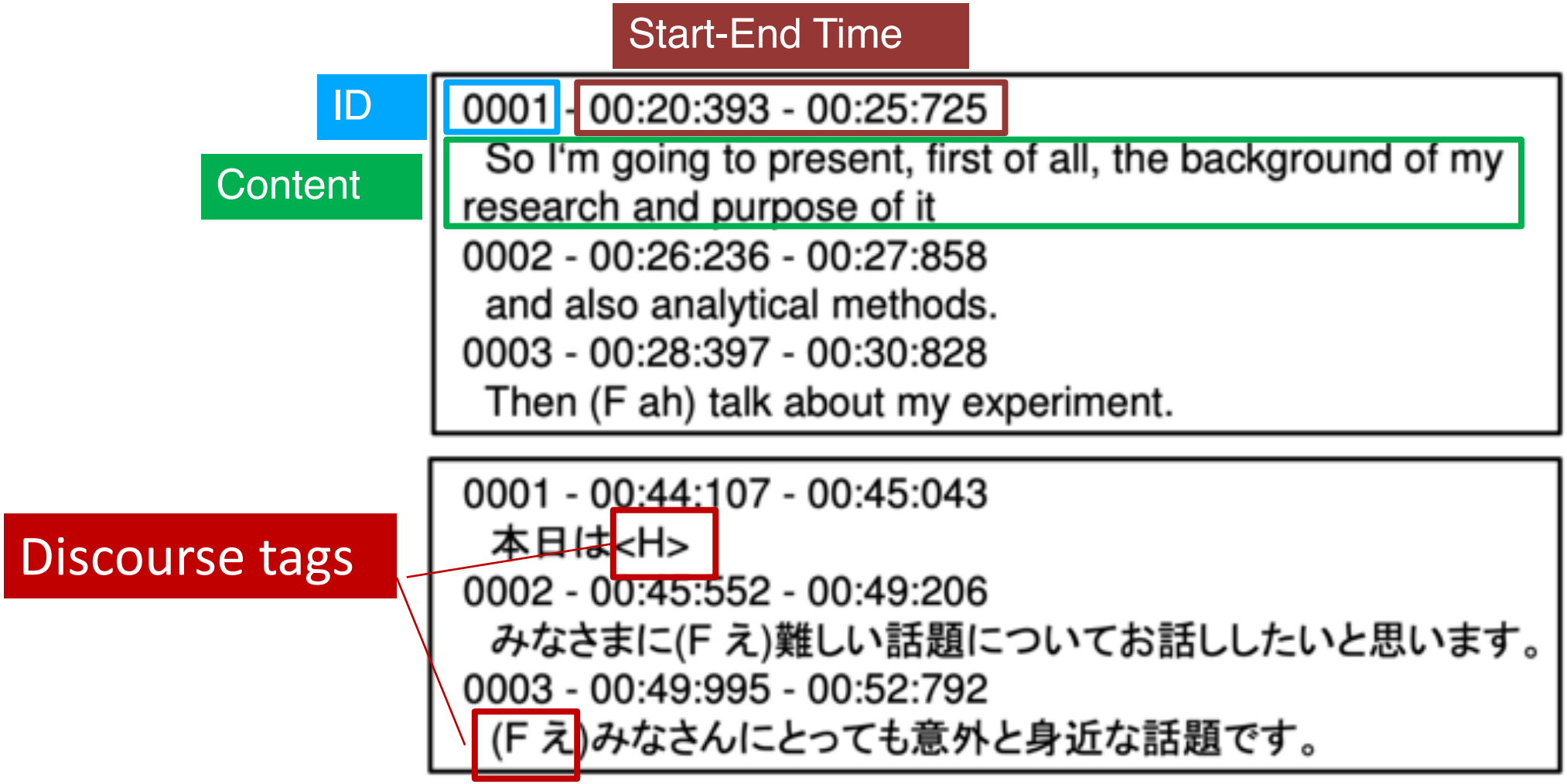
NAIST Corpus

Data	Domain	Format	Lang	Number	Minutes (avg.)	Words (avg.)
TED (S rank)	Lectures	Video	English	46	558 (12.1)	98,034 (2,131)
TED (A, B rank)	Lectures	Video	English	34	415 (12.2)	70,228 (2,066)
CSJ	Lectures	Voice	Japanese	30	326 (10.9)	85,042 (2,835)
CNN	News	Voice	English	8	27 (3.4)	4,639 (580)
NHK	News	Voice	Japanese	10	16 (1.8)	4,121 (412)



NAIST Corpus

Example of a transcript in English and Japanese



NAIST Corpus

Example of comparing the translation and simultaneous interpretation

Source	but this understates the seriousness of this particular problem because it doesn't show the thickness of the ice	
Reference (translator)	しかし / もっと深刻な / 問題 / というのは / 実は / 氷河の厚さなのです *but / more serious / problem / is / in fact / the thickness of the ice*	
Reference (S rank)	しかし / これ本当は / もっと深刻で / 氷の厚さまでは / 見せてないんですね *but / this is really / more serious and / the thickness of the ice / it isn't shown*	15 years
Reference (A rank)	この / 本当に / 問題に / なっているのは / 氷の厚さです *this / real / problem / becoming is / the thickness of the ice*	4 years
Reference (B rank)	この / 問題は *this / problem is*	1 year

CIAIR Simultaneous Interpretation Corpus

Language	English Japanese
Domain	Monologue Speech: economics, history, culture, etc. Dialogue Speech: Travel conversation (airports and hotels)
Total Length	Monologue Speech (Speaker): 21.5 hours Dialogue Speech (Speaker): 56 hours
Link:	http://shachi.org/resources/3270

CIAIR: Center for Integrated Acoustic Information Research of Nagoya University [Tohyama et al., 2004]

CIAIR Simultaneous Interpretation Corpus

Time 05’106’’-09’158’’ 09’558’’-13’654’’

Lecture’s Utterance	The theme for this speech is going to be			east coast America versus West coast America		
Interpreter’s utterance		次の	テーマですが	これが	アメリカの東海岸対	西海岸

Time

06’232’’-07’103’’ 07’344’’-08’387’’ 09’048’’-09’555’’ 10’199’’-12’568’’ 12’900’’-14’471’’

CIAIR Simultaneous Interpretation Corpus

Monologue Speech		No. of words	No. of utterance	Recording time (min)
Speaker	English	90249	8422	695
	Japanese	84278	6529	597
	Total	174527	14951	1292
Interpreter	E-J	266050	25507	1639
	J-E	127991	16083	1255
	Total	394041	41590	2904
Sum Total		568568	56541	4196

CIAIR: Center for Integrated Acoustic Information Research of Nagoya University

CIAIR Simultaneous Interpretation Corpus

Dialogue Speech		No. of words	No. of utterance	Recording time (min)
Speaker	English	107850	14223	1678
	Japanese	106258	16485	1678
	Total	214108	30708	3356
Interpreter	E-J	116776	15286	1678
	J-E	91743	13719	1678
	Total	208519	29005	3356
Sum Total		422627	59713	6712

CIAIR: Center for Integrated Acoustic Information Research of Nagoya University

EPIC: European Parliament Interpreting Corpus

Language	English, Italian, Spanish
Domain	public domain
Source Lang. Material	Europe by Satellite (EbS) TV channel
Total Words	357 speeches (18 hours, 177K words)
Link:	https://corpora.dipintra.it

[Claudio Bendazzoli & Annalisa Sandrelli, 2005]

EPIC: European Parliament Interpreting Corpus

Original Speeches (en, it, es)

Simultaneously Interpreted Speeches

sub-corpus	n. of speeches	total word count	% of EPIC
Org-en	81	42705	24%
Org-it	17	6765	3.8%
Org-es	21	14468	8.2%
Int-it-en	17	6708	3.8%
Int-es-en	21	12995	7.3%
Int-en-it	81	35765	20.1%
Int-es-it	21	12833	7.2%
Int-en-es	81	38435	21.6%
Int-it-es	17	7073	4%
TOTAL	**357**	**177748**	**100%**

[Claudio Bendazzoli & Annalisa Sandrelli, 2005]

MuST-C: a Multilingual Speech Translation Corpus

Language	English – De, Es, Fr, It, Nl, Pt, Ro, Ru
Domain	public domain, business, science, entertainment, etc.
Source Lang. Material	TED Talks
Total Words	385 ~ 504 hours per language
Link:	https://ict.fbk.eu/must-c/

Gangi, Roldano Cattoni, Luisa Bentivogli, Matteo Negri and Marco Turchi. MuST-C: a Multilingual Speech Translation Corpus, NAACL-HLT 2019

MuST-C: a Multilingual Speech Translation Corpus

Tgt	#Talk	#Sent	Hours	src w	tgt w
De	2,093	234K	408	4.3M	4.0M
Es	2,564	270K	504	5.3M	5.1M
Fr	2,510	280K	492	5.2M	5.4M
It	2,374	258K	465	4.9M	4.6M
Nl	2,267	253K	442	4.7M	4.3M
Pt	2,050	211K	385	4.0M	3.8M
Ro	2,216	240K	432	4.6M	4.3M
Ru	2,498	270K	489	5.1M	4.3M

Gangi, Roldano Cattoni, Luisa Bentivogli, Matteo Negri and Marco Turchi. MuST-C: a Multilingual Speech Translation Corpus, NAACL-HLT 2019

BSTC: Baidu Simultaneous Translation Corpus

Language	Chinese-English
Domain	science, technology, economy, culture, art, etc.
Source Lang. Material	Chinese talks
Total Words	68 hours (237 talks)
Link:	https://ai.baidu.com/broad/introduction?dataset=bstc

BSTC: Baidu Simultaneous Translation Corpus

	talks	sentences	Characters / words		Hours
			Chinese	English	
Training set	215	37901	1,028,538	524,395	64.71
Dev set	16	956	26,059	13,277	1.58
Test set	6	975	25,832	12,724	1.46

BSTC: Baidu Simultaneous Translation Corpus

Training samples

Field	Content
Audio	
ASR	那么我们今天呢，就希望从一个20年的AI工作者来说，如何从专业的角度去解读一下，我们现在究竟发生了什么事情？他的权势金生。
Transcript	那么我们今天呢就希望，从一个二十年的AI工作者来说，如何从专业的角度去解读一下，我们现在究竟发生了什么事情，它的前世今生。
Translation	So today, as one who has been working on AI for twenty years, I wish I could give you a professional interpretation of what exactly is going on, its origin, history, characteristic, and where it is going.

BSTC: Baidu Simultaneous Translation Corpus

Test Set: 3 interpreters to interpret 6 lectures,
simulating real interpreting scenario

Lectures ID	Domain	Length
1	Art	15‘
2	AI	15‘
3	Art	19’
4	Story	11’
5	Big Data	14’
6	AI	10’

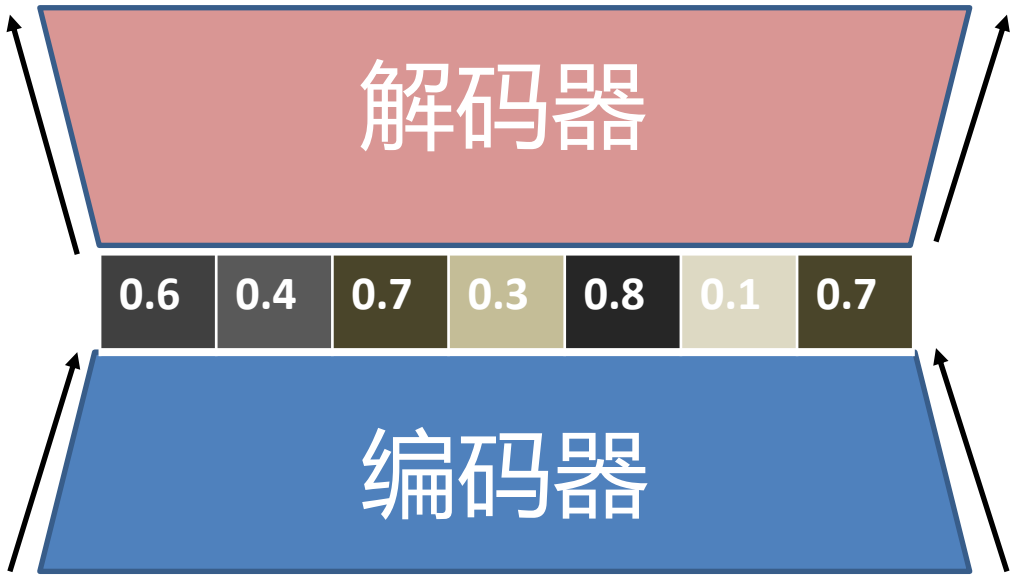
语音容错翻译

词语、音节联合建模

Let’ s meet at 6 o'clock in the hotel DaTang

词语、音节联合建模

Let’ s meet at 6 o'clock in the hotel lobby



文本向量　音节向量

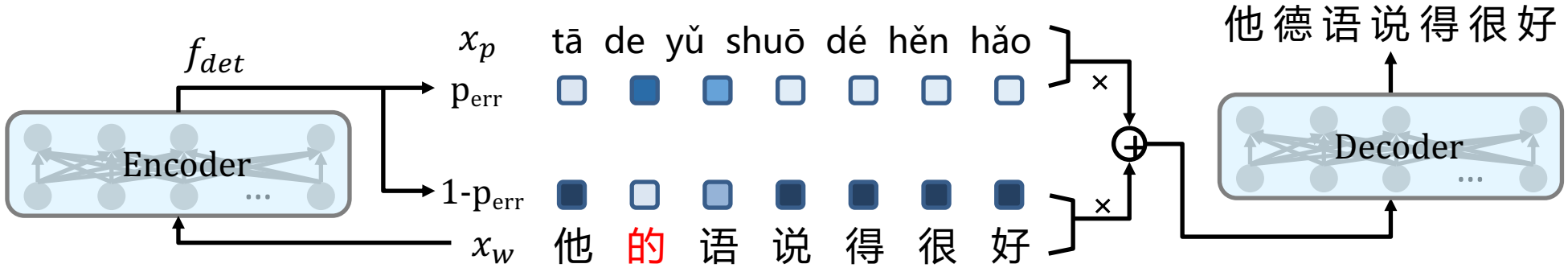
$$\pi([a, \psi(a)]) = (1-\beta) * \pi(a) + \beta * \pi(\psi(a))$$

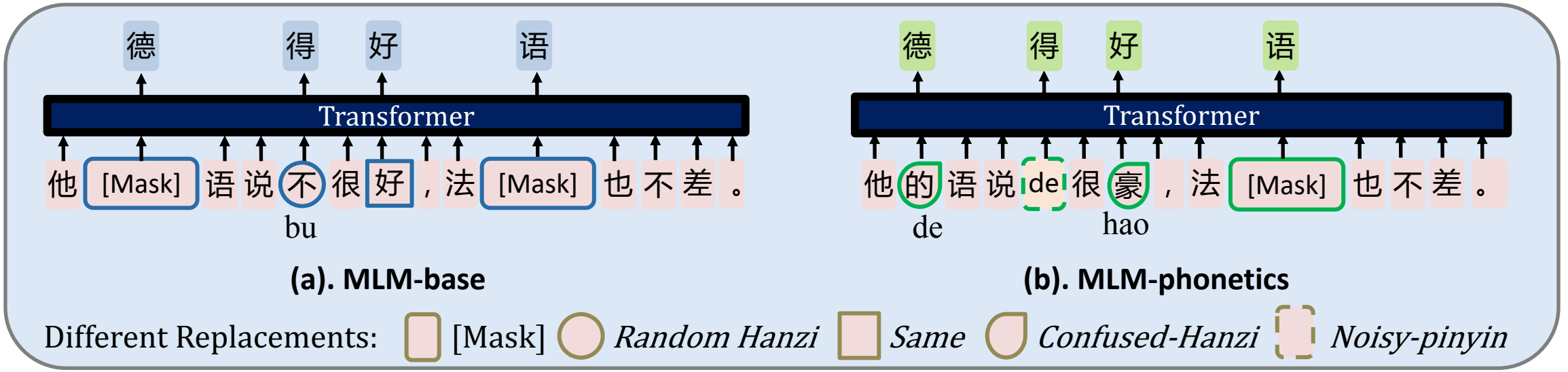
六点　在　酒店　大唐　见面

liudian zai　jiudian　datang　jianmian

Liu et al., Robust Neural Machine Translation with Joint Textual and Phonetic Embedding, ACL 2019

语音识别纠错

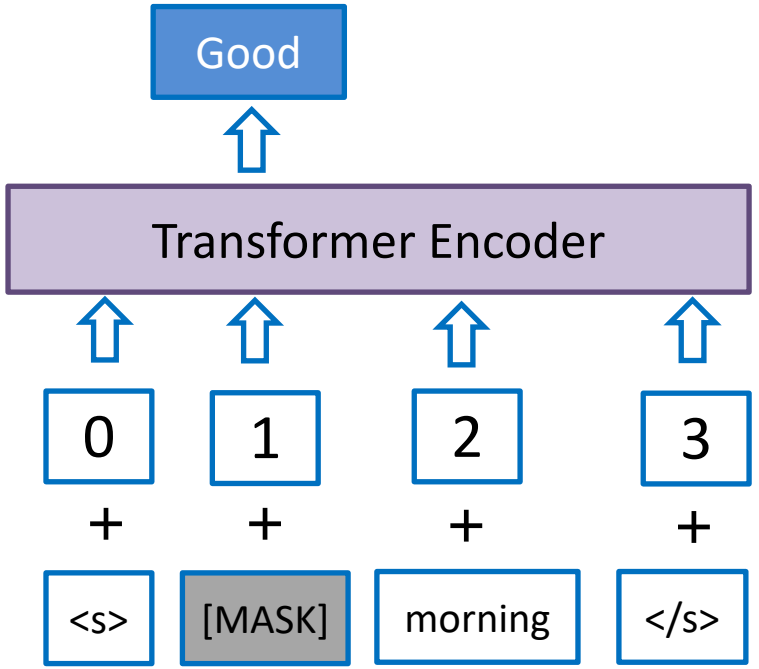




Zhang et al., Correcting Chinese Spelling Errors with Phonetic Pre-training, Findings of ACL 2021

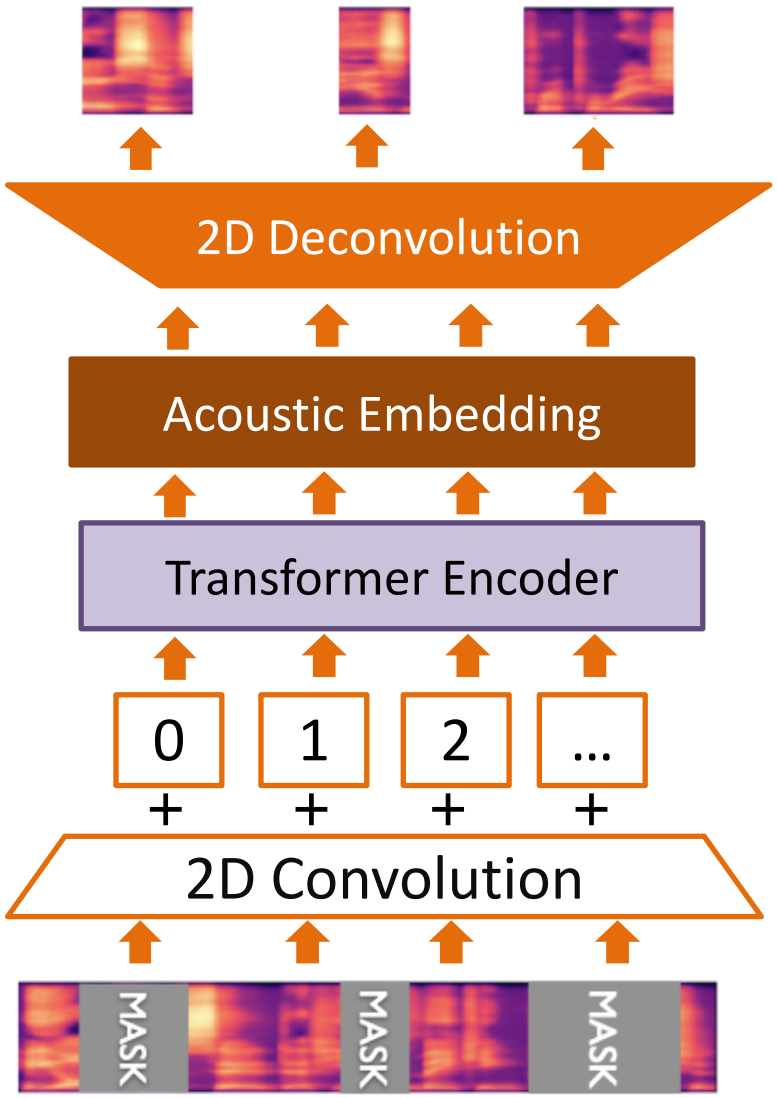
多模态信息融合

语音、语言联合预训练模型

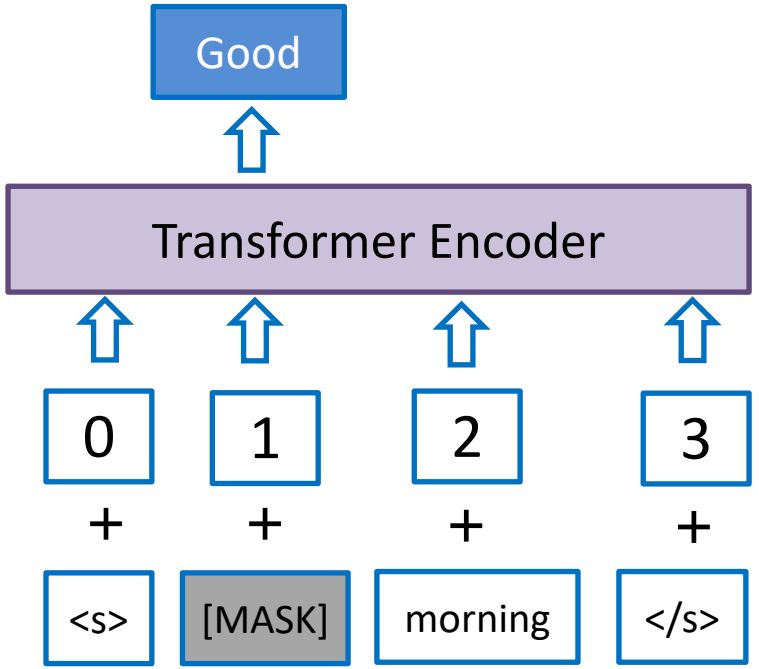


Masked Language Model

语音、语言联合预训练模型

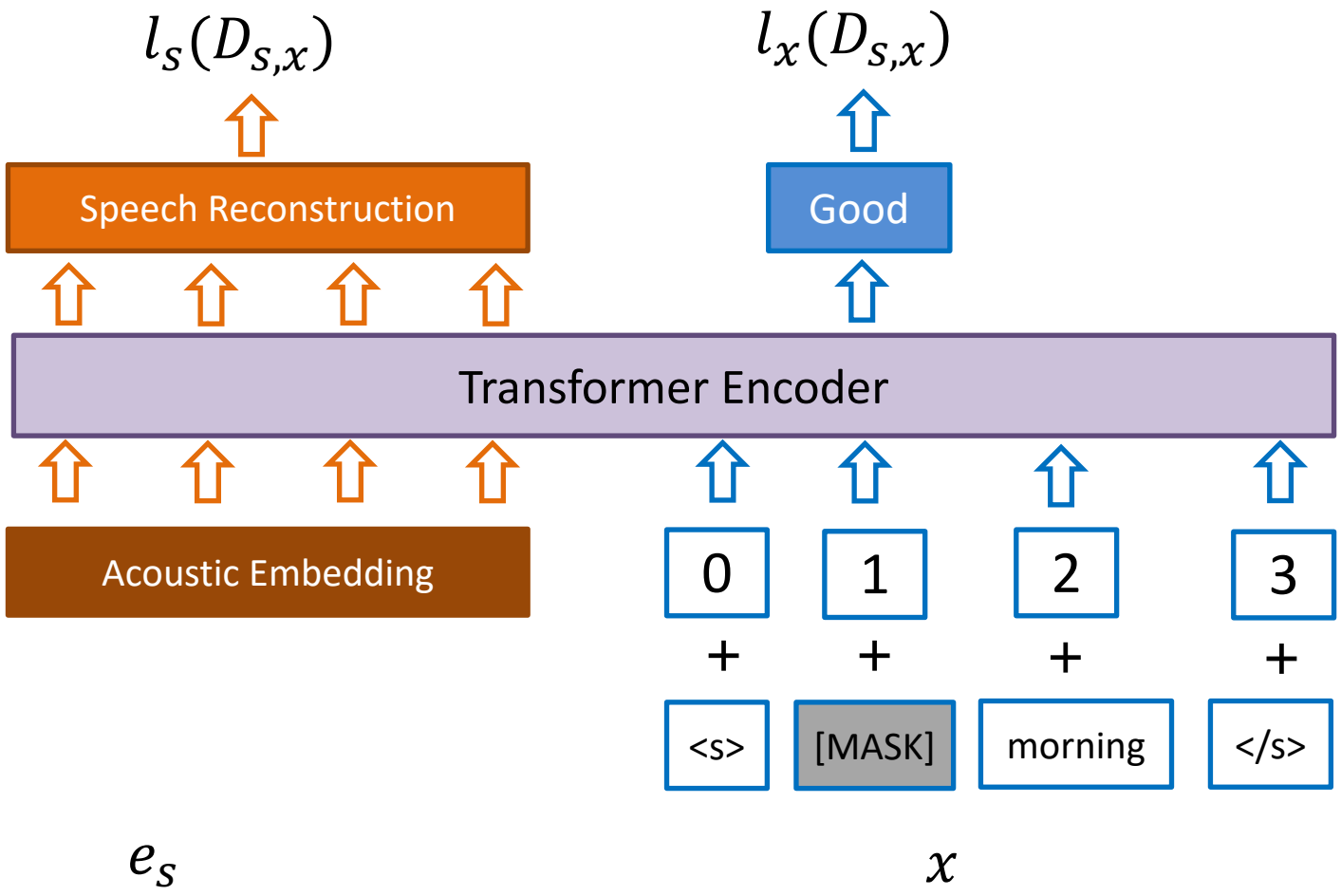


Masked Acoustic Model



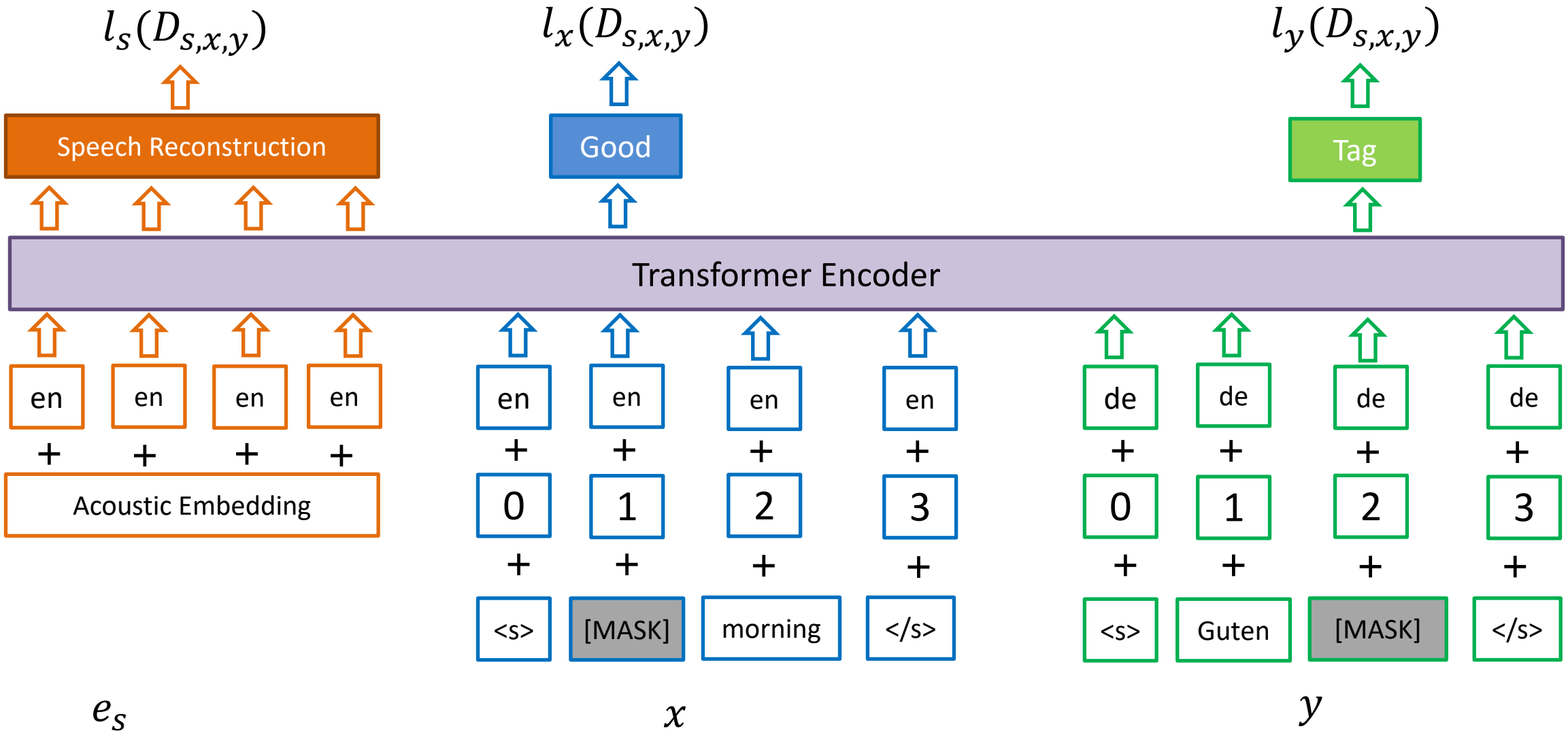
Masked Language Model

语音、语言联合预训练模型



Fused Acoustic and Text Masked Language Model (FAT-MLM)

语音、语言联合预训练模型

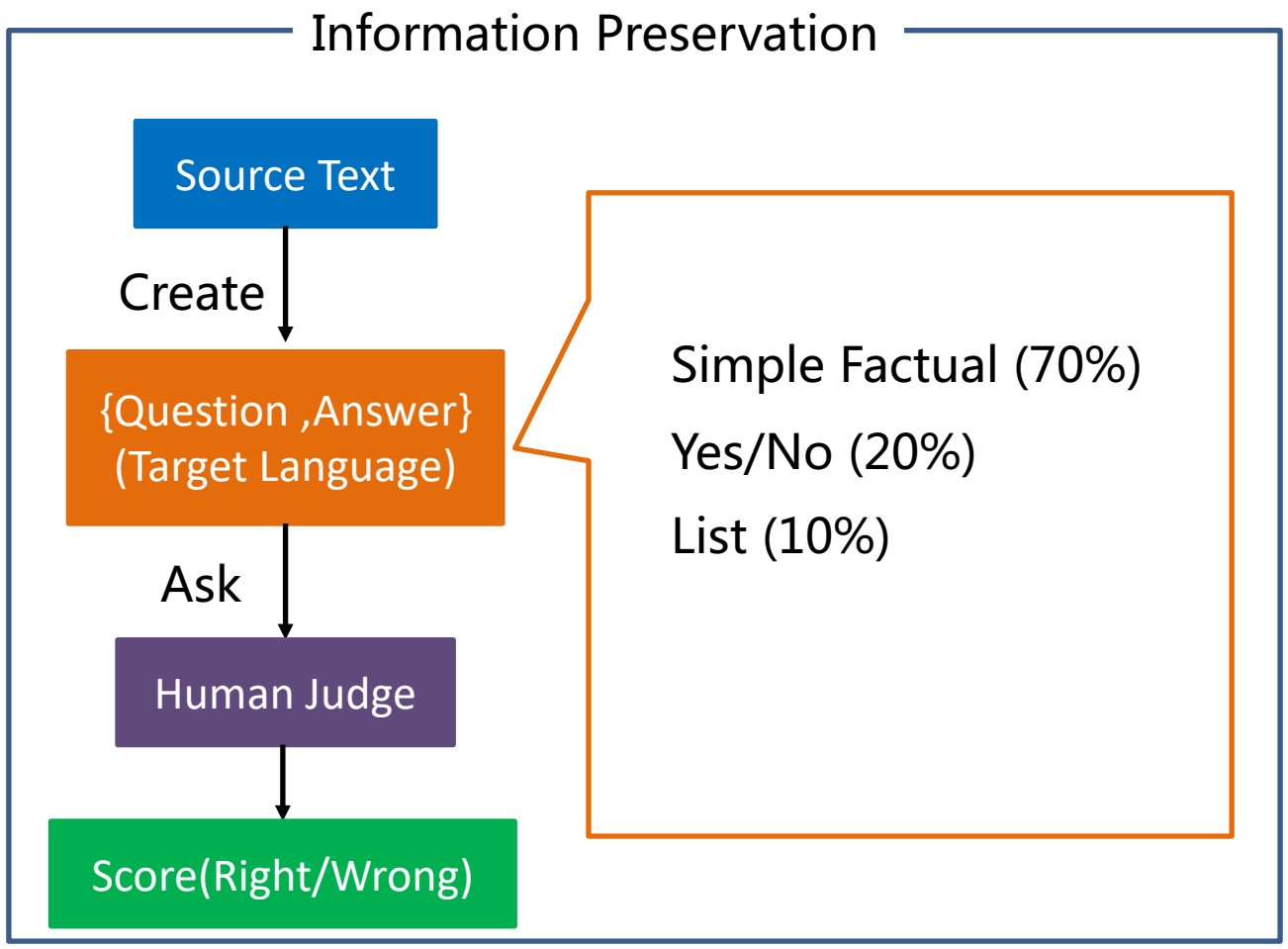


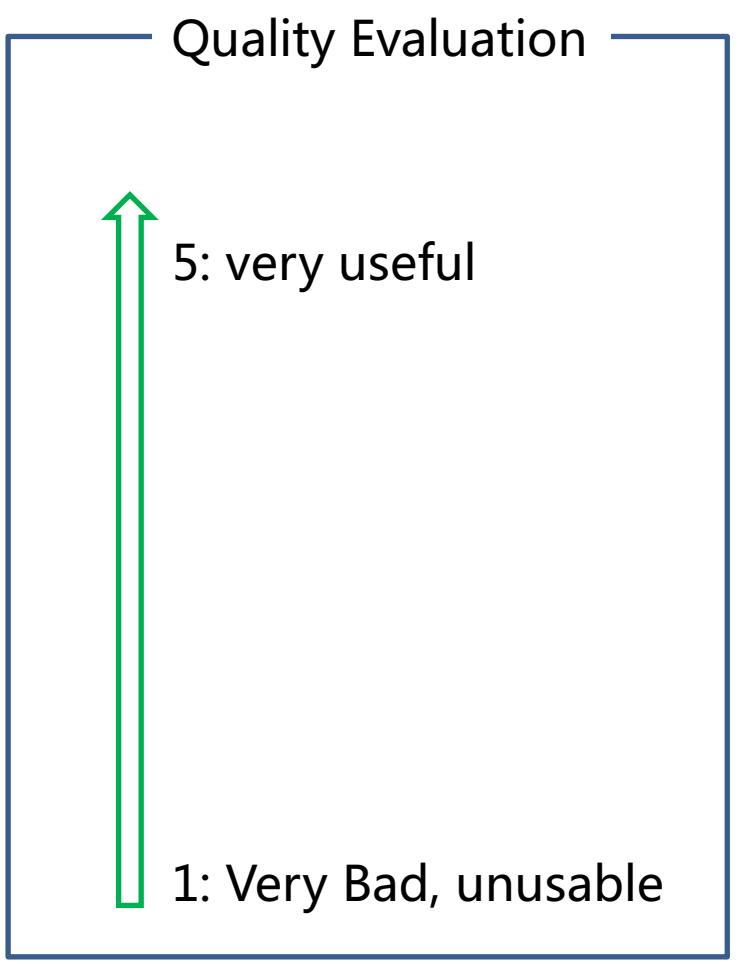
Fused Acoustic and Text Speech Translation (FAT-ST)

同传评价

基于阅读理解的端到端评价

针对同传场景，问题和答案设计较为关键，人力成本高、周期长

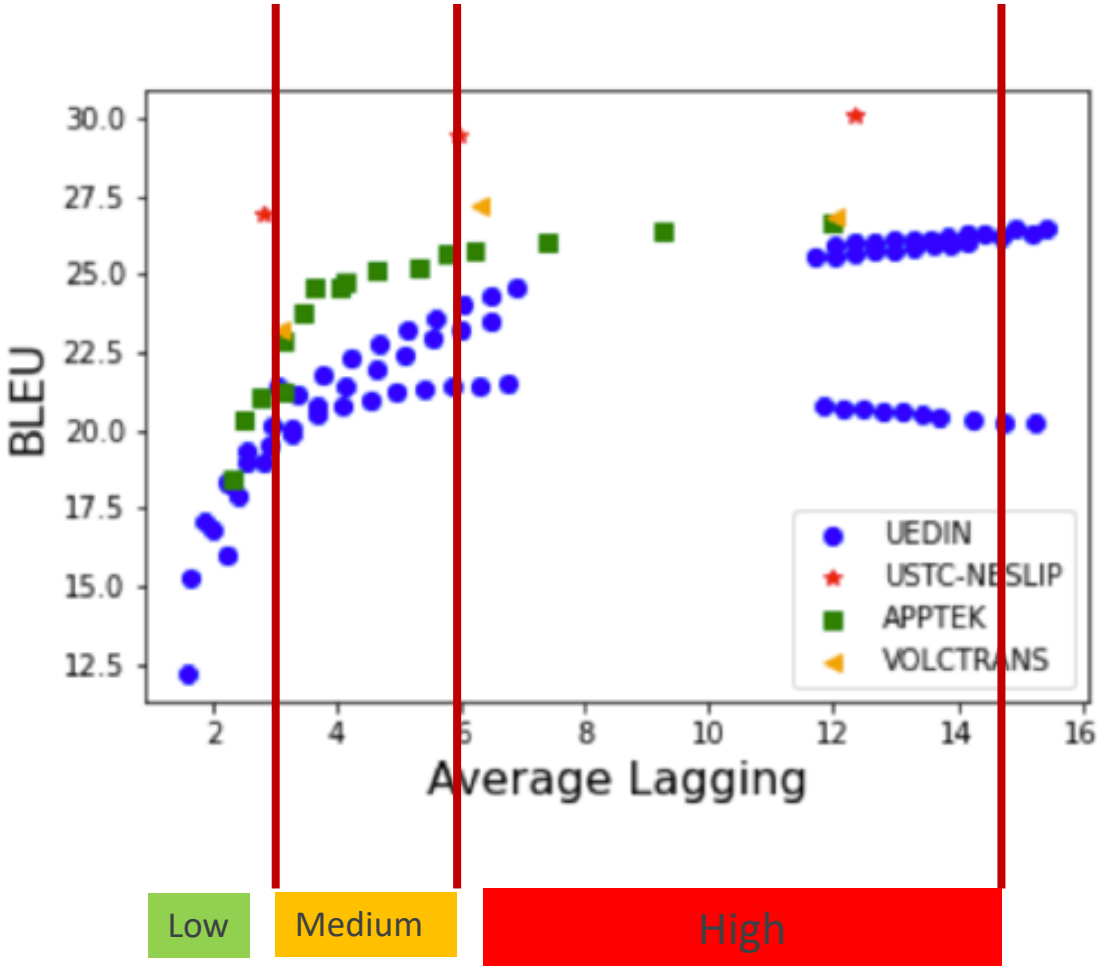




Hamon O, Fugen C, Mostefa D, et al. End-to-End Evaluation in Simultaneous Translation[C]//Proceedings of the 12th Conference of the European Chapter of the ACL, 2009:345—353.

如何兼顾质量与时延？

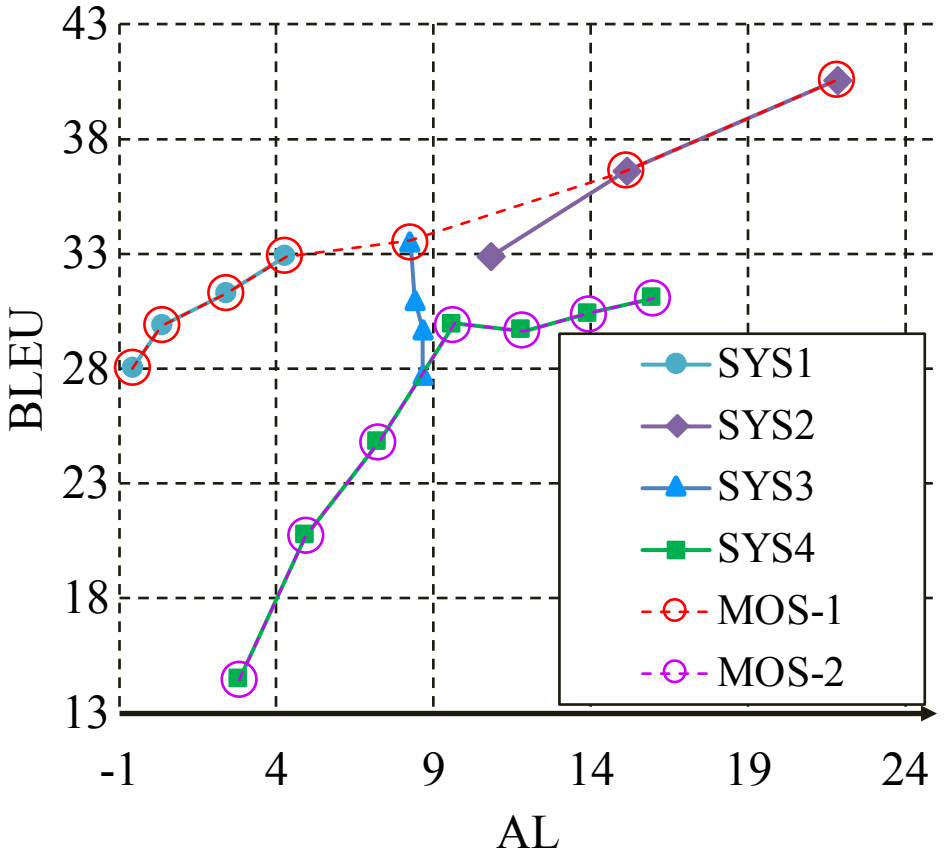
时延分为3个等级，分别评估翻译质量



https://iwslt.org/2021/simultaneous

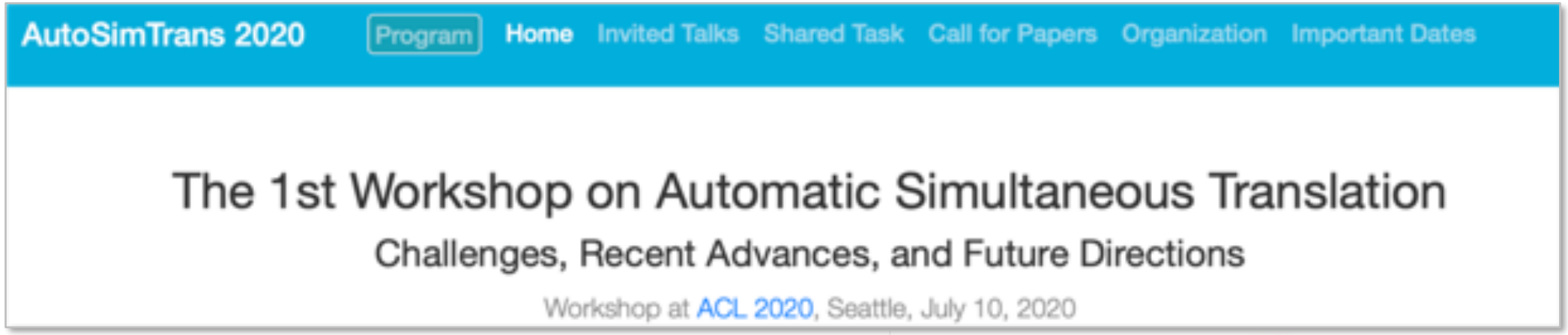
如何兼顾质量与时延？

构建凸包集合，统计各系统落在凸包上的点的个数



https://autosimtrans.github.io/shared#evaluation

同传研讨会



Baidu 百度 Google facebook Penn University of Pennsylvania 清華大學 Tsinghua University

Baidu 百度 Google HUAWEI UGA Université Grenoble Alpes

提纲

- 同声传译
- 主要挑战
- 主流方法
  - 级联模型
  - 端到端模型
- 数据、鲁棒性模型及评价方法
- 产品及应用
- 总结及展望

字幕式（语音到文字，看同传）

主流产品形式

部署方便

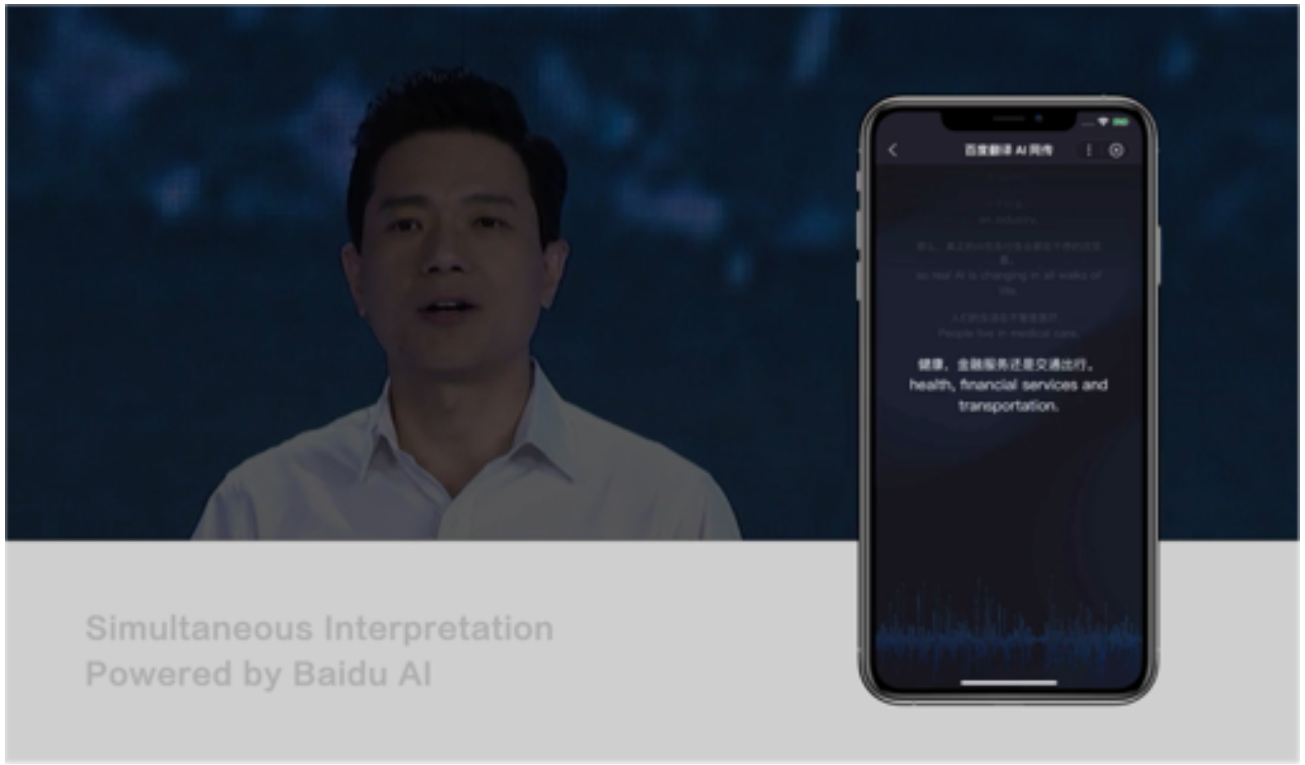
无需额外硬件

多语言展现困难

分散观众注意力看字幕

放大识别与翻译错误

语音式（语音到语音，听同传）



多语言扩展方便

沉浸式体验

缓解错误（如同音字）

需要接收装置（如手机）

语音合成，增加时延

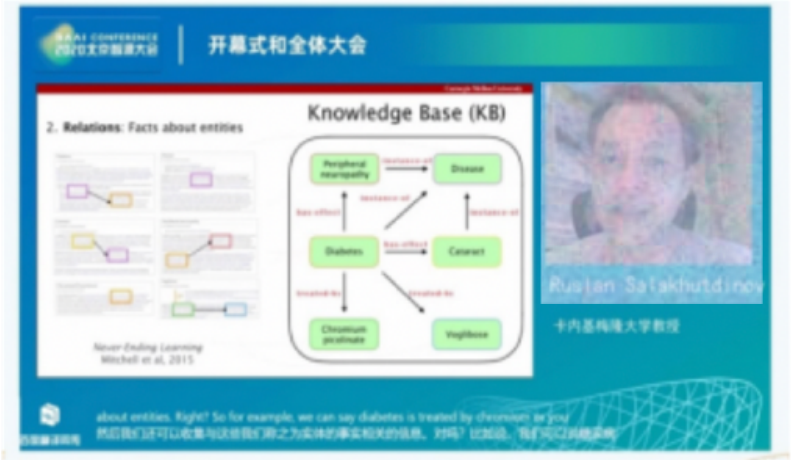
同传应用


CCTV 13
新 闻
上海
格鲁吉亚驻华大使 阿尔赤·卡岚第
这将成为全球学习的案例
11月8日 06:10


协同发展，重塑未来
如其来的新冠肺炎疫情，反映出我们个体在面临巨大的外部冲击时的脆弱性。现在全球累计确诊人
here, hold this seminar, congratulations. In addition , to all the guests, I would like to express my heartfelt thanks for
attending this conference. So the sudden Covid 19 epidemic, reflecting the vulnerability


AI EXPO
2020全球人工智能产品应用博览会


Boolan


开幕式和全体大会
Knowledge Base (KB)
2. Relations: Facts about entities
Never-Ending Learning
Mitchell et al, 2015
Ruslan Salakhutdinov
about entities. Right? So for example, we can say diabetes is treated by chromium or you

总结与展望

有一天，当你在人民大会堂和世界各国友人聚会的时候，你会发现，无论哪个国家的人在台上讲话，与会者都能从耳机里听到自己国家的语言，同时你会发现，在耳机里做翻译的不是人，而是我们的“万能翻译博士”，因为“博士”的语调不像人那样委婉自如。

--《机器翻译浅说》 1964

总结与展望

- **模型**
  - 鲁棒的同传模型
  - 端到端同传模型
  - 融合视觉信息的多模态同传
- **数据**
  - 构建更大规模的数据集
- **评价**
  - 设计符合同传场景的评价指标

谢 谢！