

第11章 机器翻译

宗成庆

中国科学院自动化研究所

cqzong@nlpr.ia.ac.cn

本章内容



1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读

1. 技术产生与发展

◆ 概念

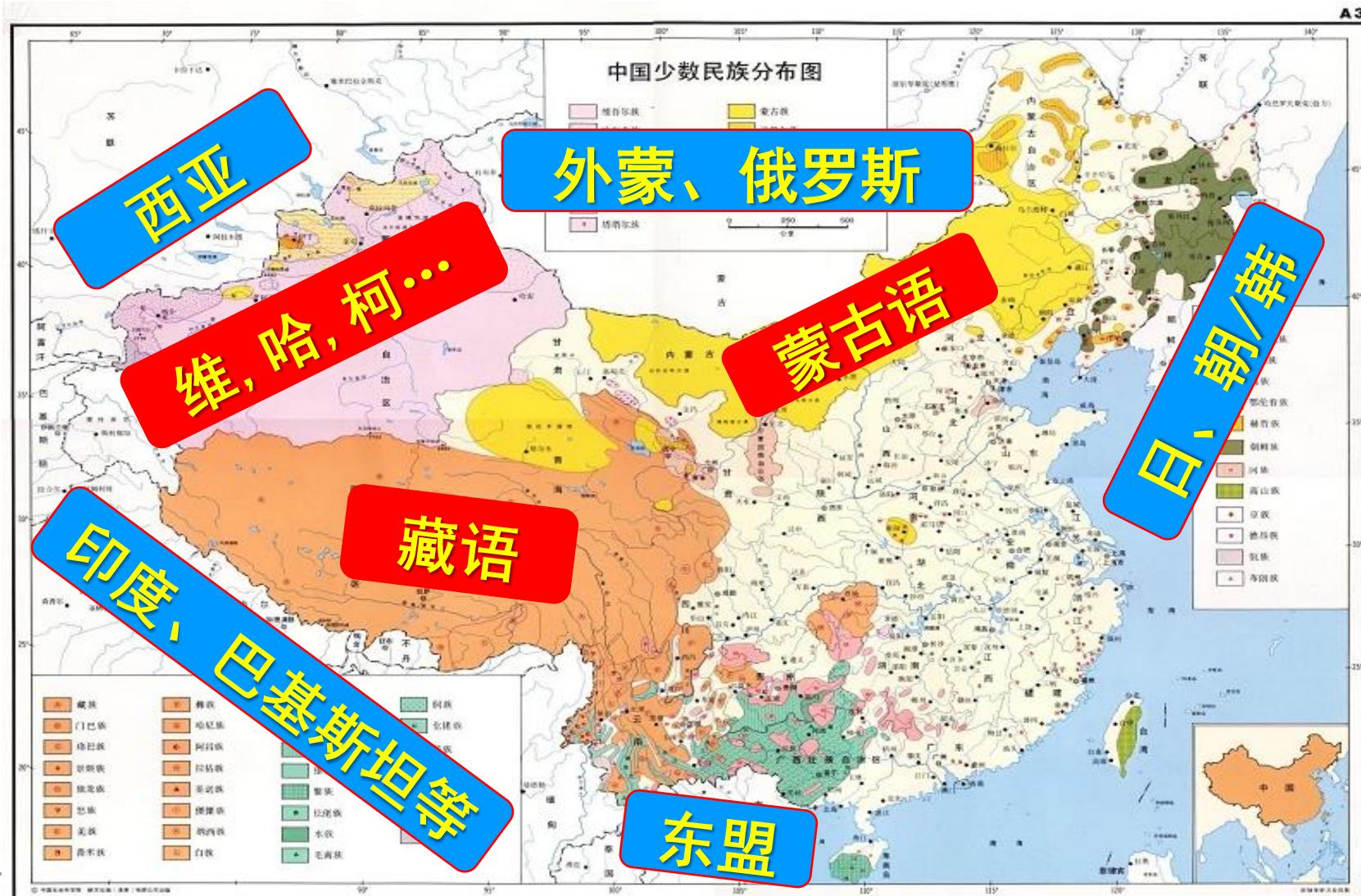
机器翻译(machine translation, MT) 是用计算机把一种语言翻译成另一种语言的一门学科和技术。

被翻译的语言称为源语言(source language)， 翻译的结果语言称为目标语言(target language)。



1. 技术产生与发展

◆ 巨大的应用市场



1. 技术产生与发展

64个国家和地区

44亿人口

50多种语言

一
带
一
路

出境游人数破亿，

前20个出境游目的地有12种语言

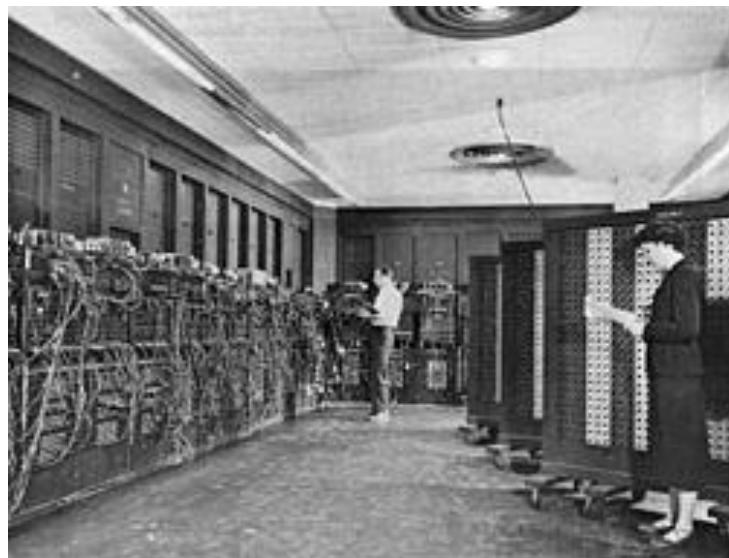
签署的合作协议涉及110多种语言。



丝绸之路经济带
21世纪海上丝绸之路

1. 技术产生与发展

◆技术诞生



1946年，世界上第一台计算机ENIAC诞生。



Warren Weaver

- ◆ 信息论先驱
- ◆ 1920至1932年威斯康星大学数学教授
- ◆ 1932至1955年担任Rockefeller Institute自然科学部主任



A. D. Booth

- ◆ 数学物理学家
- ◆ 1947年3月至9月在普林斯顿大学参与John von Neumann研究组，后来曾在伦敦大学工作

1. 技术产生与发展



[Reproduced by permission of the Rockefeller Foundation Archives]

March 4, 1947

Dear Norbert:

I was terribly sorry, when in Cambridge recently, that I got unavoidably held up by several unexpected jobs, and did not get a chance to see you.

One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.

I wondered if it were unthinkable to design a computer which would translate

Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Have you ever thought about this? As a linguist and expert on computers, do you think it is worth thinking about?

Cordially,

**July 1949:
“Translation” 备忘录**



诺伯特·维纳 (Norbert Wiener) (1894年11月26日～1964年3月18日)

Professor Norbert Wiener
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

WW:AEB

1. 技术产生与发展

Weaver 的两个基本观点：

- (1) 翻译类似于解读密码的过程：当我阅读一篇用中文写的文章时，我可以说这篇文章实际上是用英文写的，只不过它用另外一种奇怪的符号编了码，当我阅读时，我是在进行解码；
- (2) 原文和译文“说的是同样的事情”，因此，当把语言A翻译为语言B时，就意味着，从语言A出发，经过某一“通用语言(universal language)”或“中间语言(interlingua)”，然后转换为语言B，这种“通用语言”或“中间语言”可以假定是全人类共同的。

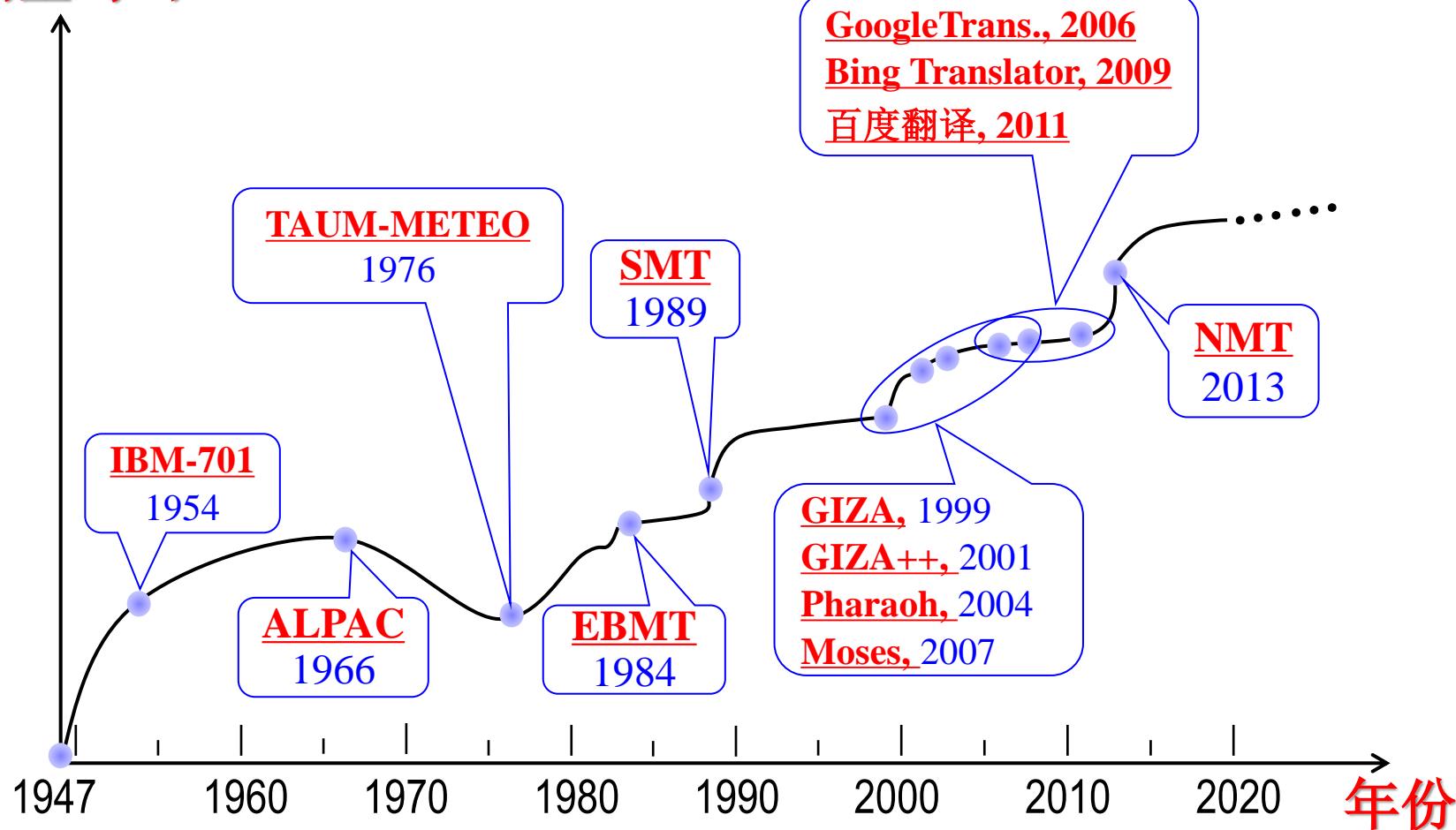
1. 技术产生与发展

- 自W. Weaver 和 A. D. Booth 提出机器翻译(machine translation, MT)概念后，美国和英国的学术界对机器翻译产生了浓厚的兴趣，并得到了实业界的支持。
- 1954年 Georgetown 大学在 IBM 协助下，用IBM-701计算机实现了世界上第一个 MT 系统，实现俄译英翻译，1954年1月该系统在纽约公开演示。系统只有250条俄语词汇，6 条语法规则，可以翻译简单的俄语句子。
- 随后10 多年里，机器翻译研究在国际上出现热潮。
- 1964年，美国科学院成立语言自动处理咨询委员会(Automatic Language Processing Advisory Committee, ALPAC)，调查机器翻译的研究情况，并于1966年11月公布了一个题为“语言与机器”的调查报告，简称ALPAC 报告。计算语言学术语首次正式出现在官方发布的学术报告里。

1. 技术产生与发展

◆ 技术发展

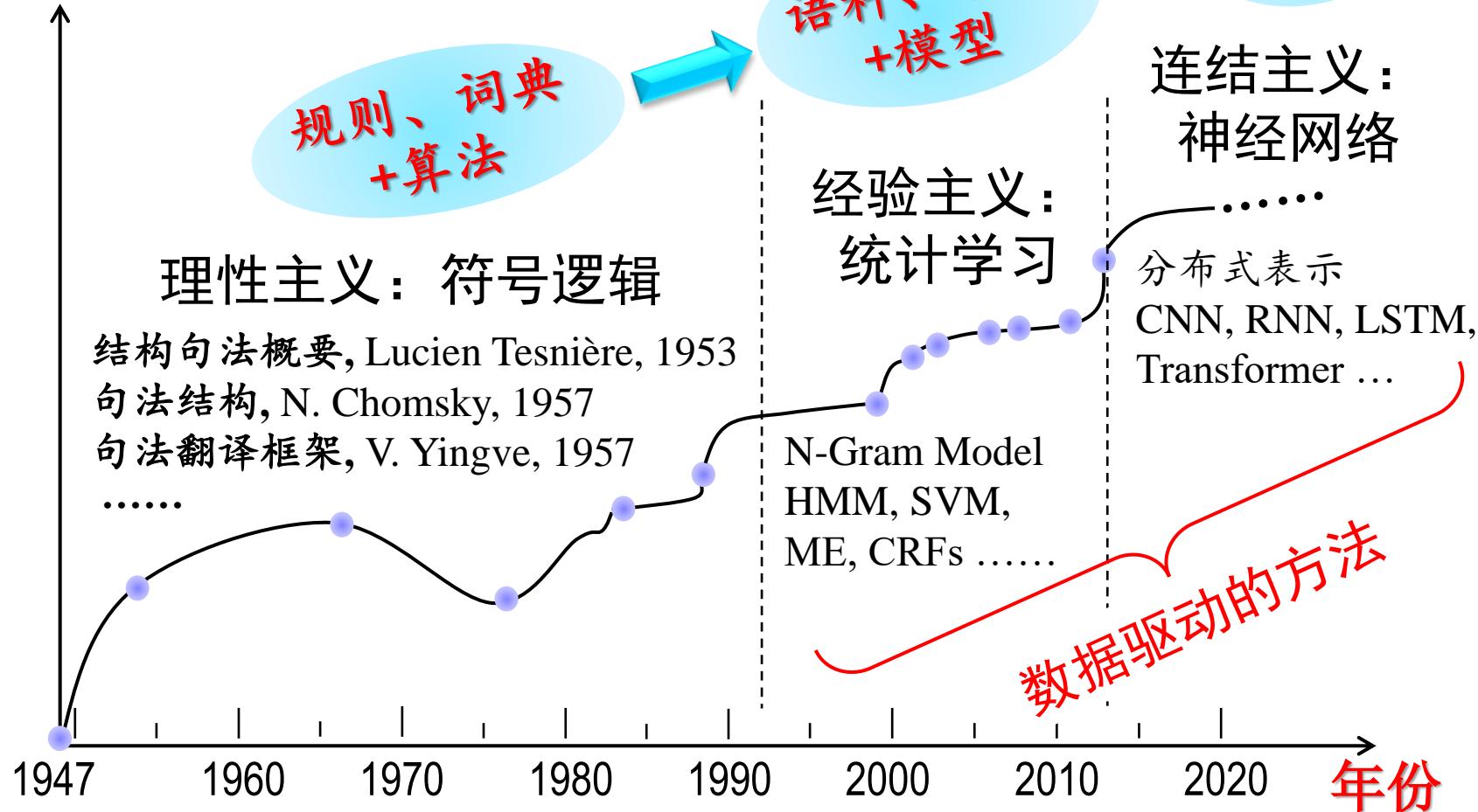
期望/水平



1. 技术产生与发展

◆ 技术发展

期望/水平



本章内容

1. 技术产生与发展
-  2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读

2. 问题与挑战

◆自然语言处理中的所有问题在机器翻译中都会遇到

- 各个语言层面普遍存在的不确定性，各种歧义 ...
- 语言现象的不可预测性，大量存在的非规范语言现象、新语言现象、术语、缩略语 ...
- 语言在不断变化和发展，数据样本永远是不充分的
- 知识表示的复杂性，常识、背景知识、专业知识 ...

◆文化差异导致不同语言的概念不对等

What's wrong/ the matter with you?

爬山：climb the mountains/ hike the mountains

馒头，韭菜盒子

本章内容

1. 技术产生与发展
2. 问题与挑战
-  3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读

3. 机器翻译方法

- ◆ 基于模板的直接转换法
- ◆ 基于规则的翻译方法
- ◆ 基于中间语言的翻译方法
- ◆ 基于语料库的翻译方法
 - 基于事例的机器翻译
 - 统计机器翻译
 - 神经机器翻译

3. 机器翻译方法

◆ 基于模板的直接转换法 (template-based direct translation)

从源语言句子的表层出发，将单词、短语或句子直接置换为目标语言译文，必要时进行简单的词序调整。对原文句子的分析仅满足于特定译文生成的需要。这类翻译系统一般针对某一个特定的语言对，将分析与生成、语言数据、文法和规则与程序等都融合在一起。例如：

I like Mary. → 我喜欢玛丽。

X like Y → X 喜欢 Y。

(1) 今天我想吃面包/ 今天我想吃食堂/ 今天我想吃大碗

Today I would like to eat bread. ✓ canteen / big bowl ✗

(2) 学英语/ 学钢琴

study English/✓ study piano ✗

(3) 写文章/ 写黑板/ 写大仿 ?

3. 机器翻译方法

◆ 基于规则的方法(rule-based MT)

1957年美国学者V. Yingve在《句法翻译框架》(Framework for Syntactic Translation)一文中提出了对源语言和目标语言均进行适当描述、把翻译机制与语法分开、用规则描述语法的实现思想，这就是基于规则的翻译方法。

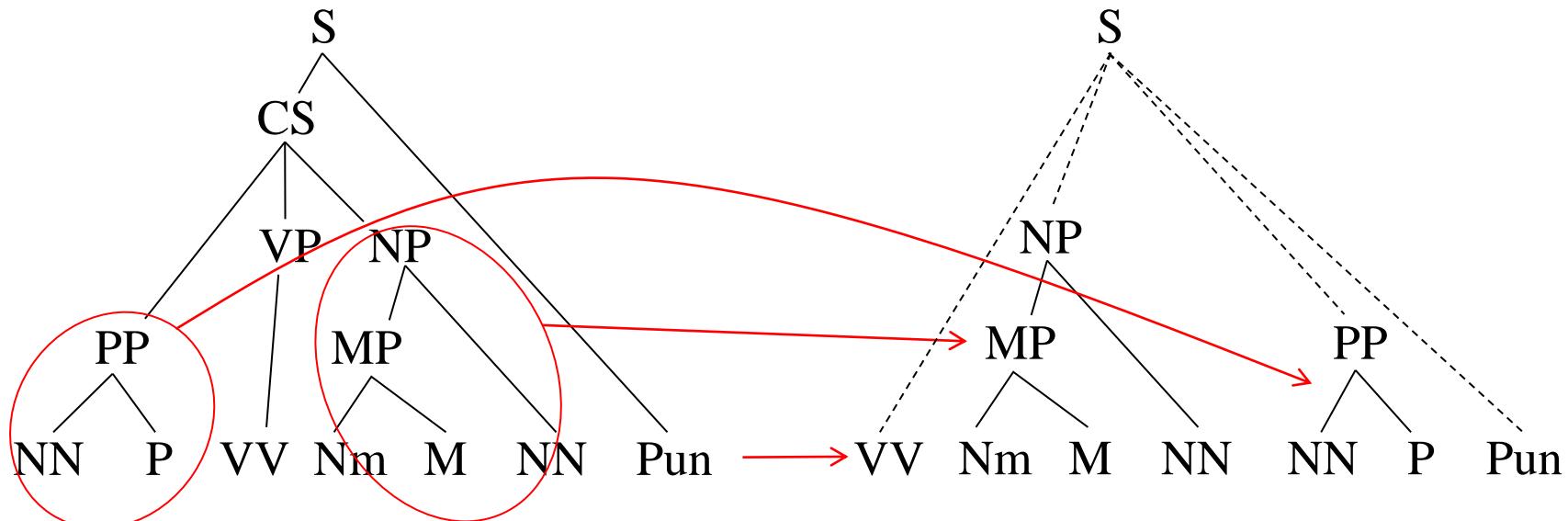
6个基本步骤：

- (a) 对源语言句子进行词法分析
- (b) 对源语言句子进行句法/语义分析
- (c) 源语言句子结构到译文结构的转换
- (d) 译文句法结构生成
- (e) 源语言词汇到译文词汇的转换
- (f) 译文词法选择与生成

3. 机器翻译方法

给定源语言句子： 桌子上有一本书。

- (a) 分词与词性标注：桌子/NN 上/P 有/VV 一/Num 本/M 书/NN 。 /Pun
 (b) 句法结构分析： (c) 结构转换； (d) 译文结构生成：



- (e) 词汇转换； (f) 译文生成：

There is a book on the desk.

#有: have/ there be/...
#一: one
#书: book
.....

3. 机器翻译方法

由于基于规则的翻译方法执行过程为“独立分析—相关转换—独立生成”，因此，又称基于转换的翻译方法(transfer-based MT)。

● 方法评价：

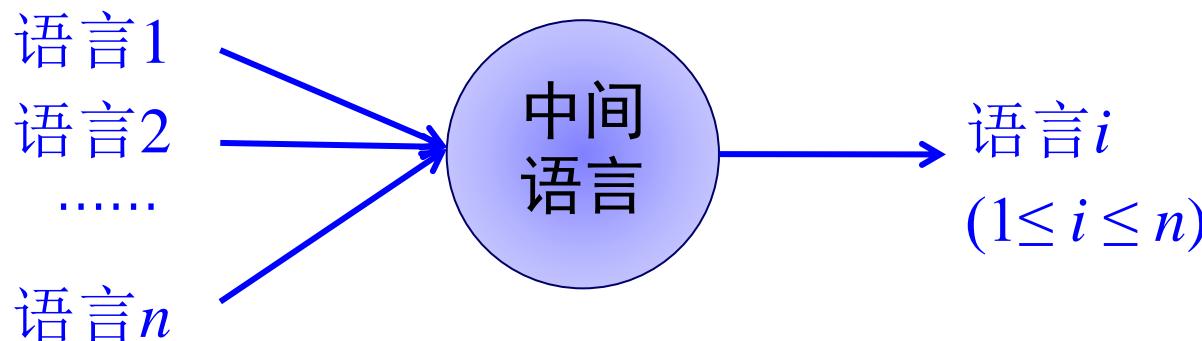
优点：可以较好地保持原文的结构，产生的译文结构与源文的结构关系密切，尤其对于语言现象已知的或句法结构规范的源语言语句具有较强的处理能力和较好的翻译效果。

弱点：规则一般由人工编写，工作量大，主观性强，一致性难以保障，不利于系统扩充，对非规范语言现象缺乏相应的处理能力。

3. 机器翻译方法

◆ 基于中间语言的翻译方法(interlingua-based MT)

- **方法:** 输入语句→中间语言→ 翻译结果
- **代表系统:** JANUS (CMU) 早期版本
 - ★ 源语言解析器
 - ★ 比较准确的中间语言(interlingua)
 - ★ 目标语言生成器(target language generator)





3. 机器翻译方法

关于中间语言的定义

- 国际先进语音翻译研究联盟(C-STAR)定义的中间转换格式(Interchange Format, IF)
- 日本东京联合国大学(United Nations University) 提出的通用网络语言(Universal Networking Language, UNL)

●方法评价：

优点：中间语言的设计可以不考虑具体的翻译语言对，因此，该方法尤其适合多语言之间的互译。

弱点：如何定义和设计中间语言的表达方式，以及如何维护并不是一件容易的事情，中间语言在语义表达的准确性、完整性等很多方面，都面临若干困难。

3. 机器翻译方法

◆ 基于事例(实例)的翻译方法(example-based)

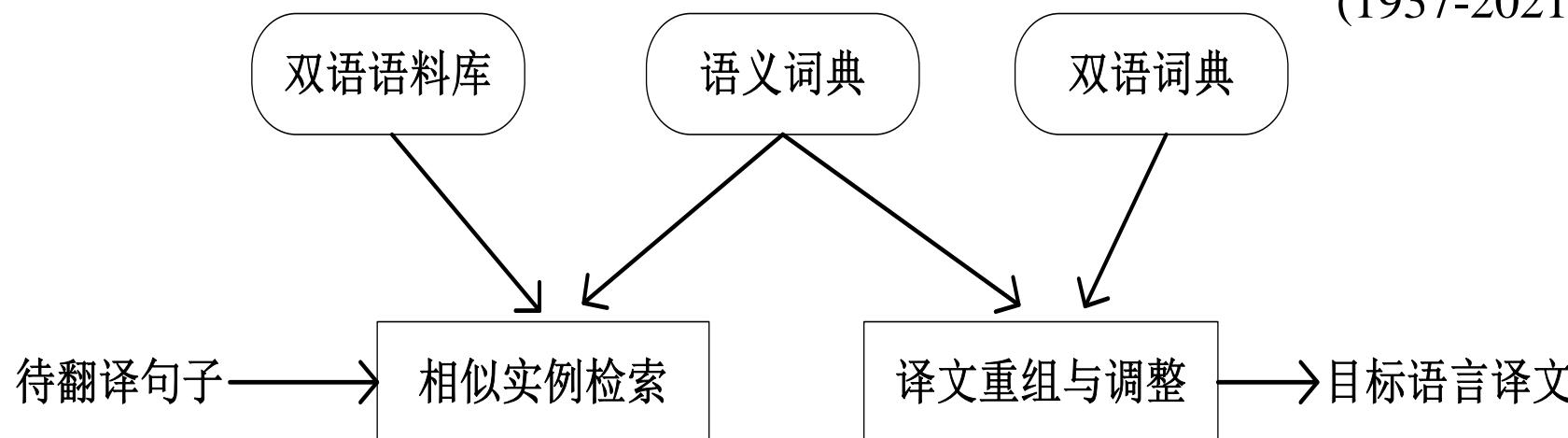
1984年由日本学者长尾真提出。

• **方法：** 输入语句 → 与事例相似度比较

→ 翻译结果

• **资源：** 大规模事例库

• **代表系统：** ATR-MATRIX (ATR, Japan)



3. 机器翻译方法

● 方法评价：

优点：不要求源语言句子必须符合语法规规定，翻译机制一般不需要对源语言句子做深入分析。

弱点：两个不同的句子之间的相似性（包括结构相似性和语义相似性）往往难以把握，尤其在口语中，句子结构一般比较松散，成分冗余和成分省略都较严重，这更增加了分析句子与事例句子的比较难度。另外，系统往往难以处理事例库中没有记录的陌生的语言现象，而且当事例库达到一定规模时，其事例检索的效率较低。

3. 机器翻译方法

◆统计机器翻译方法

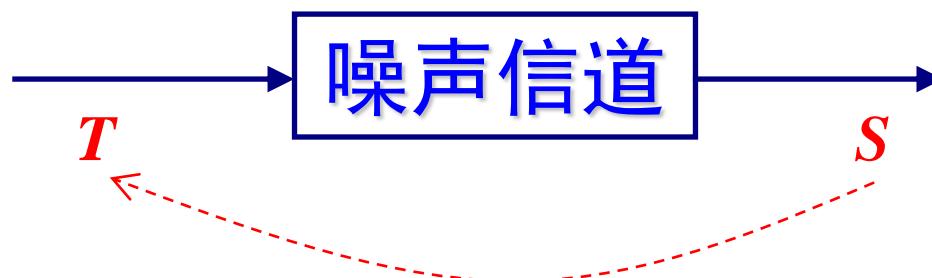
● 基本思想

- 源自1947年 W. Weaver 提出的“解读密码”思想。
- 1990年IBM的Peter F. Brown 等在*Computational Linguistics* 上发表的论文“统计机器翻译方法” [Brown et al., 1990] 和1993年他们发表在该杂志的“统计机器翻译的数学：参数估计” [Brown et al., 1993]两篇文章奠定了统计机器翻译的理论基础。
- 基于统计方法实现的*Candide* 翻译系统[Berger et al.,1994] 在ARPA组织的机器翻译评测中表现良好。

3. 机器翻译方法

● 噪声信道模型

一种语言 T 由于经过一个噪声信道而发生变形，从而在信道的另一端呈现为另一种语言 S (信道意义上的输出，翻译意义上的源语言)。翻译问题实际上就是如何根据观察到的 S ，恢复最为可能的 T 问题。这种观点认为，任何一种语言的任何一个句子都有可能是另外一种语言中的某个句子的译文，只是可能有大有小 [Brown et. al, 1990]。



3. 机器翻译方法

源语言句子: $S = s_1^m \equiv s_1 s_2 \cdots s_m$

目标语言句子: $T = t_1^l \equiv t_1 t_2 \cdots t_l$

$$p(T | S) = \frac{p(T) \times p(S | T)}{p(S)}$$

$$\hat{T} = \arg \max_T p(T) \times p(S | T)$$

语言模型

Language model, LM

翻译模型

Translation model, TM

3. 机器翻译方法

实现搜索过程的模块称为解码器(decoder)。



统计翻译中的三个关键问题：

- (1) 估计语言模型概率 $p(T)$;
- (2) 估计翻译概率 $p(S|T)$;
- (3) 快速有效地搜索 T 使得 $p(T) \times p(S | T)$ 最大。

3. 机器翻译方法

- 语言模型概率 $p(T)$

给定句子: $t_1^l = t_1 t_2 \cdots t_l$

句子概率: $p(t_1^l) = p(t_1) \times p(t_2 | t_1) \times \cdots \times p(t_l | t_1 t_2 \cdots t_{l-1})$

n-gram 问题, 不再赘述。

3. 机器翻译方法

merkezdiki dölet apparatliri bilen jaylardiki dölet apparatlinining xizmet hoquqi merkezning bir tutash rehberlikide jaylarning teshebbuskarliqi we aktipliqini toluq jari qildurush prinsipi boyiche ayridi.

中央和地方的国家机构职权的划分，遵循在中央的统一领导下，充分发挥地方的主动性、积极性的原则。

madda jungxua xelq jumhuriyitide hemme millet bapbarawer.
中华人民共和国各民族一律平等。

herqandaq milletni kemsitish we ëzishni men'i qilidu, milletler
ittipaqliqini buzidighan we milliy bölgünchilik qilidighan qilmishlarni
men'i qilidu.

禁止对任何民族的歧视和压迫，禁止破坏民族团结和制造民族分裂
的行为。

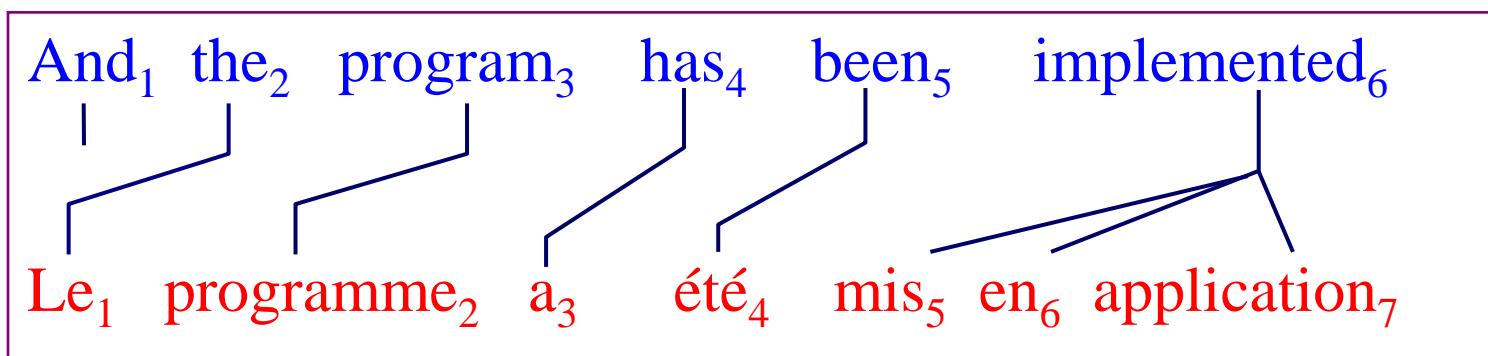
.....

3. 机器翻译方法

● 翻译模型概率 $p(S|T)$

假设英语与法语的翻译对：

(Le programme a été mis en application | And the (1) program (2)
has (3) been(4) implemented (5, 6, 7))



我们可以假定所有的单词对 (s_j, t_i) 之间存在着对应关系。那么，用来刻画这些对应关系的模型叫做**对位模型** (alignment model)。

3. 机器翻译方法

翻译句对($S|T$) 的似然率可以通过条件概率 $p(S, A|T)$ 获得:

$$p(S | T) = \sum_A p(S, A | T)$$

按照约定, 源语言句子 $S = s_1^m \equiv s_1 s_2 \cdots s_m$

有 m 个单词, 目标语言句子 $T = t_1^l \equiv t_1 t_2 \cdots t_l$

有 l 个单词, 对位序列表示成: $A = a_1^m = a_1 a_2 \cdots a_m$

其中, a_j ($j=1..m$)的取值范围为0到 l 之间的整数, 如果源语言中的第 j 个词与目标语言中的第 i 个词对齐, 那么, $a_j = i$, 如果没有词与它对齐, 则 $a_j = 0$ 。在前面的例子中, 法语句子的第一个单词 Le 与英语句子中的第2个单词 the 对齐, 那么, $a_1 = 2$ 。

3. 机器翻译方法

不失一般性：

$$p(S, A | T) = p(m | T) \prod_{j=1}^m p(a_j | a_1^{j-1}, s_1^{j-1}, m, T) \times p(s_j | a_1^j, s_1^{j-1}, m, T)$$

实际上， $p(S, A | T)$ 可以写成多种形式的条件概率的乘积，上式只是其中的一种。在该式的基础上，IBM 的研究人员通过采用不同的假设条件得到了5个翻译模型，分别称作 IBM 翻译模型1、2、3、4 和 5。

该模型以词为单位建立，因此称为**基于词的翻译模型** (word-based translation model)。

3. 机器翻译方法





3. 机器翻译方法

➤ 基于短语的翻译模型(phrase-based TM)

将一个法语句子 f 翻译成英语句子 e ：

$$\arg \max_e p(e | f) = \arg \max_e p(f | e) p(e)$$

考虑到位置的调整：

$$e_{best} = \arg \max_e p(f | e) p(e) \omega^{length(e)}$$

ω 为长度因子，一般 $\omega>1$ ，倾向于较长的输出。

3. 机器翻译方法

注意：

这里所说的短语指一个连续的字串(*n*-gram)，不一定是语言学中定义的短语(phrase)，如：

我想预订一个单人间。

I would like to reserve a single room.

假设 T 、 S 分别是目标语言句子和源语言句子， $h_1(T, S), \dots, h_M(T, S)$ 分别是 T 、 S 上的 M 个特征， $\lambda_1, \dots, \lambda_M$ 是这些特征分别对应的 M 个权值。对于给定的源语言句子 S ，其最佳译文 T 可用下公式表示：

$$\begin{aligned}\hat{e} &= \arg \max_T \{ p(T | S) \} \\ &= \arg \max_T \left\{ \sum_{m=1}^M \lambda_m h_m(T, S) \right\}\end{aligned}$$

基于最大熵的方法选择特征，最小错误率训练，确定权重 λ 。

3. 机器翻译方法

常用的特征有(以汉英翻译为例):

- ✓ 短语翻译概率: $p(ph_t|ph_s)$
- ✓ 词汇化的短语翻译概率: $p(w_{tk1}|w_{s1}) \times p(w_{tk2}|w_{s2}) \dots$
- ✓ 反向的短语翻译概率: $p(ph_s|ph_t)$
- ✓ 反向的词汇化短语翻译概率: $p(w_{sk1}|w_{t1}) \times p(w_{sk2}|w_{t2}) \dots$
- ✓ 基于 n -gram 的英语语言模型: $p(e)$
- ✓ 英语句子长度惩罚: $\omega^{length(e)}$, $\omega > 1$
- ✓ 汉语短语个数惩罚
-

3. 机器翻译方法

以汉英翻译为例：

我	不	喜欢	喝	咖啡
I	no	like	drink	coffee
me	not	like to	drinking	
	do not			drinking coffee
I	do not	like to drink		
I	do not like to			
...	...			

3. 机器翻译方法

所有翻译候选被存放在一个表里：

我: (1, 1), I, 0.813

我: (1, 1), me, 0.187

不: (1, 2), I do not, 0.435

不: (1, 2), I am not, 0.418

不: (1, 2), I never, 0.147

我不喜欢: (1, 3), I do not like to, 0.532

... ...

不: (2, 2), no, 0.352

不: (2, 2), not, 0.355

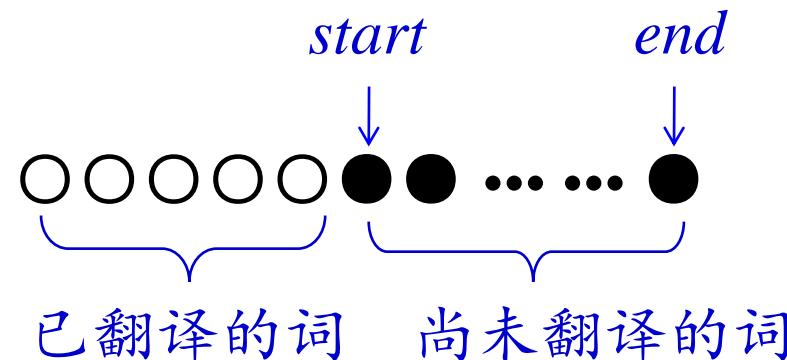
不: (2, 2), do not, 0.293

3. 机器翻译方法

● 解码器：启发式搜索

• 代价估算

- 选择候选翻译
- 剪枝



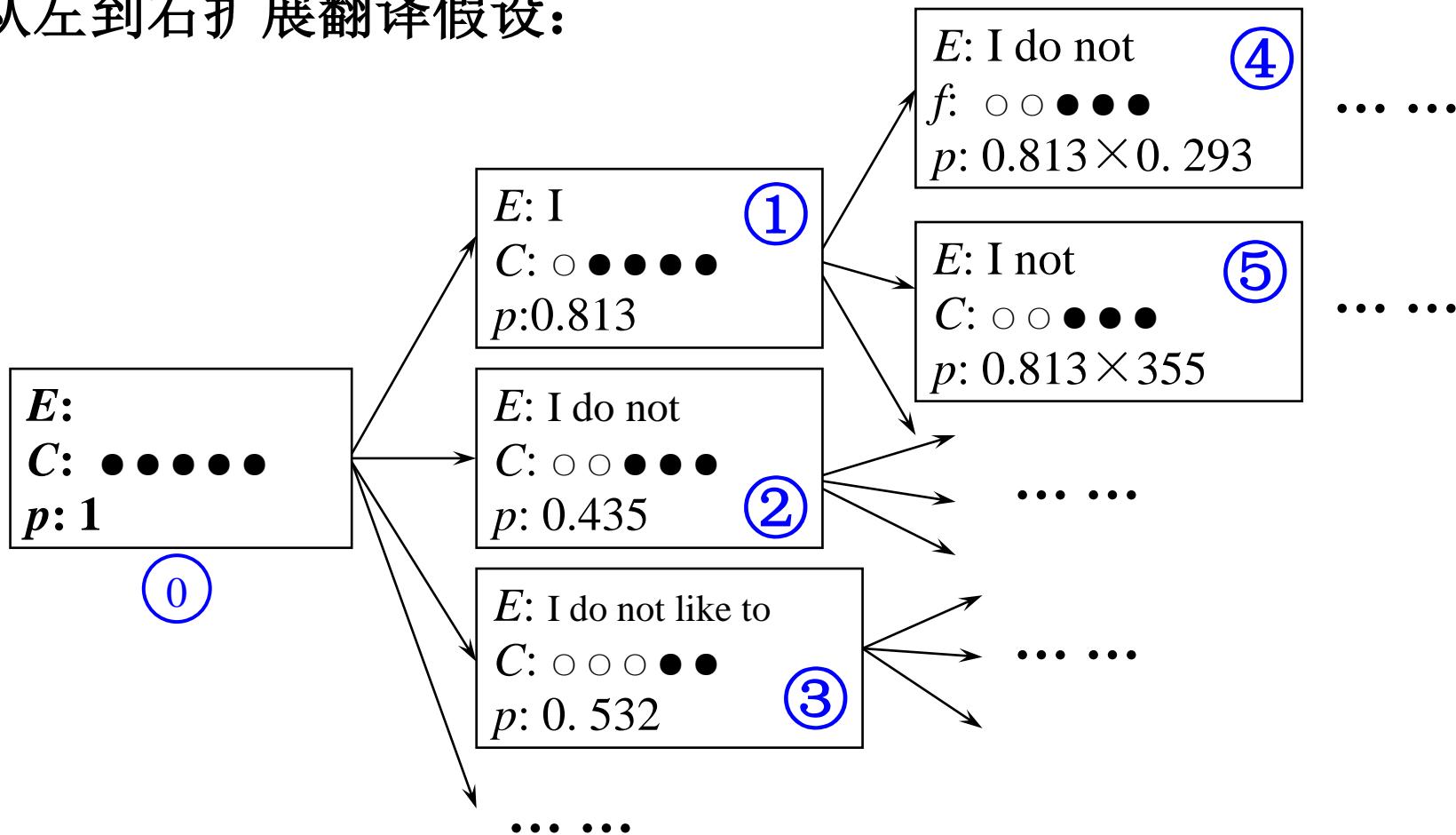
$f = \text{已翻译词所耗费的代价} + \text{未翻译部分的估算代价}$

★ 已翻译词所耗费的代价：已翻译词的 n -gram 概率

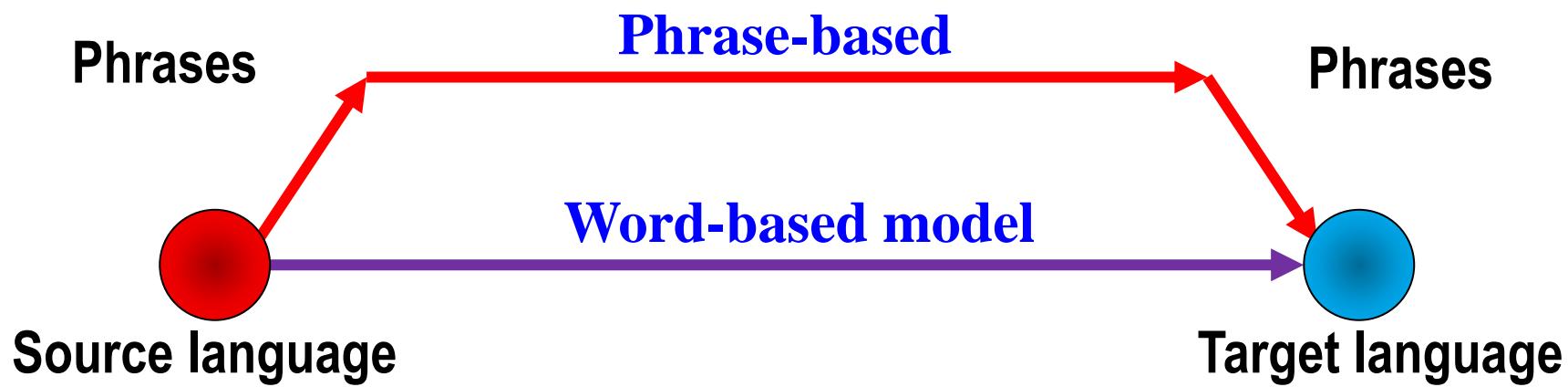
★ 翻译未来词估计需要的代价：最大概率或其他因素

3. 机器翻译方法

从左到右扩展翻译假设：



3. 机器翻译方法



3. 机器翻译方法

● 开源工具

(1) 语料预处理工具

- 工具：EGYPT- TokenizeE.perl.tmpl
<http://www.clsp.jhu.edu/ws99/projects/mt/toolkit/>
- 汉语分词工具 — 隆重推荐 Urheen：
<http://www.nlpr.ia.ac.cn/cip/software.htm>

(2) 语言模型：

- SRI LM: <http://www.speech.sri.com/projects/srilm/>
- CMU-Cambridge LM
<http://mi.eng.cam.ac.uk/~prc14/toolkit.html>

3. 机器翻译方法

(3) 词对齐工具: GIZA++

<http://www.fjoch.com/GIZA++.html>

(4) 解码器:

- Pharaoh / Moses
- ReWrite: <http://www.isi.edu/licensed-sw/rewrite-decoder/>

(5) EM 最小错误率参数训练工具

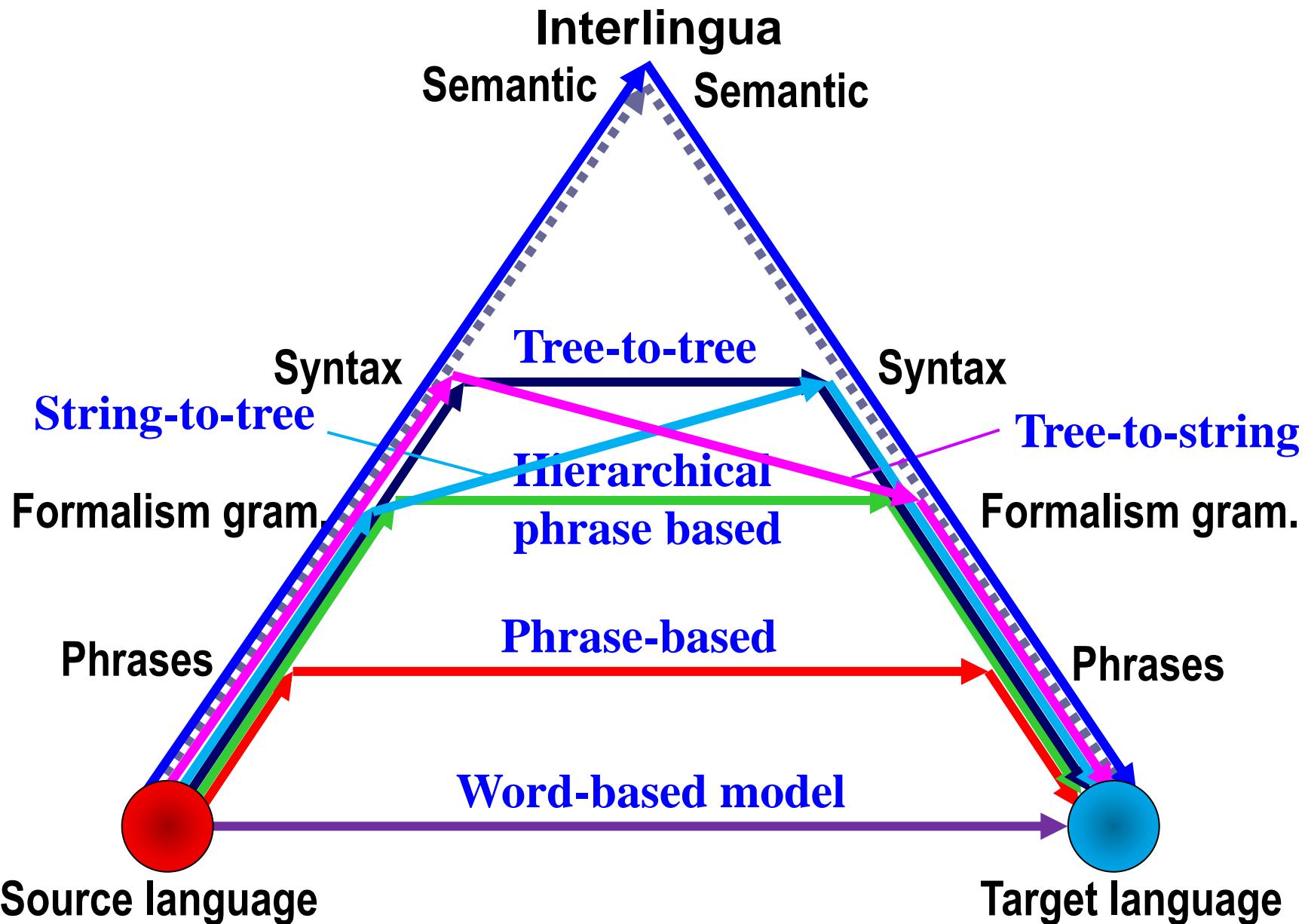
<http://www.cs.cmu/ashish/mer.html>

(6) “摩西” (Moses) 解码器

<http://www.statmt.org/moses/>

3. 机器翻译方法

- 基于层次短语的翻译模型
- 树翻译模型
 - 树到树的翻译模型
 - 树到串的翻译模型
 - 串到树的翻译模型



本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 习题



4. 神经机器翻译

◆模型提出

基于神经网络的机器翻译 (Neural Network based MT, NNMT)
或称神经机器翻译(Neural Machine Translation, NMT), 2013

代表论文:

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. *Proc. of EMNLP 2013*

Ilya Sutskever, Oriol Vinyals, Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *Proc. of NIPS 2014*

Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. *Proc. of ICLR 2015*



4. 神经机器翻译

● 基本原理

给定源语言句子: $C = c_1^l \equiv c_1 c_2 \cdots c_l$

将其翻译成目标语言句子: $E = e_1^m \equiv e_1 e_2 \cdots e_m$

词汇向量化表示: Word2Vec

$$P(e_i) \approx P(e_i | e_1 \cdots e_{i-1}, C)$$

目标函数: $L = \sum_i \log(P(e_i | C))$

本课程绪论

4. 神经机器翻译

- 端到端的框架(end-to-end, E2E)

- 编码器(encoder)-解码器(decoder)
- 基于循环神经网络编解码

$$\mathbf{h}_t = \tanh(\mathbf{W}\mathbf{x}_t + \mathbf{U}\mathbf{h}_{t-1} + \mathbf{b})$$

权重1 权重2 偏置3
↓ ↓ ↓
 \mathbf{W} \mathbf{U} \mathbf{b}

\uparrow \uparrow \uparrow
 t 时刻隐层向量 t 时刻和 $t-1$ 时刻的隐层向量

4. 神经机器翻译

◆ 系统实现

● 基础资源

（桌子上有一本书。

There is a book on the desk.

（今天是星期三。

Today is Wednesday.

（明天他将来北京。

Tomorrow he will come to Beijing.

.....



4. 神经机器翻译

● 训练过程

Step 1: 参数初始化

参数①: 词向量

源语言: 桌子 = $\begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$ 上 = $\begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$ 有 = $\begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$ 一 = $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ 本 = $\begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$ 书 = $\begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$...

目标语言: there = $\begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix}$ is = $\begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ a = $\begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$ book = $\begin{bmatrix} 0.4 \\ 0.3 \end{bmatrix}$ on = $\begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix}$ the = $\begin{bmatrix} -0.1 \\ -0.2 \end{bmatrix}$

$\text{desk} = \begin{bmatrix} -0.2 \\ 0.1 \end{bmatrix}$	$\text{START} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ ↑ 句子开始的标记	$\text{EOS} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ↑ 句子结束的标记	...
---	---	---	-----

源语言和目标语言的单词很多，实际实现系统时会设置3~8万左右(单词)，其它词记作UNK。



4. 神经机器翻译

参数②：神经网络的参数(RNN)

$$\mathbf{h}_t = \tanh(\mathbf{W}\mathbf{x}_t + \mathbf{U}\mathbf{h}_{t-1} + \mathbf{b})$$

- 编码器参数：

$$\mathbf{W}_s = \begin{bmatrix} 0.3 & 0 \\ 0 & -0.3 \end{bmatrix} \quad \mathbf{U}_s = \begin{bmatrix} 0.1 & -0.1 \\ 0.1 & -0.1 \end{bmatrix} \quad \mathbf{b}_s = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$$

- 解码器参数：

$$\mathbf{W}_t = \begin{bmatrix} -0.3 & 0 \\ 0 & 0.3 \end{bmatrix} \quad \mathbf{U}_t = \begin{bmatrix} -0.1 & 0.1 \\ -0.1 & 0.1 \end{bmatrix} \quad \mathbf{b}_t = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}$$

4. 神经机器翻译

Step 2: 对源语言句子编码: 桌子上有一本书

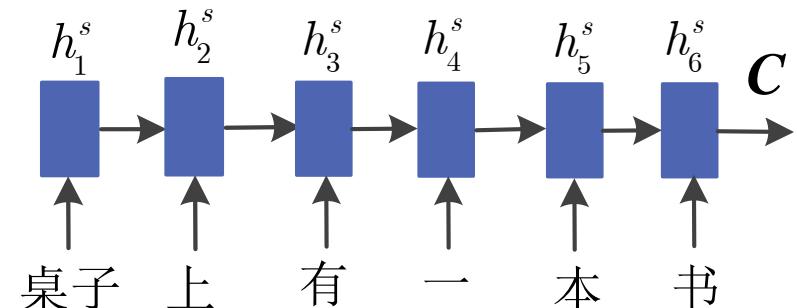
$$\mathbf{h}_1^s = \tanh(\mathbf{W}_s \times \text{桌子} + \mathbf{U}_s \times \mathbf{0} + \mathbf{b}_s) = \begin{bmatrix} 0.337 \\ 0.572 \end{bmatrix}$$

$$\mathbf{h}_2^s = \tanh(\mathbf{W}_s \times \text{上} + \mathbf{U}_s \times \mathbf{h}_1 + \mathbf{b}_s) = \begin{bmatrix} 0.315 \\ 0.729 \end{bmatrix}$$

... ...

$$\mathbf{h}_6^s = \tanh(\mathbf{W}_s \times \text{书} + \mathbf{U}_s \times \mathbf{h}_5 + \mathbf{b}_s) = \begin{bmatrix} 0.009 \\ 0.543 \end{bmatrix}$$

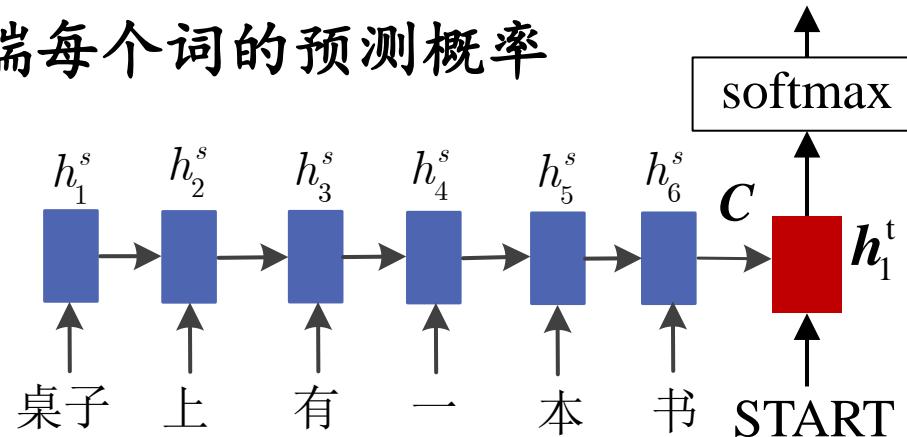
$$\mathbf{C} = \mathbf{h}_6^s = \begin{bmatrix} 0.009 \\ 0.543 \end{bmatrix}$$



4. 神经机器翻译

Step 3: 解码: 计算目标端每个词的预测概率

①时刻 $t=1$



$$\mathbf{h}_1^t = \tanh(\mathbf{W}_t \times \text{START} + \mathbf{U}_s \times \mathbf{C} + \mathbf{b}_t) = \begin{bmatrix} 0.693 \\ 0.248 \end{bmatrix}$$

$$score_1 = \text{softmax} \left(\begin{array}{l} h_1^t \times \text{the} \\ \dots \\ h_1^t \times \text{there} \\ \dots \\ h_1^t \times \text{book} \\ h_1^t \times \text{EOS} \end{array} \right) = \begin{array}{l} 0.332 \\ \dots \\ 0.195 \\ \dots \\ 0.254 \\ 0.124 \end{array}$$

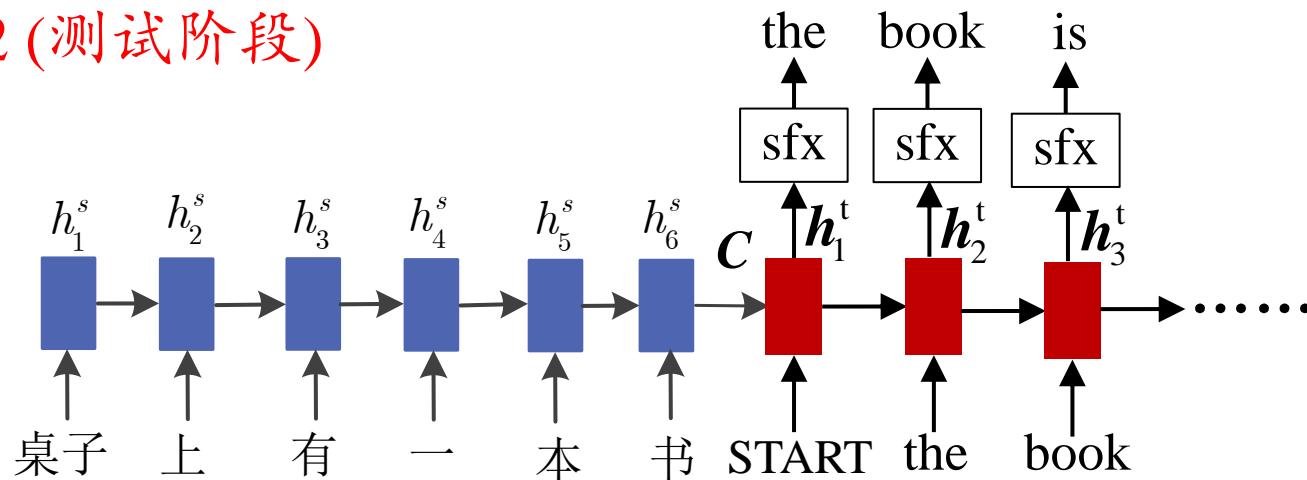
预测概率最高(the)

标准答案(there)

在训练阶段: 调试参数, 让 there 的预测概率最大。**在测试阶段:** 从预测概率分布中选择概率最高的单词, 作为下一时刻的输入。

4. 神经机器翻译

②时刻 $t=2$ (测试阶段)



$$\mathbf{h}_2^t = \tanh(\mathbf{W}_t \times \text{the} + \mathbf{U}_s \times \mathbf{h}_1^t + \mathbf{b}_t) = \begin{bmatrix} 0.693 \\ 0.248 \end{bmatrix}$$

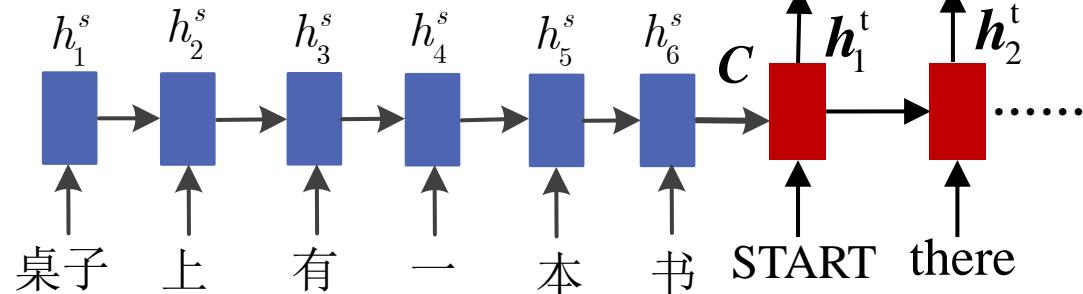
$$score_2 = \text{softmax} \begin{pmatrix} \mathbf{h}_2^t \times \text{the} \\ \dots \\ \mathbf{h}_2^t \times \text{there} \\ \dots \\ \mathbf{h}_2^t \times \text{book} \\ \mathbf{h}_2^t \times \text{EOS} \end{pmatrix} = \begin{pmatrix} 0.123 \\ \dots \\ 0.134 \\ \dots \\ \boxed{0.494} \\ 0.024 \end{pmatrix}$$

重复上述过程，直到预测出 EOS 为止。如果 EOS 不出现，通常设定最大长度。

← 预测概率最高(book)

4. 神经机器翻译

说明：在训练阶段，利用梯度下降法最大化标准答案的预测概率，并用标准答案作为下一时刻的输入。(例子中第一个输出的标准答案为there)



$$\mathbf{h}_2^t = \tanh(\mathbf{W}_t \times \text{there} + \mathbf{U}_s \times \mathbf{h}_1^t + \mathbf{b}_t) = \begin{bmatrix} 0.631 \\ 0.203 \end{bmatrix}$$

$$\begin{aligned} score_2 &= \text{softmax} \begin{pmatrix} \mathbf{h}_2^t \times \text{the} \\ \dots \\ \mathbf{h}_2^t \times \text{there} \\ \mathbf{h}_2^t \times \text{is} \\ \dots \\ \mathbf{h}_2^t \times \text{book} \\ \mathbf{h}_2^t \times \text{EOS} \end{pmatrix} = \begin{pmatrix} 0.223 \\ \dots \\ 0.034 \\ \boxed{0.147} \\ \dots \\ 0.094 \\ 0.024 \end{pmatrix} \end{aligned}$$

• 标准答案 (is)

- 每次选择标准答案作为下一时刻的输入，直到标准答案结束；
- 最大化预测概率： $\max P(Y|X, \theta)$ 。

$$\begin{aligned} P(Y | X, \theta) &= \text{score}_1 \times \text{score}_2 \times \dots \times \text{score}_8 \\ &= 0.195 \times 0.147 \times \dots \times 0.386 = 0.029 \end{aligned}$$

4. 神经机器翻译

● 问题分析

- 存在的问题：仅利用语义向量 C 表征源语言的全部信息，当源语言句子较长时， C 所编码的信息有限。
- 改进方式：**注意力机制(attention)**
基本思想：在每个解码时刻，模型从源端中选择当前时刻解码所需要的语义信息。

4. 神经机器翻译

◆ 注意力机制的使用

● Step A 注意力权重计算

(a) 对“桌子”的注意力值

$$a_{(1,1)} = (h_1^t)^T h_1^s = 0.1$$

(b) 对“上”的注意力值

$$a_{(1,2)} = (h_1^t)^T h_2^s = 0.01$$

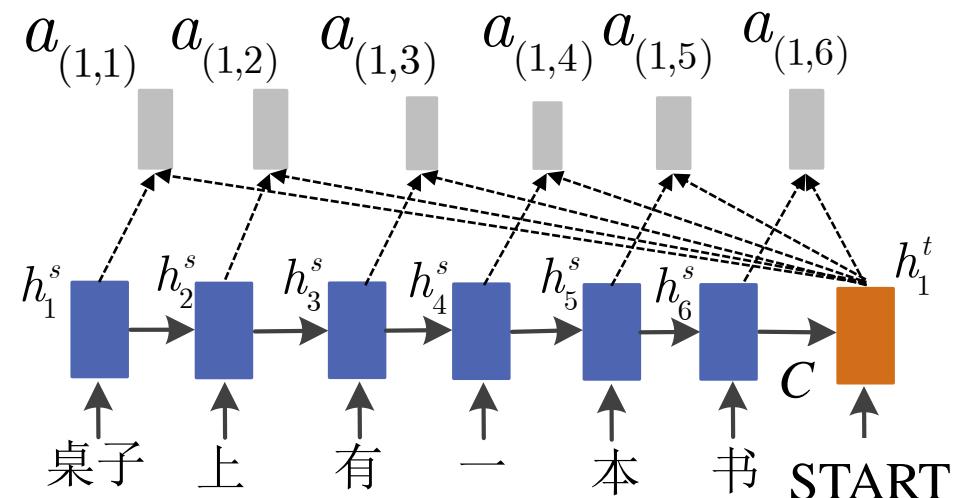
(c) 对“有”的注意力值

$$a_{(1,3)} = (h_1^t)^T h_3^s = 0.2$$

.....

(d) 对“书”的注意力值

$$a_{(1,6)} = (h_1^t)^T h_6^s = 0.11$$



在此基础上，还需要对源语言
注意力权重进行归一化。

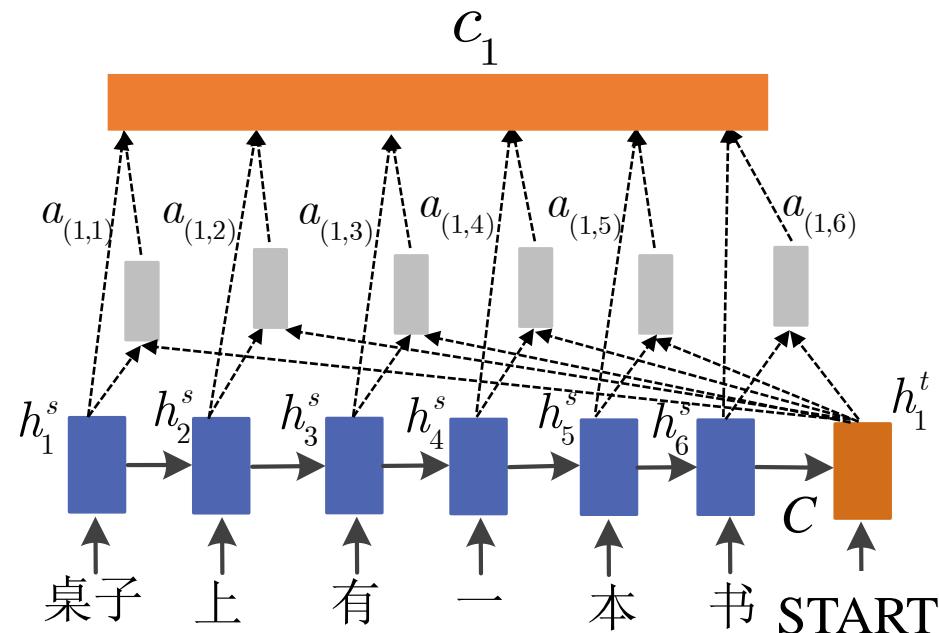
4. 神经机器翻译

● Step B 语义向量计算

注意力权重 对应的隐层向量

$$c_1 = \sum a_{(i,j)} h_j$$

$$\begin{aligned} c_1 &= a_{(1,1)} h_1^s + a_{(1,2)} h_2^s + \dots + a_{(1,6)} h_6^s \\ &= 0.1 \times \begin{bmatrix} 0.337 \\ 0.572 \end{bmatrix} + 0.01 \times \begin{bmatrix} 0.315 \\ 0.729 \end{bmatrix} + \\ &\dots + 0.11 \times \begin{bmatrix} 0.009 \\ 0.543 \end{bmatrix} \\ &= \begin{bmatrix} 0.237 \\ 0.762 \end{bmatrix} \end{aligned}$$



4. 神经机器翻译

● Step C 融合语义向量和目标端隐层状态

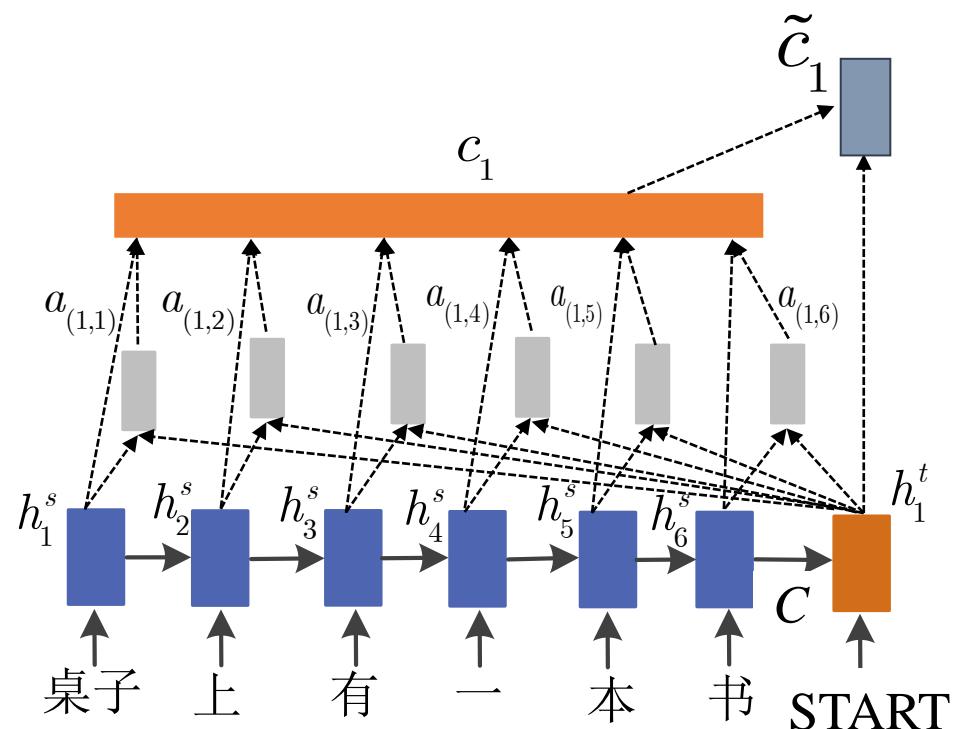
源语言语
义信息

目标语言
隐层状态

$$\tilde{c}_1 = \tanh(Wc_1 + Uh_1^t)$$

$$= \tanh\left(\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \times \begin{bmatrix} 0.237 \\ 0.762 \end{bmatrix} + \begin{bmatrix} 0 & 0.1 \\ 0.1 & 0 \end{bmatrix} \times \begin{bmatrix} 0.693 \\ 0.824 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 0.117 \\ 0.232 \end{bmatrix}$$

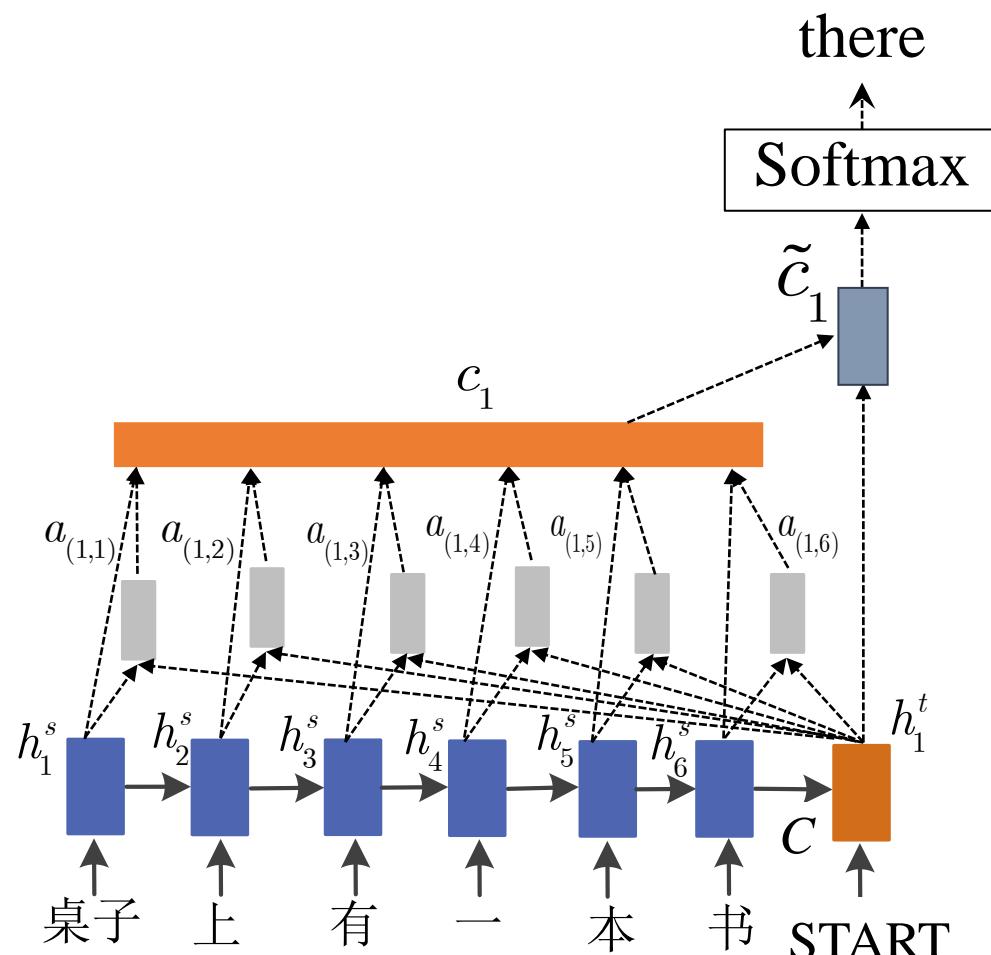


4. 神经机器翻译

● Step D 预测单词

$$score_2 = \text{softmax} \left(\begin{array}{l} \tilde{c}_1 \times \text{the} \\ \dots \\ \tilde{c}_1 \times \text{there} \\ \dots \\ \tilde{c}_1 \times \text{book} \\ \tilde{c}_1 \times \text{EOS} \end{array} \right)$$

重复上述过程，直到标准译文结束(训练)或者输出EOS(测试)。实际系统中会保留top- n 候选。



4. 神经机器翻译

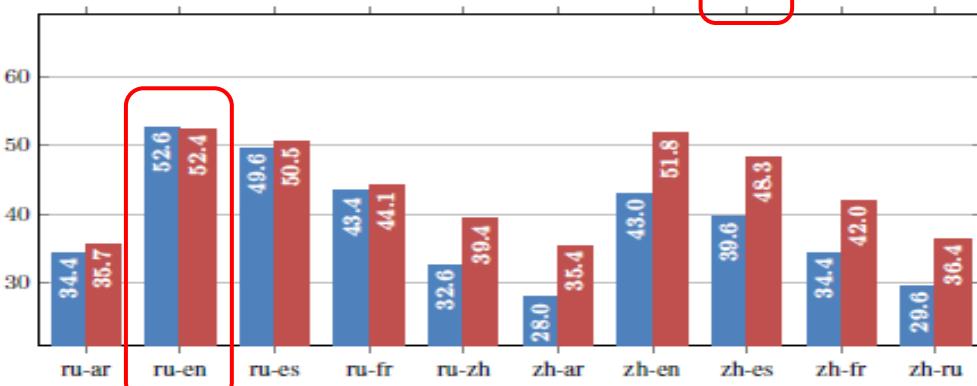
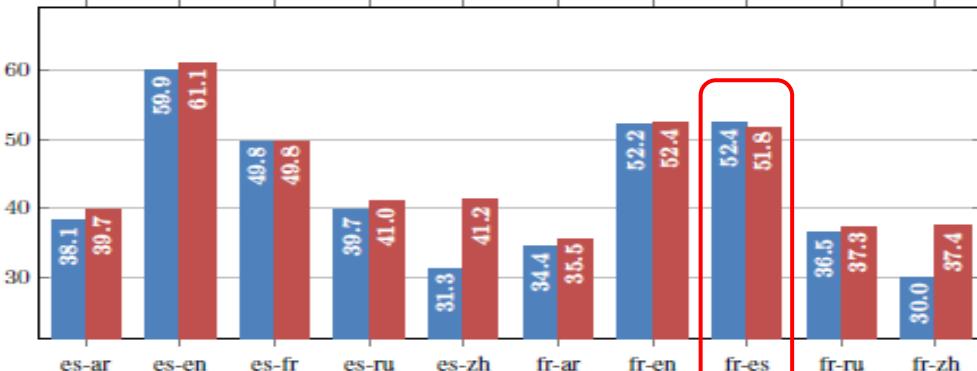
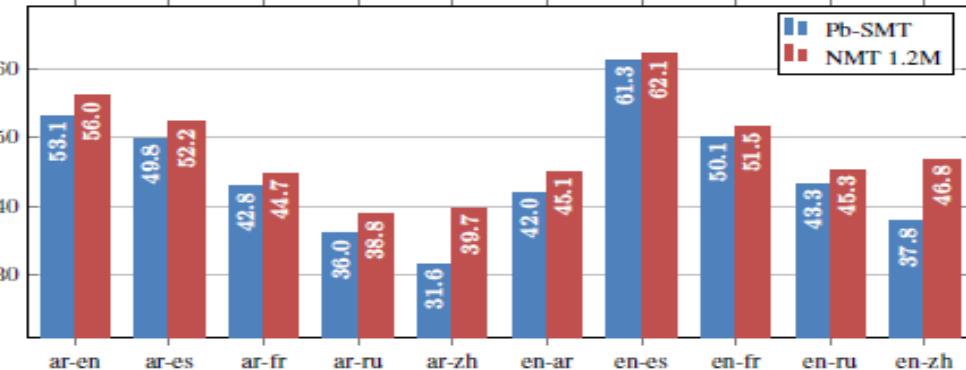
● 开源工具

- <https://github.com/tensorflow/tensor2tensor>
(Google 发布，用Tensorflow 实现，仅支持 Transformer。)
- <https://github.com/pytorch/fairseq>
(Facebook 发布，用 Pytorch 实现，包括RNN-NMT，CNN-NMT 和 Transformer 等。)
- 两个“轻量化的”开源工具：
 - <https://github.com/THUNLP-MT/THUMT> (清华大学发布，用 Tensorflow/ Pytorch/ Theano 实现。)
 - <https://github.com/ZNLP> (中科院自动化所发布，用 Tensorflow/ C /C++ 实现。)

4. 神经机器翻译

◆ SMT vs. NMT

在对比实验时，除了个别语言对，神经机器翻译全面超越统计翻译方法。

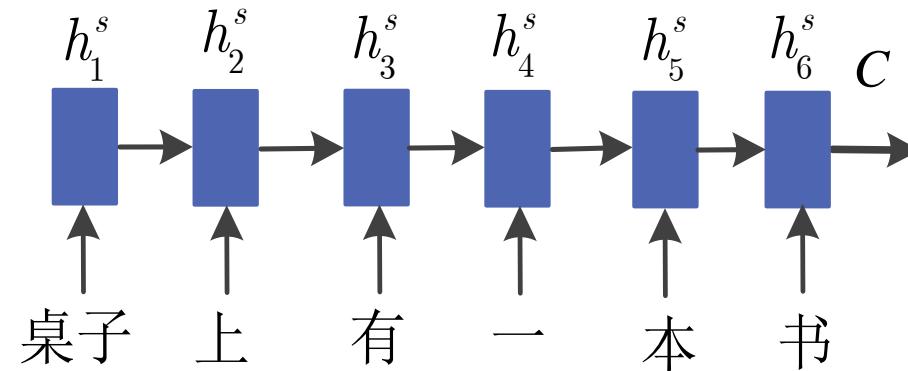


Marcin Junczys-Dowmunt,
Tomasz Dwojak and Hieu Ho
ang, [2016](#). Is Neural Machine
Translation Ready for
Deployment? A Case Study on
30 Translation Directions.
<https://arxiv.org/pdf/1610.01108.pdf>

4. 神经机器翻译

◆ 神经翻译模型的改进

- 基于卷积神经网络的机器翻译方法(CNMT, 2017/07)
- 基于自注意力机制的机器翻译方法(Transformer, 2017/12)
 - 在循环神经网络中，词汇间的语义关系需要逐词传递，远程依赖关系的学习能力不足。
 - 在循环神经网络中，前面单词编码结束后，才能编码后续单词，无法实现并行处理。



4. 神经机器翻译

A. Vaswani et al. [Attention Is All You Need.](#)

Proc. of NIPS'2017, pp 6000–6010

为了缓解上述问题，谷歌提出了基于自注意力机制的机器翻译模型，并将其命名为Transformer。

Transformer 依然遵循**编码器-解码器**的框架，但其基本单元与之前的方法不同，采用了自注意力机制。

(见第6章内容)

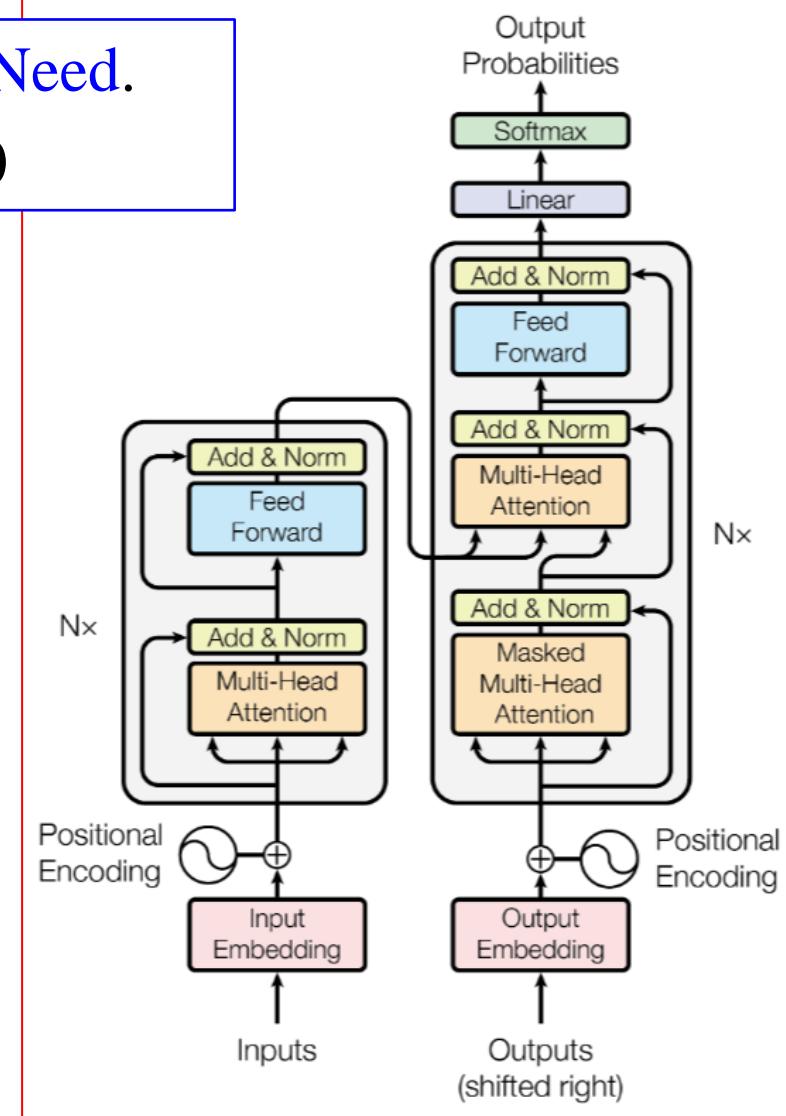
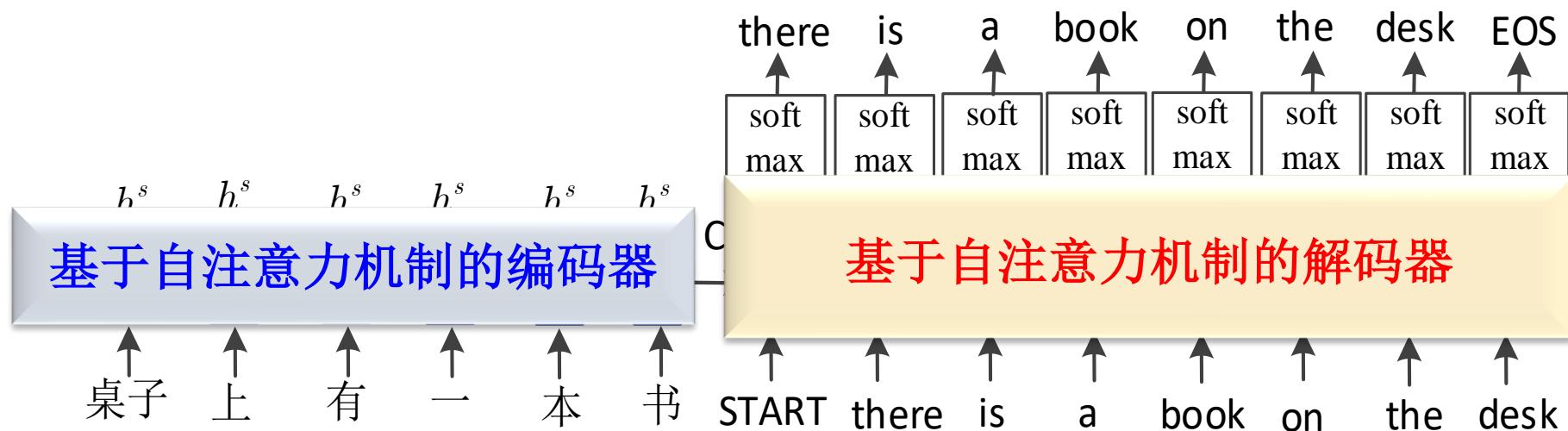


Figure 1: The Transformer - model architecture.

4. 神经机器翻译

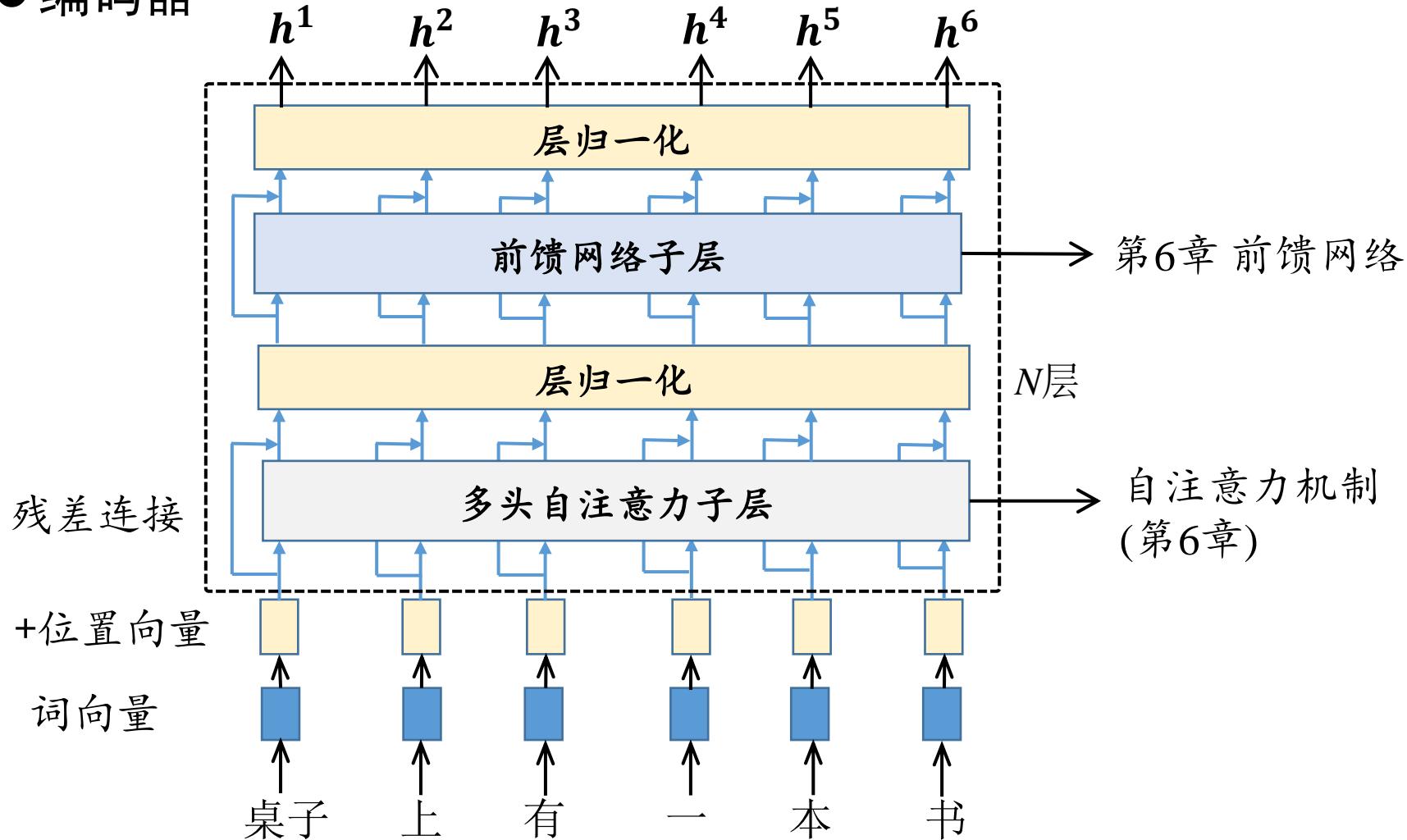
◆ Transformer 框架

● 基本思想



4. 神经机器翻译

● 编码器



4. 神经机器翻译

➤ 位置向量的确定：

- (1) 与词向量一样，随机初始化，然后根据训练数据进行优化；
- (2) 固定：

$$PE_{pos} = \begin{bmatrix} y_0 \\ \vdots \\ y_i \\ \vdots \\ y_{d-1} \end{bmatrix} \quad d \text{ (如: } d=500\text{)}$$

其中， pos 是位置， d 是总的维度数。 $2i$ 或 $2i+1$ 分别表示第 $2i$ 或者 $2i+1$ 维元素， $0 \leq i \leq d/2-1$ 。

$$y_{2i} = \sin\left(\frac{pos}{10000^{(2i/d)}}\right)$$

$$y_{2i+1} = \cos\left(\frac{pos}{10000^{(2i/d)}}\right)$$

桌子 上 有 一 本 书
1 2 3 4 5 6

$$pos = 1, 2, \dots, 6$$

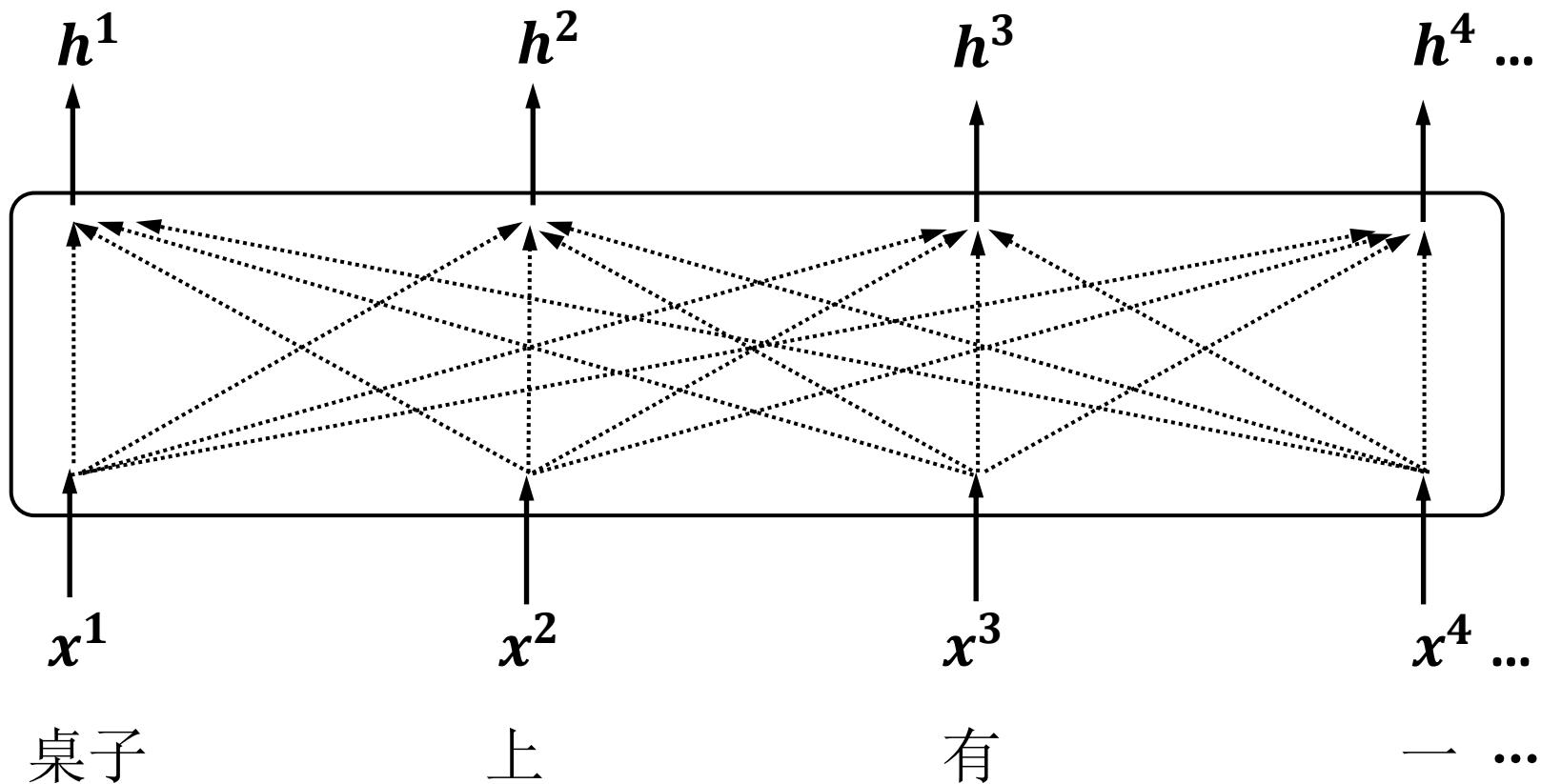
4. 神经机器翻译

● 编码器

- 编码器由 N 层组成，每一层包括2个子层：
 - ①多头自注意力子层；
 - ②前馈网络子层。
- 每个子层之间采用了**残差连接**(residual connection)和**层归一化**(layer normalization)方法。
- 多头自注意力和前馈网络。（参考第6章）

4. 神经机器翻译

➤ 编码器的结构：



4. 神经机器翻译

➤ 残差连接

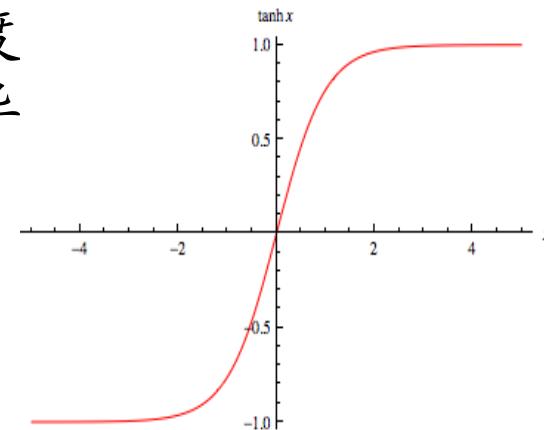
- 目的：残差网络是深度学习中的一个重要方法，能够将误差信号不经过中间权重矩阵变换直接传播到低层，缓解**梯度弥散(gradient diffusion)**问题。

使用反向传播算法时，随着传播深度的增加，梯度的变化幅度可能会减小，导致权重更新非常缓慢，不能有效地学习。

$$z^{(l)} = W^{(l)} \cdot \alpha^{(l-1)} + b^{(l)}$$

$$\alpha^{(l)} = f_l(z^{(l)})$$

$$x = \alpha^{(0)} \rightarrow z^{(1)} \rightarrow \alpha^{(1)} \rightarrow z^{(2)} \rightarrow \dots \rightarrow \alpha^{(L-1)} \rightarrow z^{(L)} \rightarrow \alpha^{(L)} = \phi(x; W, b))$$



(见第3章)

4. 神经机器翻译

- 解决方法：

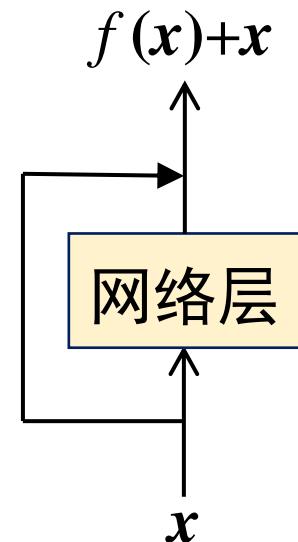
① **正向计算时**，将输入 x 与网络层变换 $f(x)$ 后直接相加，作为该层的最终输出结果，如图所示；

② **反向传播时**，误差对输入 x 的导数 $\frac{\partial E}{\partial x}$ 将增加一个常数项1，从而将误差对输出 $h(x)$ 的导数 $\frac{\partial E}{\partial h}$ 直接传递给输入 x ，以缓解梯度弥散问题。

误差 E 对输入 x 的导数 $\rightarrow \frac{\partial E}{\partial x} = \frac{\partial E}{\partial h} \times \frac{\partial h}{\partial x}$ 常数项1

$$= \frac{\partial E}{\partial h} \times \left(\frac{\partial f}{\partial x} + 1 \right) = \boxed{\frac{\partial E}{\partial h} \times \frac{\partial f}{\partial x}} + \boxed{\frac{\partial E}{\partial h}}$$

经过网络层后的导数连续
乘法可能会导致梯度弥散

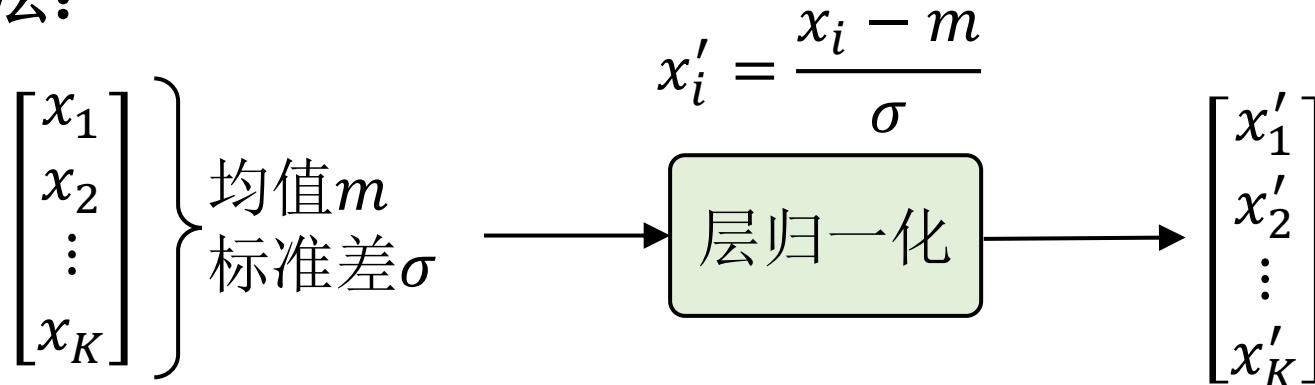


将误差对输出导数直接传递给输入 x 。

4. 神经机器翻译

➤ 层归一化

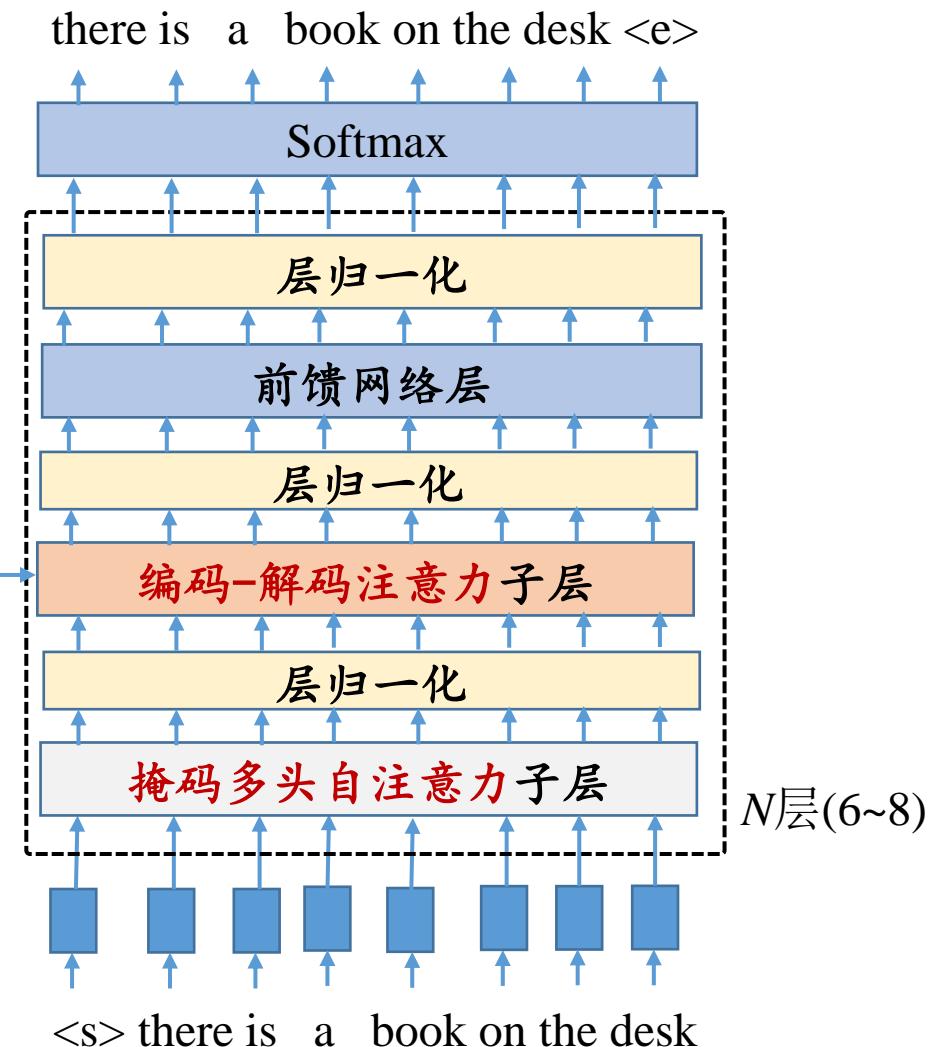
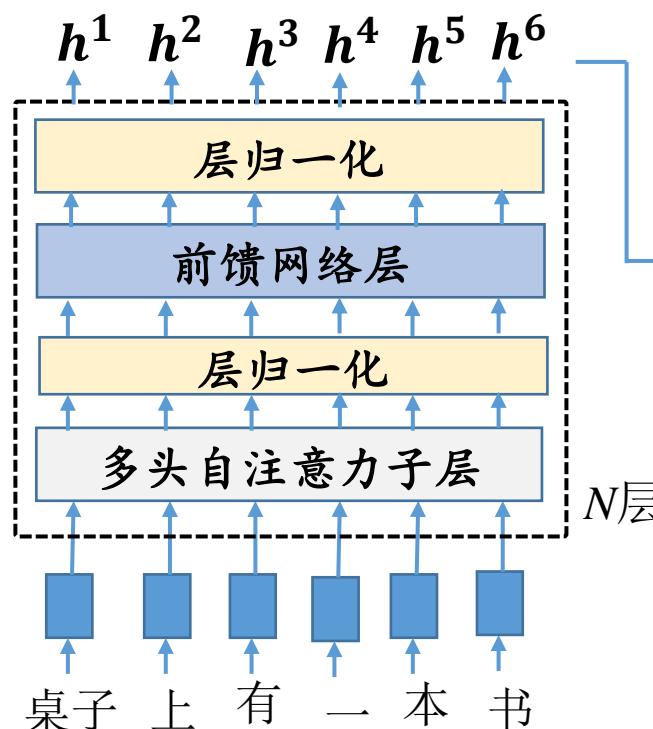
- 目的：使得网络中每层输入数据的分布相对稳定，加速模型学习速度。
- 方法：



Jimmy Lei Ba, Jamie Ryan Kiros, Geoffrey E. Hinton. 2016. [Layer normalization](#). *arXiv preprint arXiv:1607.06450*, 2016.

4. 神经机器翻译

● 解码器



注: 残差连接未标明

4. 神经机器翻译

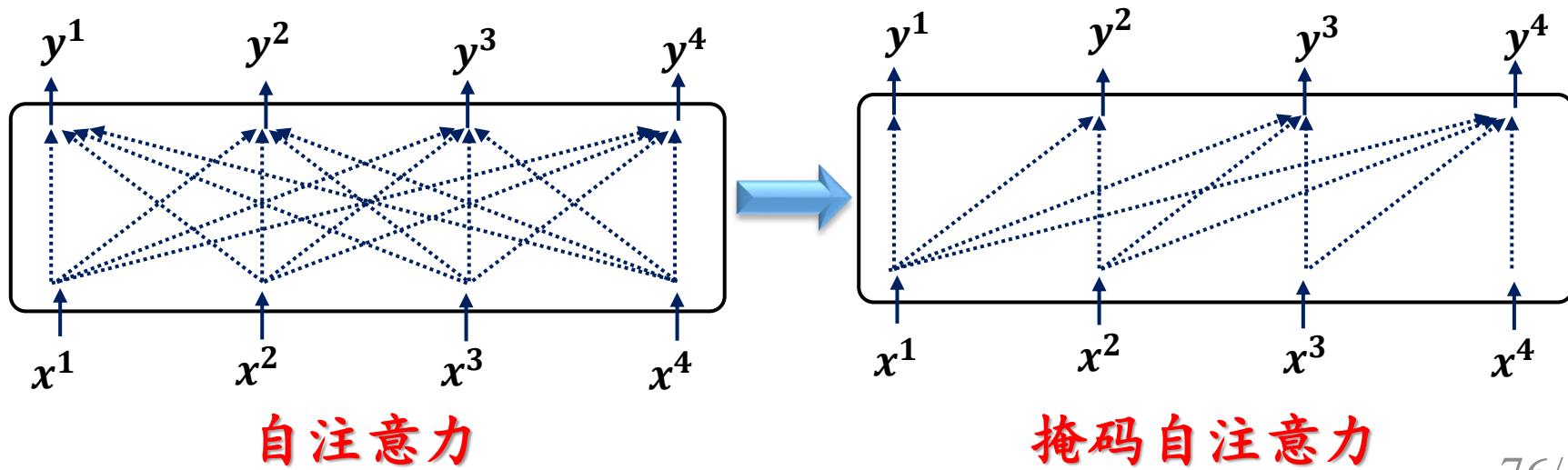
● 解码器

- 解码器由 N 层组成，每一层包括3个子层：
 - ①掩码多头自注意力子层；
 - ②编码-解码注意力子层；
 - ③前馈网络子层。
- 每个子层之间同样采用了**残差连接**(residual connection)和**层归一化**(layer normalization)方法。

4. 神经机器翻译

● 解码器

- ✧ **掩码多头自注意力子层**是指在对于某一个目标端单词而言，只与**之前**单词做注意力计算，即把后面的单词掩盖。
- ✧ **测试和训练阶段的不同：**训练阶段，标准译文是已知的，而在测试阶段，模型仅已知之前的单词（模型自身预测得到）。
- ✧ 为了保证模型在训练阶段和测试阶段保持一致，因此在测试阶段，需要将后续的单词掩盖。



4. 神经机器翻译

● 解码器

- ◆ 编码-解码注意力子层是指在对于某一个目标端单词而言，只与源端单词的隐层状态做注意力计算。
- ◆ 编码-解码注意力子层能够在解码过程中利用源语言信息生成单词，具体实现方法：

注意力计算中查询向量(q)来自目标端单词的隐层状态，而键向量(k)和值向量(v)来自源端隐层状态。

4. 神经机器翻译

- Transformer 框架的优点：

- (1) 自注意力机制能够直接建模任意两个词汇之间的语义关系，提升对于词汇间远程依赖关系的建模能力。
- (2) 模型在训练阶段，可以同时对所有源端单词和目标端单词进行编码和解码，提升模型并行计算能力。

到目前，Transformer不仅用于机器翻译任务，而且在自然语言处理的其它任务、计算机视觉和图像处理、语音识别等很多任务中均有重要应用。

4. 神经机器翻译

◆ 数据驱动方法的评价：

优点: 不需要对源语言句子进行深层次分析，甚至对源语言没有任何基本知识，只要有足够多的双语言句对就可以建立一个机器翻译系统；一般而言，神经机器翻译的译文流畅性较好。

弱点: 对于某些语言对来说，收集高质量大规模的双语句对并不是一件容易的事情；机器翻译过程和译文结果缺乏解释性；对于复杂句子、生僻词汇、指代等问题仍缺乏得利的处理手段。

本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
-  5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读

5. 译文质量评估

◆译文评估方法

- 主观评测：(1)流畅度；(2)充分性；(3)语义保持性。
- 客观评测(自动评价)
 - 句子错误率
 - 单词错误率 (mWER)
 - 位置无关的单词错误率 (mPER)
 - METEOR: 综合考虑译文词汇的准确率、召回率、F值等
 - BLEU评价指标
 - NIST 评价指标
 -

5. 译文质量评估

◆ BLEU评价方法 [Papineni, 2002]

— BiLingual Evaluation Understudy

基本思想:

将机器翻译产生的候选译文与人翻译的多个参考译文相比
较，越接近，候选译文的正确率越高。

实现方法:

统计同时出现在系统译文和参考译文中的 n 元词的个数，最
后把匹配到的 n 元词的数目除以系统译文的单词数目，得到评
测结果。

5. 译文质量评估

例如：系统译文： the cat is sitting on the mat

参考译文1： the cat is on the mat

参考译文2： there is a cat on the mat

如果 $n=1$ ，系统译文中共有7个词，其中6个词出现在答案中。

那么，得分为： $6/7 = 0.86$ 。

如果 $n=2$ ，系统译文中的2元词组序列（bi-gram）共有6个： the cat, cat is, is sitting, sitting on, on the, the mat，其中4个出现在参考答案中，因此得分为： $4/6 = 0.67$ 。

问题①：随着 n 值的增大，评价得分值几乎成指数级下降。



5. 译文质量评估

如果系统译文为: the cat

the the the the the 得分: 5/5

参考译文1: the cat is on the mat

参考译文2: there is a cat on the mat

不管 $n=1$, 还是 $n=2$, 得分都是 1。

问题②: 译文越短的确定性翻译, 得分越高。

修正: 采用 n 元语法精度的对数加权平均值, 取精度的几何平均值, n 值最大为4。同时对长度较短的译文得分进行惩罚。

5. 译文质量评估

$$BLEU = BP \times \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

长度过短句子的惩罚因子

最大语法的阶数，实际取4。

$$w_n = 1/N$$

出现在答案译文中的 n 元词语接续组占候选译文中 n 元词语接续组总数的比例。

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

c 为系统译文中单词的个数， r 为答案译文中与 c 最接近的译文单词个数。

BLEU 分值范围：0 ~ 1，分值越高表示译文质量越好，分值越小，译文质量越差。

5. 译文质量评估

例如：系统译文： the cat is sitting on the mat

参考译文1： the cat is on the mat

参考译文2： there is a cat on the mat

假设 $N = 3$ ，权重 $w_i = 1/N = 1/3 (i = 1, 2, 3)$ 。

当 $n = 1$ 时， $p_1 = 6/7 = 0.86$ 。

当 $n = 2$ 时， $p_2 = 4/6 = 0.67$ 。

当 $n = 3$ 时，译文中共有 5 个 3 元组： the cat is, sitting on the, cat is sitting, on the mat, is sitting on, 其中 2 个出现在答案中，所以， $p_3 = 2/5 = 0.40$ 。

$$BLEU = 1 \times \exp[(\ln 0.86 + \ln 0.67 + \ln 0.40)/3]$$

$$= \exp[(\ln 0.23048)/3]$$

$$= 0.61$$

系统译文的长度
 $c = 7$, 参考译文中与系统译文最接近的长度 $r = 7$ 。

$c = r$, 所以
 $BP = e^0 = 1$ 。

$$BLEU = BP \times \exp\left(\sum_{n=1}^N w_n \ln p_n\right)$$

5. 译文质量评估

◆ 影响BLEU值的诸多因素

- 训练语料与测试样本之间的差异性
- 训练语料和测试样本的规模大小
- N 的取值
- 参考译文的个数
- 汉语（译文）统计单位，以“字”还是以“词”为单位？

5. 译文质量评估

- ◆人工对专业翻译水平的评价方法
- 全国大学生英语考试四级（CET-4）评分标准

分数档 (满分15分)	评分标准
13-15分	译文准确表达了原文的意思。用词贴切，行文流畅，基本上无语言错误，仅个别小错。
10-12分	译文基本上表达了原文的意思。文字通顺、连贯，无重大语言错误。
7-9分	译文勉强表达了原文的意思。用词欠准确，语言错误相当多，其中有些是严重语言错误。
4-6分	译文仅表达了一部分原文的意思。用词不准确，有相当多的严重语言错误。
1-3分	译文支离破碎。除个别词语或句子，绝大部分文字没有表达原文意思。
0分	未作答，或只有几个孤立的词，或译文与原文毫不相关。

5. 译文质量评估

● 全国翻译专业资格考试（CATTI）评分标准

➤ 二级笔译实务评分标准

等级	汉译英/英译汉 (满分25分)	评分标准
高 (优秀)	20—25分	<ul style="list-style-type: none"> 译文忠实于原文，能准确译出80%以上的重要考点； 基本无语法错误，基本无拼写错误，标点符号使用正确； 准确流畅，文字简练，用词恰当。
中 (及格)	15—19分	<ul style="list-style-type: none"> 译文基本忠实于原文，能准确译出60%以上的重要考点； 有少量语法错误、拼写错误和标点符号错误； 文字表达基本正确，用词基本恰当。
低 (不及格)	0—14分	<ul style="list-style-type: none"> 译文错译、漏译严重，只能译出60%以下的重要考点； 有较多语法错误、拼写错误和标点符号错误； 句型结构受原文影响严重，用词不当。

5. 译文质量评估

➤ 三级笔译实务评分标准

等级	汉译英/英译汉 (满分50分)	评分标准
高 (优秀)	40—50分	<ul style="list-style-type: none"> ① 译文忠实于原文，能准确译出80%以上的重点考点； ② 基本无语法错误，基本无拼写错误，标点符号使用正确； ③ 准确流畅，文字简练，用词恰当。
中 (及格)	30—39分	<ul style="list-style-type: none"> ① 译文基本忠实于原文，能准确译出60%以上的重点考点； ② 有少量语法错误、拼写错误和标点符号错误； ③ 文字表达基本正确，用词基本恰当。
低 (不及格)	0—29分	<ul style="list-style-type: none"> ① 译文错译、漏译严重，只能译出60%以下的重点考点； ② 有较多语法错误、拼写错误和标点符号错误； ③ 句型结构受原文影响严重，用词不当。

5. 译文质量评估

● 人工与BLEU的对比结论

- 整体而言，BLEU值与专家评分正相关，BLEU值越高，译文质量好的可能性越大，在英译翻译上表现明显；
- 不同等级/题型的译文BLEU值区间覆盖不同，相同BLEU值区间下的专家评分不同、通过率也不同。难以建立能够覆盖所有类型译文的BLEU值/区间与人工评分（译文质量高低）之间的直接映射；
- BLEU 较高的系统译文与人工评价差异较大的可能性约高；
- 如果人工给出的参考译文文学色彩较浓，或者意译成分较大的话，系统译文的BLEU值较低。

5. 译文质量评估

◆ 国际评测活动

- DARPA 资助、NIST组织的机器翻译评测(简称 NIST 评测), 2002年6月首次评估, 几乎每年一次, 组织了很多次。
- 欧盟资助的机器翻译评测 WMT(Workshop on Machine Translation), 2006年起, 几乎每年一次, 主要针对欧洲语言之间的翻译, 后来增加了汉语。
- 全国机器翻译评测 (CWMT: China Workshop on MT) 2007, 2008, 2009, 2011, 2013, 2015, 2017, 2018。2019年起更名为 CCMT (China Conference on MT)。
- IWSLT(International Workshop on Spoken Language Translation) 口语翻译评测, 2004 年起, 每年一次。

5. 译文质量评估

➤ IWSLT'2018:

- <https://sites.google.com/site/iwsltevaluation2018/>
- TED corpus (English - German): Technology, Entertainment, Design

➤ IWSLT'2019:

- <https://sites.google.com/view/iwslt-evaluation-2019/speech-translation?authuser=0>
- TED corpus (English - German)
- How2 corpus (<https://github.com/srvk/how2-dataset>) (English - Portuguese): 多模态语料，包含80000个视频(2000小时)，英文字幕和摘要，其中300小时包含葡萄牙语翻译。

➤ IWSLT'2020: <http://iwslt.org/doku.php?id=evaluation>

➤ IWSLT'2021: <https://iwslt.org/2021/>

➤ IWSLT'2022: <https://iwslt.org/2022/>

5. 译文质量评估

● CCMT (CWMT):

- 2021: <http://sc.cipsc.org.cn/mt/conference/2021/>
- 2022: <http://sc.cipsc.org.cn/mt/conference/2022/>

本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读





6. 语音翻译原理

◆概念

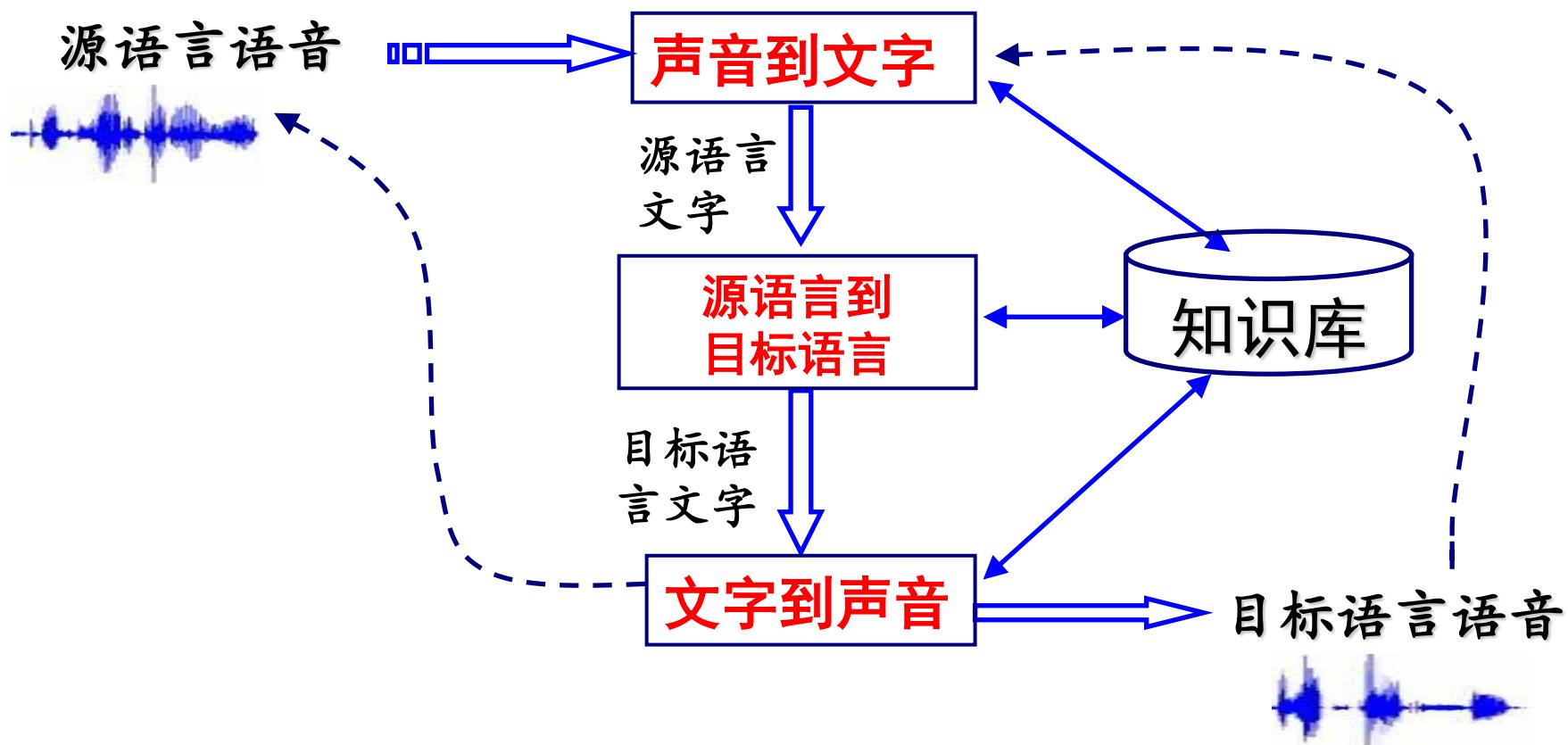
- **语音翻译**(speech-to-speech translation) 就是用计算机系统实现持不同语言的说话人之间的话语翻译的过程。
- 由于语音识别技术和口语理解技术的提高，语音翻译研究越来越倾向于探索真实对话场景下自然口语风格的话语翻译技术，因此，语音翻译一般又称 **口语翻译**(spoken language translation, SLT)，或**对话翻译**(spoken dialogue translation)。
- **里程碑：**1989年美国CMU开发的 SpeechTrans 实验系统，实现医生与病人对话的英日翻译

6. 语音翻译原理

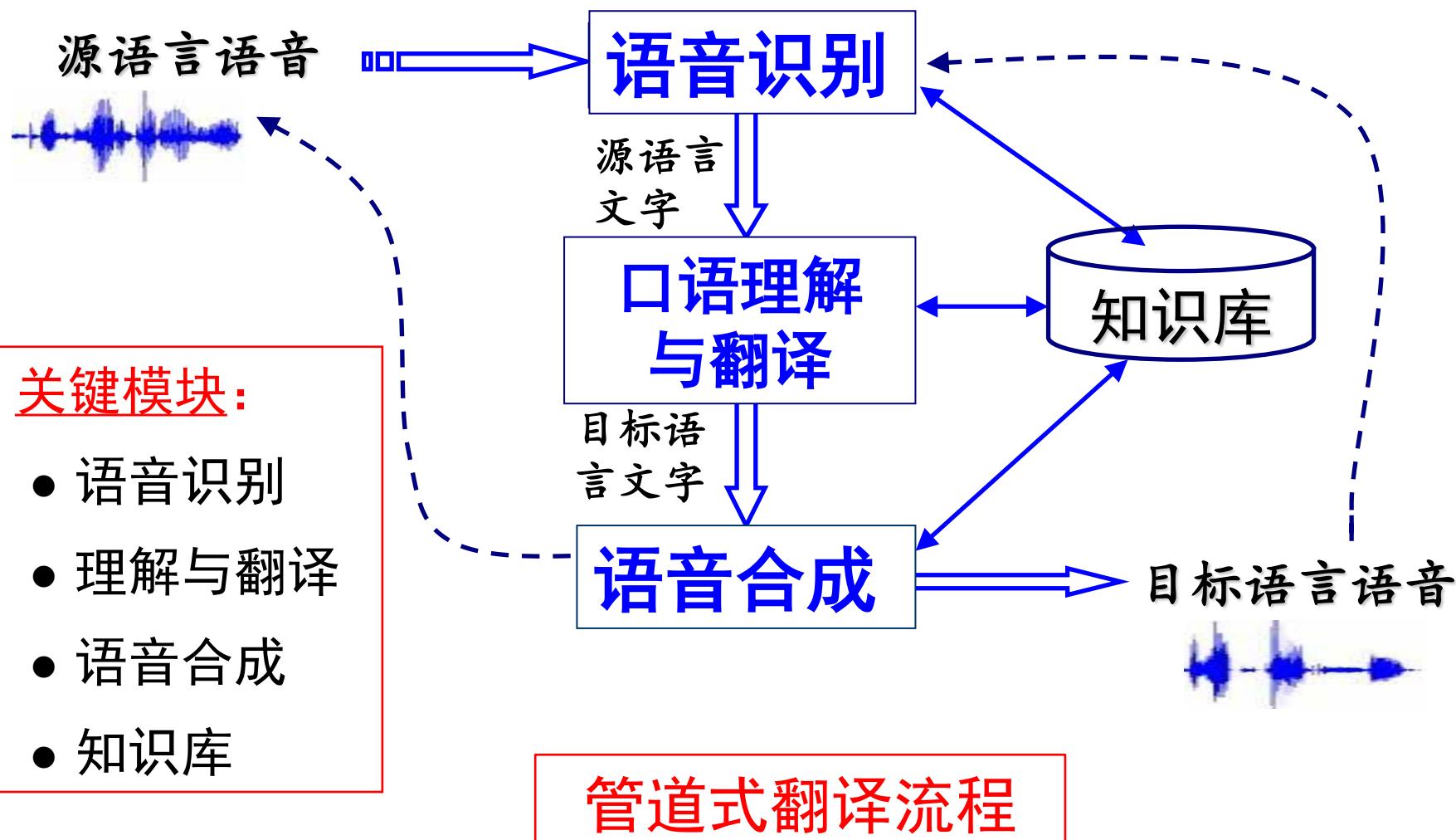
◆ 原理



6. 语音翻译原理



6. 语音翻译原理



6. 语音翻译原理

◆ 挑战

系统面对的是复杂多变的口语输入

- ▶ 重复 ▶ 冗余 ▶ 省略 ▶ 修正 ▶ 词序颠倒
- ▶ 长时间停顿等 ▶ 无标点

例子：那个 我想问一下 那个 你们这里 那里 有没有那个
房间 就是单人间 噢 不 双人间 便宜点儿的 星期三住

系统工作环境复杂

- ▶ 周围环境噪音 ▶ 语音传输过程中产生的非语声信号

系统翻译机制处理的是含有错误信息的字符串输入

- ▶ 含有错误的字或词

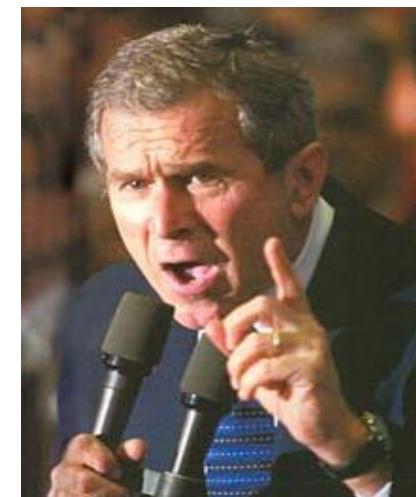
6. 语音翻译原理

系统尚无法有效地获得和利用对话过程中的非语言信息

- ▶ 语气
- ▶ 手势或动作
- ▶ 表情

人们希望系统输出自然流畅的口语语音

- ▶ 语音合成器需要模拟对话双方话语的韵律特征



6. 语音翻译原理

◆ 口语翻译与书面语翻译对比

对比要素	口语翻译	书面语翻译
句长	8个词/句， 1.8字/词	22.5个词/句， 2.5字/词
句子结构	简单	复杂
规范性	不规范，省略、重复、修正等现象普遍存在	较规范
用词	简单，常用语大约2~3万词，少用成语	词汇量大，大约5万左右的普通词，5万以上专用词。
噪声	语音识别产生噪声	无噪声
副语言信息	重要	无

6. 语音翻译原理

◆ 端到端的翻译流程

5.1 关于目前语音翻译的技术路线

在目前的语音翻译系统中，都是首先将源语言的语音信号转换成文字，然后再对文字进行分析、转换、生成，最后将译文转换成语音信号输出，如图 2 中的 L2。而实际上人在进行口语翻译时，是直接从语音到语音的，尤其当翻译者对源语音比较熟悉时，根本不需要从音到字、再从字到音的转换过程，如图 2 中的 L1。

那么，在目前的语音翻译系统中，采用 L2 处理过程显然是一种不得已而为之的处理方案，而这种方案是在计算机硬、软件技术（包括内存、运算速度和知识表示方法等）受到一定局限的情况下诞生的，可以设想，当计算机的内存和运算速度足够大，语言知识包括语音知识在内的语义知识、语用知识和语境等知识得到充分表达和利用时，直接实现从语音到语音的翻译是完全有可能做到的事情。

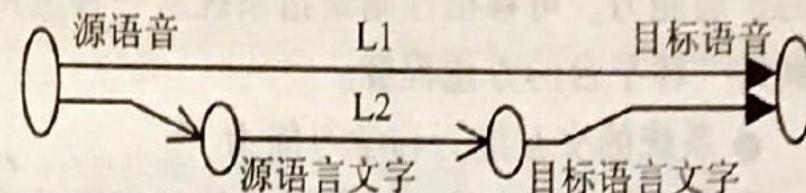


图 2 口语翻译技术路线

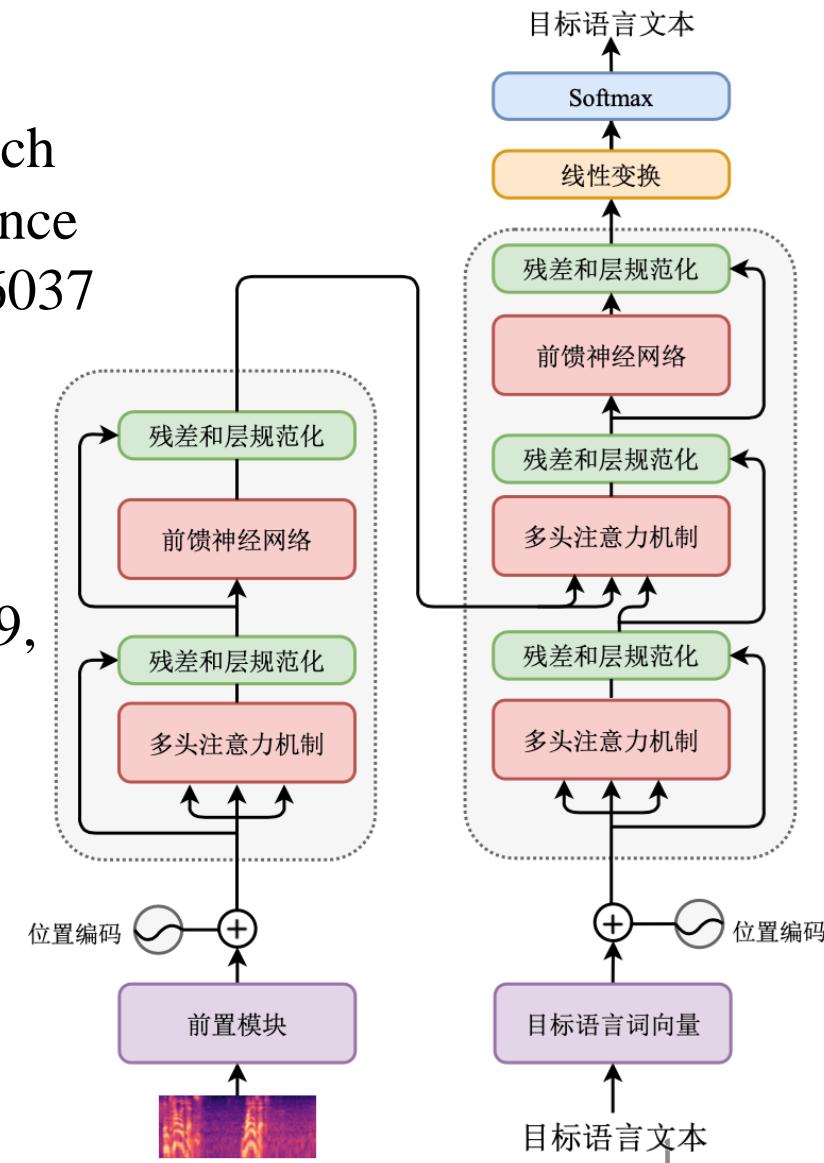
6. 语音翻译原理

● 从语音到语音的翻译(Google)

Ye Jia et al. 2019. Direct speech-to-speech translation with a sequence-to-sequence model. arXiv preprint arXiv:1904.06037

● 从语音到文本的翻译

Yuchen Liu et al. End-to-End Speech Translation with Knowledge Distillation. *Proc. of Interspeech'2019*, pp.1128-1132



本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题
10. 附录：延伸阅读



7. 双语语料获取

◆ 已有的部分机器翻译用双语平行语料

- 加拿大议会会议录 (Canadian Hansards)
<http://www.isi.edu/natural-language/download/hansard/>
- 克姆尼茨英-德翻译语料库 (Chemnitz E-G Translation Corpus)
<http://www.tu-chemnitz.de/phil/english/chairs/linguist/real/independent/transcorpus/index.htm>
- 英语-挪威语平行语料库 (ENPC)
<https://www.hf.uio.no/ilos/english/services/omc/enpc/>
- 葡-英双向平行语料库 (Compara)
<http://www.linguateca.pt/COMPARA/Welcome.html>
- 香港立法委员会会议记录 (Hong Kong Hansards)
<http://catalog.ldc.upenn.edu/LDC2000T50>

◆ 机器翻译评测活动提供部分训练集和开发集

7. 双语语料获取

◆问题分析

- 手工构造平行语料库费时费力，无论是规模、语种覆盖面、时效性，还是领域平衡性等方面都不能满足处理真实文本的需求。
- 网络是一个巨大的多语种语料库，大量存在可比语料，如 Wikipedia、双语新闻，《人民日报》和 China Daily等，包含对机器翻译有帮助的平行资源，可从中抽取平行句对、平行片段及双语词汇等。



7. 双语语料获取

机器翻译（Machine Translation，经常简写为MT，俗称机翻）属于计算语言学（Computational Linguistics）的范畴，其研究借由计算机程序将文字或演说从一种自然语言翻译成另一种自然语言。简单来说，机器翻译是通过将一个自然语言的字辞取代成另一个语言的字辞。借由使用语料库的技术，可达成更加复杂的自动翻译，包含可更佳的处理不同的文法结构、词汇辨识、惯用语的对应等。

目前的机器翻译软件通常可允许针对特定领域或是专业（例如天气预报）来加以客制化，目的在于将词汇的取代缩小于该特定领域的专有名词上，以借此改进翻译的结果。这样的技术针对一些使用较正规或是较制式化陈述方式的领域来说特别有效。例如像是政府机关或是法律相关文件，这类型的文句通常皆比起一般的文句来的正式与制式化，其机器翻译的结果往往比起像是日常生活的对话这种非正式文件要来的好。

目前的一些翻译机器，例如AltaVista Babelfish，有时可以得到可以理解的翻译结果，但是想要得到较有意义的翻译结果，往往需要在输入语句时适当地编辑，以利电脑程式分析。

Machine translation, sometimes referred to by the abbreviation **MT** (not to be confused with **computer-aided translation**, **machine-aided human translation (MAHT)** or **interactive translation**) is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another.

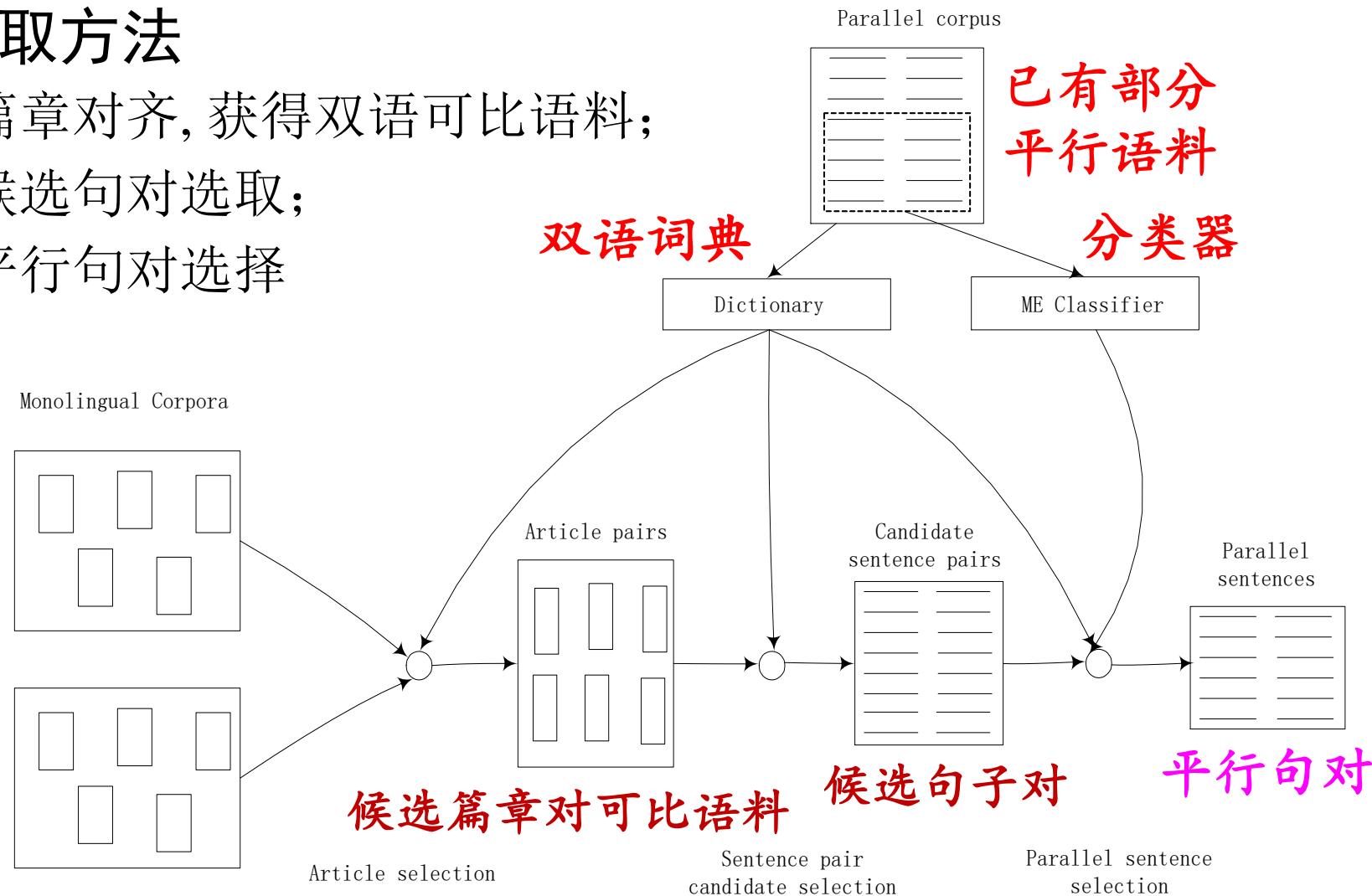
On a basic level, MT performs simple substitution of words in one natural language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed. Solving this problem with **corpus** and **statistical techniques** is a rapidly growing field that is leading to better translations, handling differences in **linguistic typology**, translation of **idioms**, and the isolation of anomalies.^[1]

Current machine translation software often allows for customization by domain or profession (such as weather reports), improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. It follows that machine translation of government and legal documents more readily produces usable output than conversation or less standardised text.

7. 双语语料获取

◆ 抽取方法

- (1) 篇章对齐，获得双语可比语料；
- (2) 候选句对选取；
- (3) 平行句对选择





7. 双语语料获取

(1) 双语可比语料的获取

➤ 候选双语网站的获取

- 指定一系列含有双语网页的网站
- 利用搜索引擎和启发式信息
- 将双语资源的获取限制在可能的网站上

<http://www.fao.org/news/story/zh/item/170600/icode/>

<http://www.fao.org/news/story/en/item/170536/icode/>

➤ 网页采集

- <http://english.peopledaily.com.cn/90883/8584635.html>
- 预先定义与语言相关的字符串，对爬虫进行限制

➤ 网页预处理

- 保留网页的标题、正文、URL信息，UTF-8编码



7. 双语语料获取

(2) 双语候选句对的选择

- 直接处理搜索空间庞大 ($P \times Q$)
 - 基于篇章的候选句对选择方法
 - 基于跨语言检索的候选句对选择方法
- **基本思想：**两个句子中的词汇至少有一部分是互译的
- **实现方法：**使用词过滤器进行过滤，根据种子词典保留翻译覆盖率在一定阈值(0.3)以上的作为候选句对。

7. 双语语料获取

例如：

① 存储目标端句子及其长度

② 给定一个源语言句子：

Bush defends his economic stimulus package.

将其映射到目标端：

布什 辩护 他的 经济 振兴 计划。

③ 搜索候选句子（利用布尔模型计算相似句子，确定候选集）



7. 双语语料获取

(3) 构建分类器，获取双语平行句对或片段

- 平行句对的识别
- 平行片段的抽取

常用的分类器特征：

- 句子长度特征/互译单词的比例
- 没有对齐单词点的单词个数所占的比例
- 单词翻译繁衍率 (fertilities) 特征
- 最长有连续对齐点/不对齐点片段的长度
- 汉英句子之间的余弦相似度
- 词对齐概率/翻译“哨兵”特征(如标点，无歧义对齐点等)
-

7. 双语语料获取

语言、文字、信仰、道德规范，形成了各个国家和民族的文化圈。

language, writing, belief, and morals have formed the culture circle of each country and nation.

.....

语言、文字 => language, writing

语言、文字、 => language, writing,

语言、文字、信仰 => language, writing, belief

语言、文字、信仰、 => language, writing, belief,

语言、文字、信仰、道德 => language, writing, belief, and morals

.....

本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
-  8. 技术现状
9. 习题
10. 附录：延伸阅读

8. 技术现状

◆ 已有大量的商用机器翻译系统

GoogleTranslate: <https://translate.google.com>

Bing Translate: www.bing.com/translator

SYSTRANet: <http://www.systranet.com/translate>

TRADOS: <https://www.trados.com.cn/>

百度在线翻译: <http://fanyi.baidu.com/>

有道在线翻译: <http://fanyi.youdao.com/>

金山快译: <http://www.suiniyi.com/s/fyjq/202848/>

腾讯翻译君: <https://fanyi.qq.com/>

搜狗翻译: <https://fanyi.sogou.com/>

.....

8. 技术现状

◆ 技术现状

文本翻译

①原文: Beijing made a third solemn **representation** to Manila and warned that it is hard to be optimistic about a territorial impasse over an island. Authorities say they have prepared for any escalation of the situation by Manila.

(<http://www.chinadaily.com.cn/>, 8 May 2012)

Google Translator 译文:

8. 技术现状

2012.5.8~

..... 技术进展是显而易见的

- ◆ 北京由第三严正交涉到马尼拉，并警告说这是很难约了一个岛领土僵局持乐观态度。当局说，他们已经为马尼拉局势的升级准备。[\(2016.5.1\)](#)
- ◆ 北京对马尼拉进行了第三次庄严的代表，并警告说，对岛上的领土僵局很难看好。当局表示，他们为马尼拉的情况升级做好了准备。[\(2017.5.15\)](#)
- ◆ 北京向马尼拉提出了第三次严正交涉，并警告说很难对一个岛屿的领土僵局持乐观态度。当局表示，他们已经为马尼拉局势升级做好了准备。[\(2018.5.20\)](#)
- ◆ 北京向马尼拉提出了第三次庄严的**代表**，并警告说，对一个岛屿的领土僵局很难保持乐观。有关当局表示，他们已经为马尼拉的局势升级做好了准备。[\(2019.5.6\)](#)
- ◆ 北京对马尼拉进行了第三次庄严的**代表**，并警告说，对一个岛屿的领土僵局不容乐观。当局说，他们已经为马尼拉局势的任何升级做好了准备。[\(2020.10.21\)](#)
- ◆ 北京第三次向马尼拉提出严正交涉，并警告说，对一个岛屿的领土僵局很难乐观。当局表示，他们已为马尼拉局势升级做好准备。[\(2021.8.24\)](#)
- ◆ 北京向马尼拉提出了第三次严正交涉，并警告说，很难对一个岛屿的领土僵局感到乐观。当局表示，他们已为马尼拉局势升级做好准备。[\(2022.5.24\)](#)



8. 技术现状

仍存在大量问题

②原文：美方提出的所谓“合作、竞争、对抗”三分法就是“障眼法”。合作是权宜之计，竞争是话语陷阱，对抗遏制是本质。有求于中方时就要求合作，在有优势的领域就脱钩制裁，为了遏制不惜冲突对抗。美国的企图注定不会得逞，在台湾问题上，中国是有底气的。

(https://mbd.baidu.com/newspage/data/landingsuper?context=%7B%22nid%22%3A%22news_8914846101828938243%22%7D&n_type=0&p_from=1)

(2021.10.2)

The so-called "cooperation, competition, and confrontation" rule of thirds proposed by the United States is the "obscure approach." Cooperation is a stopgap measure, competition is a discourse trap, and confrontation and containment are the essence. When there is a need for China, cooperation is required, and sanctions are decoupled in areas where it has advantages, in order to curb conflict and confrontation. The US attempt is doomed to fail, and China has the confidence on the Taiwan issue.

Google Translator, 2021.10.2



8. 技术现状

她穿着一件灰色的风衣，带着一副茶色的眼镜，左肩挎着一个白色的高档包，右手领着一条白色的小狗，**据说是局长的女儿。**

Google Translator, 2022.5.24

据说是只名犬 / 据说是个模特

She was wearing a gray trench coat, with a pair of tea-colored glasses, a white high-end bag slung over her left shoulder, and a white puppy in her right hand, **which is said to be the daughter of the director.**

She was wearing a gray trench coat, with a pair of brown glasses, a white high-end bag slung over her left shoulder, and a white puppy, **said to be a famous dog**, in her right hand.

She was wearing a gray trench coat, with a pair of tea-colored glasses, a white high-end bag slung over her left shoulder, and a white puppy in her right hand. **She was said to be a model.**



8. 技术现状

③原文：《习近平谈“一带一路”》收入的42篇重要文稿，深刻阐述了“一带一路”建设的重大意义、指导原则、丰富内涵、目标路径等重大问题，对于推动“一带一路”国际合作不断取得新进展，造福沿线各国人民，推动构建人类命运共同体，具有十分重要的指导意义。

The 42 important manuscripts from "Xi Jinping Talks on the "One Belt and One Road"" profoundly elaborated on the major issues of the "Belt and Road" construction, such as the great significance, guiding principles, rich connotations, and target paths. They have continuously made new achievements in promoting the "Belt and Road" international cooperation. Progress, to benefit the people of all countries along the route, and to promote the building of a community with a shared future for mankind is of very important guiding significance. (2021.10.2)

在复杂长句翻译中存在漏翻、错翻和前后不一致等问题。



8. 技术现状

④ 面对原文中的噪声常常束手无策

印尼语、马来语、菲律宾语三种语言的主流媒体中新闻文本
分别有20%、14%和10%以上的词汇存在拼写错误。

如印尼语中表示“失败”的词“gagal”，常被“ga2l”替代；
为了强调语气，有时会重复输入字符，“cemungudh”常被写为
“cemunguuuuudh”。

Arrive at tks at 0800hrs → 在08时00分到达塔克斯。

whr cldy → The weather is cloudy

The army leaves #loc1 for #loc2. → 部队离开#loc1前往#loc2。

⑤ 俚语/成语翻译仍然是一座不易攻克的堡垒

8. 技术现状

同声传译

理解，但是我想不要突破上限。不要投不上，要有个科学的态度。那么重。中美贸易战的前后是这个对。中国的。2025

shot. Have a scientific attitude. So heavy.
Before and after the Sino-US trade war. This is
right. Chinese.

8. 技术现状

with the large bilingual dictionaries, but to any pair of languages, especially those with a few bilingual resources and we think that this is accurate area for the way

需要。它们依赖于大。双语词典开始学习。
所以我们的重点世界是扩展这一领域的研究，
不仅适用于拥有大型双语词典的语言，但对于
任何一对语言，尤其是那些拥有少量双语资源

8. 技术现状

◆ 翻译总体水平

- 对于资源较为丰富的语言对上（如英汉、日汉等），专业（特定）领域的文本翻译准确率可达到80%左右(**大部分译文基本达到“信”和“达”的水平**)；新闻领域的翻译准确率跨度较大，总体而言，文本翻译的准确率约为70%左右。**(基本停留在“信”的水平，挣扎在“达”的层面。)** ($\text{BLEU} = 0.3\sim0.35$) Transformer, Captain
- 对于译文质量要求较高的翻译任务，如领导人的著作、文学名著、严肃场景下的演讲稿和对话（包括领导人讲话、外交部答记者问、学术讲座等），机器翻译系统都难以胜任。**在可预见的未来看不到机器翻译系统替代人工翻译的可能性。**
- 对于资源稀少的小语种与汉语之间的翻译，机器翻译系统只能以快速获取信息为目的帮助人们大致了解原文的主题和内容。



8. 技术现状

16/content_32646723.htm?from=groupmessage&isappinstalled=0

光明日报

时政

国际

时评

理论

文化

科技

教育



首页 > 光明日报 > 正文

机器翻译的梦想与现实

来源：光明网-《光明日报》 2019-03-16 04:54

作者：宗成庆（中国科学院自动化研究所研究员）

主持人语：

近年来，机器翻译等语言智能技术迅猛发展，并开始在舆论中对人工外语服务形成冲击。一时间，语言智能技术落地、机器翻译替代人工、外语教育存废、外语教学改革等，成为社会热议的话题。本期聚焦“语言智能与外语教育协同发展”，邀请多领域专家学者共同探究，外语教育如何将语言智能的挑战作为转型的契机更好发展，怎样理性认识语言智能技术的发展未来，全力适应人机共存的语言生活形态，努力过好智能时代的外语生活，从而推动语言智能和外语教育协同发展，共同进步，共同助力人类命运共同体的形成。

——李宇明

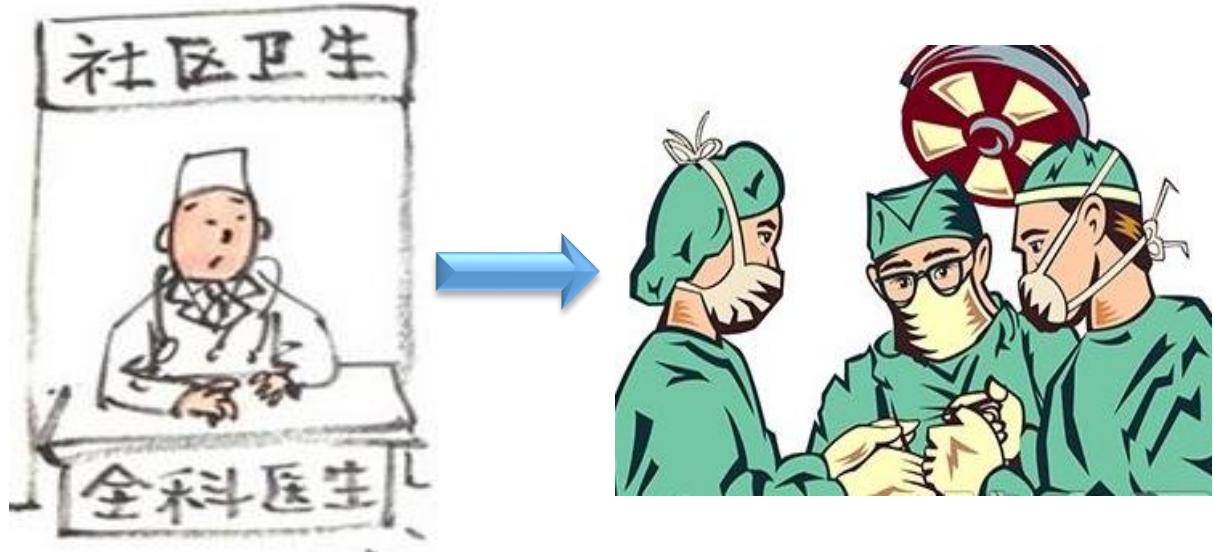
光明网：http://share.gmw.cn/news/2019-03/16/content_32646723.htm



8. 技术现状

◆未来的走向

- 通用翻译系统
- 专用翻译系统



我的系统

■ 支持近80余种外语与汉语的互译：

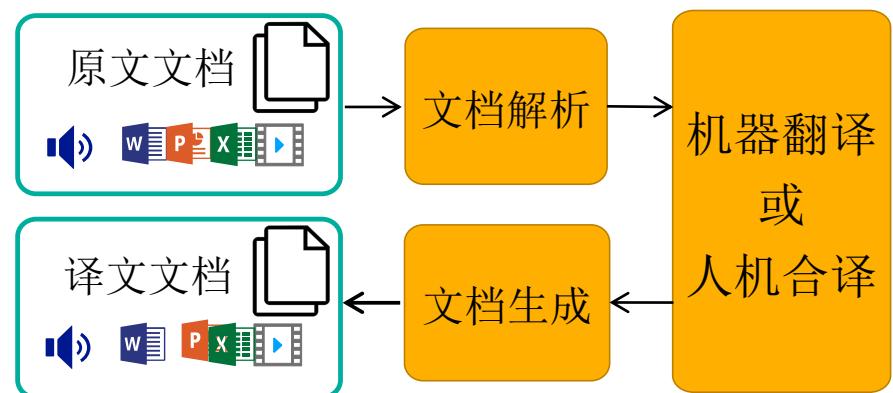
- **世界主要语种：** 英语、日语、韩语、俄语、西班牙语、阿拉伯语等主要语种
- **少数民族语种：** 蒙语、藏语、维语、哈萨克斯坦语等
- **"一带一路"语种：** 越语、缅语、泰语、老挝语、印尼语、印地语、老挝语、尼泊尔语等



飞译：www.fycat.com

■ 支持多格式的文档翻译：

- **文本文件：** txt
- **Word文件：** doc, docx
- **PDF文件：** pdf
- **电子表格：** xls,xlsx
- **图片、音频、视频等**
- **tmx, tbx, xliff等本地化格式下载**



本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
-  9. 习题
10. 附录：延伸阅读

9. 习题

1. 利用提供的中英双语数据(TED)，训练基于RNN或者基于Transformer的机器翻译模型，完成一份实验报告。**(3周时间)**

说明：

- ① 提供的中英双语数据为TED数据（数量为410K，源语言为TED2020.en-zh_cn.en，目标语言为TED2020.en-zh_cn.zh_cn）。
- ② 考虑到计算资源的限制，选择的训练数据和神经网络参数不要过大，如可以选择5K或者10K的双语数据，向量维数设置成20或30维。根据自己的需求，自行划分训练集、开发集和测试集。
- ③ 词表选择1000个左右单词即可，其余单词用<UNK>代替，隐层向量也可以设置为10维左右，层数也可适当降低。
- ④ 由于数据量和参数规模较小，测试集的翻译结果可能非常差，因此报告中主要说明训练集或开发集的损失函数。



9. 习题

⑤ 可以使用Pytorch或者TensorFlow等开源工具。如果依然觉得难度大，可以直接使用下面开源程序：

- GitHub - tensorflow/tensor2tensor: Library of deep learning models and datasets designed to make deep learning more accessible and accelerate ML research (谷歌)
- GitHub - tensorflow/nmt: TensorFlow Neural Machine Translation Tutorial (TensorFlow 教程)
- GitHub-fafacebookresearch/fairseq: Facebook AI Research Sequence-to-Sequence Toolkit written in Python. (Facebook)
- GitHub - THUNLP-MT/THUMT: An open-source neural machine translation toolkit developed by Tsinghua Natural Language Processing Group (清华)



9. 习题

2. 编写程序，实现BLEU值的计算。可以从网络上选择高质量的平行句对，利用所写的BLEU值计算程序和平行句对，计算谷歌、百度和搜狗等3个不同翻译引擎的译文BLEU值。完成一份实验报告。**(3周时间)**

说明：

为同学们提供了WMT-18的中英新闻领域测试集（源语言为newstest2018-zhen-src-ts.zh.sgm，参考译文为newstest2018-zhen-ref-ts.en.sgm），可以从数据集中选择30个测试句对，计算并比较三个不同翻译引擎的翻译BLEU值，并进行简要分析。

本章小结

◆ 机器翻译基本方法

Rule-based; Interlingua-based; Example-based; SMT; NMT

◆ 神经机器翻译

编/解码；注意力机制；Transformer

◆ 译文质量评估

人工评价；BLEU

◆ 语音翻译原理

语音识别；机器翻译；语音合成；端到端的翻译方法

◆ 语料自动获取：从互联网抓取

◆ 机器翻译现状：正在实用化，但仍存在很多问题

本章内容

1. 技术产生与发展
2. 问题与挑战
3. 机器翻译方法
4. 神经机器翻译
5. 译文质量评估
6. 语音翻译原理
7. 双语语料获取
8. 技术现状
9. 习题

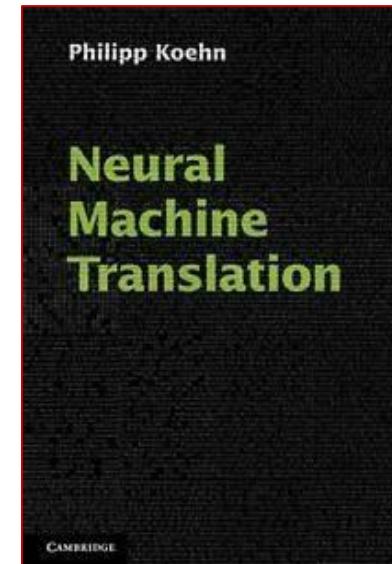
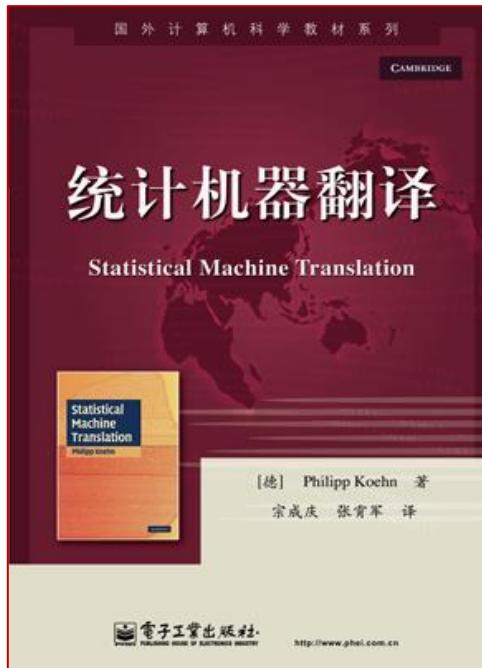
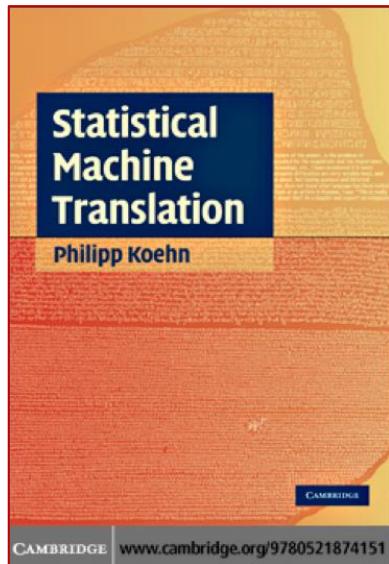


10. 附录：延伸阅读

10. 附录：延伸阅读

◆Books

- [1] Philipp Koehn, Neural Machine Translation, Cambridge University Press, 2020 (中文版：机械工业出版社，2022)
- [2] Philipp Koehn, Statistical Machine Translation, Cambridge University Press, 2010 (中文版：电子工业出版社，2012)





10. 附录：延伸阅读

1. Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio, 2015. Neural Translation by Jointly Learning to Align and Translate. <http://arxiv.org/pdf/1409.0473.pdf>
2. Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2016. Semi-Supervised Learning for Neural Machine Translation. *Proc. of ACL 2016*.
3. Kyunghyun Cho et al., 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Proc. of EMNLP 2014*.
4. Daxiang Dong, HuaWu, Wei He, Dianhai Yu and Haifeng Wang, 2015. Multi-Task Learning for Multiple Language Translation. *Proc. of ACL 2015*.
5. Xiaoqing Li, Jiajun Zhang and Chengqing Zong, 2016a. Towards Zero Unknown Word in Neural Machine Translation. *Proc. of IJCAI 2016*.
6. Xiaoqing Li, Jiajun Zhang and Chengqing Zong, 2016b. One Sentence One Model for Neural Machine Translation. <https://arxiv.org/pdf/1609.06490.pdf>
7. Minh-Thang Luong, Ilya Sutskever, Quoc Le, Oriol Vinyals and Wojciech Zaremba, 2015a. Addressing the Rare Word Problem in Neural Machine Translation. *Proc. of ACL 2015*.
8. Minh-Thang Luong, Hieu Pham and Christopher D. Manning, 2015b. Effective Approaches to Attention-based Neural Machine Translation. *Proc. of EMNLP 2015*.

10. 附录：延伸阅读

9. Minh-Thang Luong and Christopher D. Manning, 2016. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models. *Proc. of ACL 2016*.
10. Sébastien Jean, Kyunghyun Cho, Roland Memisevic and Yoshua Bengio, 2015. On Using Very Large Target Vocabulary for Neural Machine Translation. *Proc. of ACL 2016*.
11. Marcin Junczys-Dowmunt, Tomasz Dwojak and Hieu Hoang, 2016. Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions.
<https://arxiv.org/pdf/1610.01108.pdf>
12. Nal Kalchbrenner and Phil Blunsom, 2013. Recurrent Continuous Translation Models. *Proc. of EMNLP 2013*.
13. Shiqi Shen Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2016. Minimum Risk Training for Neural Machine Translation. *Proc. of ACL 2016*.
14. Rico Sennrich, Barry Haddow and Alexandra Birch, 2016. Improving Neural Machine Translation Models with Monolingual Data. *Proc. of ACL 2016*.
15. Ilya Sutskever, Oriol Vinyals and Quoc V. Le, 2014. Sequence to Sequence Learning with Neural Networks. *Proc. of NIPS 2014*.
16. Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu and Hang Li, 2016. Modeling Coverage for Neural Machine Translation. *Proc. of ACL 2016*.

10. 附录：延伸阅读

17. Mingxuan Wang, Zhengdong Lu, Hang Li and Qun Liu. Memory-enhanced Decoder for Neural Machine Translation. *Proc. of EMNLP 2016*.
18. Yonghui Wu et al., 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. <https://arxiv.org/pdf/1609.08144.pdf>
19. Biao Zhang, Deyi Xiong, Jingsong Su, Hong Duan and Min Zhang, 2016. Variational Neural Machine Translation. *Proc. of EMNLP 2016*.
20. Jiajun Zhang and Chengqing Zong. 2016a. Exploiting Source-side Monolingual Data in Neural Machine Translation. *Proc. of EMNLP 2016*.
21. Jiajun Zhang and Chengqing Zong. 2016b. Bridging Neural Machine Translation and Bilingual Dictionaries. <https://arxiv.org/pdf/1610.07272.pdf>
22. Barret Zoph et al., 2016. Multi-source Neural Machine Translation. *Proc. of NAACL 2016*.
23. Long Zhou, Wenpeng Hu, Jiajun Zhang and Chengqing Zong, 2017. Neural System Combination for Machine Translation. *Proc. of ACL 2017*.
24. Yang Zhao, Yining Wang, Jiajun Zhang and Chengqing Zong, 2018. Phrase Table as Recommendation Memory for Neural Machine Translation. *Proc. of IJCAI 2018*.
25. Jiajun Zhang et al. 2020. Synchronous bidirectional inference for neural sequence generation. *Artificial Intelligence*, 281(2020)103234, pp.1-19

10. 附录：延伸阅读

26. Yang Zhao et al. 2020. Knowledge Graphs Enhanced Neural Machine Translation. *Proc. of IJCAI-PRICAI 2020*
27. Yang Zhao et al. 2019. Addressing the Under-translation Problem from the Entropy Perspective. *Proc. AAAI'2019*, pp.451-458
28. Long Zhou, Jiajun Zhang, Chengqing Zong. 2019. Synchronous Bidirectional Neural Machine Translation, *Transactions of Association for Computational Linguistics (TACL)*, Vol. 7, pp. 91-105, 2019
29. Guoping Huang et al. 2019. Input Method for Human Translators: a Novel Approach to Integrate Machine Translation Effectively and Imperceptibly. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, Vol. 18, No. 1, Article 4, 22 pages, January 2019
30. Yining Wang et al. 2018. Three Strategies to Improve One-to-Many Multilingual Translation. *Proc. EMNLP'2018*, pp. 2955-2960
31. Yuchen Liu et al. 2019. End-to-End Speech Translation with Knowledge Distillation. *Proc. Interspeech'2019*, pp.1128-1132
32. Jingjing Xu et al. 2021. Vocabulary Learning via Optimal Transport for Neural Machine Translation. *Proc. ACL-IJCNLP'2021*

10. 附录：延伸阅读

33. Elena Voita, Rico Sennrich and Ivan Titov. 2021. Analyzing the Source and Target Contributions to Predictions in Neural Machine Translation. *Proc. ACL-IJCNLP'2021*
34. Luisa Bentivogli, Mauro Cettolo, Marco Gaido, Alina Karakanta, Alberto Martinelli, Matteo Negri and Marco Turchi. 2021. Cascade versus Direct Speech Translation: Do the Differences Still Make a Difference? *Proc. ACL-IJCNLP'2021*
35. Benjamin Marie, Atsushi Fujita and Raphael Rubino. 2021. Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers. *Proc. ACL-IJCNLP'2021*
36. Xiangpeng Wei, Heng Yu, Yue Hu, Rongxiang Weng, Weihua Luo and Rong Jin. 2022. Learning to Generalize to More: Continuous Semantic Augmentation for Neural Machine Translation. *Proc. ACL'2022*

ACL/ EMNLP/ COLING 论文集：<https://aclanthology.org/events/>

谢谢！

Thanks!

