

第11章 机器翻译



内容提要

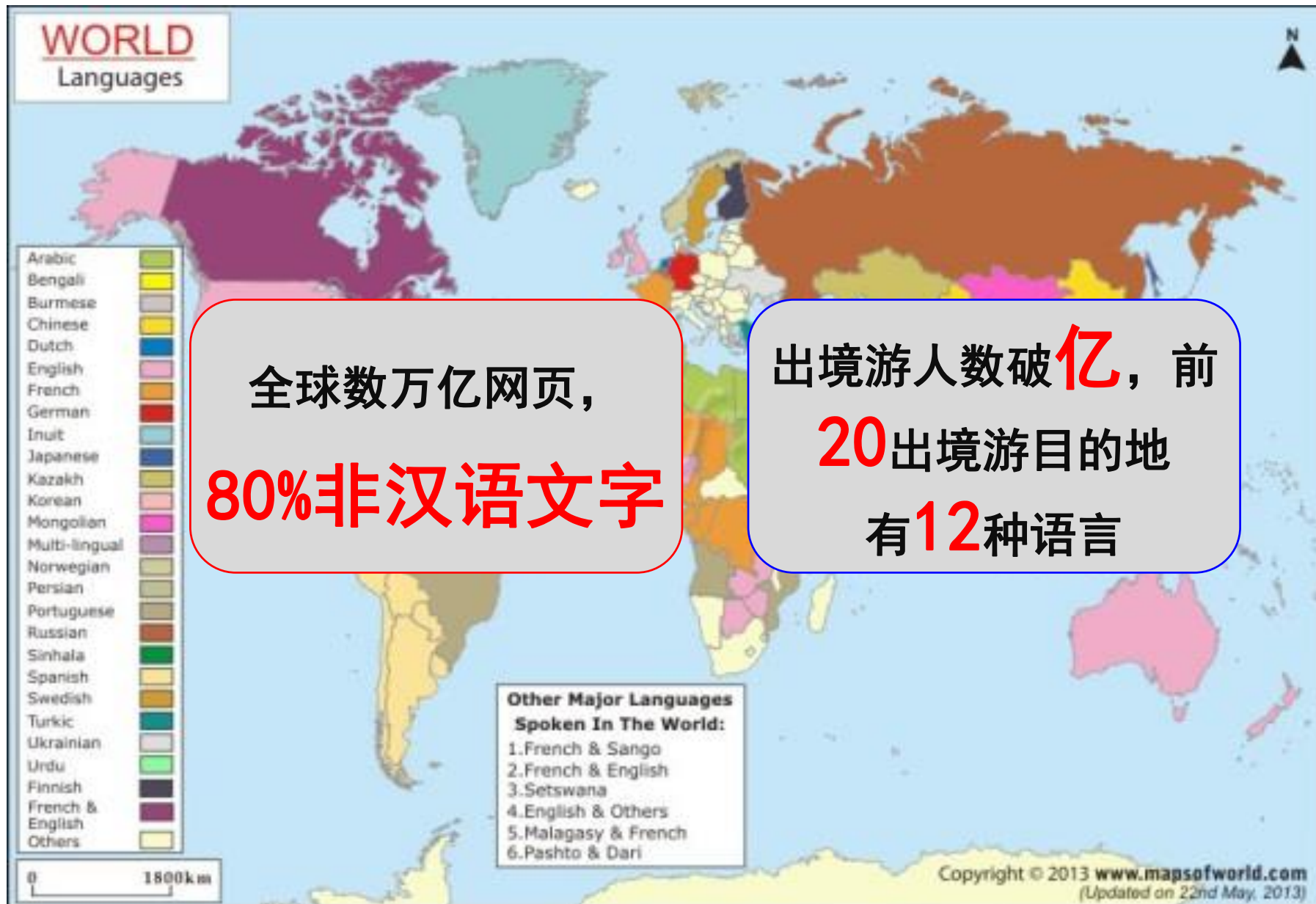
11.1 机器翻译概论

11.2 统计机器翻译



11.1 机器翻译概论

世界语言地图



中国语言地图



背景：一带一路

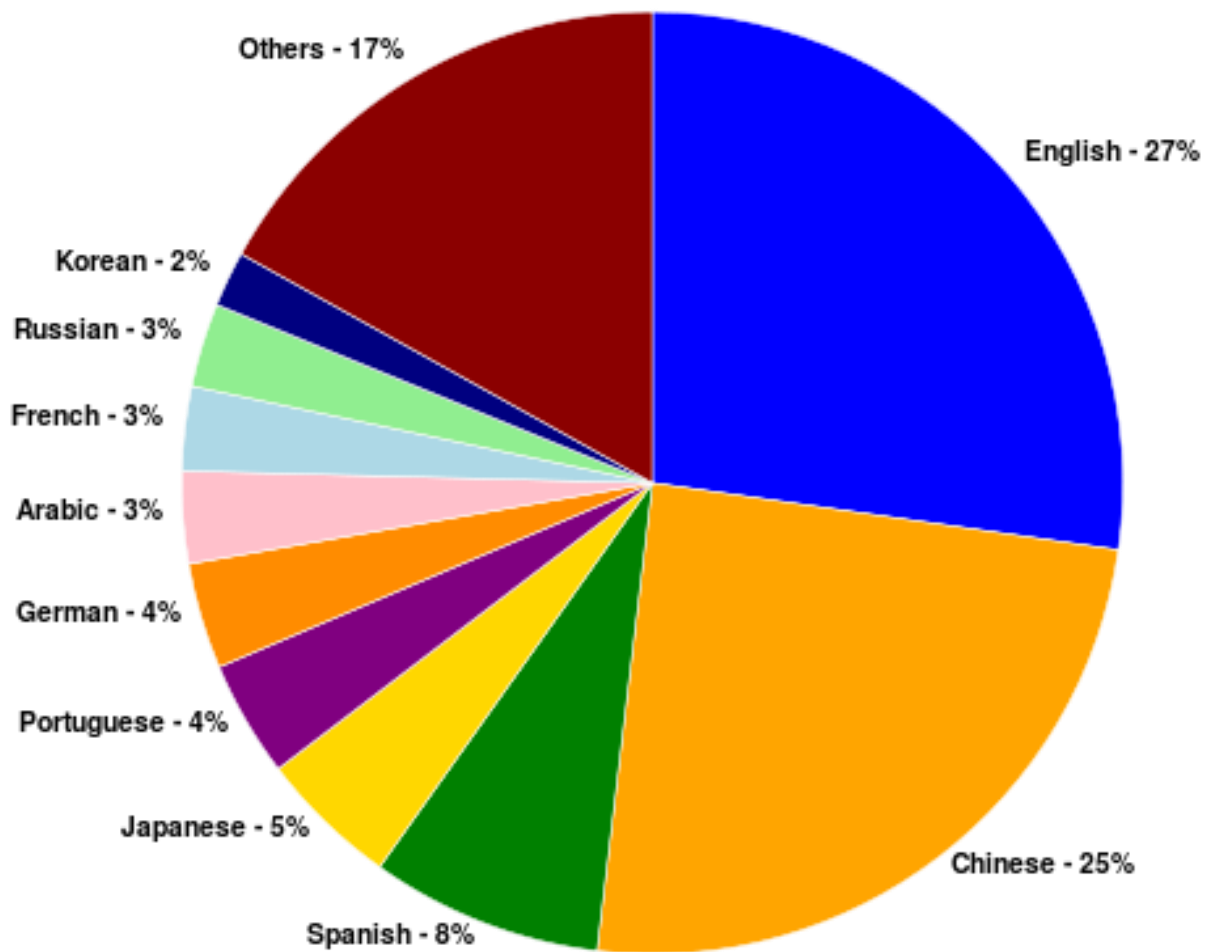
64个国家地区

44亿人口

50多种语言



互联网用户





11.1.1 引言

- ◆ 有关专家已经指出，语言障碍是21世纪国际社会全球化面临的主要困难之一；
- ◆ 机器翻译涉及语言学、计算语言学、认知科学和数学等多种学科，具有重要的科学意义；
- ◆ 机器翻译具有巨大的社会需求，不仅具有极大的经济利益，而且对于情报获取、信息安全意义重大；
- ◆ 以欧洲为例，有380多种语言，2004年5月1日以前欧盟有11种官方语言，每年为这11种语言翻译、转录文件耗费的人力费用大约 549M 欧元。2014年欧盟24种官方工作语言。

11.1.1 引言

The screenshot shows the Google Translate interface. At the top, the Google logo is visible. Below it, the word "Translate" is written in red. A large red text overlay reads "103 种语言". The interface shows a language selection menu with "English" selected. A dropdown menu is open, displaying a grid of 103 languages. The languages are arranged in columns and rows, with "English" highlighted in the second row, third column. The languages listed include: Detect language, Afrikaans, Albanian, Arabic, Armenian, Azerbaijani, Basque, Belarusian, Bengali, Bosnian, Bulgarian, Catalan, Cebuano, Chichewa, Chinese, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Filipino, Finnish, French, Galician, Georgian, German, Greek, Gujarati, Haitian Creole, Hausa, Hebrew, Hindi, Hmong, Hungarian, Icelandic, Igbo, Indonesian, Irish, Italian, Japanese, Javanese, Kannada, Kazakh, Khmer, Korean, Lao, Latin, Latvian, Lithuanian, Macedonian, Malagasy, Malay, Malayalam, Maltese, Maori, Marathi, Mongolian, Myanmar (Burmese), Nepali, Norwegian, Persian, Polish, Portuguese, Punjabi, Romanian, Russian, Serbian, Sesotho, Sinhala, Slovak, Slovenian, Somali, Spanish, Sundanese, Swahili, Swedish, Tajik, Tamil, Telugu, Thai, Turkish, Ukrainian, Urdu, Uzbek, Vietnamese, Welsh, Yiddish, Yoruba, and Zulu.

Google

Translate:



11.1.1 引言

网页 图片 视频 地图 资讯 词典 影响力 在线翻译 更多



在线翻译 手机适用 用于电脑 网站管理员适用 必应 BING 工具栏适用 帮助

源语言: 自动检测

目标语言: 简体中文



|

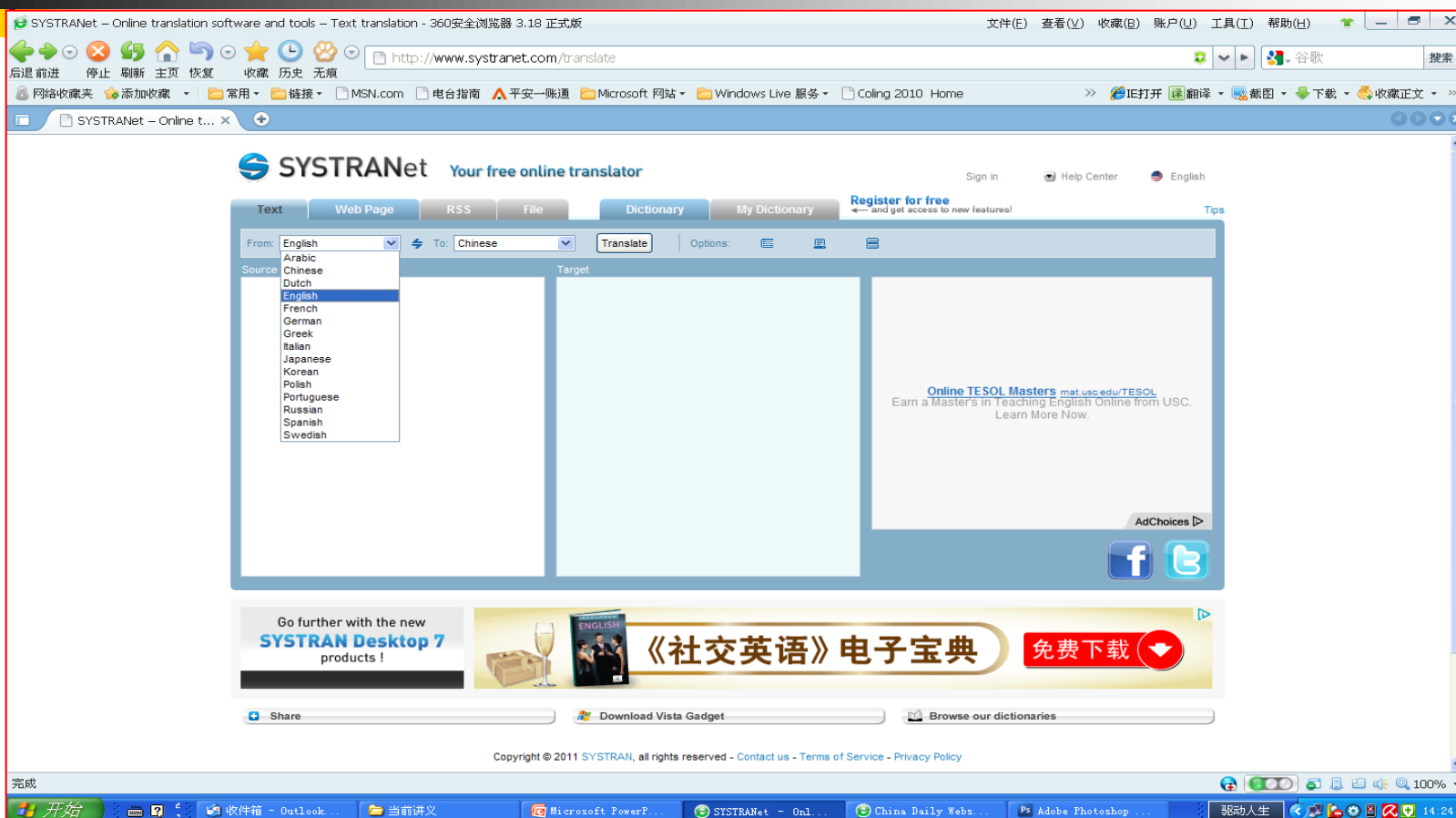
输入文本或网页URL

Bing

Translate:



11.1.1 引言



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



11.1.1 引言



官网

抗疫行动

翻译 API

同传

上新啦

200+ 语种

自动检测



中文(简体)

翻 译

人工翻译

输入文字、网址 / 粘贴图片 / 拖入文档

上传文档

百度在线翻译: <http://fanyi.baidu.com/>



11.1.1 引言

The screenshot shows the Youdao online translation website. The browser title is "在线翻译_有道 - 360安全浏览器 3.18 正式版". The address bar shows "http://fanyi.youdao.com/". The page header includes navigation links like "网易", "网页", "图片", "新闻", "购物", "音乐", "词典", "翻译", and "更多". The main content area features the Youdao logo and the text "免费、即时的多语种在线翻译". Below this is a large text input field labeled "原文/网址:". To the right of the input field are three checkmarks indicating features: "支持中英、中日、中韩、中法互译", "支持网页翻译, 在输入框直接输入网页地址即可, 查看示例", and "下载最好用的免费全能翻译软件 - 有道词典". At the bottom left, there is a language selection dropdown menu with options for Chinese, English, Japanese, Korean, French, and Vietnamese. The bottom of the page contains promotional banners for "2012 重点院校在职研究生招生" and "2012名校专升本报名(春)".

有道在线翻译: <http://fanyi.youdao.com/>

11.1.1 引言

知乎 图片 视频 明医 英文 问问 **翻译** 更多»

 搜狗翻译

自动检测




中文

翻译

请输入您要翻译的文字或网址

0 / 5000

 查询历史

暂时没有您的查询历史

不再显示查询历史

搜狗在线翻译: <http://fanyi.sougou.com/>

小牛翻译

■ 小牛word插件升级版---小牛快译 来啦！ 下载安装后可以在Word、WPS中一键完成全文翻译，110种语言任意互译！ 不但保留全部格式，还支持术语词典和辅助翻译助手，注册登录后可以使用PDF转Word功能，再也不用担心写论文，读文献啦！





11.1.2 机器翻译的 产生与发展

11.1.2 机器翻译的产生与发展

- ◆ **概念：**机器翻译 (machine translation, MT) 是用计算机把一种语言 (源语言, source language) 翻译成另一种语言 (目标语言, target language) 的一门学科和技术。



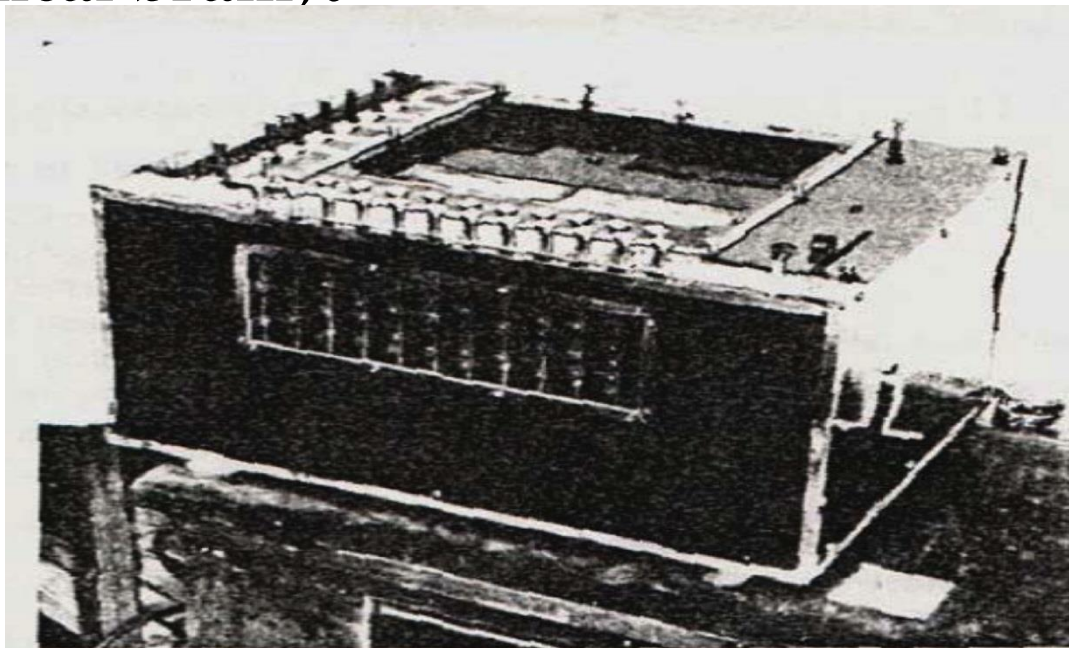
11.1.2 机器翻译的产生与发展

◆ 古希腊时期

- ◆ 17世纪: 笛卡儿(Descartes)莱布尼兹(Leibniz)试图用统一的数字代码编写词典; 17世纪中页贝克(Cave Beck)等人出版类似的词典。

11.1.2 机器翻译的产生与发展

- ◆ **1930s**: 亚美尼亚法国工程师阿尔楚尼(G. B. Arsouni)提出了用机器来进行语言翻译的想法, 并在1933年7月22日获得了一项“翻译机”的专利, 叫做机器脑(mechanical brain)。



11.1.2 机器翻译的产生与发展

- ◆ **1933年**，前苏联发明家特洛扬斯基设计了用机械方法把一种语言翻译成为另一种语言的机器。
- ◆ **1946年**，世界上第一台电子计算机 ENIAC 诞生以后，英国工程师 A. D. Booth 和美国洛克菲勒基金会 (Rockefeller Foundation) 副总裁 W. Weaver 提出了利用计算机进行机器翻译的设想。
- ◆ **1947年3月**，W. Weaver 给 N. Wiener 写信，讨论机器翻译问题，但遭到了 N. Wiener 的反对。
- ◆ **1949年**，W. Weaver 发表了以 ‘Translation’ 为题目的备忘录，正式提出机器翻译问题。

11.1.2 机器翻译的产生与发展



W. Weaver

**翻译类似于解读密码的过程：
当我阅读一篇用俄语写的文章时，
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。--1949**

11.1.2 机器翻译的产生与发展



W. Weaver

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。

11.1.2 机器翻译的产生与发展

- ◆ 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，MT研究在国际上出现热潮。

草创时期

11.1.2 机器翻译的产生与发展

- ◆ 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”的报告，简称 ALPAC 报告，宣称：“在目前给机器翻译以大力支持还没有多少理由”，“机器翻译遇到了难以克服的语义障碍 (semantic barrier)”。从此，机器翻译研究在世界范围内进入低迷状态。

11.1.2 机器翻译的产生与发展

- ◆ 1970 ~1976年，法国、日本、加拿大等国，仍坚持机器翻译研究，而且，在这一阶段语法与算法分开，机器翻译研究开始复苏。代表系统：法国 GETA，IMAG-CLIPS 开发的 ARIANE-78 系统。

复苏阶段

11.1.2 机器翻译的产生与发展

◆ 1976~至今, **繁荣时期**

- **加拿大蒙特利尔大学**与加拿大联邦政府翻译局联合开发的实用机器翻译系统 **TAUM-METEO**, 用于天气预报翻译。每小时可以翻译6 ~ 30万个词; 每天翻译**1500—2000**篇天气预报资料, 并通过电视、报纸等立即公布;
- **1978年**欧共体启动多语言机器翻译计划;
- **1982 ~ 1986**, 日本在提出第五代机的同时, 研究日英双向机器翻译系统 **Mu** 和亚洲多语言机器翻译 (日语、汉语、印尼语、马来西亚语、泰国语);
- **1990年**, **IBM** 提出统计机器翻译模型, 机器翻译研究进入了一个空前辉煌的繁荣时期。

11.1.2 机器翻译的产生与发展

◆ 1990~1999年,

- **统计机器翻译** 由于仅依赖双语平行语料，大大降低了进入机器翻译研究的门槛；但因为数学模型过于复杂，很难实现。

◆ 1999~现在,

- **翻译开源工具** GIZA、Pharaoh等开源项目发布
- **翻译自动评价方法** BLEU提出
- **参数训练方法** MERT提出

11.1.2 机器翻译的产生与发展

◆ 2014~现在,

突破期

- 基于深度学习的统计机器翻译 (Devlin et al., 2014)
- 神经网络机器翻译 (Sutskever et al., 2014, Bahdanau et al., 2014)



Yoshua Bengio



Kyunghyun Cho



11.1.3 机器翻译的困难

11.1.3 机器翻译的困难

- ◆ 自然语言中普遍存在的歧义和未知现象
 - 句法结构歧义/词汇歧义/语用歧义 ...
 - 新的词汇、术语、结构、语义 ...
- ◆ 机器翻译不仅仅是字符串的转换
 - 不同语言之间文化的差异
 - 现有方法无法表示和利用世界知识和常识
- ◆ 机器翻译的解不唯一，而且始终存在的人为的标准

几乎自然语言处理中的所有问题在机器翻译中都会遇到。

11.1.3 机器翻译的困难

黛玉自在枕上感念宝钗……又听见窗外竹梢焦叶之上，雨声淅沥，消寒透幕，不觉又滴下泪来。

(《红楼梦》第45回)

As she lay there alone, Dai-Yu's thoughts turned to Bao-chai ...Then she listened to the insistent rustle of the rain on the bamboos and plantains outside her window. The coldness penetrated the curtains of her bed. Almost without noticing it she had begun to cry.

— 文学翻译家 David Hawkes

摘自冯志伟著《机器翻译研究》，2004



11.1.3 机器翻译的困难

[Google Translator](http://translate.google.com/) (<http://translate.google.com/>, 2017.4.16):

**Daiyu comfortable pillow on the feeling of Chai
Chai and heard the bamboo leaves above the
leaves above the rain, rain pattering, cold through the
curtain, feel and dripping tears.**

机器不是人，更不是文学翻译家！

11.1.3 机器翻译的困难

[Baidu Translator \(2022.4.12\):](#)

逆向翻译：黛玉独自一人躺在那里，思绪转向宝钗……然后她听到窗外竹子和芭蕉上持续不断的雨声。寒冷穿透了她床上的窗帘。她几乎没注意到就哭了起来。

黛玉自在枕上感念宝钗……又听见窗外竹梢蕉叶之上，雨声淅沥，消寒透幕，不觉又滴下泪来。

(《红楼梦》第45回)

机器不是人，更不是文学翻译家！

11.1.3 机器翻译的困难

食品或菜单名的翻译：

(1) 红烧狮子头： **Fried Lion's Head**

→ **Braised pork balls**

(2) 夫妻肺片： **Husband and wife's lung slices**

→ **Fuqifeipian/ Spicy beef**

(3) 童子鸡： **Young lad chicken**

Boy chicken without sexual life

→ **Spring chicken/ Broiler chicken**

11.1.3 机器翻译的困难

例句: **We do chicken right.**

- (1) 我们做鸡是对的。
- (2) 我们做鸡正点耶。
- (3) 我们就是做鸡的。
- (4) 我们有做鸡的权利。
- (5) 我们只做鸡的右边。
- (6) 我们可以做鸡, 对吧!
- (7) 我们行使了鸡的权利。
- (8) 我们只做右边的鸡。
- (9) 我们主张鸡权。
- (10) 我们公正地做鸡!
- (11) 我们要把鸡打成右派。
- (12) 我们做鸡肉权利。
- (13) 我们还是做鸡好。
- (14) 我们用正确的方法炸鸡。
- (15) 我们做鸡有理!
- (16) 我们让鸡向右看齐。
- (17) 我们肯定是鸡, 对!
- (18) 我们做的鸡才是正宗。

..... **几十种翻译!**

11.1.3 机器翻译的困难



11.1.3 机器翻译的困难



11.1.3 机器翻译的困难





11.1.4 机器翻译研究现状

11.1.4 机器翻译研究现状

◆ 若干翻译系统已实用化或接近实用化

- Google translator (<http://translate.google.com>)
- Systran (<http://www.systransoft.com>)
- TAUM-METEO
- 百度在线翻译/ 华建英汉翻译系统/ 金山词霸 等

◆ 仍面临若干问题

- 理论模型和方法
- 实现技术问题
- 用户认识问题

应彻底消除对机器翻译的误解
让机器翻译系统翻译诗歌、散文和小说等高难度文艺作品是不现实的，也是不可能的。

11.1.4 机器翻译研究现状

◆ 新闻翻译

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)

Results given by Google Translator:

- 北京做第三严正交涉到马尼拉，并警告说这是很难约领土僵局的一个岛屿乐观。当局说，他们已经准备了马尼拉的情况有任何升级。[\(Google, 2015.4.28\)](#)
- 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。[\(Google, 2016.5.1\)](#)
- 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。[\(2017.4.16\)](#)
- 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。[\(Google, 2018.5.15\)](#)
- 北京对马尼拉进行了第三次庄严代表，并警告说，很难对一个岛屿的领土僵局持乐观态度。当局说，他们已经为马尼拉局势的任何升级做好了准备。[\(Google, 2019.4.9\)](#)
- 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。[\(Google, 2021.5.11\)](#)
- 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。[\(Google, 2022.4.12\)](#)

- 北京对马尼拉进行了第三次庄严代表，并警告说，很难对一个岛屿的领土僵局持乐观态度。当局说，他们已经为马尼拉局势的任何升级做好了准备。[\(Baidu, 2018.5.15\)](#)
- 北京第三次向马尼拉提出严正交涉，并警告说，很难对一个岛屿的领土僵局感到乐观。当局说，他们已经为马尼拉局势的任何升级做好了准备。[\(Baidu, 2021.5.11\)](#)
- 北京第三次向马尼拉提出严正交涉，并警告说，很难对一个岛屿的领土僵局感到乐观。当局表示，他们已经为马尼拉局势的任何升级做好了准备。[\(Baidu, 2022.4.12\)](#)

11.1.4 机器翻译研究现状

◆ 基本观点

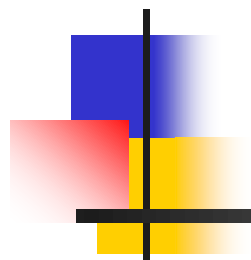
在机器翻译研究中实现人机共生 (man-machine symbiosis), 人机互助比追求完全自动的高质量的翻译 (Full Automatic High Quality Translation, FAHQT) 更现实、更切合实际[Hutchins, 1995]。

我们需要的是计算机帮助人类完成某些翻译工作, 而不是完全替代人, 人与机器翻译系统之间应该是互补的关系, 而不是相互竞争[Hutchins, 2001]。

11.1.4 机器翻译研究现状

机器翻译研究在过去的五十多年曲折发展经历中，无论是它给人们带来的希望还是失望我们都必须客观地看到，机器翻译作为一个科学问题在被学术界不断深入研究的同时，企业家们已经从市场上获得了相应的利润。

- ✧ 机器翻译还不成熟(state-of-the-art)，需要的是人与系统的配合，而不是有意为难，辅助机器翻译可以大大减轻人的负担。
- ✧ “信、达、雅”是人类翻译追求的目标，计算机在这方面永远都不会替代人。



11.1.5 基本翻译方法



11.1.5 基本翻译方法

- ◆ 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
 - 基于事例的翻译方法
 - 统计翻译方法
 - 神经网络机器翻译

11.1.5 基本翻译方法

◆ 直接转换法

从源语言句子的表层出发，将单词、短语或句子**直接置换**成目标语言译文，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. → Me(I) gusta(like) Maria(Mary).

X like Y → Y X gusta



11.1.5 基本翻译方法

◆ 基于规则的翻译方法(Rule-based)

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。



11.1.5 基本翻译方法

基于规则的翻译过程分成6个步骤：

- (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**

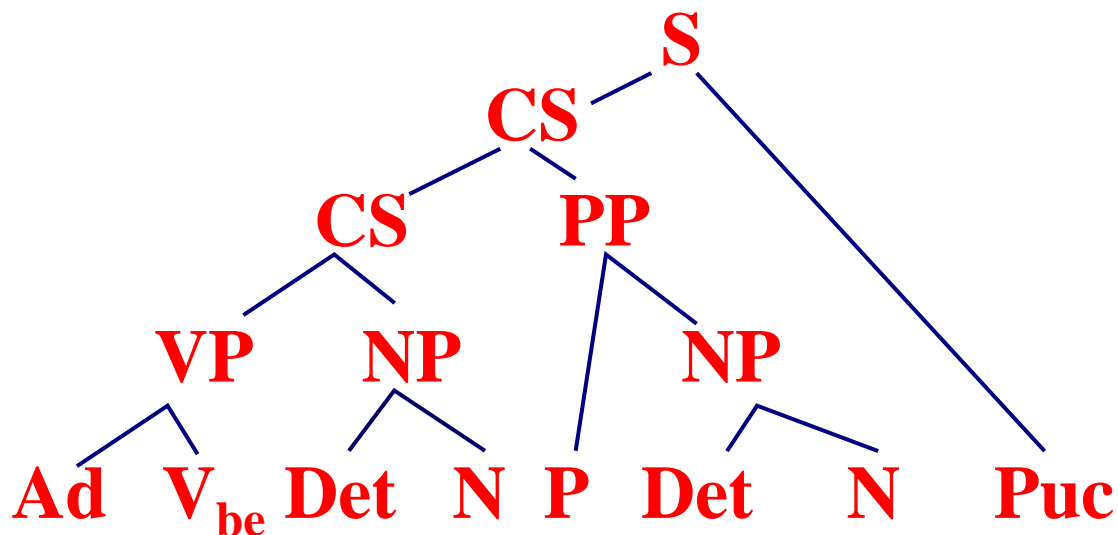
11.1.5 基本翻译方法

给定源语言句子：There is a book on the desk.

词法分析：

There/**Ad** is/**V_{be}** a/**Det** book/**N** on/**P** the/**Det** desk/**N** ./**Puc**

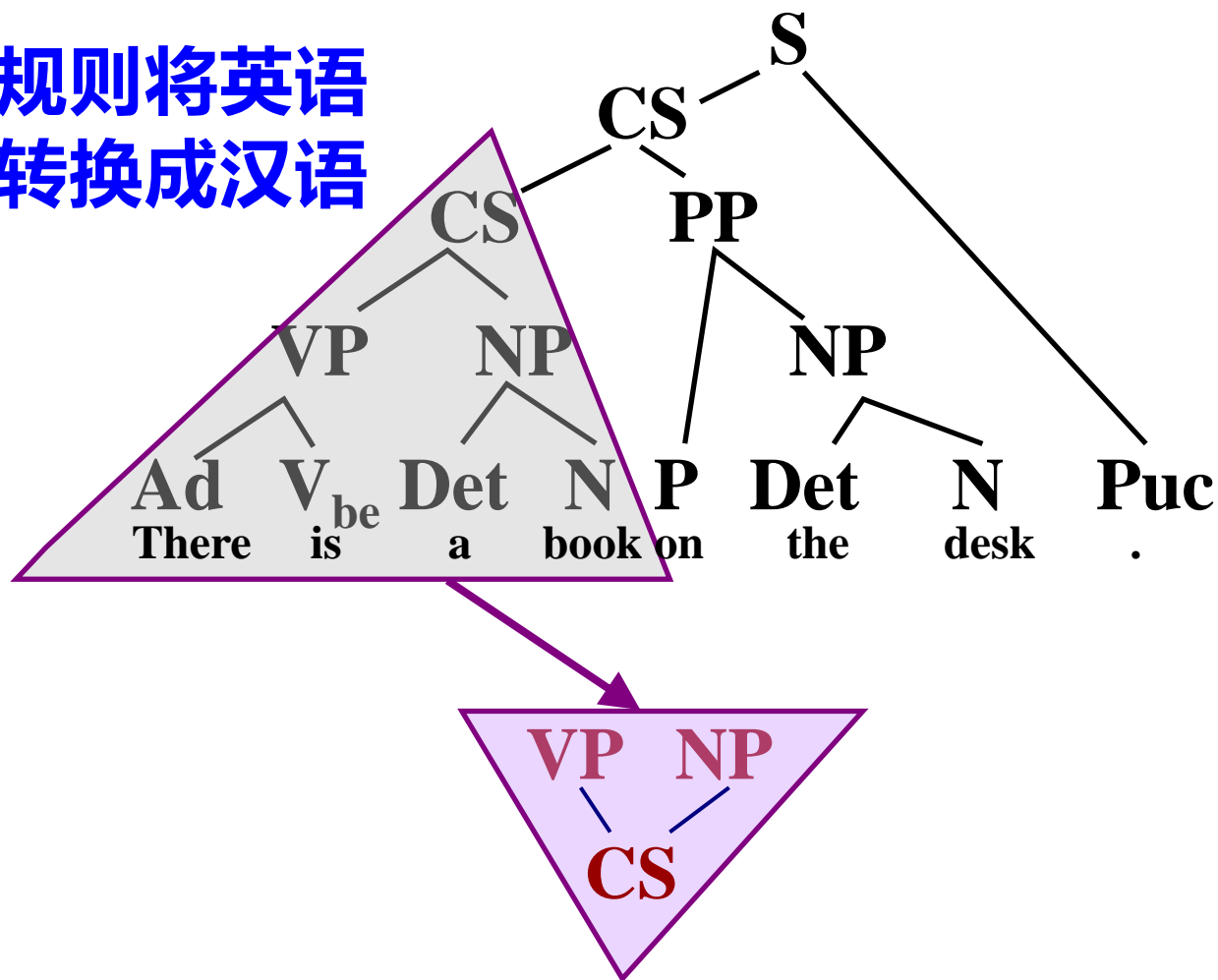
利用句法规则进行句法结构分析：



动词短语 (verb phrase, VP)
名词短语 (noun phrase, NP)
介词短语 (preposition, PP)
2类连词 (conjunction, 分别记作: CC, CS)

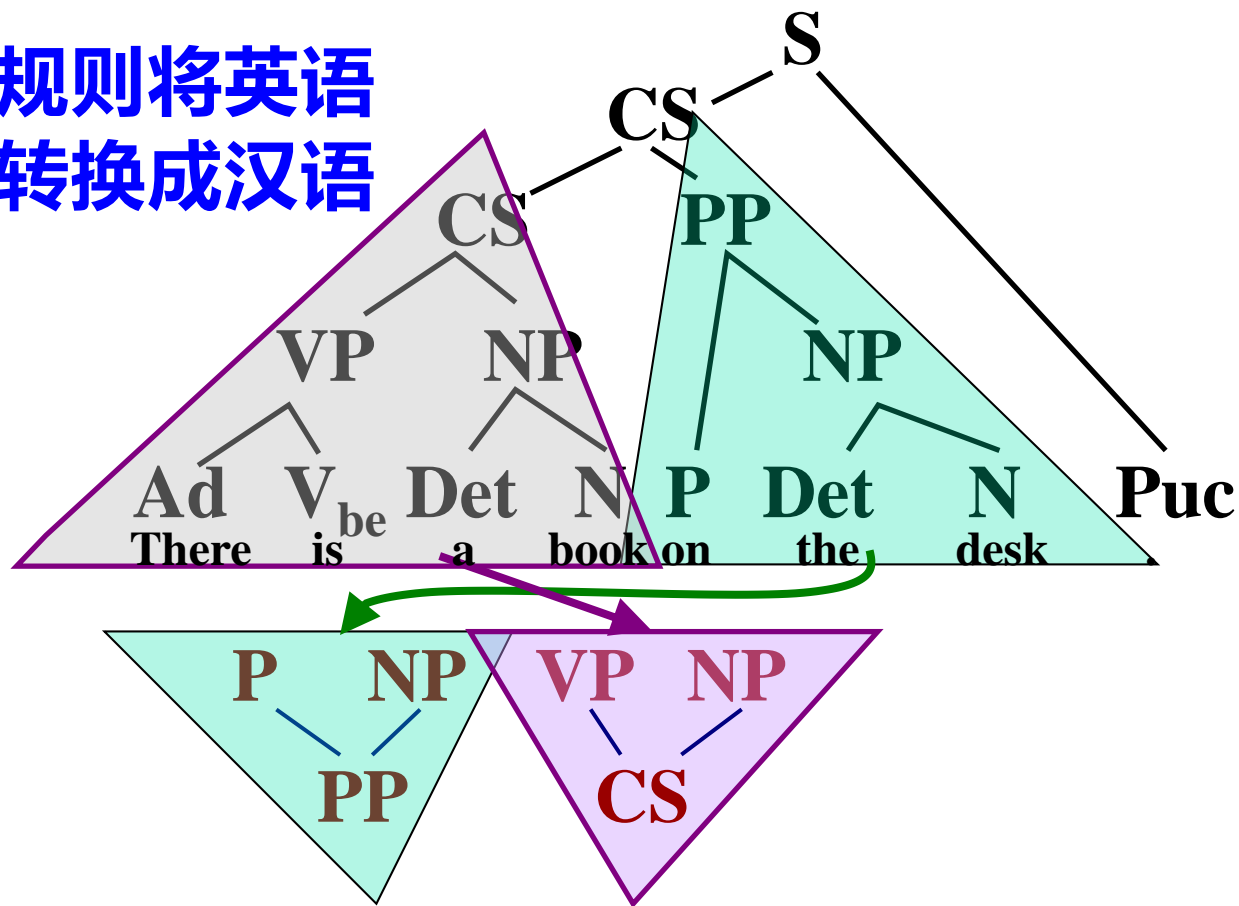
基本研究方法

✧ 利用转换规则将英语句子结构转换成汉语句子结构



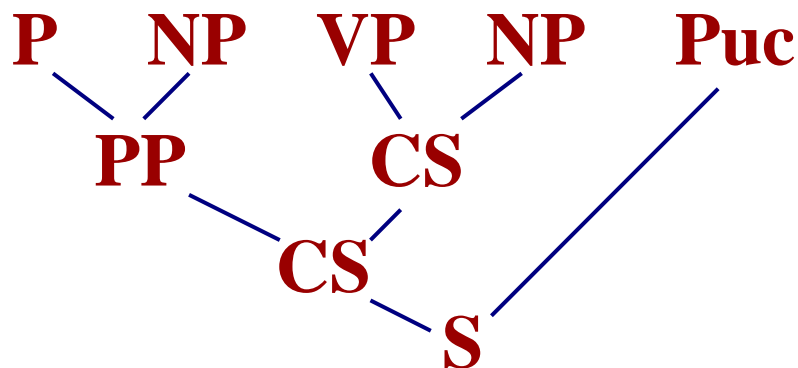
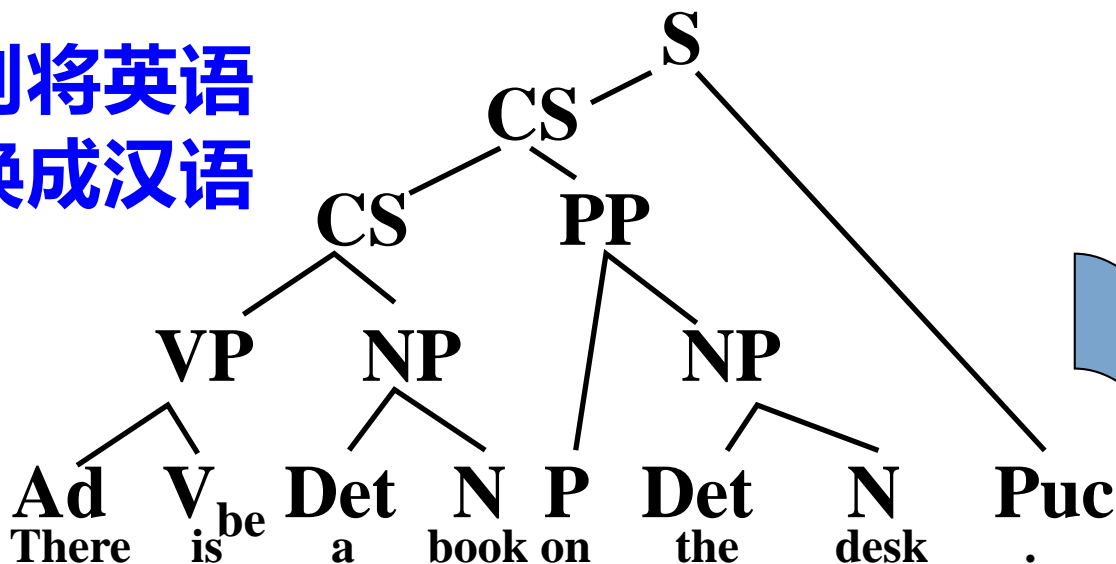
基本研究方法

利用转换规则将英语句子结构转换成汉语句子结构



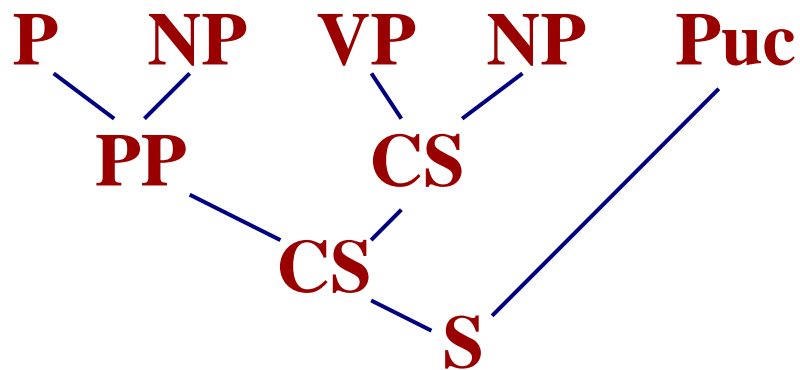
基本研究方法

利用转换规则将英语句子结构转换成汉语句子结构



基本研究方法

✧ 根据转换后的句子结构，利用词典和生成规则生成翻译的结果句子



#a, Det, 一

#book, N, 书; V, 预订

#desk, N, 桌子

#on, P, 在 X 上

#There be, V, 有

输出译文:

在桌子上有一本书。



11.1.5 基本翻译方法

由于基于规则的翻译方法执行过程为：

“独立分析 - 独立生成 - 相关转换”

因此，又称**基于转换的翻译方法**。

其代表系统是法国格勒诺布尔(Grenoble)机器翻译研究所(GETA)开发的ARIANE翻译系统。

1976年加拿大蒙特利尔大学与加拿大联邦翻译局联合开发的实用性机器翻译系统 TAU-METEO：天气预报信息服务。



11.1.5 基本翻译方法

对基于规则的翻译方法的评价：

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。

11.1.5 基本翻译方法

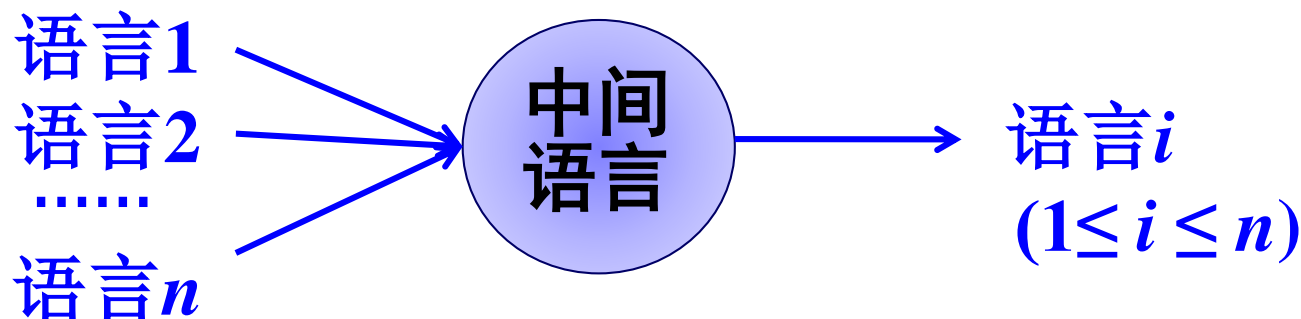
◆ 基于中间语言的翻译方法(Interlingua-based)

- 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)





11.1.5 基本翻译方法

◆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**)



11.1.5 基本翻译方法

对基于中间语言的翻译方法评价：

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。

11.1.5 基本翻译方法

◆ 基于事例(实例)的翻译方法 (Example-based)

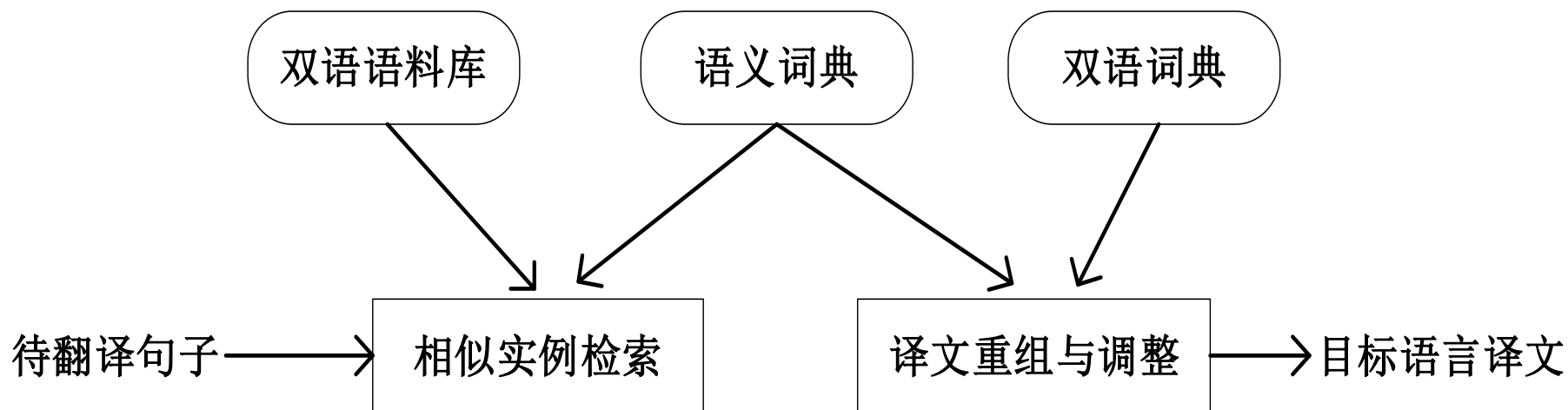
1984年由日本学者长尾真提出。



1936年出生，1959年毕业于京都大学工学系电子工学专业，1961年京都大学研究生院硕士课程毕业，1966年从京都大学获得工学博士称号，1973年担任京都大学教授，1997年担任京都大学校长，2004年担任信息通信研究机构(NICT)理事长，2007年4月起担任日本国立国会图书馆馆长。研究开发的业绩涉及自然语言处理、图像处理、信息工学、智能信息学等多个领域。2003年获得ACL终生成就奖。

11.1.5 基本翻译方法

- **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)



11.1.5 基本翻译方法

对基于实例的翻译方法评价：

优点： 不要求源语言句子必须符合语法规则，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。

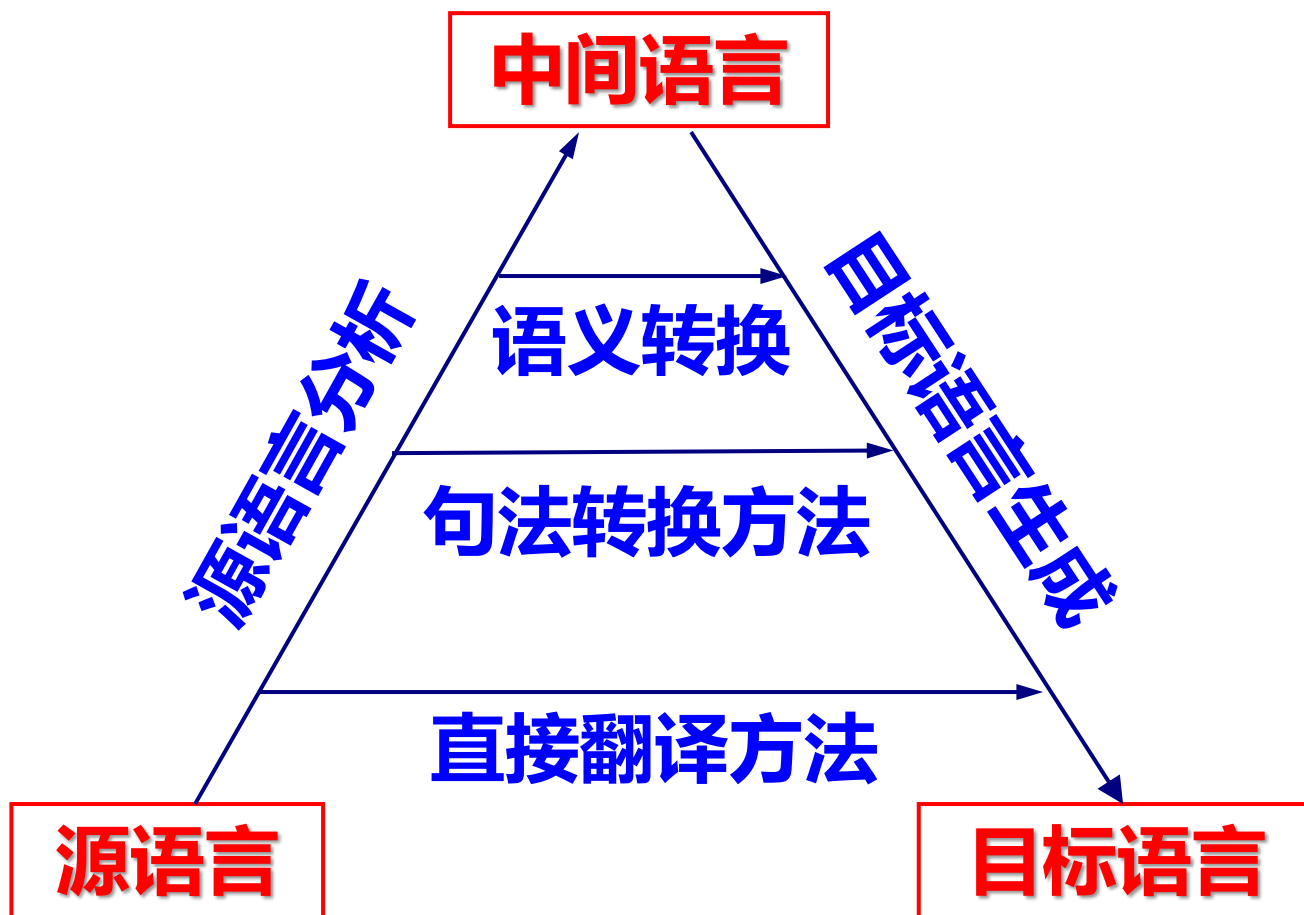


11.1.5 基本翻译方法

◆ 其它翻译方法

- 基于记忆的(memory-based)翻译方法
- 统计翻译方法(statistical method)
- 基于多引擎的翻译方法(multi-engine)
- 基于神经网络(neural network)的翻译方法

11.1.5 基本翻译方法





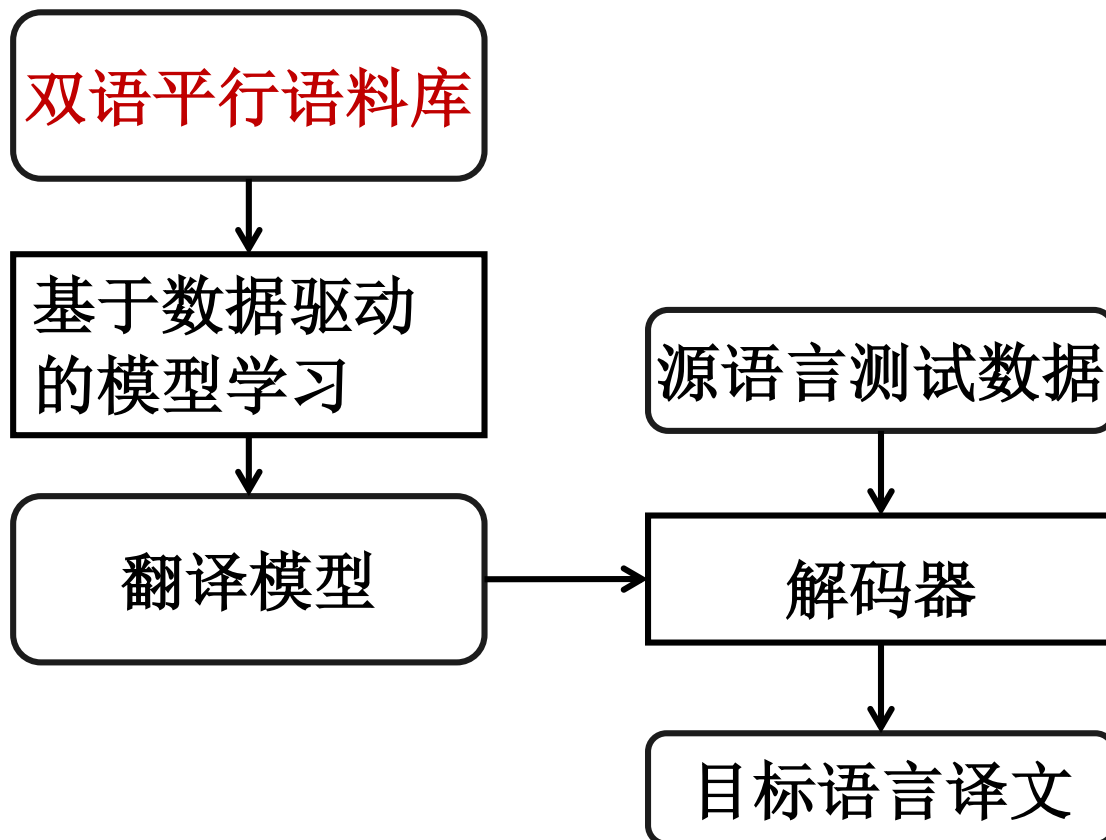
本讲小结

- ◆ 机器翻译的产生与发展
- ◆ 机器翻译研究现状
- ◆ 机器翻译基本方法
 - 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
 - 基于事例的翻译方法



11.2 统计机器翻译

统计翻译的思想



Garcia and associates .

Garcia y asociados .

Carlos Garcia has three associates .

Carlos Garcia tiene tres asociados .

his associates are not strong .

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

the clients and the associates are enemies .

los clientes y los asociados son enemigos .

the company has three groups .

la empresa tiene tres grupos .

its groups are in Europe .

sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the small groups are not modern .

los grupos pequenos no son modernos .

Garcia and associates .

Garcia y asociados .

Carlos Garcia has three associates .

Carlos Garcia tiene tres asociados .

his associates are not strong .

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

the clients and the associates are enemies .

los clientes y los asociados son enemigos .

the company has three groups .

la empresa tiene tres grupos .

its groups are in Europe .

sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the small groups are not modern .

los grupos pequenos no son modernos .

11.2.1 统计翻译的诞生

- 1990年IBM的Peter F. Brown 等人在*Computational Linguistics* 上发表论文“统计机器翻译方法” [Brown, 1990]; 1993年他们在该杂志发表论文“统计机器翻译的数学：参数估计” [Brown, 1993], 两篇文章奠定了统计机器翻译的理论基础。



左: **Robert Mercer**

右: **Peter F. Brown**

11.2.1 统计翻译基本原理

◆ 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]。



11.2.1 统计翻译基本原理

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

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

贝叶斯公式: $P(T|S) = \frac{P(T) \times P(S|T)}{P(S)}$

$$T' = \operatorname{argmax}_T P(T) \times P(S|T)$$

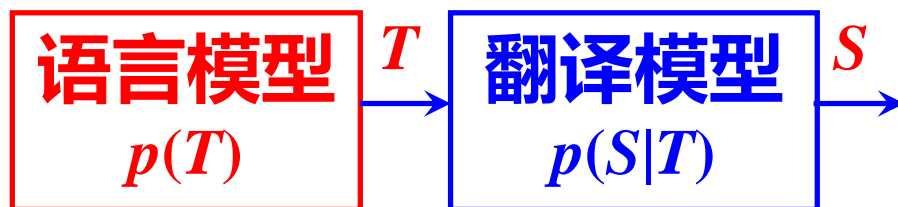
语言模型

Language model, LM

翻译模型

Translation model, TM

11.2.1 统计翻译基本原理



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

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。

11.2.1 统计翻译基本原理

◆ 估计语言模型概率 $p(T)$

给定句子: $T = t_1^l = t_1 t_2 \cdots t_l$

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

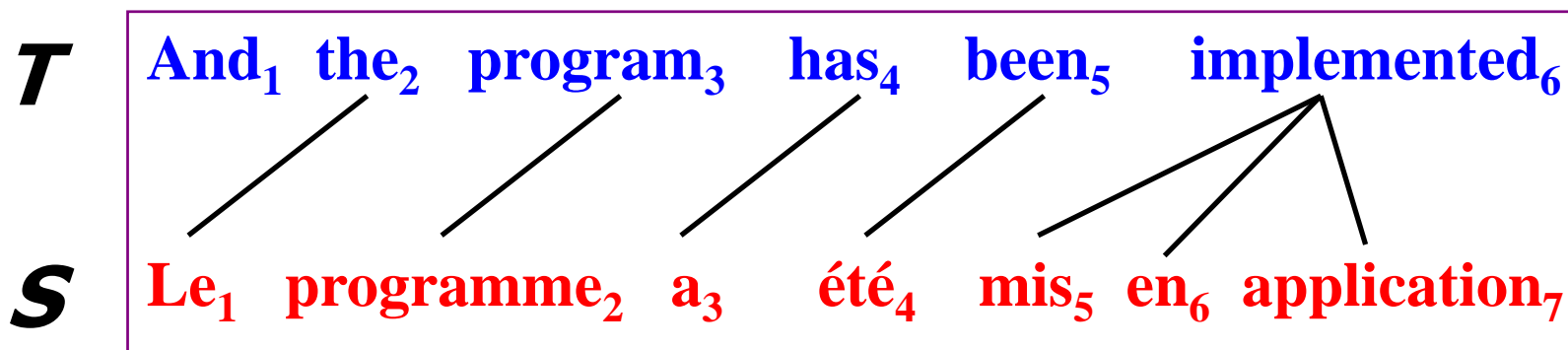
n -gram 问题, 不再赘述。

11.2.1 统计翻译基本原理

◆ 翻译概率 $p(S|T)$ 的计算

关键问题是怎样定义目标语言句子中的词与源语言句子中的词之间的对应关系。

假设英语与法语的翻译对：



11.2.1 统计翻译基本原理

不妨，我们用 $\mathcal{A}(S, T)$ 表示源语言句子 S 与目标语言句子 T 之间所有**对位关系**的集合。在目标语言句子 T 的长度（单词的个数）为 l ，源语言句子 S 的长度为 m 的情况下， T 和 S 的单词之间有 $2^{l \times m}$ 种不同的**对应关系**。

$$|\mathcal{A}(S, T)| = 2^{l \times m} \quad A(S, T) \in \mathcal{A}(S, T)$$

用来刻画这些对应关系 $A(S, T)$ 的模型叫做**对位模型** (alignment model)。

11.2.1 统计翻译基本原理

将对位模型 A 视为隐含变量，则：

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

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

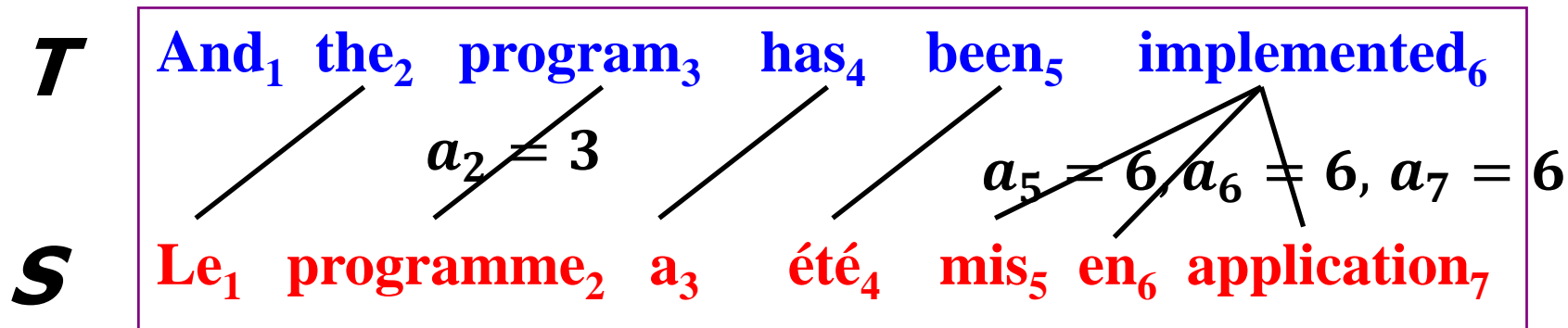
目标语言句子 $T = t_1^l = t_1 t_2 \cdots t_l$ 有 l 个单词

每一种对位序列表示成：

$$A = a_1^m = a_1 a_2 \cdots a_m \quad a_j \in [0, 1, \dots, l]$$

$a_j=i$ 表示从S的第 j 个词 到T的第 i 个词 的对位关系

11.2.1 统计翻译基本原理



$$l = 6 \quad m = 7$$

$$a_1 = 2, a_2 = 3, a_3 = 4, a_4 = 5, a_5 = 6, a_6 = 6, a_7 = 6$$

$$A = a_1^m = (2, 3, 4, 5, 6, 6, 6)$$

11.2.1 统计翻译基本原理

◆ 翻译概率 $P(S|T)$ 的计算

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

→ $P(S, A|T)$?

$$P(S, A|T) = P(m|T) \times P(A|T, m) \times P(S|T, A, m)$$

对位模型

词汇翻译模型

先以概率 $P(m|T)$ 选择一个长度 m ，然后以概率 $P(A|T, m)$ 生成对位序列 A ，然后以概率 $P(S|T, A, m)$ 生成源序列 S

11.2.1 统计翻译基本原理

$$P(S, A|T) = p(m|T) \times P(A|T, m) \times P(S|T, A, m)$$
$$= p(m|T) \prod_{j=1}^m p(a_j | \alpha_1^{j-1}, s_1^{j-1}, m, T) \times p(s_j | \alpha_1^j, s_1^{j-1}, m, T)$$

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。在上式的基础上，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。



11.2.2 IBM 翻译模型1

11.2.2 IBM 翻译模型1

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

翻译模型1：

(1) 假设 $\varepsilon \equiv p(m|T)$ 是一个较小的常量；

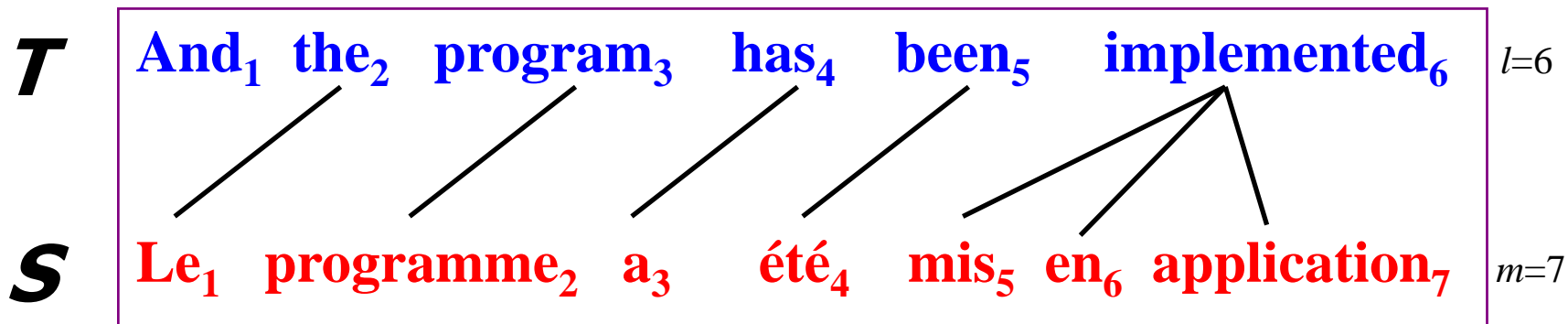
(2) 假设 $a_j \sim \text{uniform}(0, 1, 2, \dots, l)$ $p(a_j | a_1^{j-1}, s_1^{j-1}, m, T) = \frac{1}{l+1}$

(3) 假设 $s_j \sim \text{Categorical}(\theta_{t_{a_j}})$ $p(s_j | a_1^j, s_1^{j-1}, m, T) = p(s_j | t_{a_j})$

11.2.2 IBM 翻译模型1

$$\begin{aligned} P(S, A|T) &= p(m|T) \prod_{j=1}^m p(a_j | a_1^{j-1}, s_1^{j-1}, m, T) \times p(s_j | a_1^j, s_1^{j-1}, m, T) \\ &= \varepsilon \prod_{j=1}^m \frac{1}{l+1} \times p(s_j | t_{a_j}) \\ &= \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m p(s_j | t_{a_j}) \end{aligned}$$

11.2.2 IBM 翻译模型1



$$P(S, A|T) = \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m p(s_j | t_{a_j})$$

$$\frac{\varepsilon}{(6+1)^7} \times [p(\text{Le}|\text{the}) \times \dots \times p(\text{application}|\text{implemented})]$$

11.2.2 IBM 翻译模型1

$$P(S, A|T) = \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m p(s_j|t_{a_j})$$

$$\begin{aligned} P(S|T) &= \sum_A P(S, A|T) = \sum_A \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m p(s_j|t_{a_j}) \\ &= \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m p(s_j|t_{a_j}) \end{aligned}$$

如何训练？

11.2.2 IBM 翻译模型1

$$\operatorname{argmax} P(S|T)$$

$$\text{w.r.t. } \sum_s p(s|t) = 1$$



$$h(p, \lambda) = P(S|T) - \sum_t \lambda_t \left(\sum_s p(s|t) - 1 \right)$$

$$= \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m p(s_j | t_{a_j}) - \sum_t \lambda_t \left(\sum_s p(s|t) - 1 \right)$$

11.2.2 IBM 翻译模型1

$$h(p, \lambda) = \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m p(s_j | t_{a_j}) - \sum_t \lambda_t \left(\sum_s p(s|t) - 1 \right)$$

$$\frac{\partial h(p, \lambda)}{\partial p(s|t)} = 0$$



$$\frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \sum_{j=1}^m \delta(s = s_j) \delta(t = t_{a_j}) \frac{1}{p(s|t)} \prod_{k=1}^m p(s_k | t_{a_k}) - \lambda_t = 0$$

11.2.2 IBM 翻译模型1

$$\frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \sum_{j=1}^m \delta(s = s_j) \delta(t = t_{a_j}) \frac{1}{p(s|t)} \prod_{k=1}^m p(s_k | t_{a_k}) - \lambda_t = 0$$



$$p(s|t) =$$

单词 s_j 出现在源语言句子 S 中的次数

$$\frac{1}{\lambda_t} \times \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \sum_{j=1}^m \delta(s = s_j) \delta(t = t_{a_j}) \prod_{k=1}^m p(s_k | t_{a_k})$$

对位A中 t 连接到 s 的次数

11.2.2 IBM 翻译模型1

$$p(s|t) = \frac{1}{\lambda_t} \times \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \sum_{j=1}^m \delta(s = s_j) \delta(t = t_{a_j}) \prod_{k=1}^m p(s_k | t_{a_k})$$

EM参数训练思路：给翻译概率一个任意的初始估计值，我们可以计算出等式右边的值，并可以利用这个值作为新的 $p(s|t)$ 的估计值

$$p(s|t) = \frac{1}{\lambda_t} \sum_A P(S, A | T) \sum_{j=1}^m \delta(s = s_j) \delta(t = t_{a_j})$$

对位A中 t 连接到 s 的次数

$\sum_{j=1}^m \delta(s = s_j)$ 单词 s 出现在源语言句子 S 中的次数

$\sum_{i=0}^l \delta(t = t_i)$ 单词 t 出现在目标语言句子 T 中的次数

11.2.2 IBM 翻译模型1

忽略详细的数学推导，IBM 翻译模型1表示为如下等式：

$$\begin{aligned} P(S|T) &= \frac{\varepsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m p(s_j|t_{a_j}) \\ &= \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m \sum_{a_j=0}^l p(s_j|t_{a_j}) \end{aligned}$$

11.2.2 IBM 翻译模型1

$$\begin{aligned} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m p(s_j | t_{a_j}) &= \sum_{a_m=0}^l \cdots \sum_{a_1=0}^l \prod_{j=1}^m p(s_j | t_{a_j}) \\ &= \sum_{a_m=0}^l \cdots \sum_{a_2=0}^l \left\{ \sum_{a_1=0}^l p(s_1 | t_{a_1}) \prod_{j=2}^m p(s_j | t_{a_j}) \right\} \\ &= \sum_{a_m=0}^l \cdots \sum_{a_2=0}^l \left\{ \prod_{j=2}^m p(s_j | t_{a_j}) \left\{ \sum_{a_1=0}^l p(s_1 | t_{a_1}) \right\} \right\} \\ &= \sum_{a_1=0}^l p(s_1 | t_{a_1}) \sum_{a_m=0}^l \cdots \sum_{a_2=0}^l \left\{ \prod_{j=2}^m p(s_j | t_{a_j}) \right\} \end{aligned}$$

11.2.2 IBM 翻译模型1

根据IBM翻译模型1，由英语句子 e (T)生成法语句子 f (S) 的实现过程：

- (1) 根据概率分布为法语句子 f 选择一个长度 m ；
- (2) 对于每一个 $j = 1, 2, \dots, m$ ，根据均匀分布原则从 $0, 1, \dots, l$ 中选择一个值给 a_j ；
- (3) 对于每一个 $j = 1, 2, \dots, m$ ，根据概率 $p(f_j | e_{a_j})$ 选择一个法语单词 f_j 。

11.2.2 IBM 翻译模型1

e And₁ the₂ program₃ has₄ been₅ implemented₆ $l=6$

f f₁ f₂ f₃ f₄ f₅ f₆ f₇ $m=7$

$$\varepsilon \equiv p(m|T)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

$$\frac{1}{l+1}$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f Le₁ programme₂ a₃ été₄ mis₅ en₆ application₇

$$p(s_j|t_{a_j})$$





11.2.3 IBM 翻译模型2

11.2.3 IBM 翻译模型2

在IBM 模型2中，除了假定概率 $P(a_j | a_1^{j-1}, s_1^{j-1}, m, T)$ 依赖于位置 j 、对位关系 a_j 和源语言句子长度 m 以及目标语言句子长度 l 以外，另外两个假设与IBM模型1中的假设一样。

引入了对位概率(alignment probabilities)的概念：

$$a(a_j | j, m, l) = P(a_j | a_1^{j-1}, s_1^{j-1}, m, l)$$

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

11.2.3 IBM 翻译模型2

对于每一个三元组 (j, m, l) ，对位概率满足如下约束条件：

$$\sum_{i=0}^l a(i | j, m, l) = 1$$

类似于IBM模型1的推导，得到模型2：

$$p(S | T) = \varepsilon \prod_{j=1}^m \sum_{i=0}^l p(s_j | t_i) \times a(i | j, m, l) \quad (7)$$

如果对位概率设为常数，IBM 模型2退化为模型1，即模型1是模型2的特例。

模型1

$$p(S | T) = \frac{\varepsilon}{(l+1)^m} \prod_{j=1}^m \sum_{i=0}^l p(s_j | t_i)$$

11.2.3 IBM 翻译模型2

根据 IBM模型2，由英语句子 e 生成法语句子 f 的实现过程：

- (1) 根据概率分布为法语句子 f 选择一个长度 m ；
- (2) 对于每一个 $j = 1, 2, \dots, m$ ，根据概率分布 $a(a_j | j, l, m)$ 从 $0, 1, \dots, l$ 中选择一个值给 a_j ；
- (3) 对于每一个 $j = 1, 2, \dots, m$ ，根据概率选择一个法语单词 f_j 。

11.2.3 IBM 翻译模型2

e_m And₁ the₂ program₃ has₄ been₅ implemented₆ $l=6$

f_l f_1 f_2 f_3 f_4 f_5 f_6 f_7 $m=7$

$$\varepsilon \equiv p(m|T)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f_1 f_2 f_3 f_4 f_5 f_6 f_7

$$p(a_j|j, m, l)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f Le₁ programme₂ a₃ été₄ mis₅ en₆ application₇

$$p(s_j|t_{a_j})$$

11.2.3 IBM 翻译模型1-2

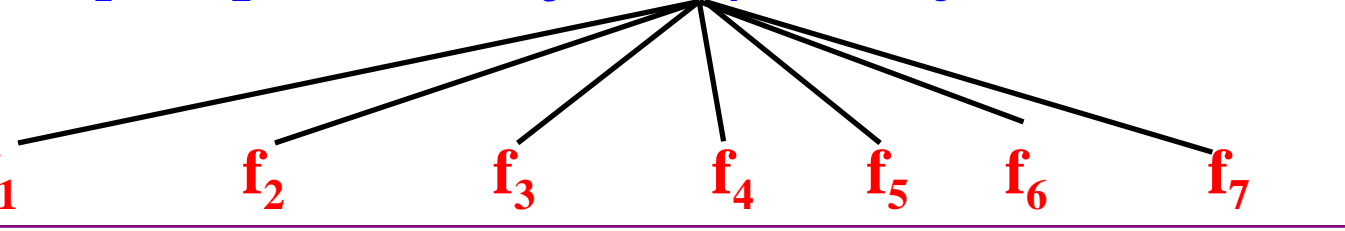
e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

$\varepsilon \equiv$
 $p(m|T)$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇



如何避免一个目标语言单词生成过多的源语言单词？



11.2.4 IBM 翻译模型3

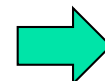
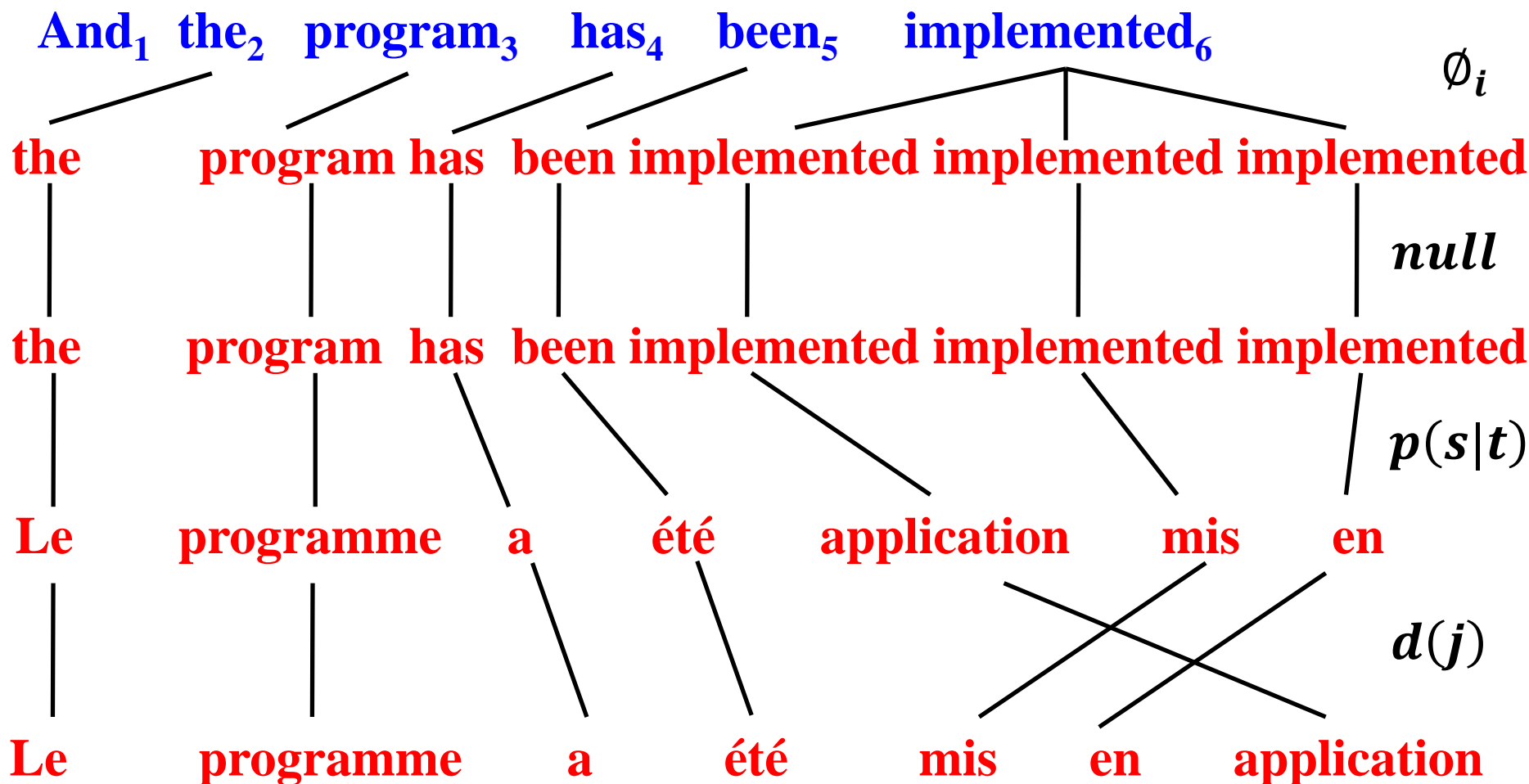
11.2.4 IBM 翻译模型3

对目标语言单词生成源语言单词的数目进行建模？

定义：在随机选择对位关系的情况下，与目标语言句子中的单词 t 对应的源语言句子中的单词数目是一个随机变量，不妨记做 Φ_t ，我们称该变量为单词 t 的繁衍能力或产出率(fertility)。一个具体的取值记做： ϕ_t

实际上，所谓的繁衍能力就是目标语言单词与源语言单词之间一对多的关系。

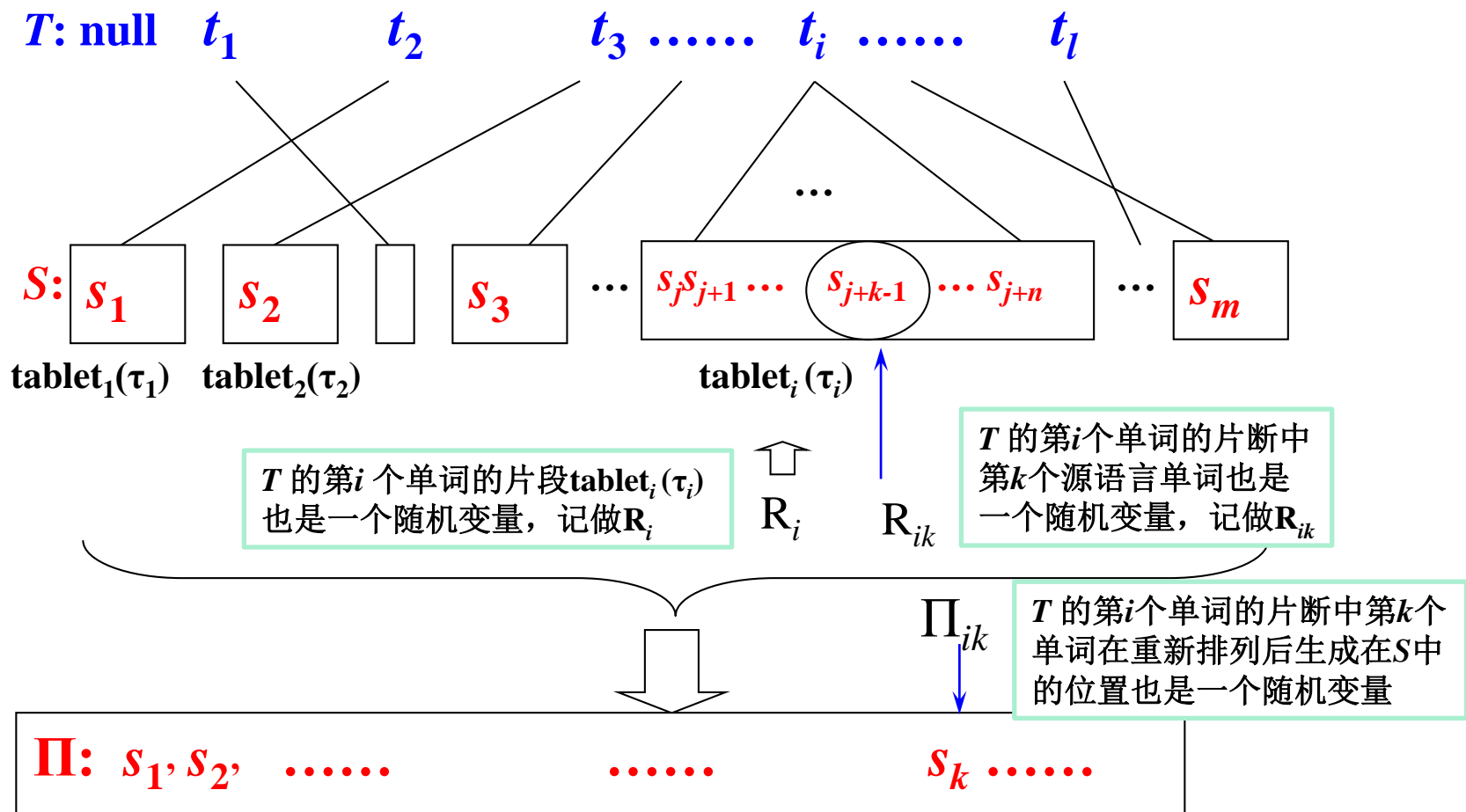
11.2.4 IBM 翻译模型3



11.2.4 IBM 翻译模型3

定义：假设给定一个目标语言句子 T ， T 中的每一个单词 t 在源语言句子中可能有若干个词与之对应，源语言句子中所有与 t 对位的单词列表我们称之为 t 的一个**片断(tablet)**，这个片断可能为**空**。一个目标语言句子 T 的所有片断的集合是一个随机变量，我们称之为 T 的**片断集(tableau)**，记做符号 \mathbf{R} 。 T 的第 i 个单词的片段也是一个随机变量，记做 \mathbf{R}_i ，那么， T 的第 i 个单词的片断中第 k 个源语言单词也是一个随机变量，记做 \mathbf{R}_{ik} 。

11.2.4 IBM 翻译模型3



11.2.4 IBM 翻译模型3

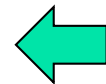
片段集 τ (\mathbf{R} 的一个具体取值) 和单词排列 π ($\mathbf{\Pi}$ 的一个具体取值, 即 τ 中单词的一种排列方式) 的联合似然率为:

$$p(\tau, \pi | T) = \prod_{i=1}^l p(\phi_i | \phi_1^{i-1}, T) \times p(\phi_0 | \phi_1^l, T) \times \prod_{i=0}^l \prod_{k=1}^{\phi_i} p(\tau_{ik} | \tau_{i1}^{k-1}, \tau_0^{i-1}, \phi_0^l, T) \times \prod_{i=1}^l \prod_{k=1}^{\phi_i} p(\pi_{ik} | \pi_{i1}^{k-1}, \pi_1^{i-1}, \tau_0^l, \phi_0^l, T) \times \prod_{k=1}^{\phi_0} p(\pi_{0k} | \pi_{01}^{k-1}, \pi_1^l, \tau_0^l, \phi_0^l, T) \quad (8)$$

繁衍概率
(fertility prob.)

翻译概率
(tran. prob.)

位变概率
(distortion prob.)



11.2.4 IBM 翻译模型3

假设：

(1) 对于1到 l 中的每一个 i ，概率 $p(\phi_i | \phi_1^{i-1}, T)$ 仅依赖于 ϕ_i 和 t_i ，记作： $n(\phi | t_i) \equiv p(\phi | \phi_1^{i-1}, T)$

(2) 对于所有的 i ，概率 $p(\tau_{ik} | \tau_{i1}^{k-1}, \tau_0^{i-1}, \phi_0^l, T)$ 只依赖于 τ_{ik} 和 t_i ，记作： $p(s | t_i) \equiv p(R_{ik} = s | \tau_{i1}^{k-1}, \tau_0^{k-1}, \phi_0^l, T)$

(3) 对于1到 l 中的每一个 i ， $p(\pi_{ik} | \pi_{i1}^{k-1}, \pi_1^{i-1}, \tau_0^l, \phi_0^l, T)$ 只依赖于 π_{ik} ， i ， m 和 l 。位置概率记作：

$$d(j | i, m, l) \equiv p(\Pi_{ik} = j | \pi_{i1}^{k-1}, \pi_1^{i-1}, \tau_0^l, \phi_0^l, T)$$

11.2.4 IBM 翻译模型3

设想 τ_1^l 中每个写字板中的一组词都存在一个额外词 (即这个词在对齐时对空), 假设这个额外词出现的概率为 p_1 。另外, 由于 $\phi_0 + \phi_1 + \dots + \phi_l = m$, 因此,

$$\begin{aligned} p(S|T) &= \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l p(S, A|T) \\ &= \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \binom{m-\phi_0}{\phi_0} (1-p_1)_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i | t_i) \\ &\quad \times \prod_{j=1}^m p(s_j | t_{a_j}) d(j | a_j, m, l) \end{aligned} \quad (9)$$

11.2.4 IBM 翻译模型3

其中,

$$\sum_s p(s | t) = 1$$

$$\sum_j d(j | i, m, l) = 1$$

$$\sum_\varphi n(\varphi | t) = 1$$

估计这些参数和 p_1 。

11.2.4 IBM 翻译模型3

根据 IBM模型3, 一个英语句子 e 翻译成法语句子 f 的工作过程如下:

- (1) 对于英语句子中的每一个单词 e , 选择一个产出率 ϕ , 其概率为 $n(\phi|e)$;
- (2) 对于所有单词的产出率求和, 得到 $m\text{-prime}$;
- (3) 按照下面的方式构造一个新的英语单词串: 删除产出率为0的单词, 复制产出率为1的单词, 复制两遍产出率为2的单词, 依此类推;
- (4) 在这 $m\text{-prime}$ 个单词的每一个后面, 决定是否插入一个空单词NULL, 插入的概率为 p_1 , 不插入的概率为 p_0 ;

11.2.4 IBM 翻译模型3

- (5) 设 ϕ_0 为插入空单词 NULL 的个数;
- (6) 设 m 为目前总单词的个数: $m - \text{prime} + \phi_0$;
- (7) 根据概率表 $p(f|e)$ 将每一个单词 e 替换为法语单词 f ;
- (8) 对不是由空单词 NULL 产生的每一个法语单词, 根据概率表 $d(j | i, l, m)$ 赋予一个位置。这里 j 是法语单词在法语句子中的位置, i 是产生当前这个法语单词的英语单词在其句子中的位置, l 是英语句子的长度, m 是法语句子的长度;
- (9) 如果任何一个法语句子的位置被多重登录(含一个以上的单词), 则失败返回;



11.2.4 IBM 翻译模型3

- (10) 给空单词NULL产生的法语单词在句子中赋予一个位置，这些位置必须是没有被占领的空位置。任何一个赋值都被认为是等概率的，概率值为 $1/\phi_0$;
- (11) 读出法语单词串，其概率为上述每一步概率的乘积，按概率大小输出结果。



11.2.5 IBM 模型4和模型5

11.2.5 IBM 模型4和模型5

- ◆ **模型 4:** 考虑片断的中心词的概率和其他单词的位置概率。
- ◆ **模型 5:** 源语言句子单词间的相对位置。
- ◆ **请参见文献:** Peter F. Brown *et al.* 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2): 263-311.

11.2.5 IBM 模型4和模型5

模型	假设	参数训练	简评
IBM1	翻译模型仅与单词间的直译概率有关，句长概率和对齐概率都是均匀分布。	应用EM算法，从双语语料库中训练获得，可以得到全局最优参数，与初始值无关。	模型简单、易于实现，但仅考虑了单词的影响，没有考虑词序的影响。
IBM2	翻译模型和句长模型同IBM1，对位概率为0阶对齐。	应用EM算法，从双语语料库中训练获得，只能收敛到局部最优。	模型简单、易于实现，同时考虑了单词和词序的影响。
IBM3	翻译模型依赖于繁衍率模型和单词间的直译概率，对齐概率取0阶词对齐。	需要首先应用模型IBM1或IBM2对双语语料进行单词级对位，然后训练繁衍概率参数。	引入了描述单词间一对多情况的繁衍概率，参数较多，实现过程较复杂。
IBM4	翻译模型依赖于单词间的直译概率、繁衍概率、词类、语言片断中心位置和语言片断内相对位置等因素，对齐概率取1阶词对齐。	需要首先应用模型IBM1~IBM3对双语语料进行单词级对位和语言片断划分，然后训练两种位置概率参数。	不仅考虑了一对多的情况，还将语言片断作为一个整体进行考虑。参数较多、不易实现。
IBM5	翻译模型依赖于直译概率、繁衍概率、语言片断中心位置、语言片断内相对位置和对位的历史等因素。	需要在模型IBM1~ IBM4参数训练的基础上获得参数。	对IBM4进行了修正，同时考虑了当前对位信息和对位历史。模型的表现力最强，但过于复杂，实用性不强。
HMM	句长模型和翻译模型同IBM1，对齐模型为1阶对齐。	应用EM算法，从双语对照语料中训练获得。	模型简单，易于实现，考虑了词序的影响。

11.2.2 IBM 翻译模型1

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

$$\varepsilon \equiv p(m|T)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

$$\frac{1}{l+1}$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f Le₁ programme₂ a₃ été₄ mis₅ en₆ application₇

$$p(s_j|t_{a_j})$$

11.2.3 IBM 翻译模型2

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

$$\varepsilon \equiv p(m|T)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇

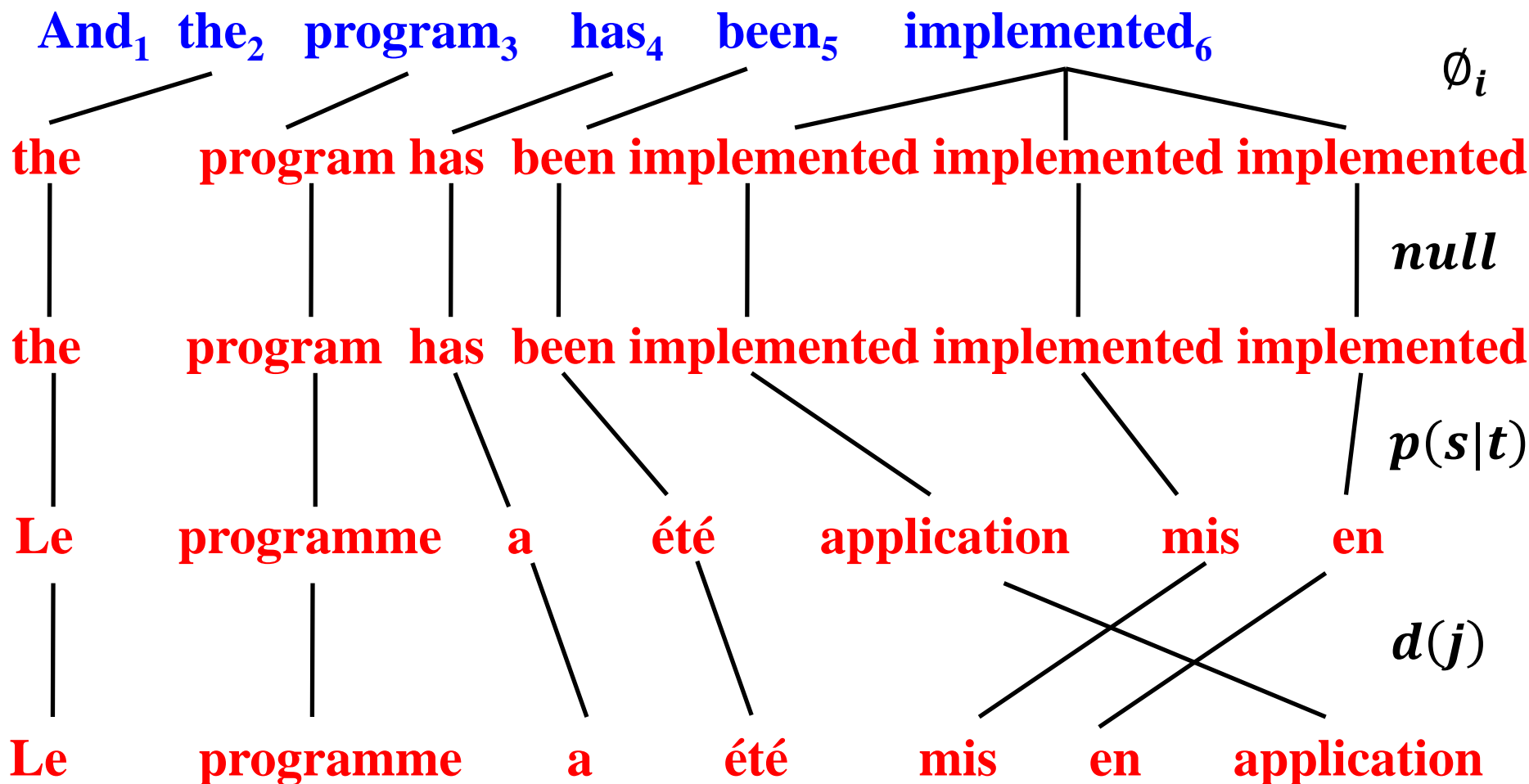
$$p(a_j|j, m, l)$$

e And₁ the₂ program₃ has₄ been₅ implemented₆

f Le₁ programme₂ a₃ été₄ mis₅ en₆ application₇

$$p(s_j|t_{a_j})$$

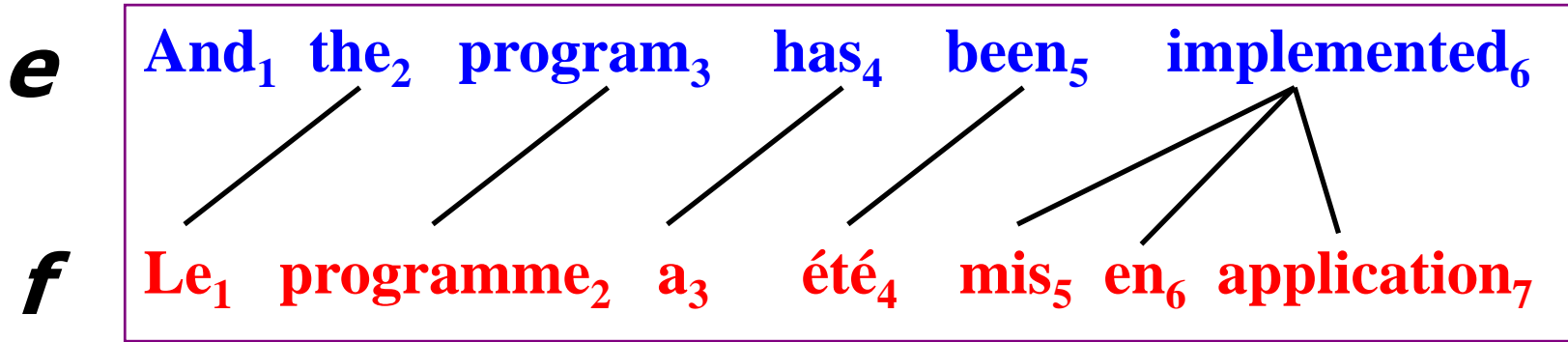
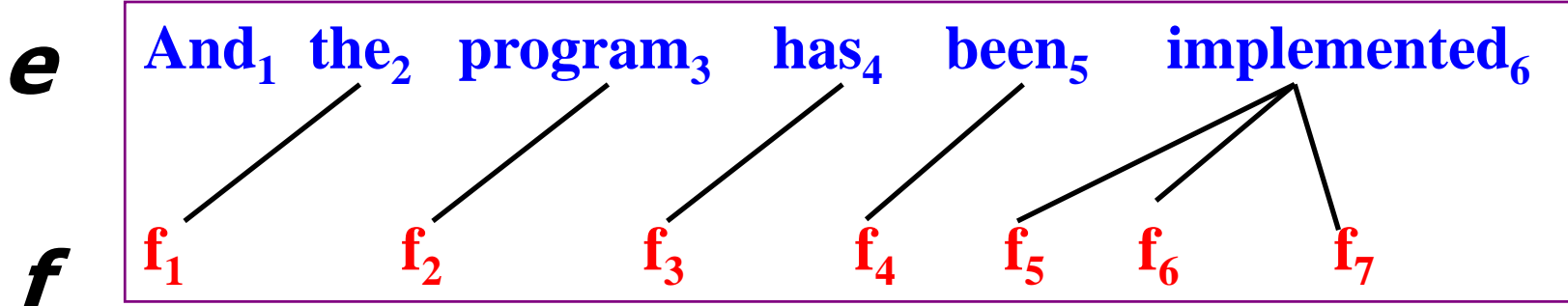
11.2.4 IBM 翻译模型3



11.2.2 IBM 翻译模型

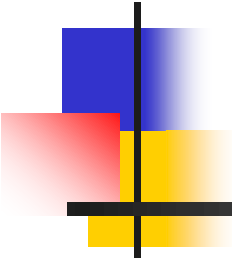
e And₁ the₂ program₃ has₄ been₅ implemented₆

f f₁ f₂ f₃ f₄ f₅ f₆ f₇



$$p(s_j | t_{a_j})$$





11.2.6 基于短语的 翻译模型

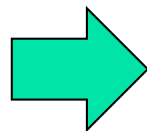
11.2.6 基于短语的翻译模型

◆生成式模型转向判别式模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_T \frac{P(T) \times P(S|T)}{P(S)} \\ &= \operatorname{argmax}_T P(T) \times P(S|T) \end{aligned}$$

生成式模型

翻译质量



≈

$$T' = \operatorname{argmax}_T P(T) \times P(T|S)$$

?

翻译质量



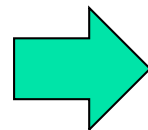
11.2.6 基于短语的翻译模型

◆生成式模型转向判别式模型

$$T' = \operatorname{argmax}_T P(T) \times P(S|T)$$

$$T' = \operatorname{argmax}_T P(T) \times P(T|S)$$

翻译质量



<

$$T' = \operatorname{argmax}_T P(T) \times P(S|T) \times P(T|S)$$

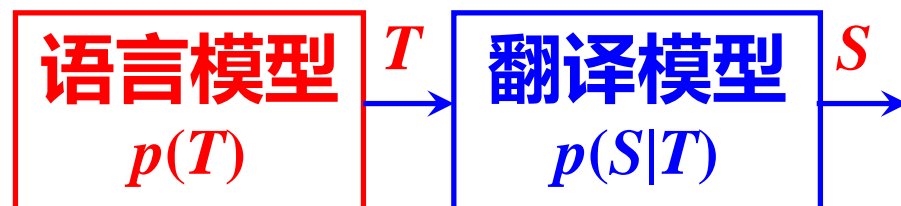
翻译质量

11.2.6 基于短语的翻译模型

◆ 基于最大熵的方法 (判别式)[Och, 2002等]

➤ 问题:

- 反向的翻译模型用噪声信道模型无法解释。



➤ 解决方法:

- 将语言模型概率、正向和反向翻译概率都视为一个特征，采用最大熵方法建模



11.2.6 基于短语的翻译模型

◆ 最大熵方法的基本思想

➤ 任务

- 对于一个随机事件，假设已经有了一组样例，我们希望建立一个统计模型来模拟这个随机事件的分布

➤ 目标

- 对于一组特征，使得统计模型在这一组特征上的模型分布与样例中的经验分布完全一致，同时不对未知事件作任何假设，即保证这个模型尽可能的“均匀”（也就是要求模型的熵值达到最大）

11.2.6 基于短语的翻译模型

- ▶ 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权值。

利用最大熵模型，给定源语言句子 S ，其最佳译文 T 可以用以下公式求得：

$$T' = \operatorname{argmax}_T P(T|S) = \operatorname{argmax}_T \frac{\exp\{\sum_1^M \lambda_m h_m(T, S)\}}{\sum_{T^*} \exp\{\sum_1^M \lambda_m h_m(T^*, S)\}}$$
$$= \operatorname{argmax}_T \left\{ \sum_1^M \lambda_m h_m(T, S) \right\}$$

11.2.6 基于短语的翻译模型

如果将两个特征分别取为 $h_1(T, S) = \log P(T)$ 和 $h_2(T, S) = \log P(S|T)$ ，并取 $\lambda_1 = \lambda_2 = 1$ ，那么，这个模型就等价于噪声信道模型，基于噪声信道模型的方法只是基于最大熵方法的一种特例。

$$T' = \operatorname{argmax}_T P(T|S) = \operatorname{argmax}_T \left\{ \sum_1^M \lambda_m h_m(T, S) \right\}$$

$$= \operatorname{argmax}_T \{ \log P(T) + \log P(S|T) \}$$

$$= \operatorname{argmax}_T \{ \log (P(T) \times P(S|T)) \}$$

$$= \operatorname{argmax}_T \{ P(T) \times P(S|T) \}$$

对数线性模型

11.2.6 基于短语的翻译模型

◆ 翻译基本单元由词转向短语

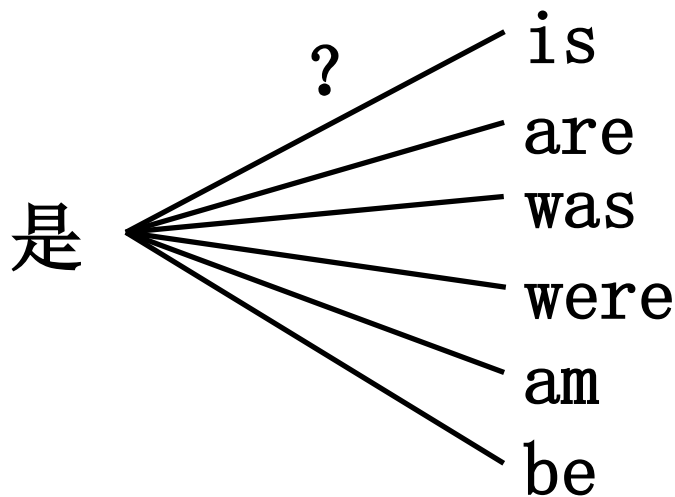
- 基于词的翻译模型的问题：
 - 很难处理词义消歧问题
 - 很难处理一对多、多对一和多对多的翻译问题

澳洲是与北韩有邦交的少数国家之一

11.2.6 基于短语的翻译模型

◆ 词义消歧问题

澳洲是与北韩有邦交的少数国家之一



在基于词的模型中，处理词义消歧问题需要充分利用上下文信息，并对上下文信息进行有效建模。

11.2.6 基于短语的翻译模型

◆ 一对多、多对一与多对多翻译问题

澳洲 是 与 北韩 有 邦交 的 少数 国家 之一

北韩 $\overset{?}{\rightarrow}$ North Korea

邦交 $\overset{?}{\rightarrow}$ the diplomatic relations

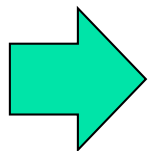
在基于词的模型中，处理上述问题同时需要准确的繁衍率模型、词汇翻译模型以及对位模型。

11.2.6 基于短语的翻译模型

◆ 以短语为基本翻译单元

澳洲是与北韩有邦交的少数国家之一

短语翻
译规则



(澳洲 是, Australia is)

(北韩, Korea North)

(邦交, the diplomatic relations)

在基于短语的模型中，直接将繁衍率信息、上下文信息以及局部对位调序信息记录在翻译规则中。

11.2.6 基于短语的翻译模型

◆ 以短语为基本翻译单元

澳洲是与北韩有邦交的少数国家之一

(澳洲是, Australia is)

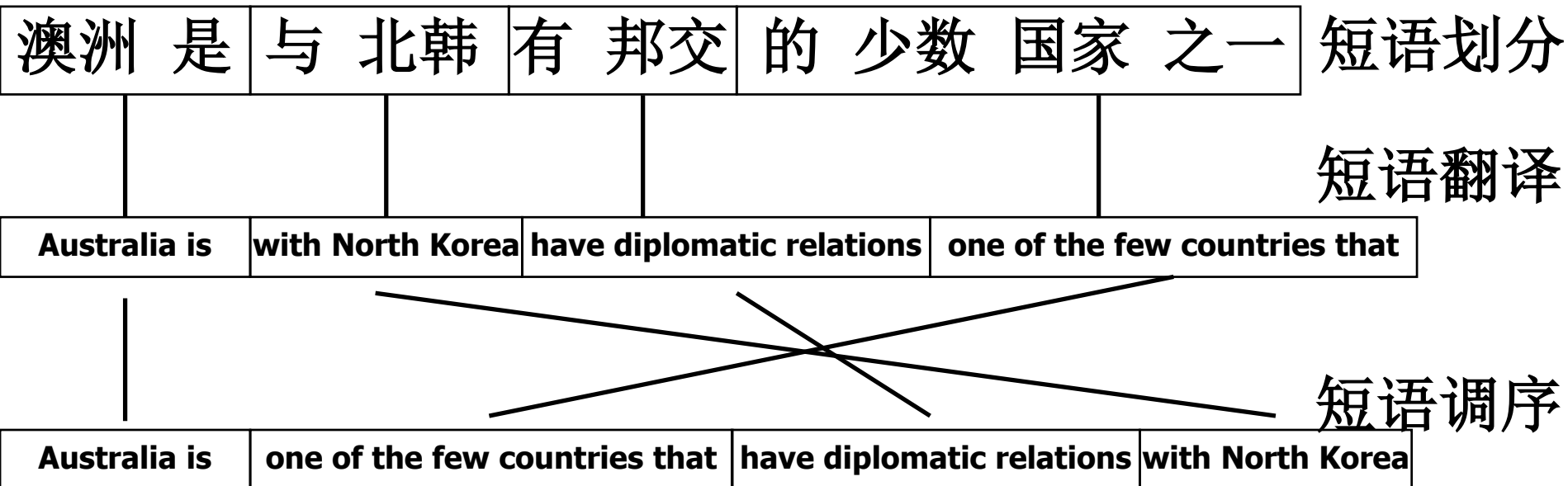
(与北韩, with North Korea)

(有邦交, have the diplomatic relations)

(的少数国家之一, one of the few countries that)

11.2.6 基于短语的翻译模型

◆ 以短语为基本翻译单元



11.2.6 基于短语的翻译模型

◆ 基于短语的翻译模型 [Koehn, 2003]



Philipp Koehn

短语：连续的词串（非句法意义）

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} P(S_1^K | S) \times P(T_1^K | S_1^K, S) \times \\ &\quad P(T_1^{K'} | T_1^K, S_1^K, S) \times P(T | T_1^{K'}, T_1^K, S_1^K, S) \end{aligned}$$

JHU、爱丁堡大学 教授

K' 表示对 K 的顺序进行调整

11.2.6 基于短语的翻译模型

➤ 注意：

这里所说的**短语**指一个连续的词串(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.

11.2.6 基于短语的翻译模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} \underbrace{P(S_1^K | S)}_{\text{短语划分模型}} \underbrace{P(T_1^K | S_1^K, S)}_{\text{短语翻译模型}} \underbrace{P(T_1^{K'} | T_1^K, S_1^K, S)}_{\text{短语调序模型}} \underbrace{P(T | T_1^{K'}, T_1^K, S_1^K, S)}_{\text{目标语言模型}} \end{aligned}$$

11.2.6 基于短语的翻译模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} \underbrace{P(S_1^K | S)}_{\downarrow} \underbrace{P(T_1^K | S_1^K, S)} \underbrace{P(T_1^{K'} | T_1^K, S_1^K, S)} \underbrace{P(T | T_1^{K'}, T_1^K, S_1^K, S)} \end{aligned}$$

短语划分模型

目标：将一个词序列如何划分为短语序列

方法：一般假设每一种短语划分方式都是等概率的

11.2.6 基于短语的翻译模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} \underbrace{P(S_1^K | S)} \underbrace{P(T_1^K | S_1^K, S)} \underbrace{P(T_1^{K'} | T_1^K, S_1^K, S)} \underbrace{P(T | T_1^{K'}, T_1^K, S_1^K, S)} \end{aligned}$$

剩下的三个核心模型：

1. 短语翻译模型： $P(T_1^K | S_1^K, S)$
2. 短语调序模型： $P(T_1^{K'} | T_1^K, S_1^K, S)$
3. 目标语言模型： $P(T | T_1^{K'}, T_1^K, S_1^K, S)$

11.2.6 基于短语的翻译模型

短语翻译模型: $P(T_1^K | S_1^K, S)$

1. 如何学习短语翻译规则
2. 如何估计短语翻译概率

双语句对词语对齐

短语翻译规则抽取

(澳洲 是, Australia is)

(与 北韩, with North Korea)

(有 邦交, have the diplomatic relations)

(的 少数 国家 之一, one of the few countries that)



11.2.6 基于短语的翻译模型

双语句对词语对齐

澳洲 是 与 北韩 有 邦交 的 少数 国家 之一

Australia is one of the few countries that have diplomatic relations with North Korea



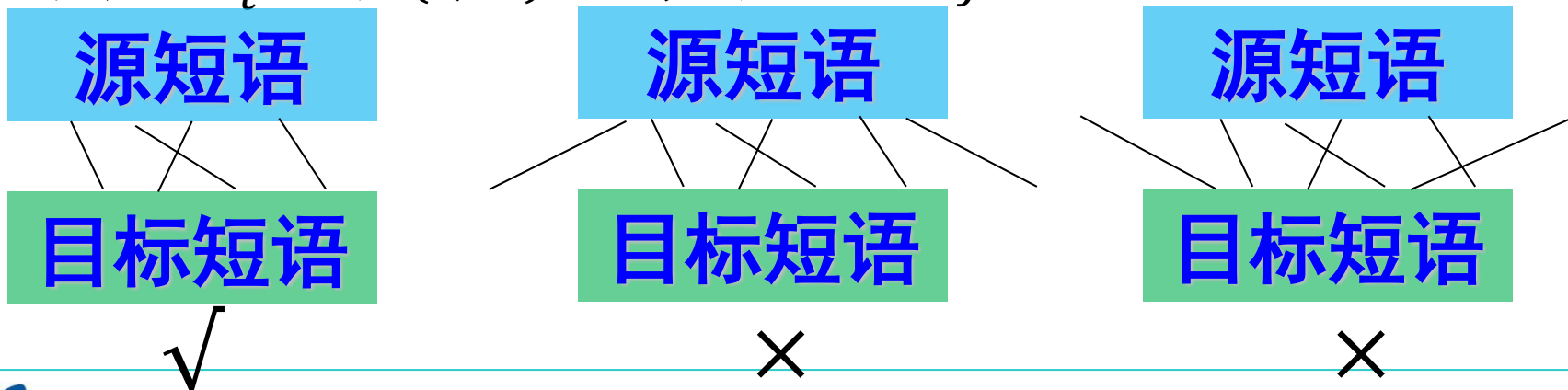
IBM model 1-5

11.2.6 基于短语的翻译模型

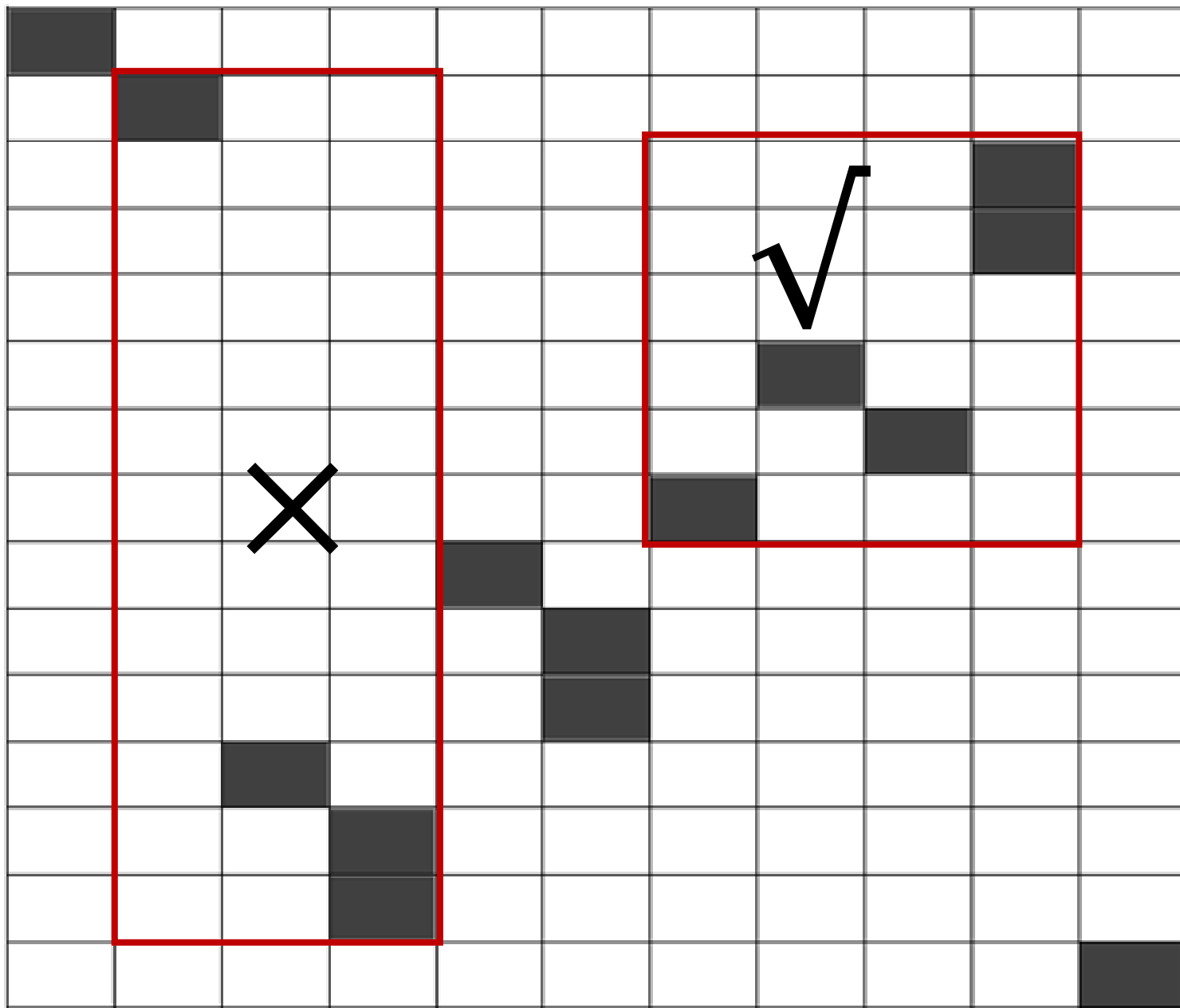
短语翻译规则抽取

算法: 对于源语言句子 S 中的任一短语 S_i^j , 根据词语对齐 A 找到目标语言句子 T 中的对齐片段 $T_{i'}^{j'}$, 若 S_i^j 与 $T_{i'}^{j'}$ 满足对齐一致性, 则 $(S_i^j, T_{i'}^{j'})$ 为一条短语翻译规则。

对齐一致性: S_i^j 中每个词 S_k , 若 $(k, k') \in A$, 则 $i' \leq k' \leq j'$, $T_{i'}^{j'}$ 中每个词 $T_{t'}$, 若 $(t, t') \in A$, 则 $i \leq t \leq j$ 。



澳 洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之 一 。



Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 是 与 北 有 邦 的 少 国 之 一 。



(澳洲 Australia)

Australia

is

one

of

the

few

countries

that

have

diplomatic

relations

with

North

Korea

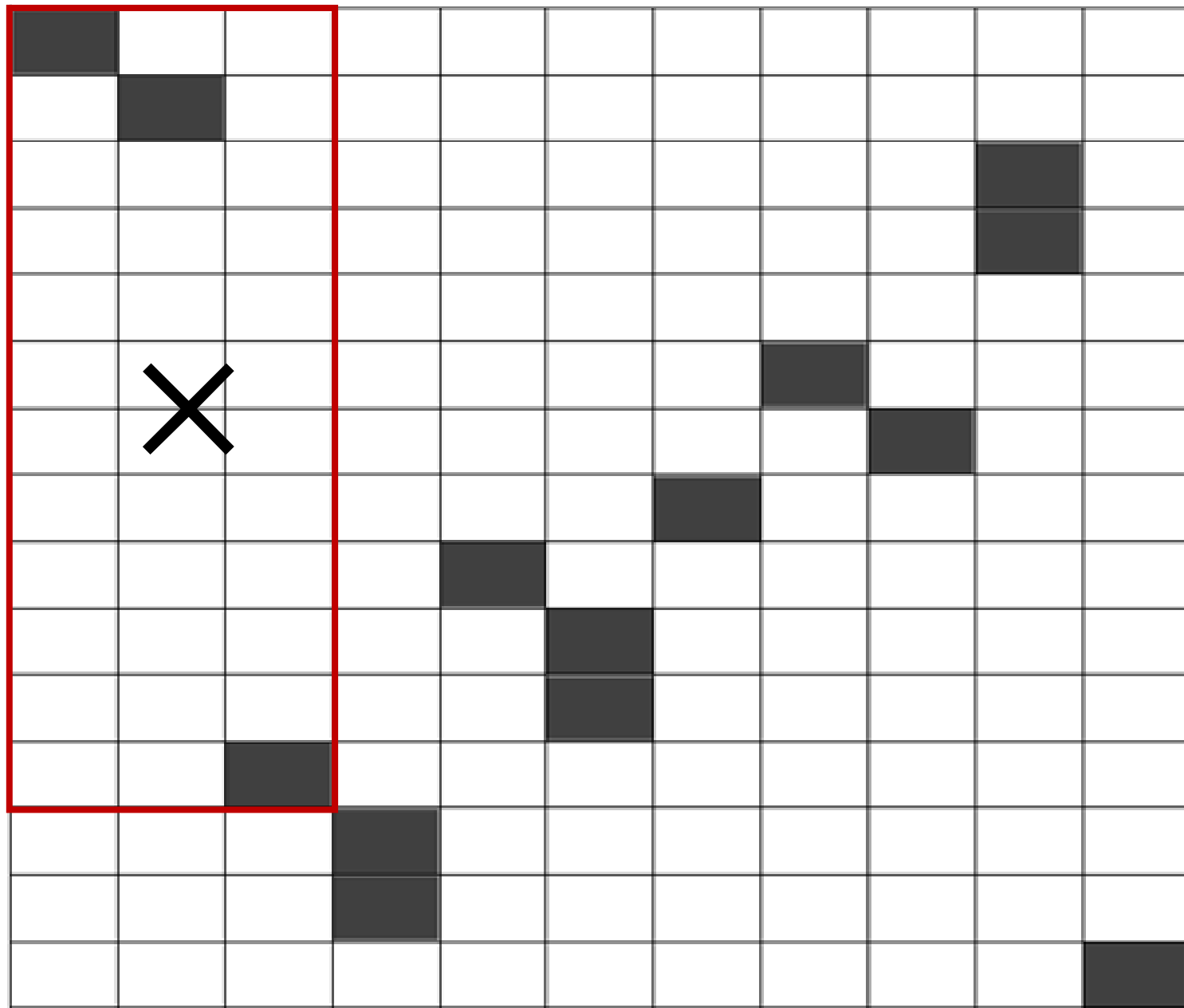
.

澳 洲 是 与 北 有 邦 的 少 国 家 之 一 。

■	■									
■	■	(澳	洲	是	Australia	is)			
								■	■	
								■	■	
							■	■		
							■	■		
				■	■					
					■	■				
		■	■							
			■	■						
			■	■						
										■

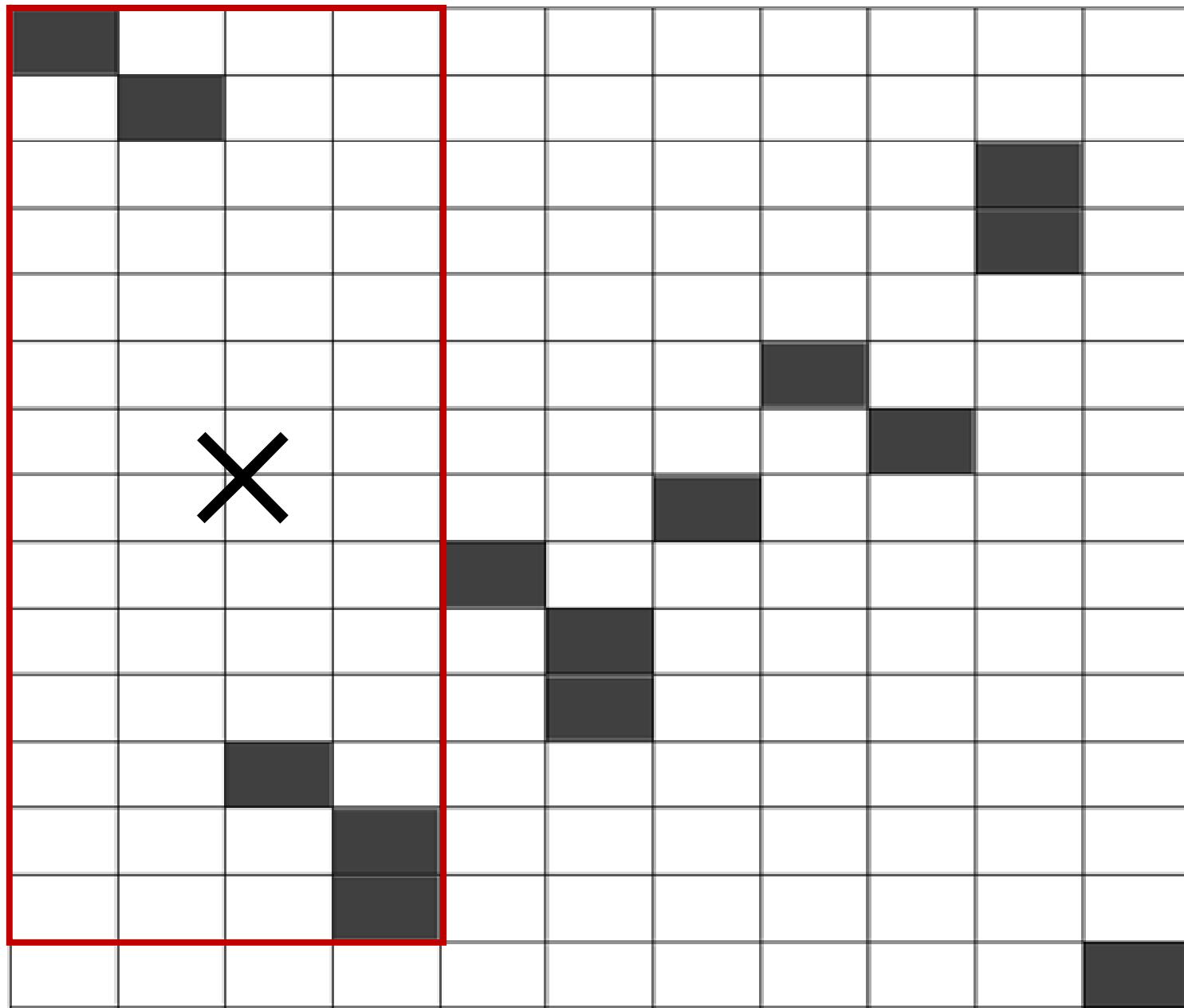
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 是 与 北 有 邦 的 少 国 家 之 一 。



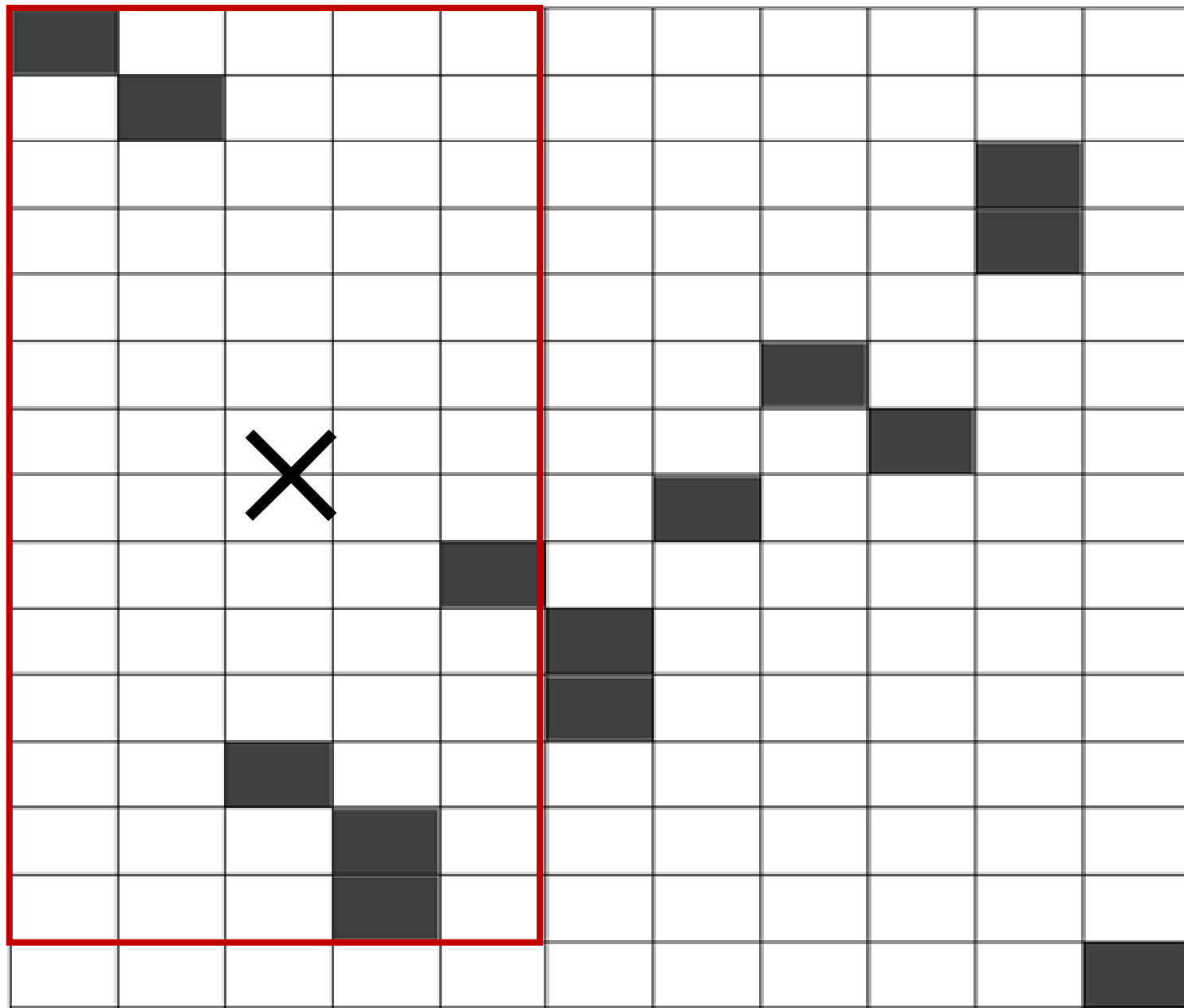
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 是 与 北 有 邦 的 少 国 家 之 一 。



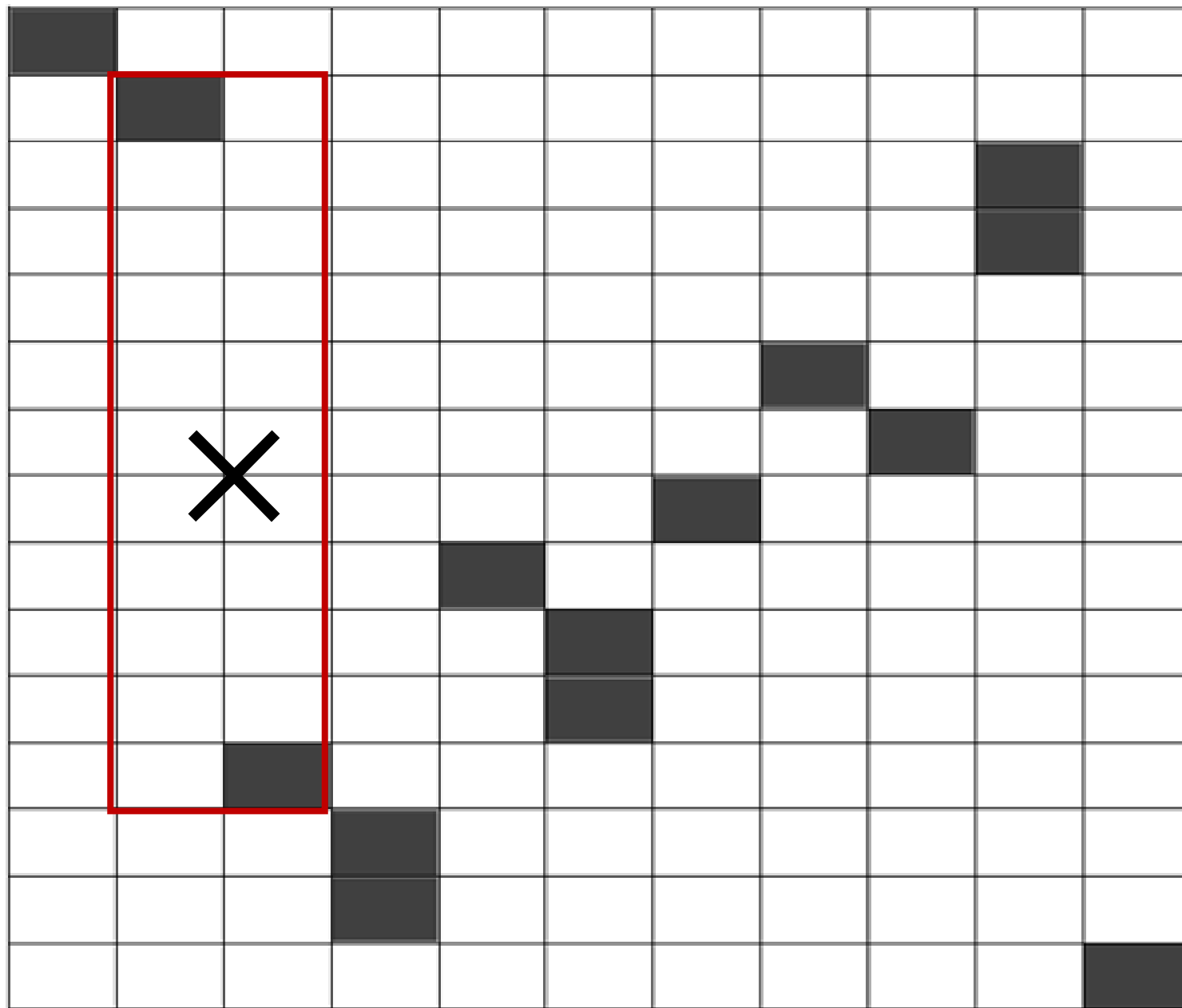
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 洲 是 与 北 有 邦 的 少 国 家 之 一 。



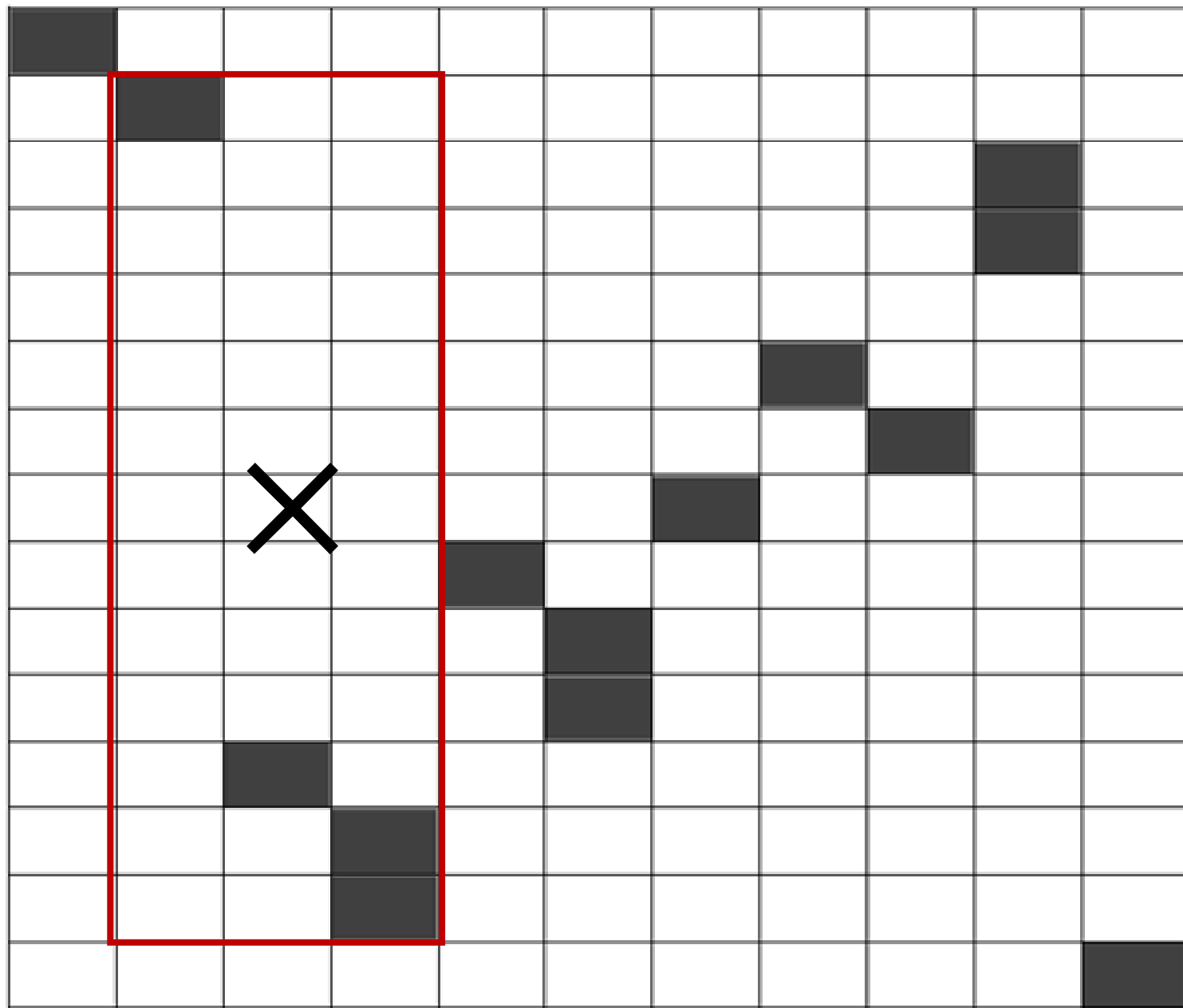
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之 一 。



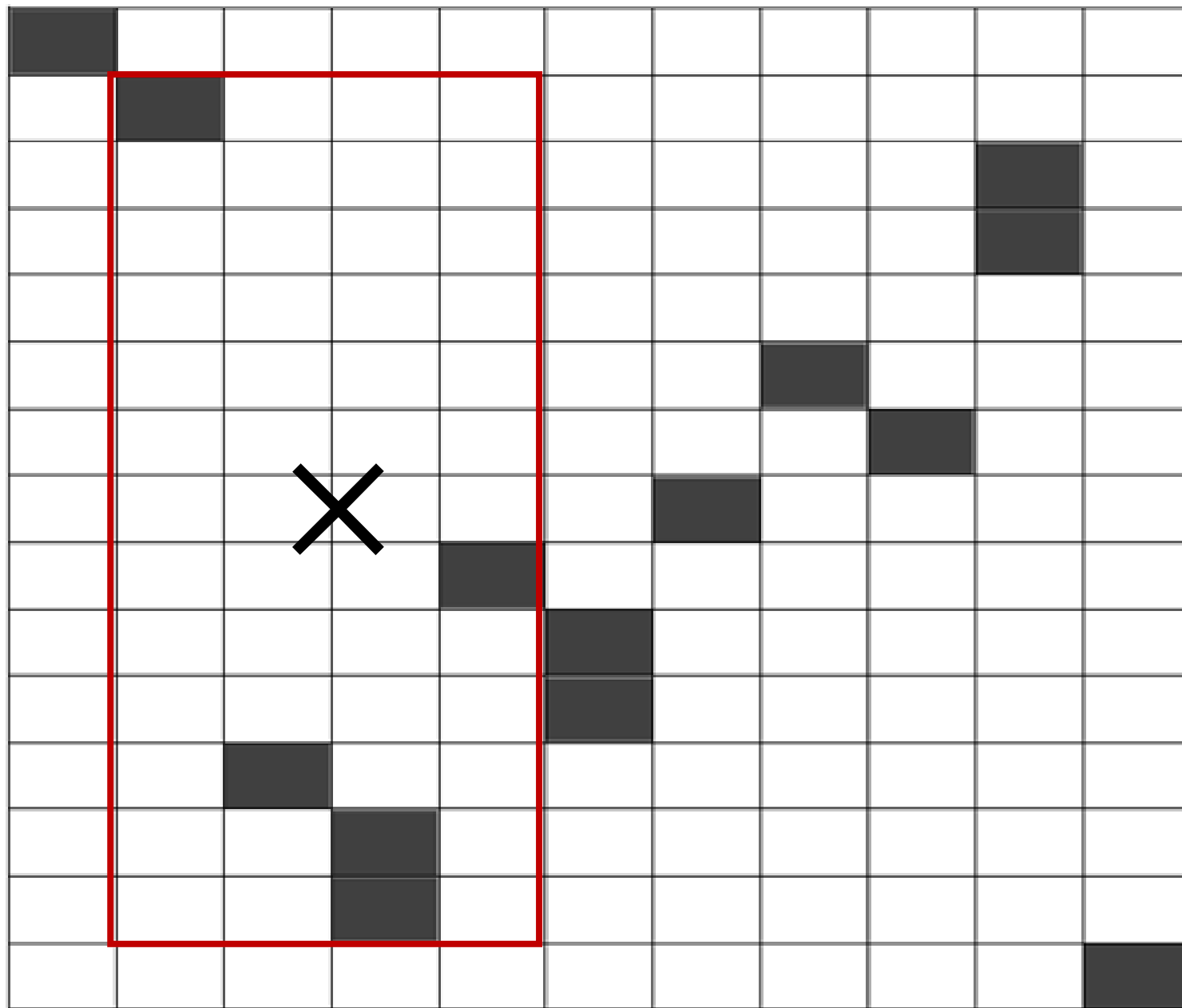
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之 一 。



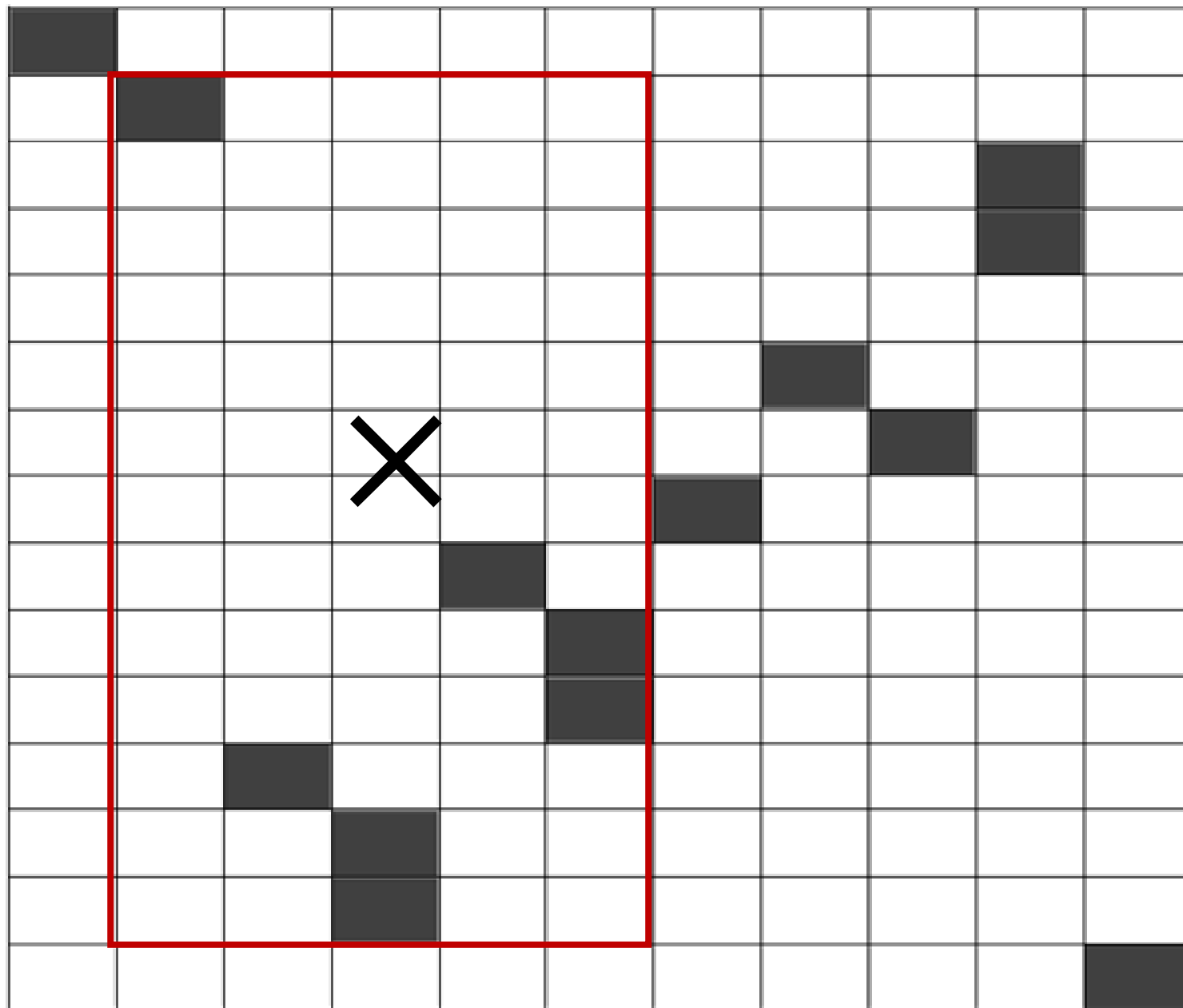
Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 洲 是 与 北 有 邦 的 少 国 家 之 一 。



Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

澳 洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之 一 。



Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

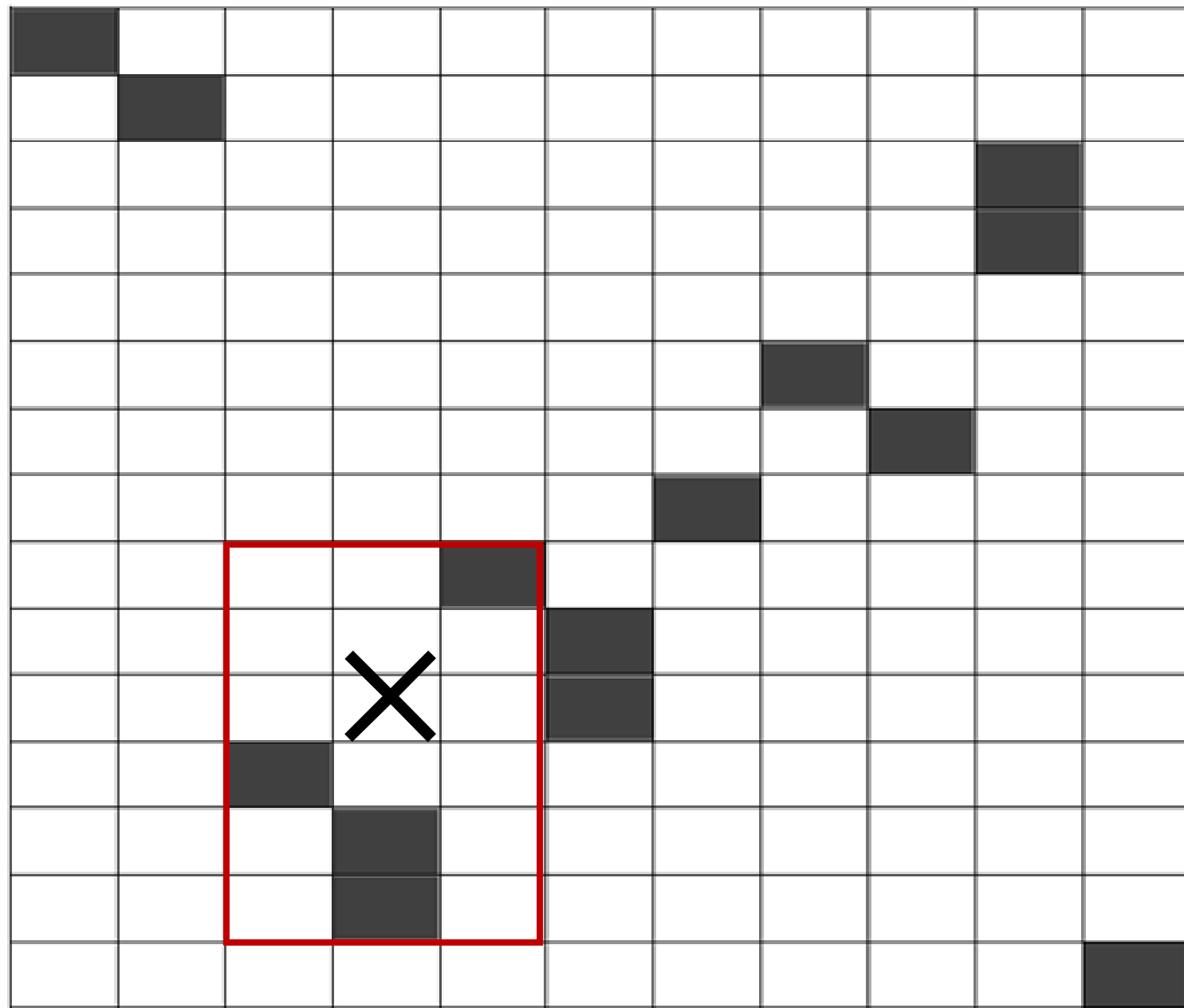
澳 是 与 北 有 邦 的 少 国 之 一 。

澳										
	是									
		与								
			北							
				有						
					邦					
						的				
							少			
								国		
									之	
									一	
										。

Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

(与 北韩 with North Korea)

澳 洲 是 与 北 有 邦 的 少 国 家 之 一 。



Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea
.

11.2.6 基于短语的翻译模型

短语翻译模型: $P(T_1^K | S_1^K, S)$

1. 如何学习短语翻译规则

2. 如何估计短语翻译概率

双语句对词语对齐

短语翻译规则抽取

(澳洲 是, Australia is)

(与 北韩, with North Korea)

(有 邦交, have the diplomatic relations)

(的 少数 国家 之一, one of the few countries that)

11.2.6 基于短语的翻译模型

短语翻译概率估计：4个翻译概率（最大似然）

1. 正向、逆向短语翻译概率 $p(t|s), p(s|t)$
2. 正向、逆向词汇化翻译概率 $p_{lex}(t|s), p_{lex}(s|t)$

11.2.6 基于短语的翻译模型

短语翻译概率估计：4个翻译概率（最大似然）

1. 正向、逆向短语翻译概率 $p(t|s), p(s|t)$
2. 正向、逆向词汇化翻译概率 $p_{lex}(t|s), p_{lex}(s|t)$

(与 北韩, with North Korea)

(与 北韩, and North Korea)

(和 北韩, with North Korea)

(与 朝鲜, with North Korea)

s

t

$$p(t|s) = \frac{1}{2}$$

$$p(s|t) = \frac{1}{3}$$

11.2.6 基于短语的翻译模型

短语翻译概率估计：4个翻译概率（最大似然）

1. 正向、逆向短语翻译概率 $p(t|s), p(s|t)$

2. 正向、逆向词汇化翻译概率 $p_{lex}(t|s), p_{lex}(s|t)$

s t
(与 北韩, with North Korea)

$$p(\text{with}|\text{与}) = 0.4$$

$$p(\text{North}|\text{北韩}) = 0.1$$

$$p(\text{Korea}|\text{北韩}) = 0.5$$

$$p_{lex}(t|s) = \prod_{j=1}^{|t|} \frac{1}{|\{i | (j, i) \in A\}|} \sum_{\forall (j, i) \in A} p(t_j | s_i) = 0.4 \times 0.1 \times 0.5 = 0.02$$

11.2.6 基于短语的翻译模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} \underbrace{P(S_1^K | S)} \underbrace{P(T_1^K | S_1^K, S)} \underbrace{P(T_1^{K'} | T_1^K, S_1^K, S)} \underbrace{P(T | T_1^{K'}, T_1^K, S_1^K, S)} \end{aligned}$$

剩下的三个核心模型：

1. 短语翻译模型： $P(T_1^K | S_1^K, S)$
2. 短语调序模型： $P(T_1^{K'} | T_1^K, S_1^K, S)$
3. 目标语言模型： $P(T | T_1^{K'}, T_1^K, S_1^K, S)$

11.2.6 基于短语的翻译模型

短语调序模型： $(T_1^{K'} | T_1^K, S_1^K, S)$

两种常用方法：

1. 距离跳转模型
2. 分类模型

澳洲	是	与	北韩	有	邦交	的	少数	国家	之一
----	---	---	----	---	----	---	----	----	----

Australia	is	one of the few countries that	have diplomatic relations	with North Korea
-----------	----	-------------------------------	---------------------------	------------------

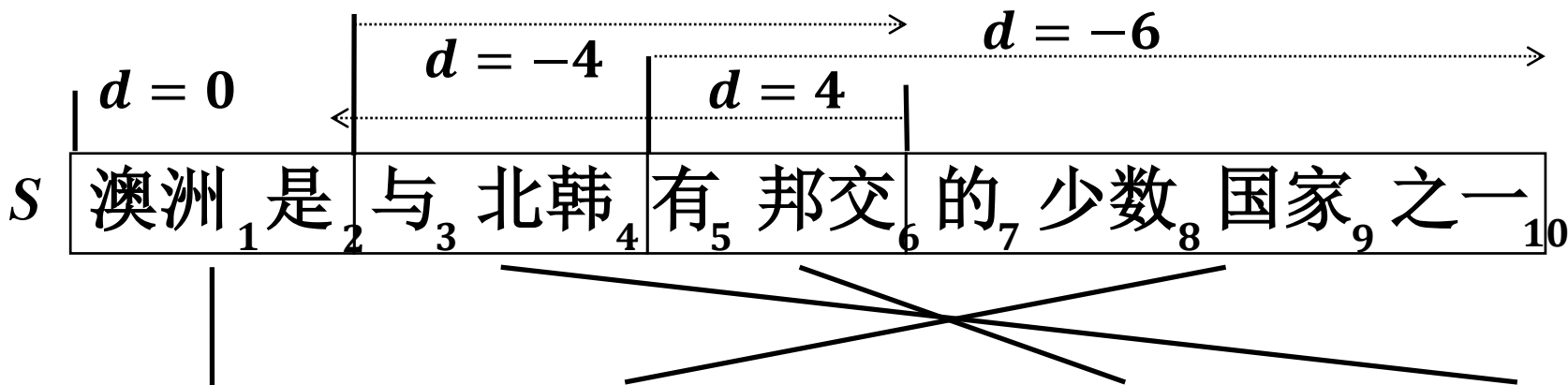
11.2.6 基于短语的翻译模型

前一个翻译的源语言 S 短语的最后一个词的下标

当前翻译的源语言 S 短语的第一个词的下标

◆ 距离跳转模型

$$d = next_{begin} - last_{end} - 1$$

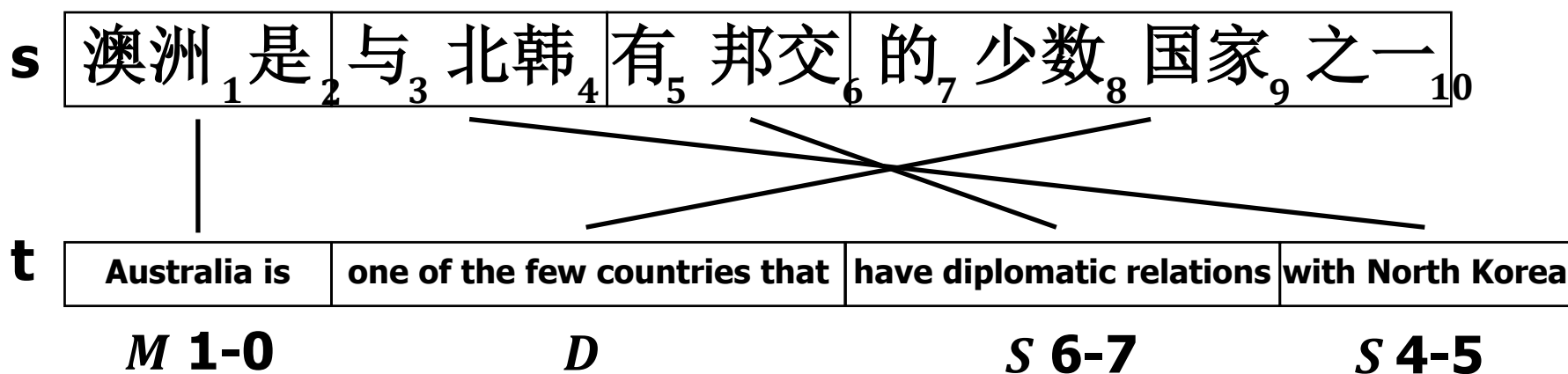


T Australia is one of the few countries that have diplomatic relations with North Korea

第 <i>i</i> 个源短语	短语跨度	距离跳转	距离
1	1-2	句子开始	0
2	7-10	跳过3-6	+4 7-2-1
3	5-6	向前跳过5-10	-6 5-10-1
4	3-4	向前跳过3-6	-4 3-6-1

11.2.6 基于短语的翻译模型

◆ 分类模型: Monotone (M) Swap (S) Discontinuous (D)



$$M: next_{begin} - last_{end} = 1$$

$$S: next_{end} - last_{begin} = -1$$

MaxEnt Classifier

$$D: next_{begin} - last_{end} \neq 1 \text{ and } next_{end} - last_{begin} \neq -1$$

11.2.6 基于短语的翻译模型

$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} \underbrace{P(S_1^K | S)} \underbrace{P(T_1^K | S_1^K, S)} \underbrace{P(T_1^{K'} | T_1^K, S_1^K, S)} \underbrace{P(T | T_1^{K'}, T_1^K, S_1^K, S)} \end{aligned}$$

剩下的三个核心模型：

1. 短语翻译模型： $P(T_1^K | S_1^K, S)$
2. 短语调序模型： $P(T_1^{K'} | T_1^K, S_1^K, S)$
3. 目标语言模型： $P(T | T_1^{K'}, T_1^K, S_1^K, S)$

11.2.6 基于短语的翻译模型

◆ 基于短语的判别式翻译模型


$$\begin{aligned} T' &= \operatorname{argmax}_T P(T|S) \\ &= \operatorname{argmax}_{T, S_1^K} P(T, S_1^K | S) \\ &= \operatorname{argmax}_{T, S_1^K, T_1^K, T_1^{K'}} P(S_1^K | S) P(T_1^K | S_1^K, S) P(T_1^{K'} | T_1^K, S_1^K, S) P(T | T_1^{K'}, T_1^K, S_1^K, S) \end{aligned}$$



$$T' = \operatorname{argmax}_T P(T|S) = \operatorname{argmax}_T \left\{ \sum_1^M \lambda_m h_m(T, S) \right\}$$

11.2.6 基于短语的翻译模型

◆ 基于短语的判别式翻译模型

$$T' = \operatorname{argmax}_T P(T|S) = \operatorname{argmax}_T \left\{ \sum_1^M \lambda_m h_m(T, S) \right\}$$


$$h_1(T, S) = \log p(t|s)$$

$$h_5(T, S) = \log P(T_1^{K'} | T_1^K, S_1^K, S)$$

$$h_2(T, S) = \log p(s|t)$$

$$h_6(T, S) = \log P(T | T_1^{K'}, T_1^K, S_1^K, S)$$

$$h_3(T, S) = \log p_{lex}(t|s)$$

$$h_7(T, S) = \log \text{len}(T)$$

$$h_4(T, S) = \log p_{lex}(s|t)$$

$$h_8(T, S) = \log \text{count}(\text{phrases}) = \log K$$

11.2.6 基于短语的翻译模型

基于短语翻译模型8特征(以汉英翻译为例):

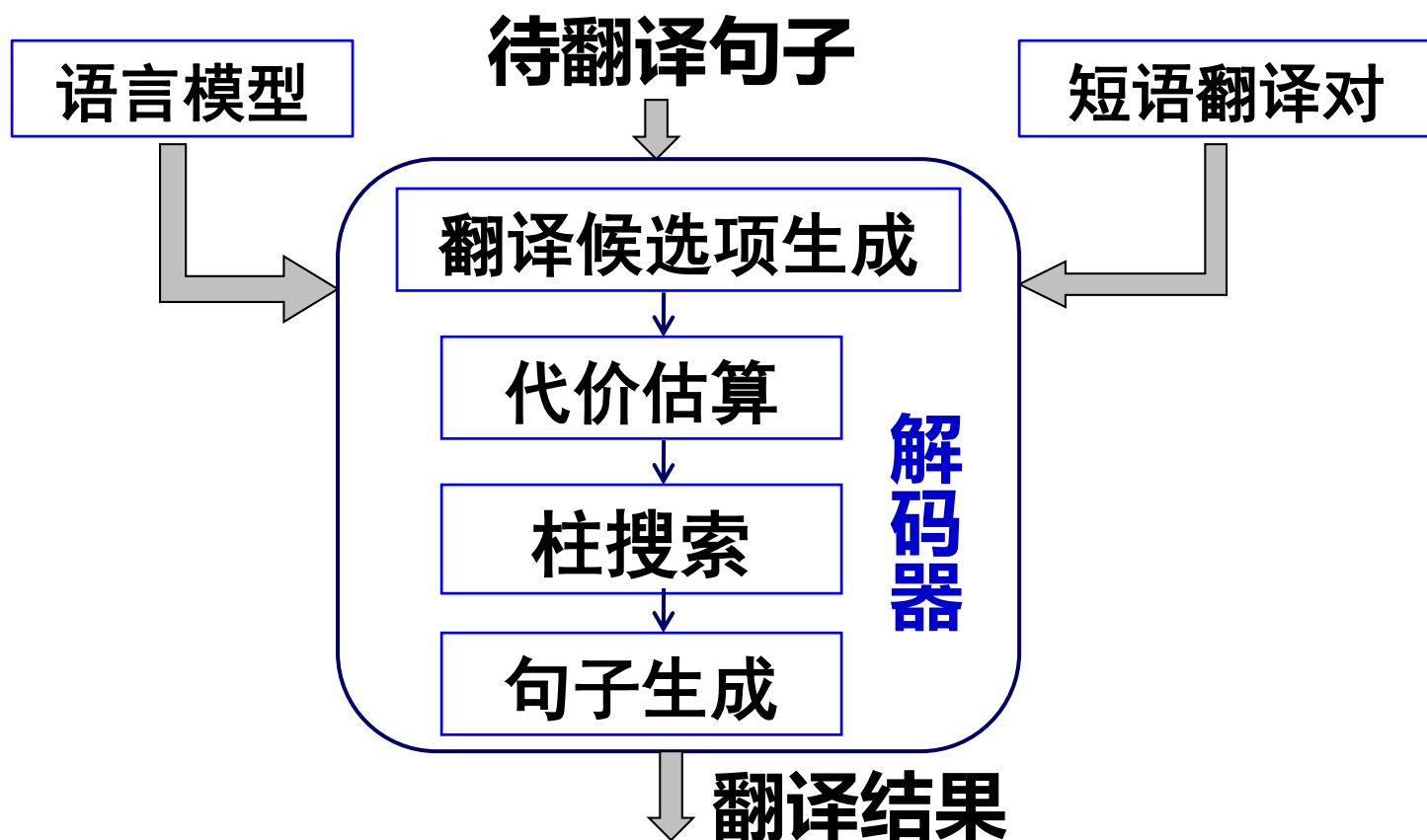
- ✓ 短语翻译概率 $\log p(e|c)$
- ✓ 词汇化的短语翻译概率 $\log p_{lex}(e|c)$
- ✓ 反向的短语翻译概率 $\log p(c|e)$
- ✓ 反向的词汇化短语翻译概率 $\log p_{lex}(c|e)$
- ✓ 短语调序模型 $\log P(E_1^{K'} | E_1^K, C_1^K, C)$
- ✓ 基于 n -gram 的英语语言模型 $\log P(E | E_1^{K'}, E_1^K, C_1^K, C)$
- ✓ 英语句子长度惩罚 $\log len(E)$
- ✓ 汉语短语个数惩罚 $\log K$



11.2.7 基于短语的翻译模型的解码算法

11.2.7 短语模型的解码算法

解码算法取决于翻译模型。在基于短语的翻译系统中：



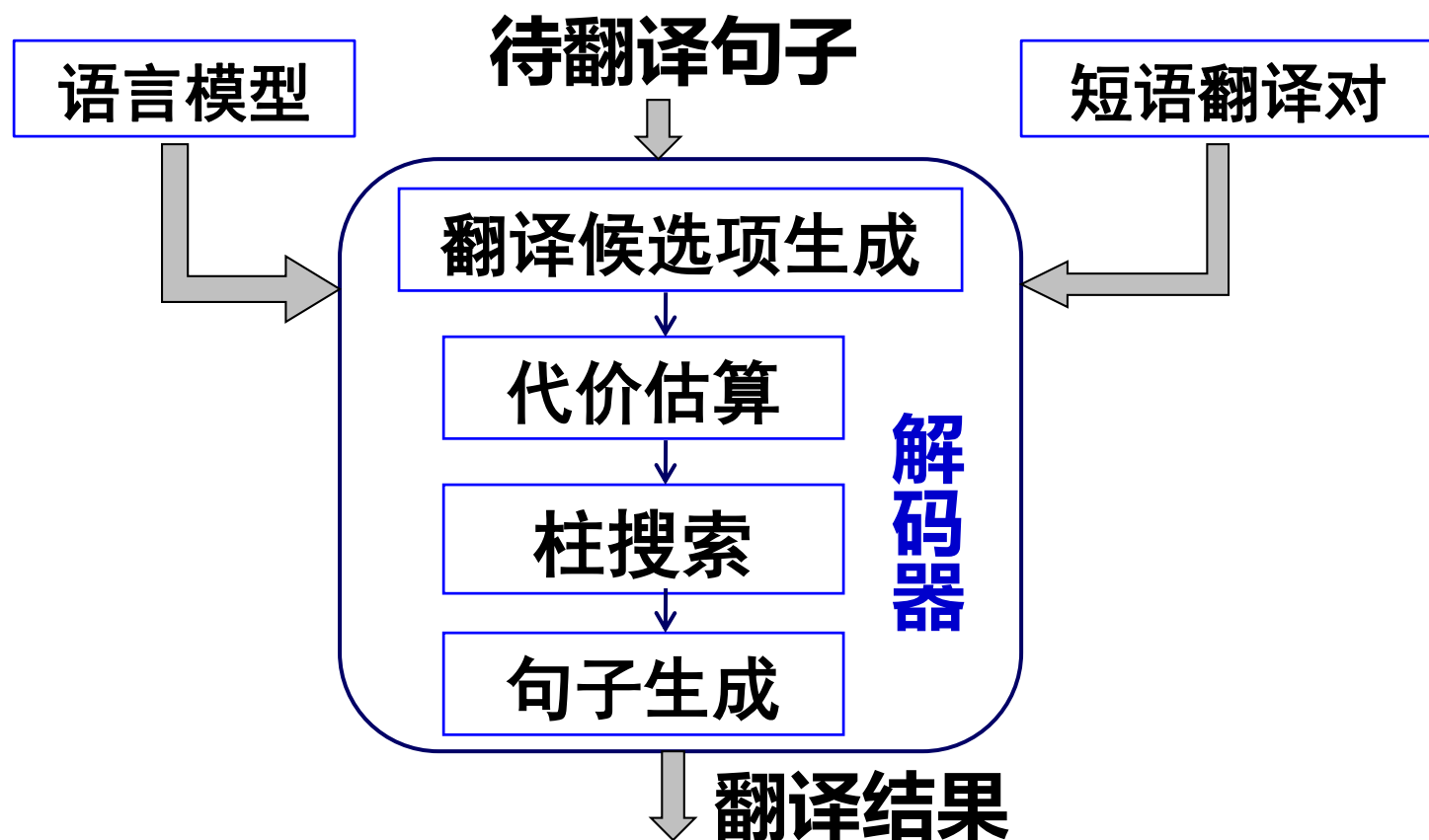
11.2.7 短语模型的解码算法

➤ 翻译候选生成（根据测试句子，搜索短语翻译规则表）

我	不	喜欢	喝	咖啡
<u>I</u>	<u>no</u>	<u>like</u>	<u>drink</u>	<u>coffee</u>
<u>me</u>	<u>not</u>	<u>like to</u>	<u>drinking</u>	<u>cafe</u>
	<u>do not</u>		<u>drinking coffee</u>	
<u>I do not</u>		<u>like to</u>	<u>drink</u>	
<u>I do not</u>		<u>like to</u>		
...	...			

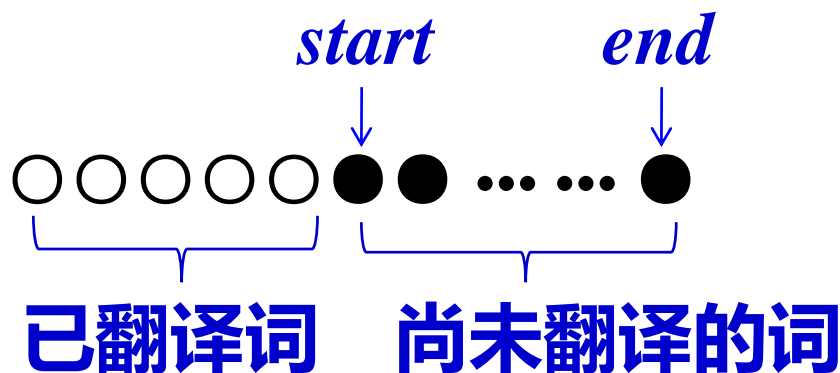
11.2.7 短语模型的解码算法

解码算法取决于翻译模型。在基于短语的翻译系统中：



11.2.7 短语模型的解码算法

➤ 代价估算



$Score =$ 已翻译词所耗费的代价 + 未翻译部分的估算代价

★ 已翻译词所耗费的代价: 已翻译词的模型概率

★ 翻译未来词估计需要的代价: 最大概率或其他因素

11.2.7 短语模型的解码算法

➤ 最大概率:

- ✓ 尚未翻译部分的正向短语翻译概率 (汉→英);
- ✓ 尚未翻译部分的逆向短语翻译概率 (英→汉);
- ✓ 尚未翻译部分的正向词汇化翻译概率 (汉→英);
- ✓ 尚未翻译部分的逆向词汇化翻译概率 (英→汉)。

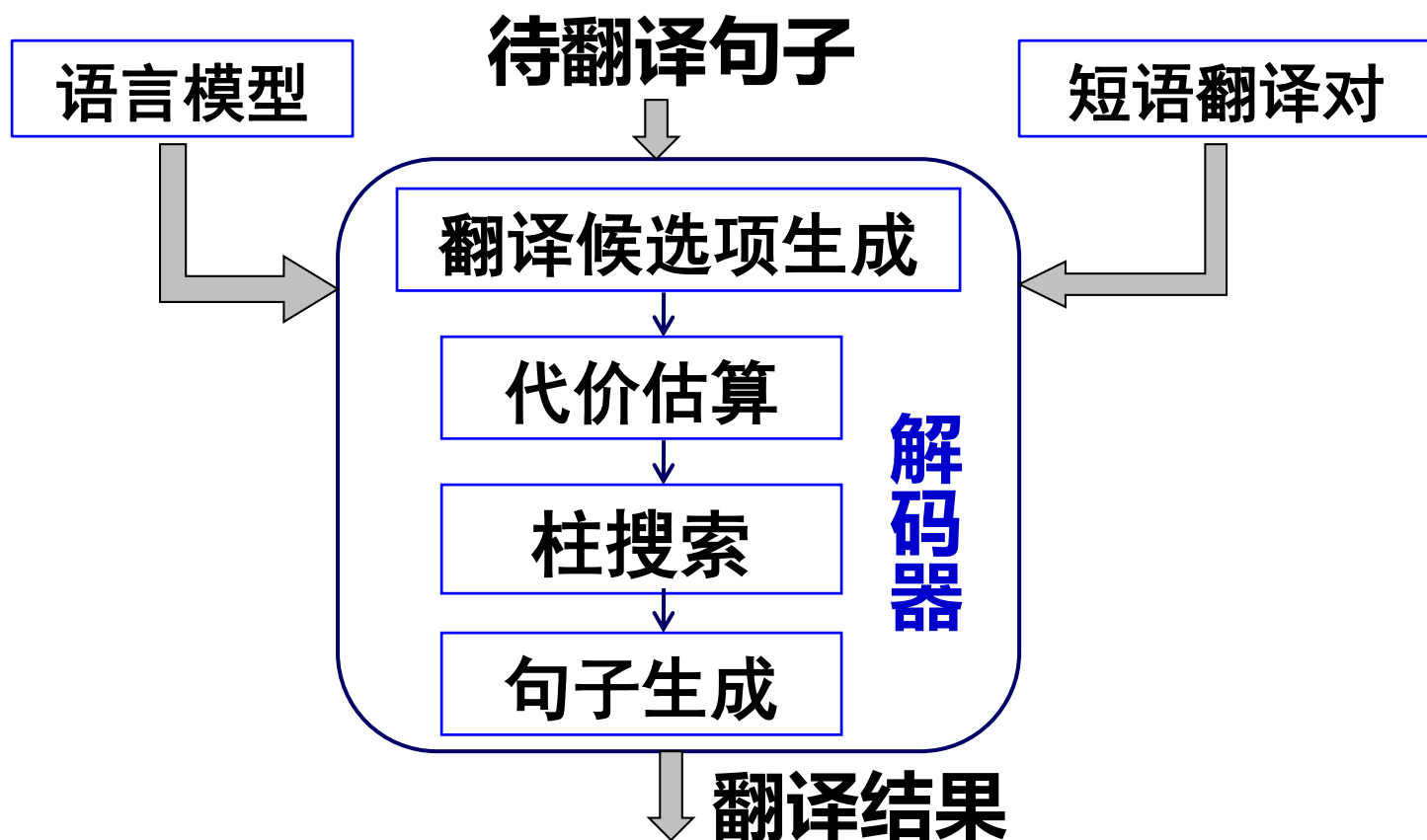
$$TP(f_{start}^{end}) = \max \sum \lambda_i \times \log(p_i(e|f))$$

➤ 其他因素:

- ✓ 译文短语的长度、语言模型概率等。

11.2.7 短语模型的解码算法

解码算法取决于翻译模型。在基于短语的翻译系统中：



11.2.7 短语模型的解码算法

➤ 柱搜索(**beam search**) [Ney, 1992; Tillmann, 2003]

基本思想：给定一个输入**句子**，生成对应的**短语序列**，每个短语对应一组**翻译候选**。短语序列按从左到右的顺序或目标短语生成的先后顺序搜索最可能的**翻译假设(hypothesis)**。

采用适当的剪枝策略。

11.2.7 短语模型的解码算法

➤ 汉英翻译例子:

我	不	喜欢	喝	咖啡
<u>I</u>	<u>no</u>	<u>like</u>	<u>drink</u>	<u>coffee</u>
<u>me</u>	<u>not</u>	<u>like to</u>	<u>drinking</u>	<u>cafe</u>
	<u>do not</u>		<u>drinking coffee</u>	
<u>I</u>	<u>do</u>	<u>not</u>	<u>like to</u>	<u>drink</u>
<u>I</u>	<u>do</u>	<u>not</u>	<u>like to</u>	
...	...			

11.2.7 短语模型的解码算法

➤ 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

我不: (1, 2), I do not, 0.435

我不: (1, 2), I am not, 0.418

我不: (1, 2), I never, 0.147

我不喜欢: (1, 3), I do not like to, 0.532

... ..

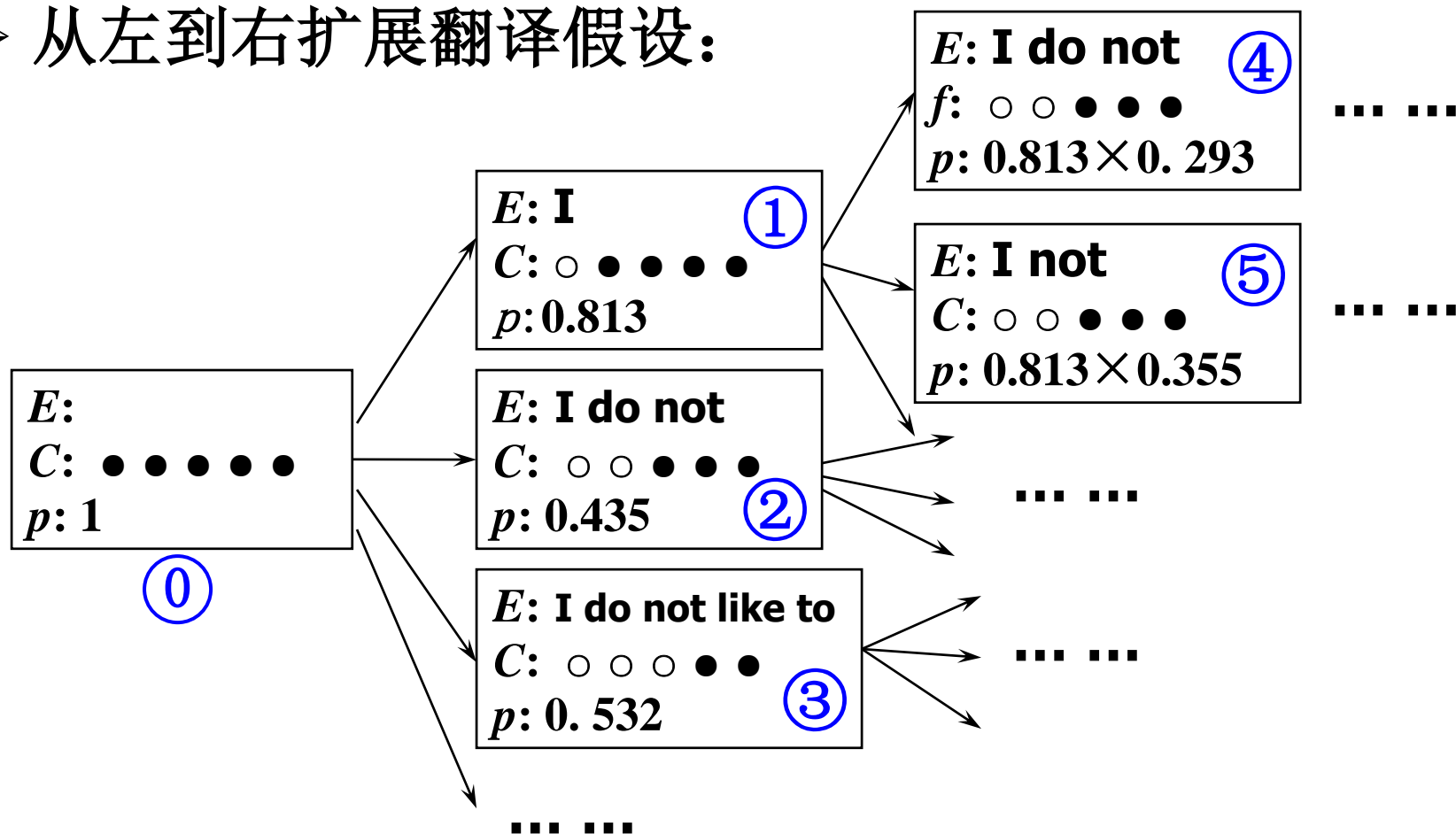
不: (2, 2), no, 0.352

不: (2, 2), not, 0.355

不: (2, 2), do not, 0.293

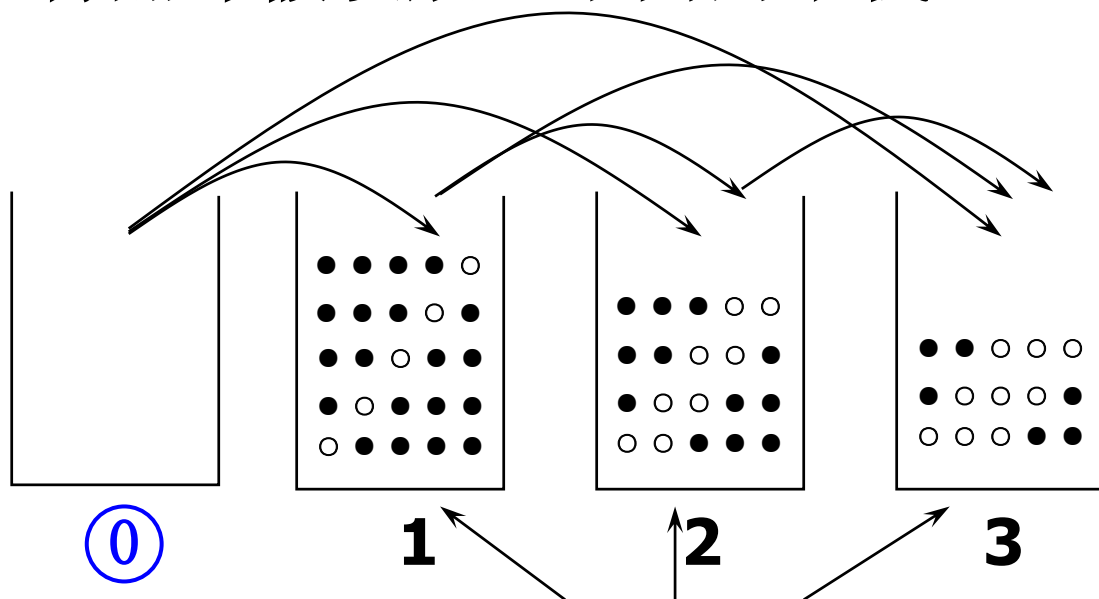
11.2.7 短语模型的解码算法

➤ 从左到右扩展翻译假设:



11.2.7 短语模型的解码算法

- 根据当前状态下被翻译(覆盖)的源语言单词的个数, 将翻译假设放入不同的堆栈:



未来代价估计保证栈内翻译假设的可比性

如果栈内的假设数太多, 采用适当的策略剪枝:

- ◆ 根据翻译阈值
- ◆ 根据柱容量
- ◆ 根据观察分值

11.2.7 短语模型的解码算法

- (1) 初始化假设栈 $\text{HypoStack}[0 \dots nf]$;
- (2) 创建初始假设 hyp_init ; 将假设压入栈 $\text{HypoStack}[0]$
- (3) 从 $i=0$ 到 $i = nf-1$ 执行如下循环:
 对于栈 $\text{HypoStack}[i]$ 中的每个假设 hx 执行如下循环 {
 对于 hx 可生成的每个新假设 new_hx 执行操作 {
 $\text{nf}[\text{new_hx}] =$ 被 new_hx 覆盖的源语言单词个数;
 把新假设 new_hx 压栈 $\text{hypoStack}[\text{nf}[\text{new_hx}]]$;
 对栈 $\text{HypoStack}[\text{nf}[\text{new_hx}]]$ 进行剪枝; } }
 }
- (4) 从栈 $\text{HypoStack}[\text{nf}]$ 中找到最好的假设 best_hyp ;
- (5) 输出产生 best_hyp 的最佳路径。

11.2.7 短语模型的解码算法

➤ 译文生成:

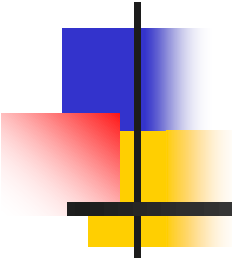
从最后一个栈中找到概率最大的 **Hypothesis**, 根据其指向父亲节点的指针向前回溯。可以产生 *n*-best 翻译结果。

➤ Pharaoh 解码器 [Koehn, 2003]

<http://www.isi.edu/licensed-sw/pharaoh/>

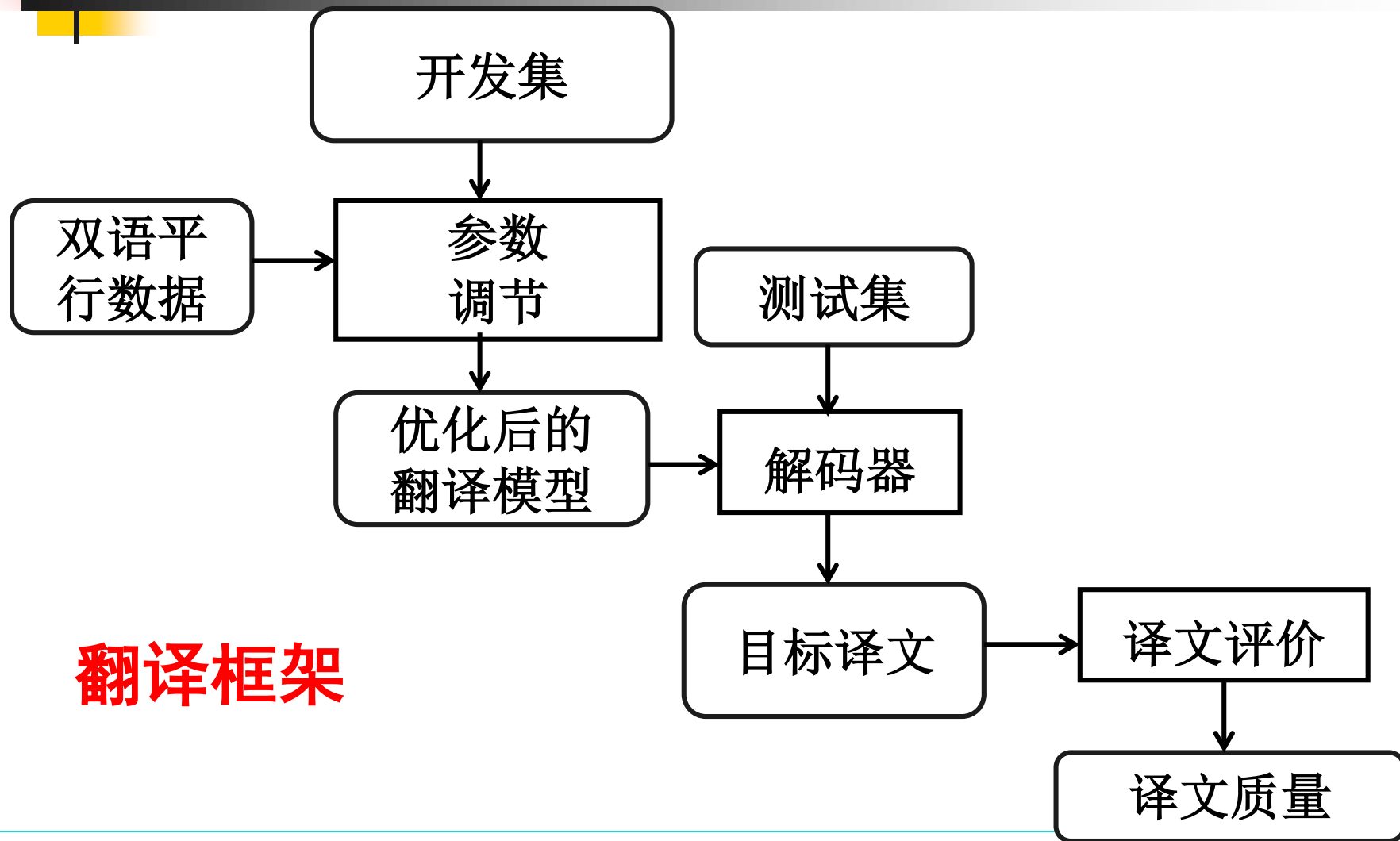
➤ Moses 解码器

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



11.2.8 基于短语模型的 SMT系统实现

11.2.8 基于短语的SMT系统实现



翻译框架

11.2.8 基于短语的SMT系统实现

以基于短语的统计翻译系统实现为例。

原文： 他 将 于 4 月 10 日 访问 美国

短语划分： 他将 于 4 月 10 日 访问 美国

翻译： He will on April 10 visit America

短语调序： He will visit America on April 10



11.2.8 基于短语的SMT系统实现

1. 可利用的工具

(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>

11.2.8 基于短语的SMT系统实现

➤ 词对齐工具: GIZA++

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

➤ 解码器:

✓ Pharaoh (法老王) / Moses (摩西)

✓ ReWrite

• <http://www.isi.edu/licensed-sw/rewrite-decoder/>

➤ EM 最小错误率参数训练工具

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

11.2.8 基于短语的SMT系统实现

2. 语料准备

- 大规模双语句子级对齐语料
- 目标语言句子

3. 语料预处理

- 汉语分词、全角转半角等
- 英语词汇化、大小写转换、标点分离

4. 词语对齐

利用词对齐工具：**GIZA++**

11.2.8 基于短语的SMT系统实现

5. 短语抽取和概率计算

银行 和 保险 公司	banks and	insurance companies
	1;	0.106; 1; 0.026
坚持 改革 开放	of reform and opening up	
	0.333; 0.001; 1; 0.014	

4个常用的概率:

(1) 英汉短语翻译概率
$$p(c|e) = \frac{N(c, e)}{\sum_{e'} N(c, e')}$$

11.2.8 基于短语的SMT系统实现

(2) 英汉词汇化的翻译概率

$$\text{lex}(c_1^J | e_1^I, a) = \prod_{j=1}^J \frac{1}{|\{i | (j, i) \in a\}|} \sum_{\forall (j, i) \in a} p(c_j | e_i)$$

(3) 汉英短语翻译概率 $p(e|c)$

(4) 汉英词汇化的翻译概率 $\text{lex}(c|e, a)$

11.2.8 基于短语的SMT系统实现

6. 语言模型训练

- 可采用 SRILM 等工具

7. 解码器实现

8. 输出后处理

- 汉语分词空格的去除
- 英语大小写转换、标点处理

11.2.8 基于短语的SMT系统实现

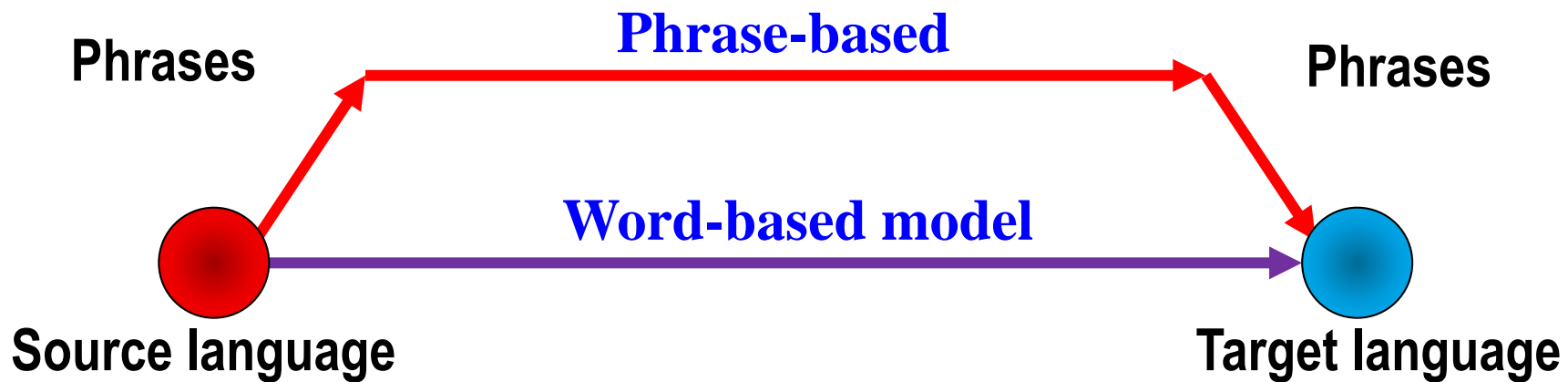
◆开源系统:

➤ “法老王” 与 “摩西”

✓ <http://www.isi.edu/licensed-sw/pharaoh>

✓ <http://www.statmt.org/moses/>

Moses系统是“法老王”的升级版，比“法老王”多了许多功能，是由英国爱丁堡大学(The University of Edinburgh)和德国亚琛工业大学(RWTH)等8家单位联合于2006年在约翰霍普金斯大学(JHU)研讨班上用6周时间开发完成的，也采用基于短语的翻译模型。





11.2.9 基于层次化短语 的翻译模型

11.2.9 基于层次化短语的翻译模型

◆ 问题提出：

- (1) 基于短语的翻译模型能够比较鲁棒地翻译较短的子串，当短语长度扩展到3个以上的单词时，翻译系统的性能提高很少，短语长度增大以后，数据稀疏问题变得非常严重。
- (2) 在很多情况下简单的短语翻译模型无法处理短语之间（尤其是长距离）的调序。
- (3) 基于短语翻译模型无法处理非连续短语翻译现象，例如（在 ... 时，when ...）

11.2.9 基于层次化短语的翻译模型



David Chiang (蒋伟) (2005) 提出了基于层次短语的翻译模型试图解决这些问题，其基本思路：不破坏基于短语的翻译方法的优势，而是利用这些优势（短语有益于实现局部词义消歧和局部语序调整），并且对翻译规则进行泛化，解决数据稀疏、短语调序以及非连续短语翻译问题。

11.2.9 基于层次化短语的翻译模型

基于层次短语的翻译过程同步进行双语解析，所使用的同步上下无关文法是从没有做任何句法信息标注的双语对照语料中自动学习获得的。

1. 层次短语翻译规则学习
2. 层次短语模型解码过程

11.2.9 基于层次化短语的翻译模型

1. 层次短语翻译规则学习

澳 洲 是 与 北 有 邦 的 少 国 之 一 。
澳 洲 是 与 北 有 邦 的 少 国 之 一 。

■											Australia
	■										is
									■		one
									■		of
											the
							■				few
								■			countries
						■					that
				■							have
					■						diplomatic
					■						relations
		■									with
			■								North
			■								Korea
										■	.

11.2.9 基于层次化短语的翻译模型

1. 层次短语翻译规则学习

(北韩, North Korea)

(邦交, the diplomatic relations)

(与北韩有邦交, have the diplomatic relations with North Korea)

(少数国家, few countries)

(与北韩有邦交的少数国家, the few countries that have the diplomatic relations with North Korea)

(与北韩有邦交的少数国家之一, one of the few countries that have the diplomatic relations with North Korea)

11.2.9 基于层次化短语的翻译模型

(北韩, North Korea)

(邦交, the diplomatic relations)

(与 北韩 有 邦交, have the diplomatic relations with North Korea)

定义层次化的短语由单词和子短语(subphrase)构成:

〈 与 ① 有 ②, have ② with ① 〉 —(a)

其中, ① 和 ② 表示两个子短语。

实际上这是一条CFG(上下文无关文法)规则, 也就是汉语句子里中类似“与①有②”表达的短语将被翻译成英语短语 ‘have ② with ①’ 。

11.2.9 基于层次化短语的翻译模型

(与 北韩 有 邦交, have the diplomatic relations with North Korea)

(少数 国家, few countries)

(与 北韩 有 邦交 的 少数 国家, the few countries that have the diplomatic relations with North Korea)

〈 ① 的 ②, the ② that ① 〉 —(b)

这条规则主要用于调整汉语句子中在名词短语**左边**的**关系从句**的翻译次序, 因为在英语句子中由**that**引起的起修饰作用的关系从句一般在名词短语的**右边**。

11.2.9 基于层次化短语的翻译模型

(与 北韩 有 邦交 的 少数 国家, the few countries that have the diplomatic relations with North Korea)

(与 北韩 有 邦交 的 少数 国家 之一, one of the few countries that have the diplomatic relations with North Korea)

如下规则是为了调整汉语词语“之一”在英语句子中的次序:

〈 ① 之一, one of ① 〉 —(c)

11.2.9 基于层次化短语的翻译模型

上面给出的层次化短语对可以被形式化地描写成如下同步CFG规则：

- (a) $X \rightarrow \langle \text{与 } X_{①} \text{ 有 } X_{②}, \text{ have } X_{②} \text{ with } X_{①} \rangle$
- (b) $X \rightarrow \langle X_{①} \text{ 的 } X_{②}, \text{ the } X_{②} \text{ that } X_{①} \rangle$
- (c) $X \rightarrow \langle X_{①} \text{ 之一}, \text{ one of } X_{①} \rangle$

其中，带数字下标的X表示被符号 ~ 关联的出现在X中的非终结符号。

这里只使用一个非终结符号X代替短语的句法类。

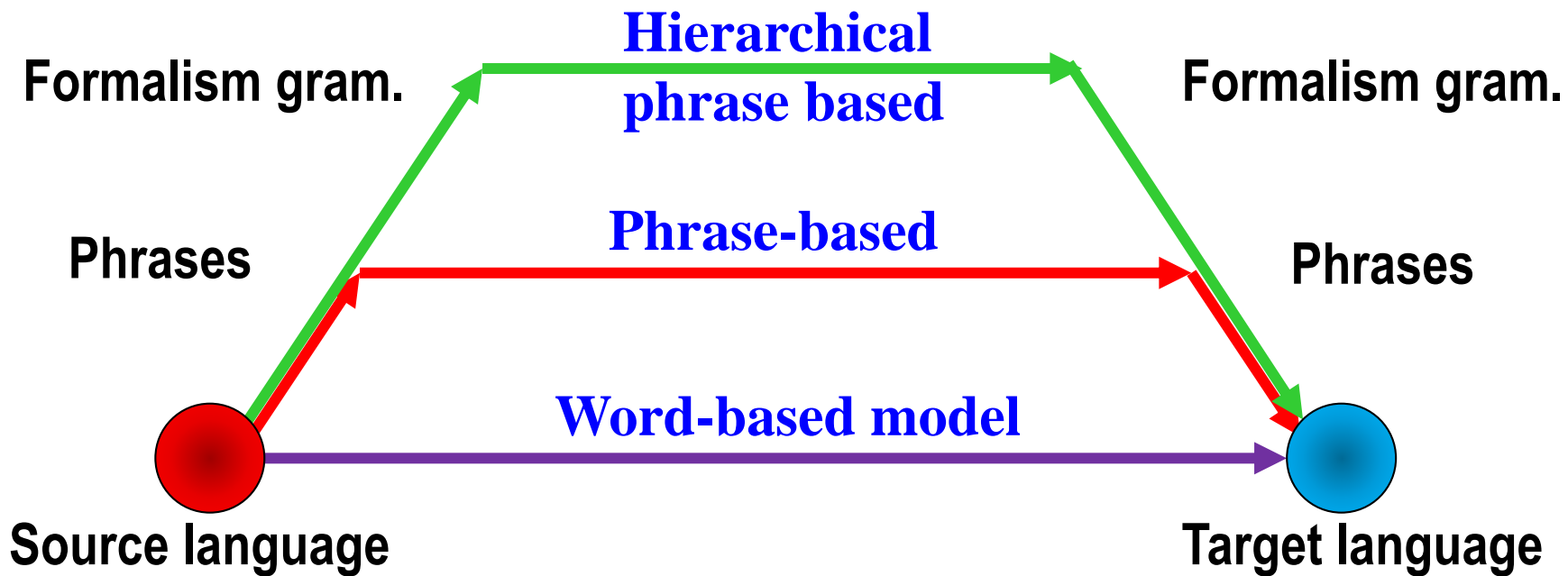
11.2.9 基于层次化短语的翻译模型

从双语语料中抽取出来的语法规则中，除了下面两条特殊的规则将X序列构成句子S以外，都只用一个非终结符号。

$$(d) \quad S \rightarrow \langle S_{①}X_{②}, S_{①}X_{②} \rangle$$

$$(e) \quad S \rightarrow \langle X_{②}, X_{②} \rangle$$

基于层次化短语的翻译模型首先利用层次化短语产生句子的局部翻译，然后，像常规的基于短语的模型一样，将这些局部的翻译顺序地连接起来，从而形成整个句子的翻译。



11.2.9 基于层次化短语的翻译模型

➤ Joshua系统

✓ <http://www.clsp.jhu.edu/wiki2/JosHUa> - JHU Open Source Architecture

Joshua 是2008年6月JHU暑期研讨班上开发的一个开源的基于层次短语的统计机器翻译系统。该系统实现了上下文无关文法所需的所有算法，并采用了基于后缀数组的文法规则抽取算法。它使得层次短语模型能够工作在大规模训练语料上。

11.2.9 基于层次化短语的翻译模型

➤ SAMT系统

✓ <http://www.cs.cmu.edu/~zollmann/samt/>

SAMT (Syntax Augmented Machine Translation) 系统由美国卡内基-梅隆大学 (CMU) 实现，其创新之处在于将基于层次短语的翻译模型思想用到了句法结构树上。该系统同时融合了以下三种短语：

- (1) 基于Moses系统抽取出来的一般短语；
- (2) 句法驱动的泛化短语；
- (3) 分层短语。



11.2.10 树翻译模型



11.2.10 树翻译模型

- ◆ 树到串的翻译模型
- ◆ 树到树的翻译模型
- ◆ 串到树的翻译模型

11.2.10 树翻译模型

◆ 问题提出：

- (1) 基于层次短语的翻译模型只使用一个非终结符 X ，过于泛化。
- (2) 基于层次短语的翻译模型在处理长距离的短语调序问题时能力有限。



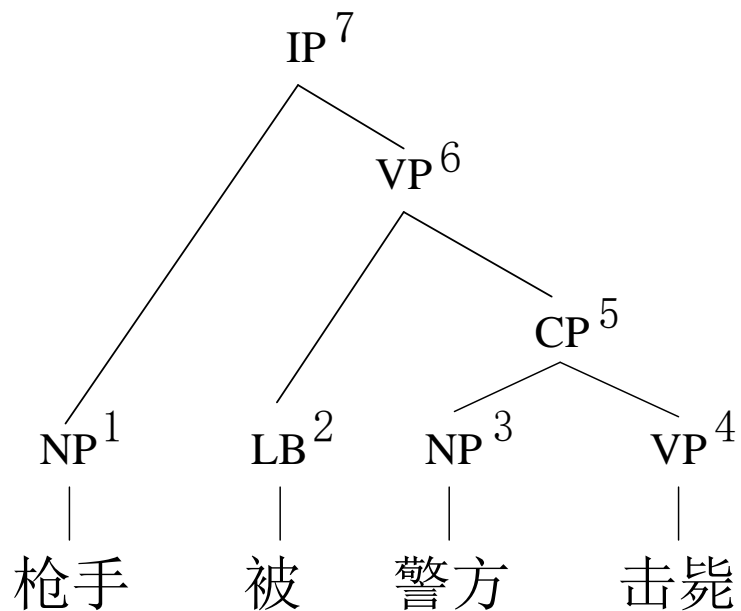
11.2.10 树翻译模型

◆ 树到串的翻译模型

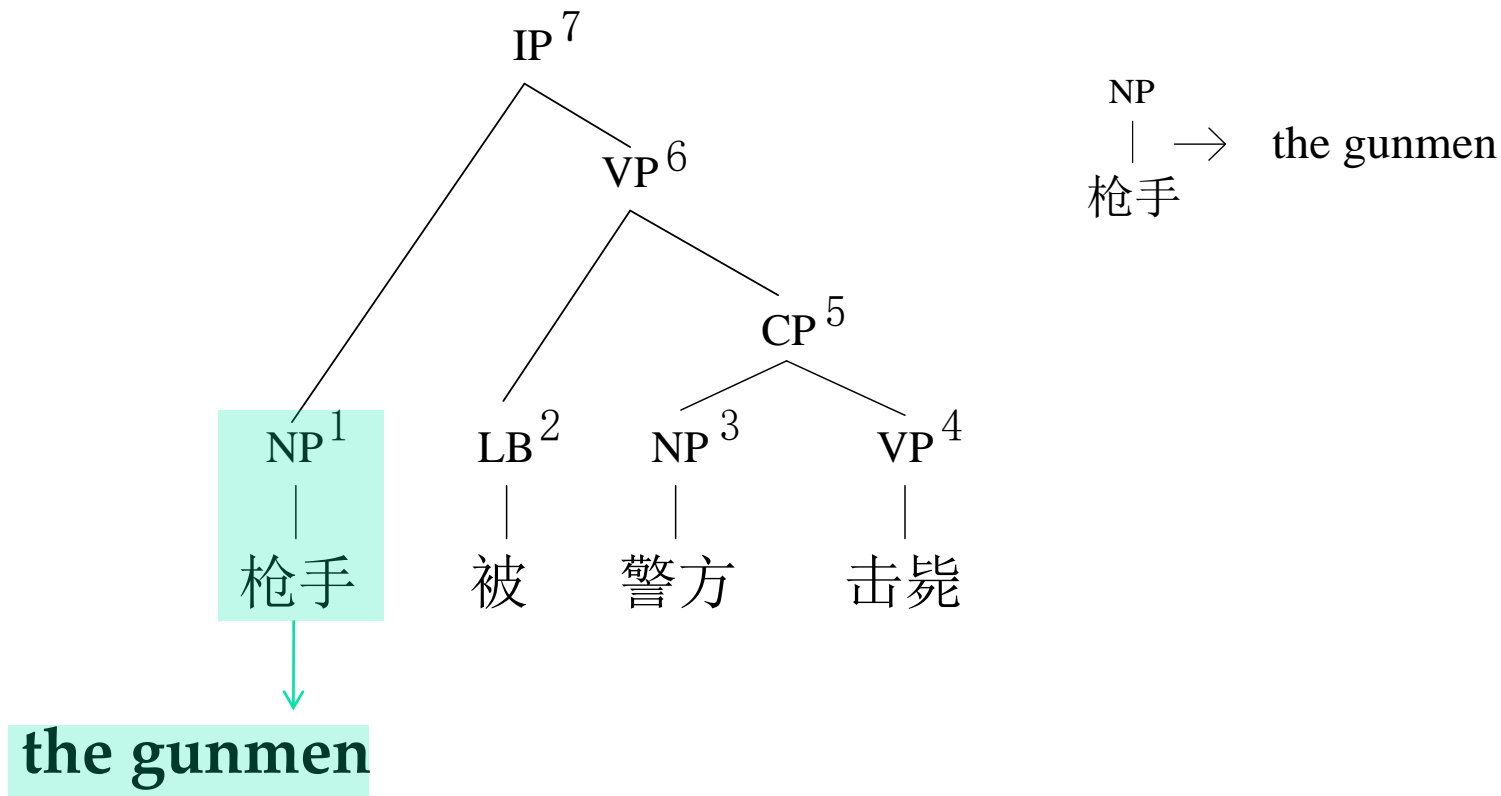
Liu et al. (2006), Huang et al. (2006) 提出了树到串的翻译模型。

- **句法分析**：将源语言句子分析为一棵句法结构树（短语结构树）
- **树到串的转变**：递归地将源语言句子的句法结构树转换为目标语言句子

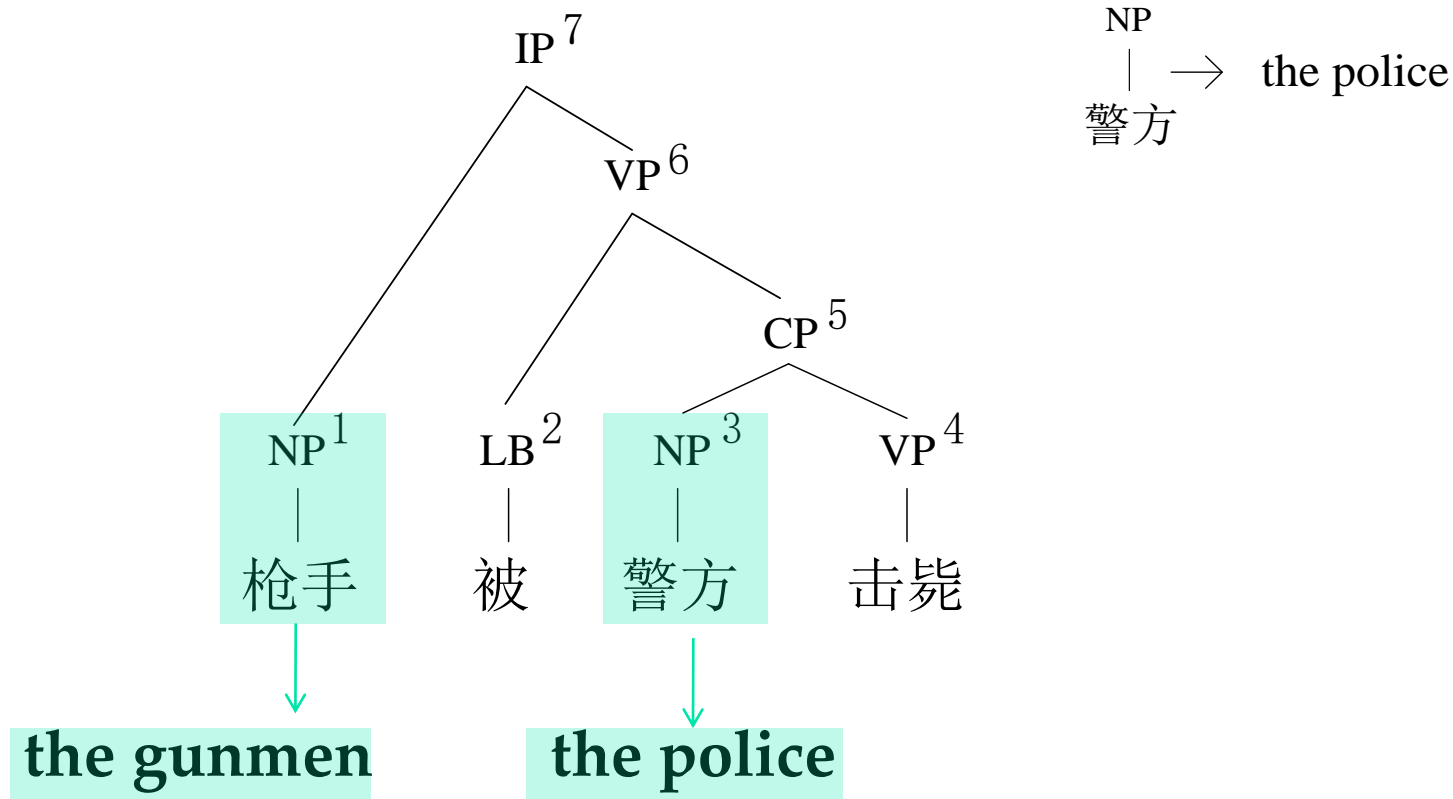
11.2.10 树翻译模型



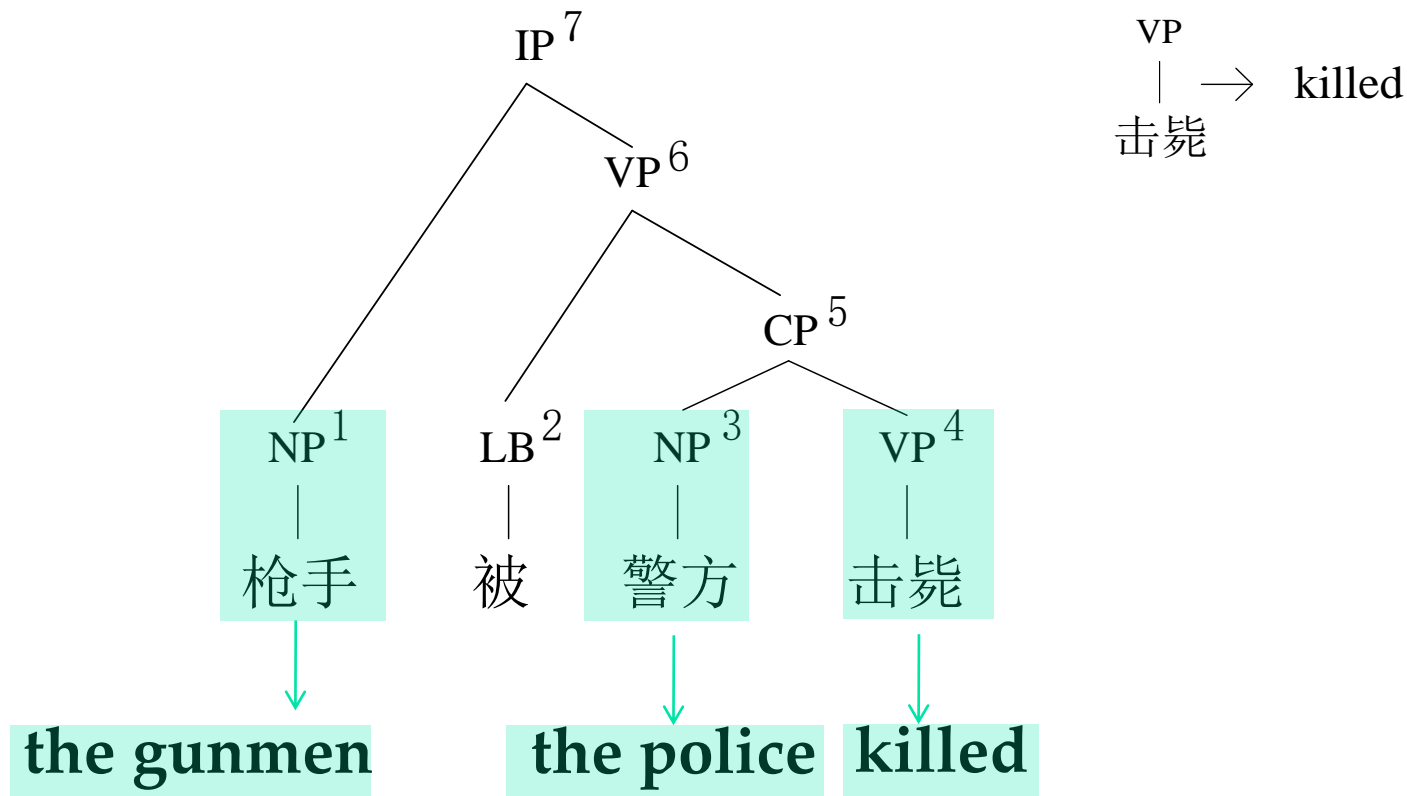
11.2.10 树翻译模型



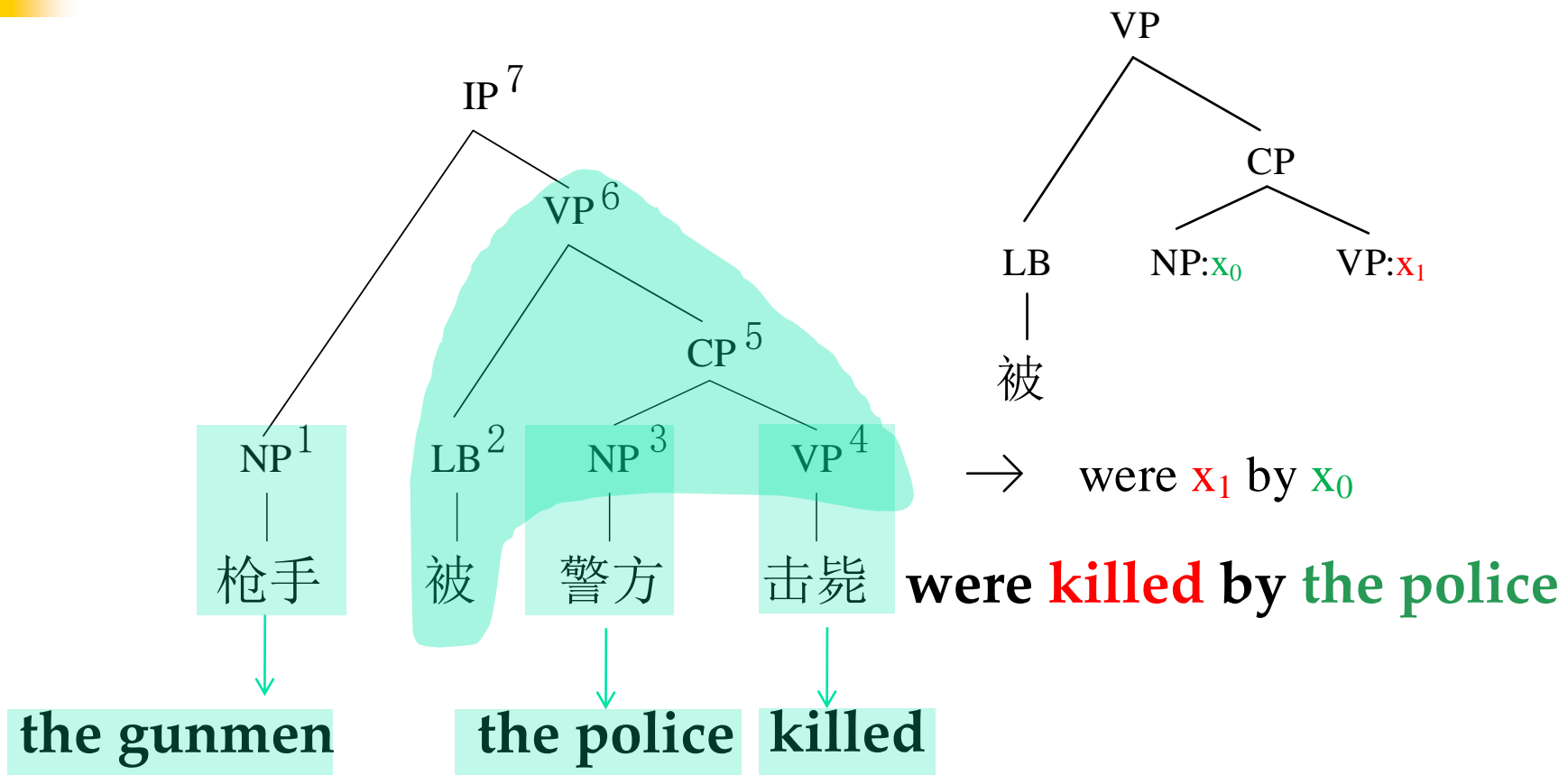
11.2.10 树翻译模型



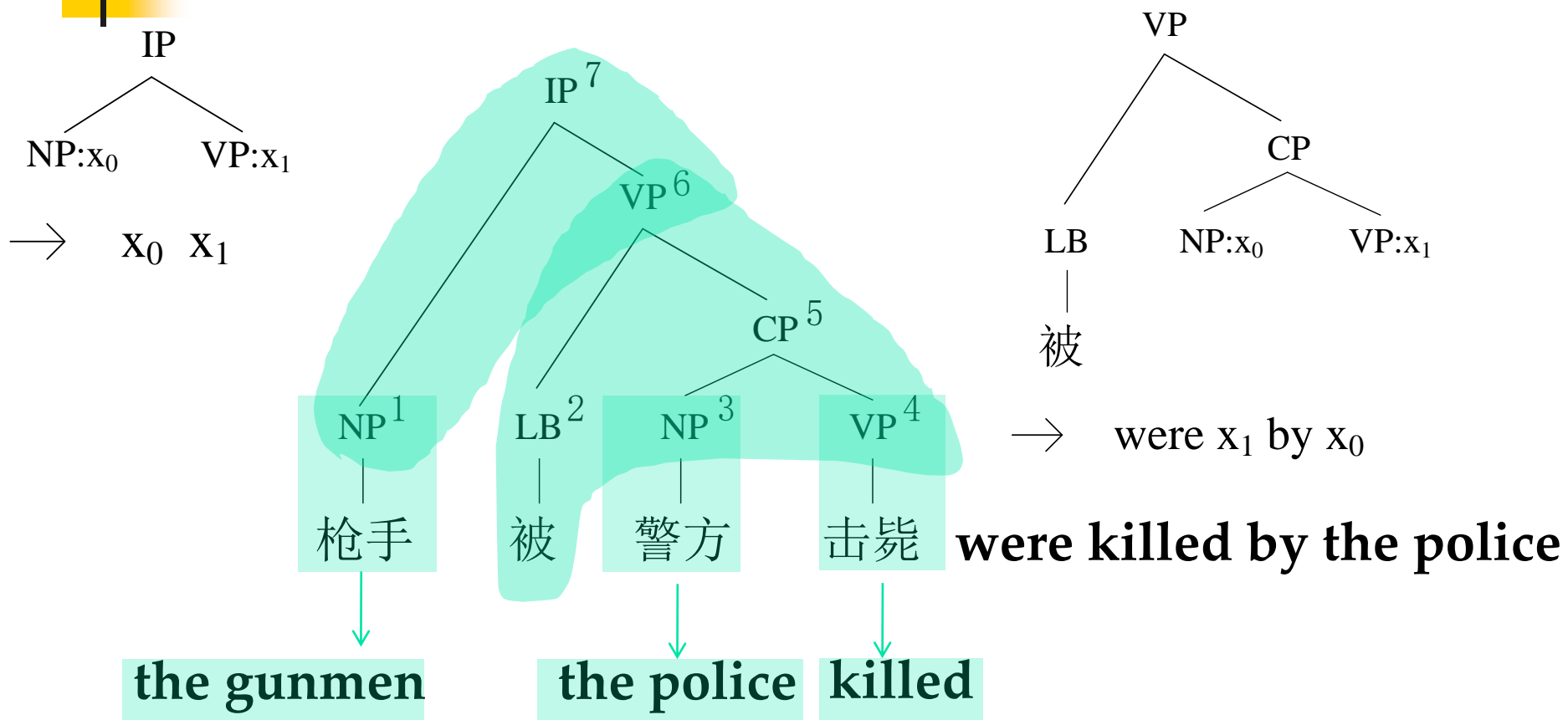
11.2.10 树翻译模型



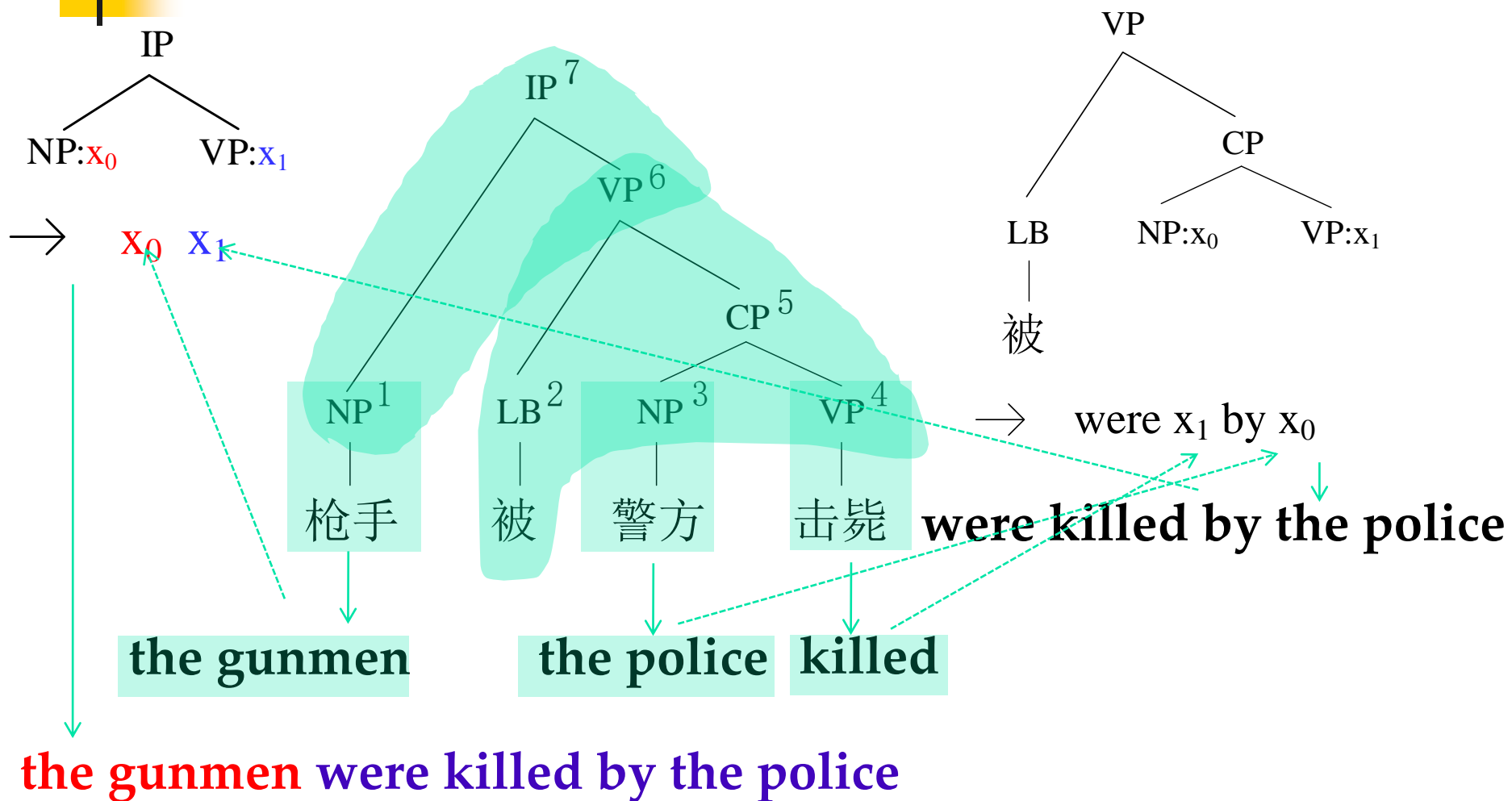
11.2.10 树翻译模型



11.2.10 树翻译模型

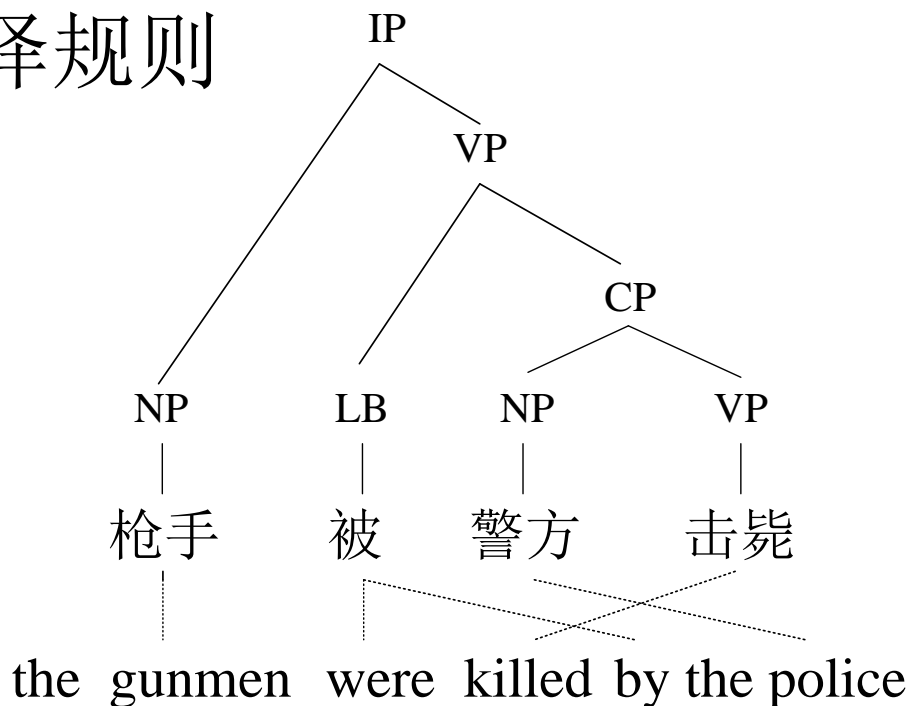


11.2.10 树翻译模型



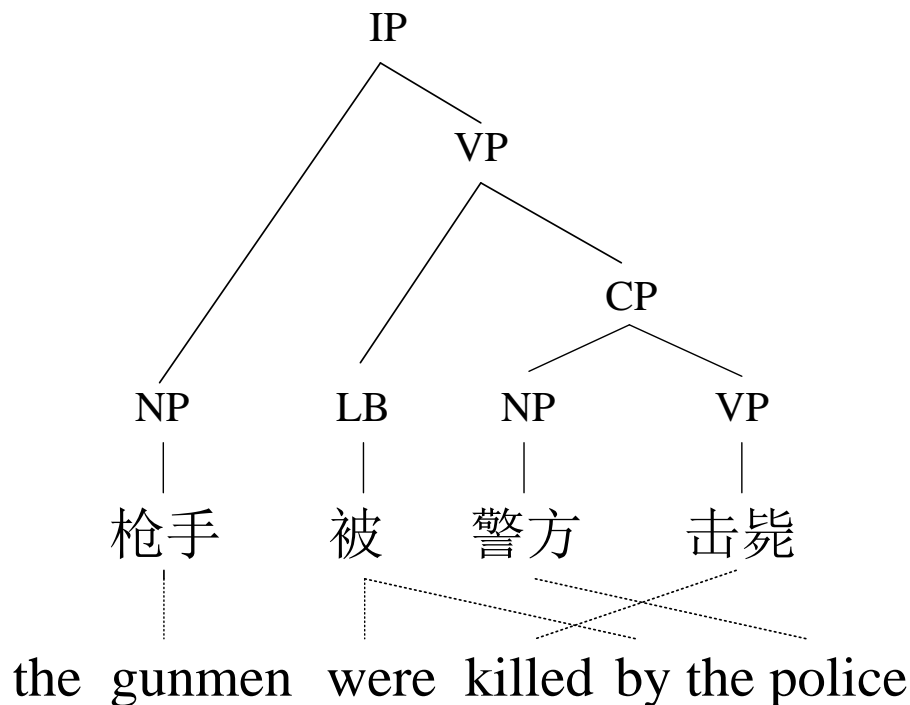
11.2.10 树翻译模型

- **树到串翻译规则抽取：** 给定源语言和目标语言的双语平行句对（经过词语对齐、源语言端句法分析），抽取满足词语对齐的树到串翻译规则

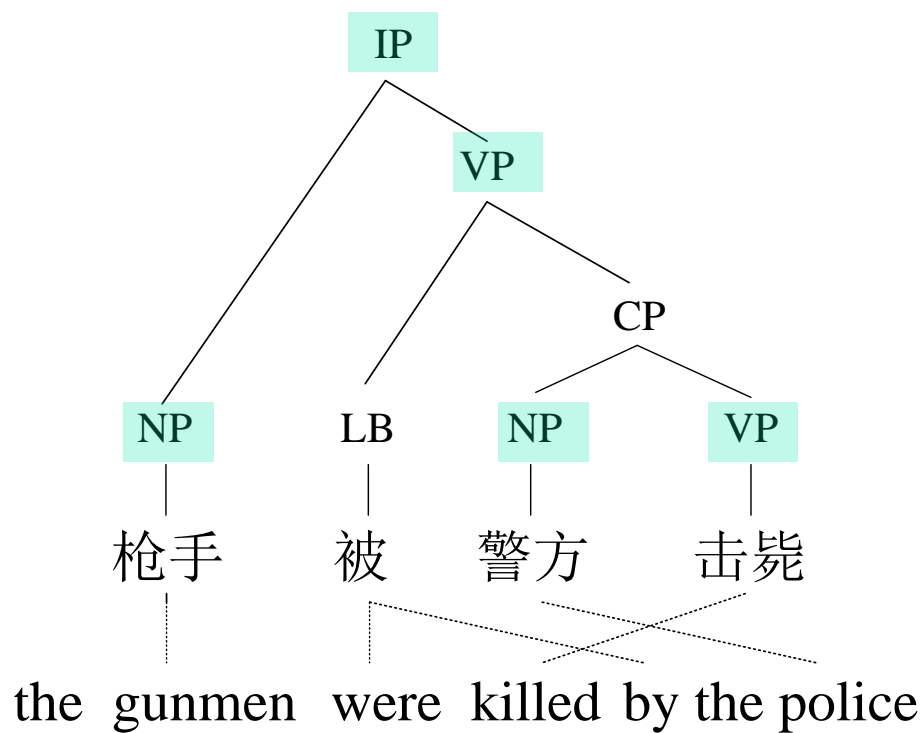


11.2.10 树翻译模型

- **确定满足词语对齐的树节点：** 源语言句法树节点所能到达的目标语言子串与该树节点覆盖的源语言子串满足词语对齐约束。

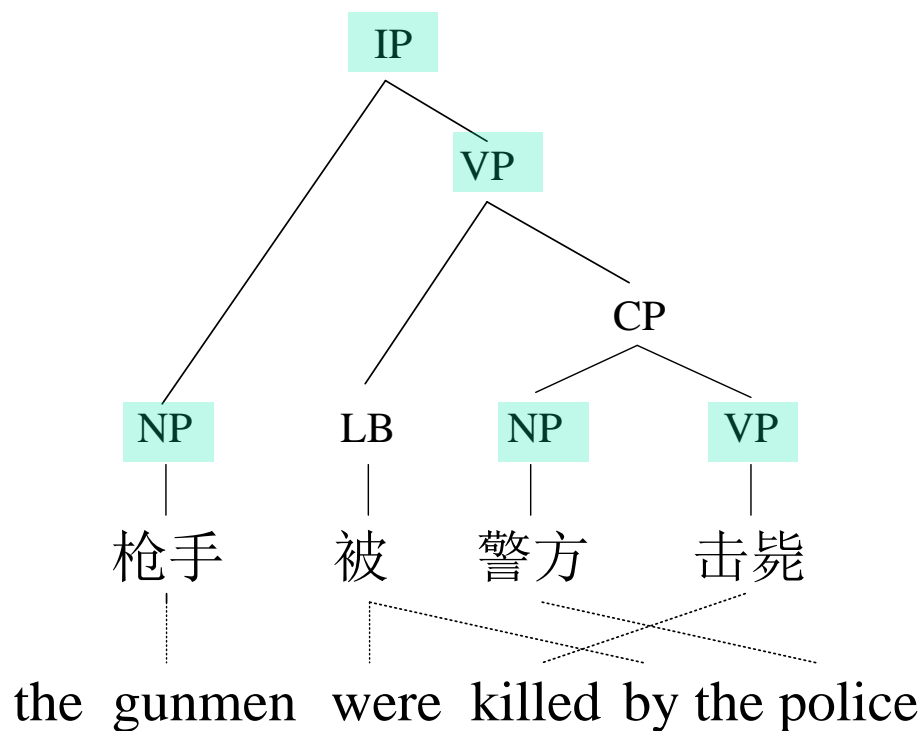


11.2.10 树翻译模型

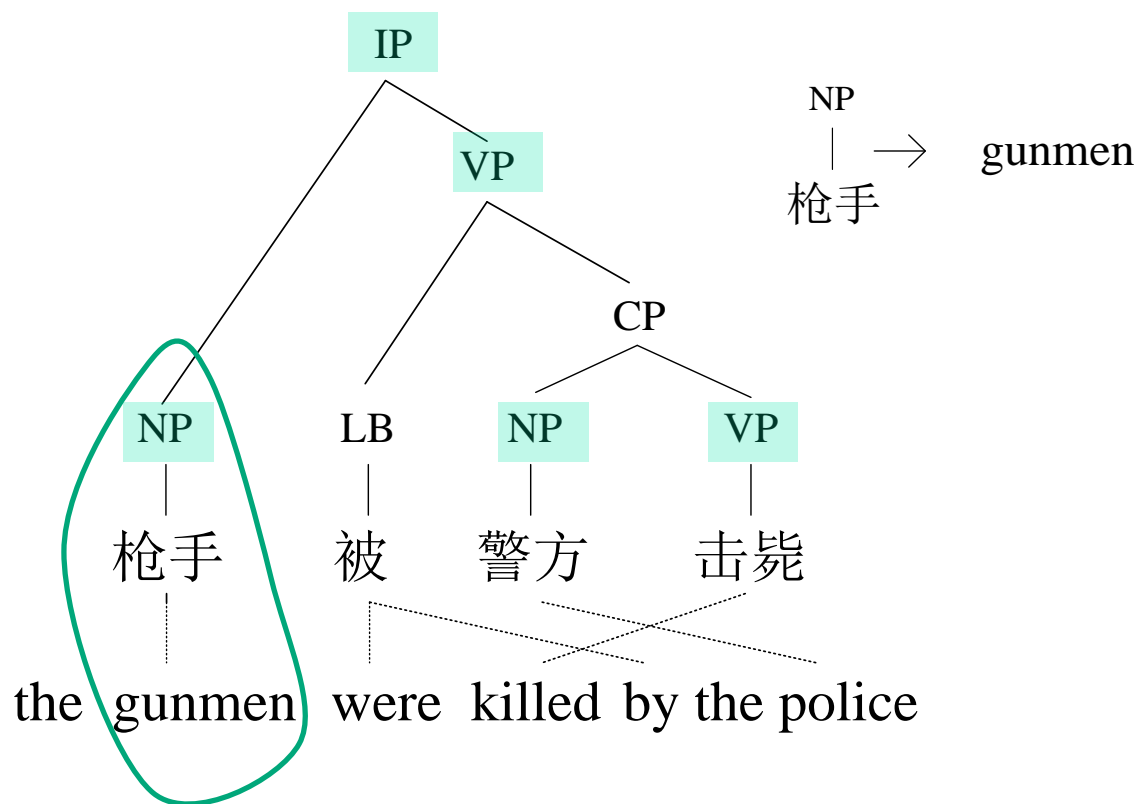


11.2.10 树翻译模型

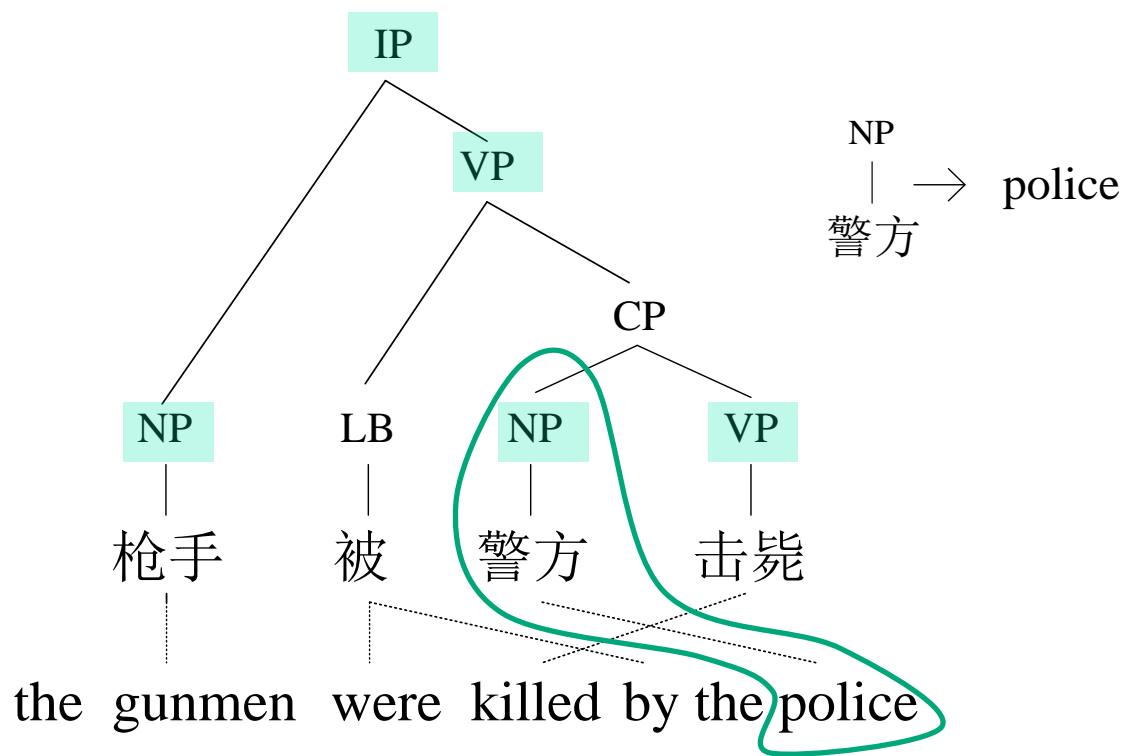
- 对于每个满足词语对齐的树节点，我们可以抽取一条最小规则。



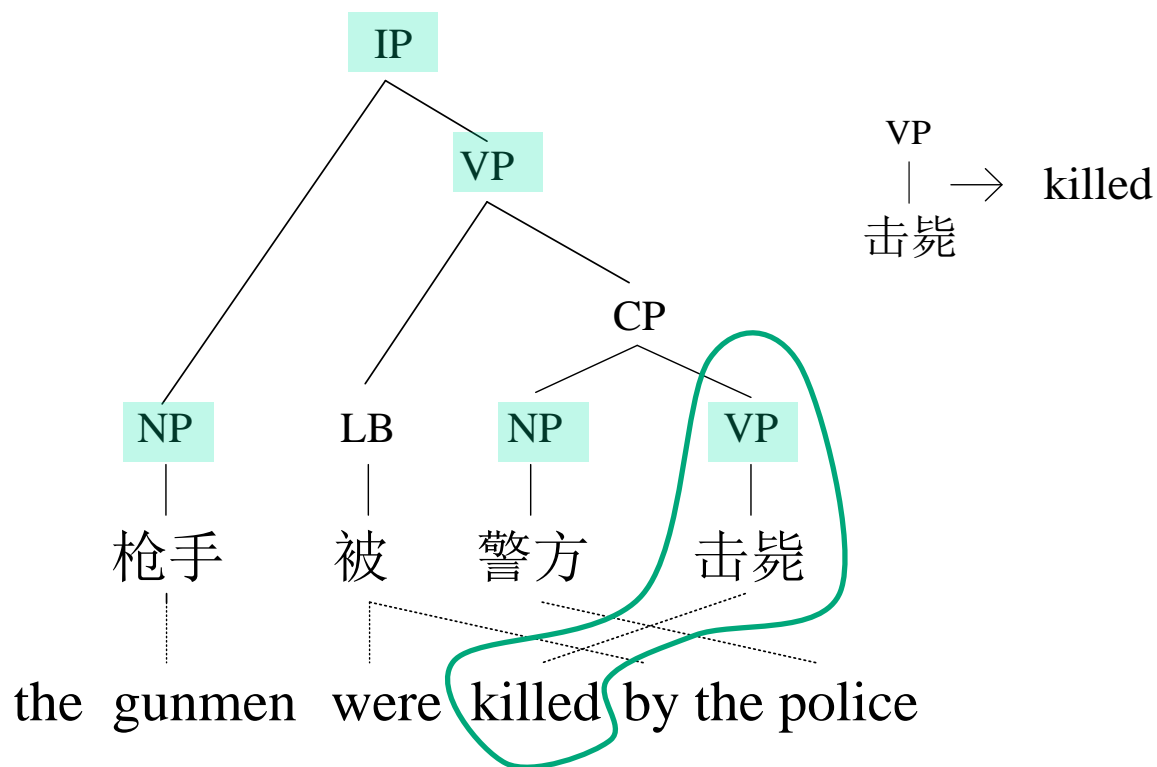
11.2.10 树翻译模型



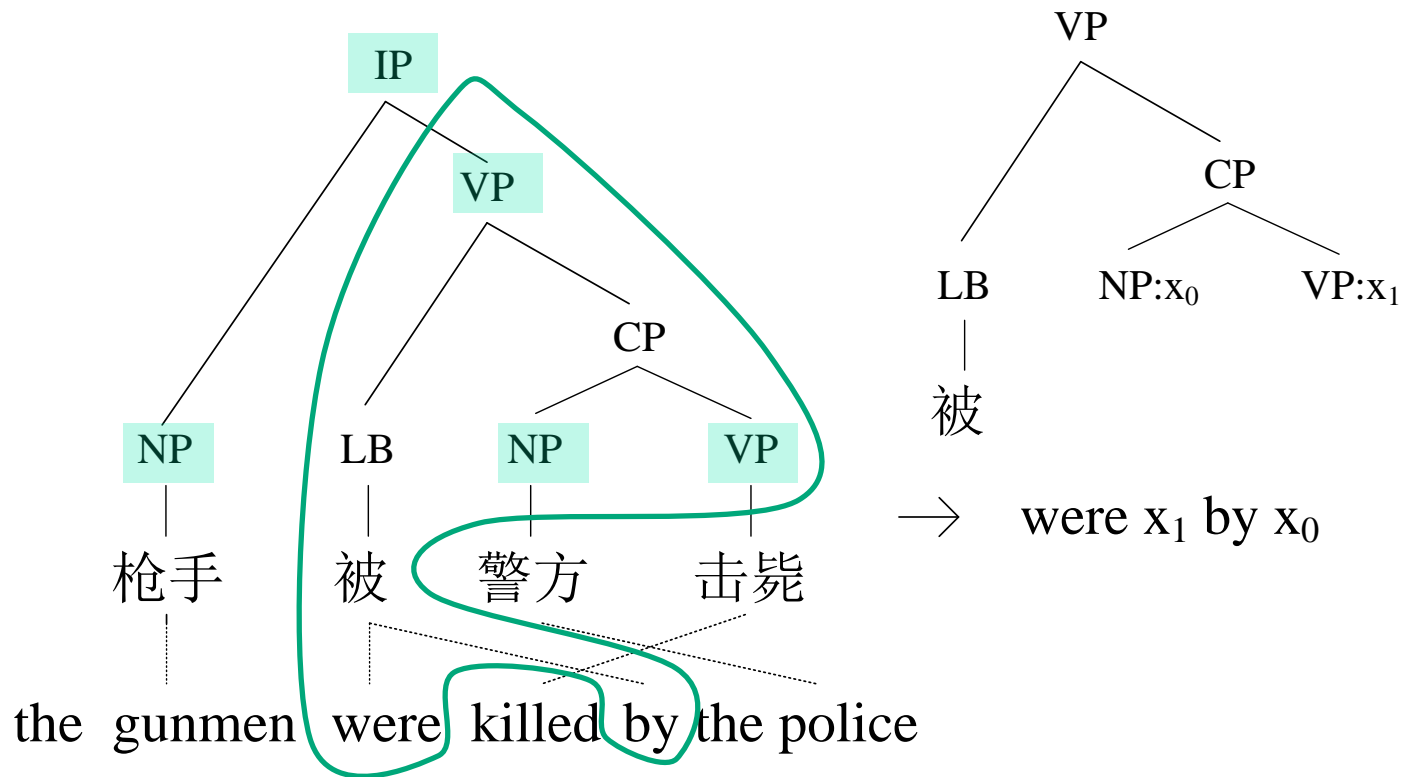
11.2.10 树翻译模型



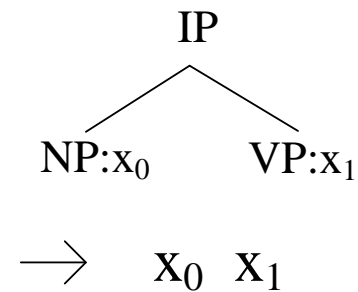
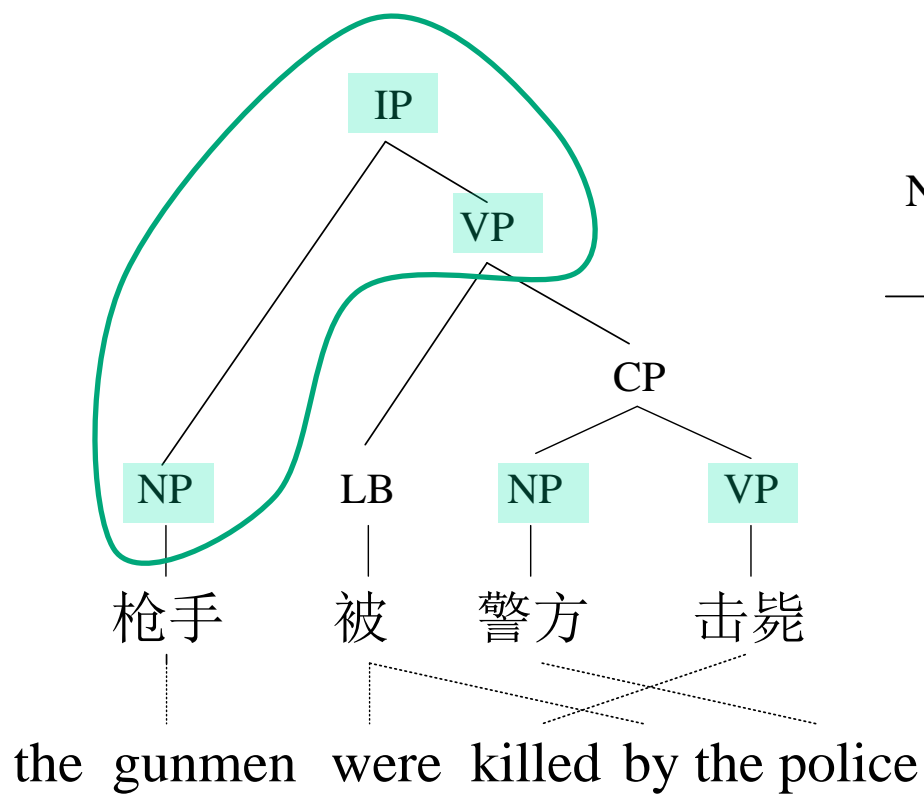
11.2.10 树翻译模型



11.2.10 树翻译模型

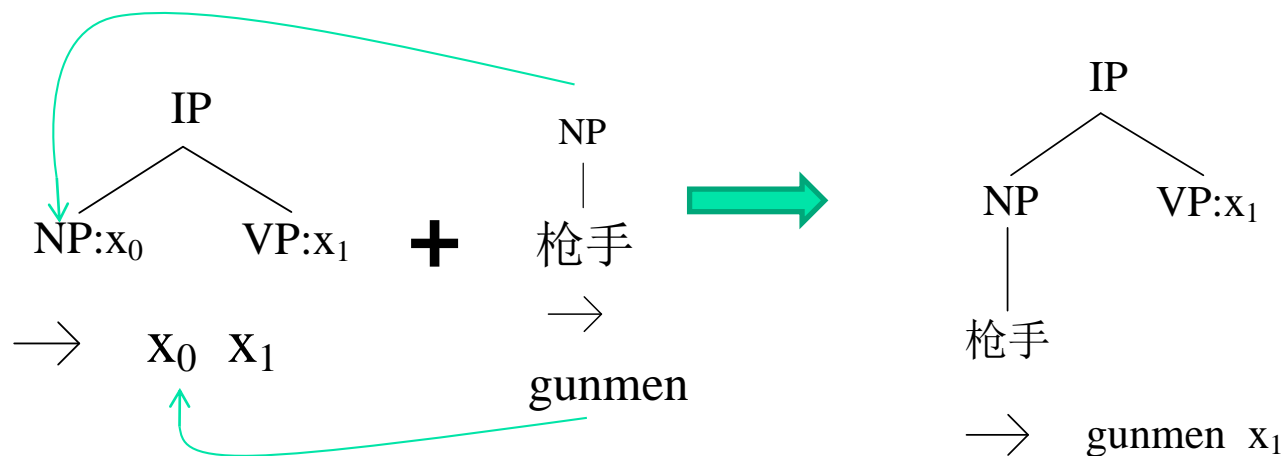


11.2.10 树翻译模型



11.2.10 树翻译模型

组合规则:





11.2.10 树翻译模型

■ 树到串模型的优势

- 搜索空间小、解码效率高
- 句法分析质量较高的前提下，翻译效果不错

■ 树到串模型的不足

- 强烈依赖于源语言句法分析的质量
- 利用源语言端句法结构精确匹配，数据稀疏严重
- 没有使用任何目标语言句法知识，无法保证目标译文符合文法



11.2.10 树翻译模型

◆ 树到树的翻译模型

Zhang et al.(2007, 2008) 提出了树到树的翻译模型。

- **句法分析：** 将源语言句子分析为一棵句法结构树（短语结构树）
- **树到树的转换：** 递归地将源语言句子的句法结构树转换为目标语言句子的句法结构树，拼接叶结点得到译文。

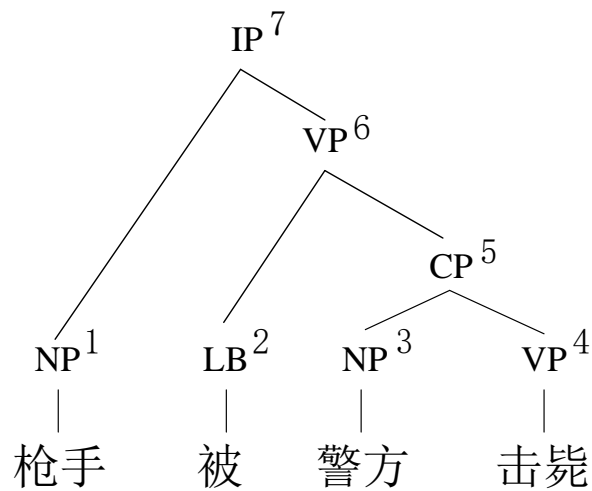


11.2.10 树翻译模型

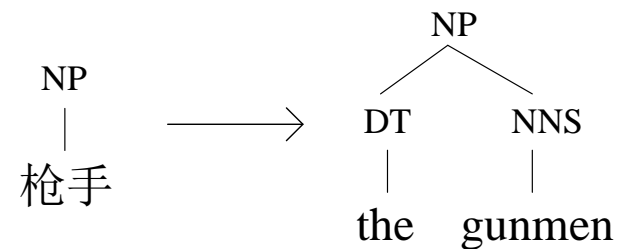
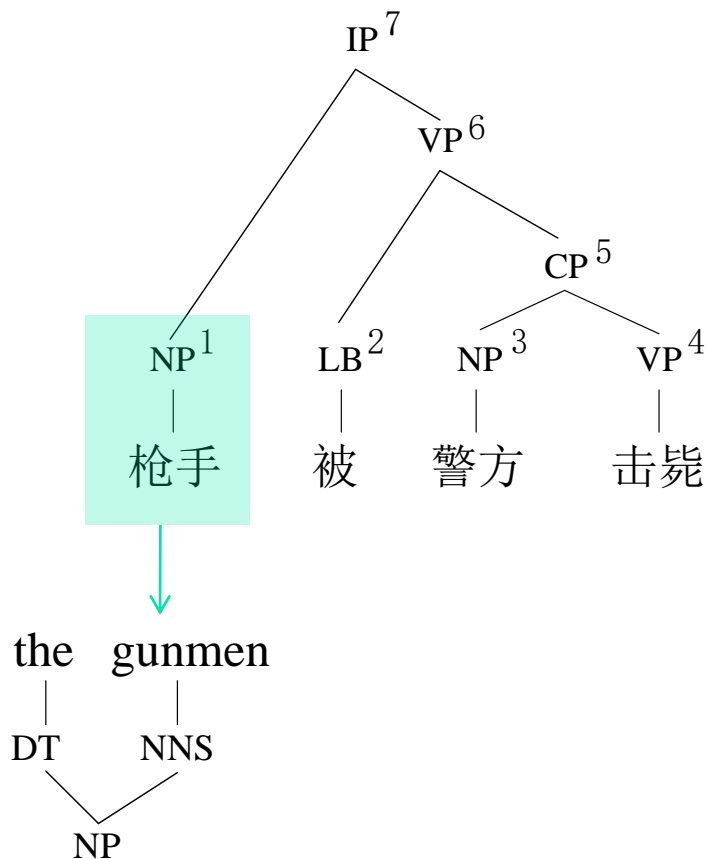
◆ 树到树翻译模型的解码

- **解码算法：**对于源语言句法结构树，自底往上或自顶往下考虑每个节点，为每个节点搜索能够匹配的**树到树翻译规则**。至所有节点匹配完毕，得到最佳译文。

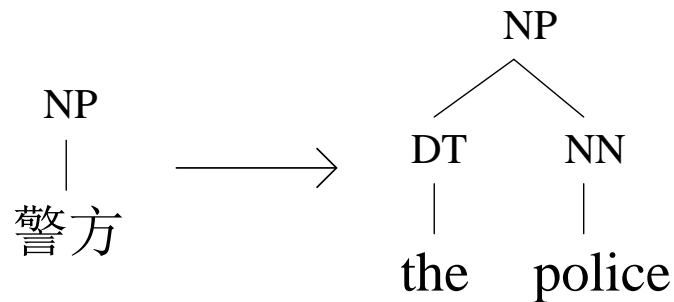
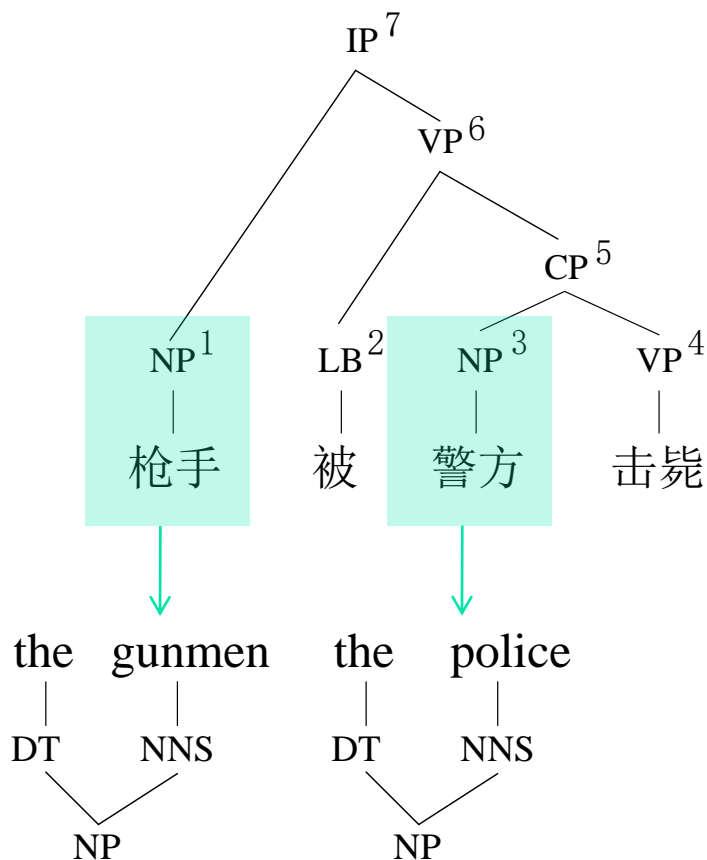
11.2.10 树翻译模型



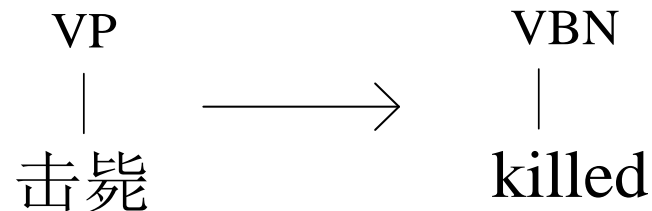
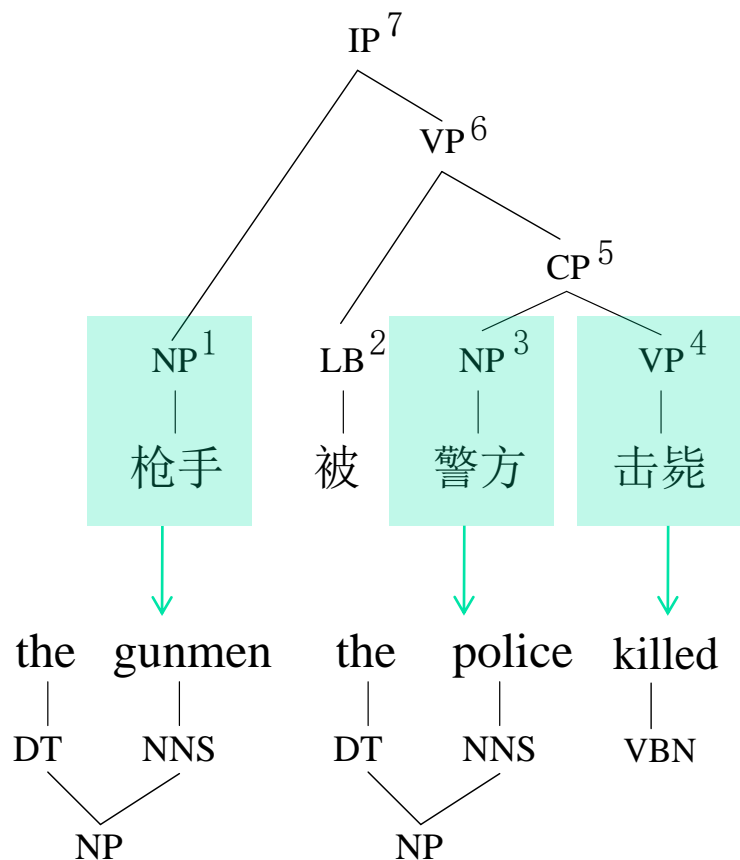
11.2.10 树翻译模型



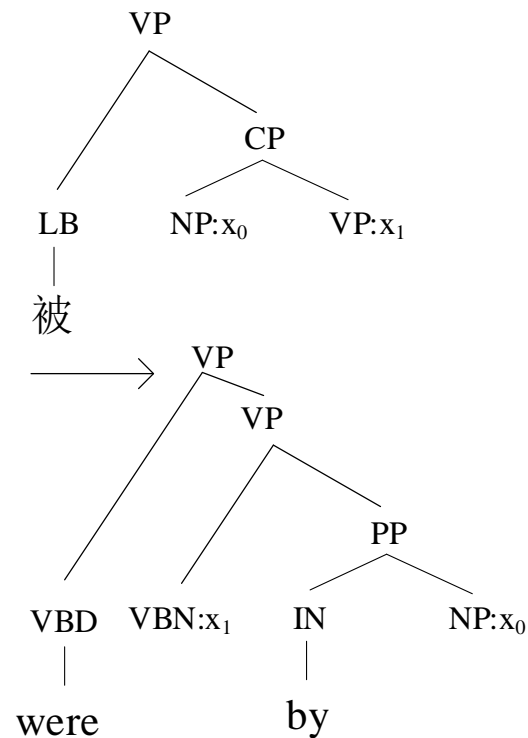
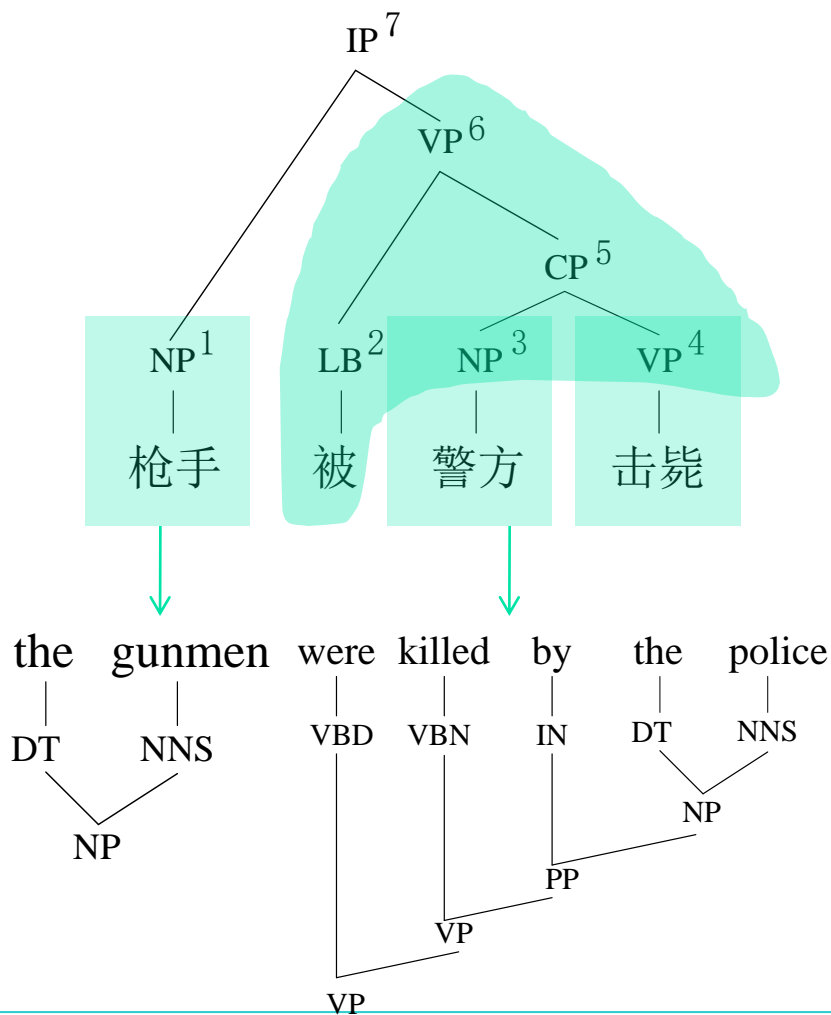
11.2.10 树翻译模型



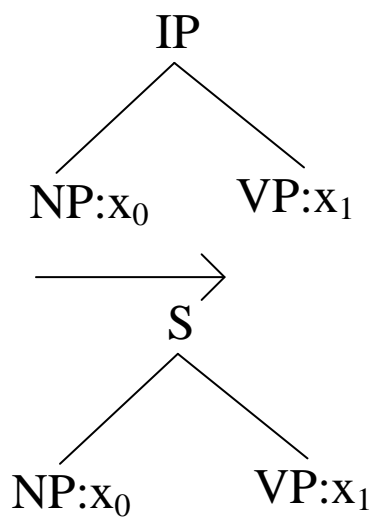
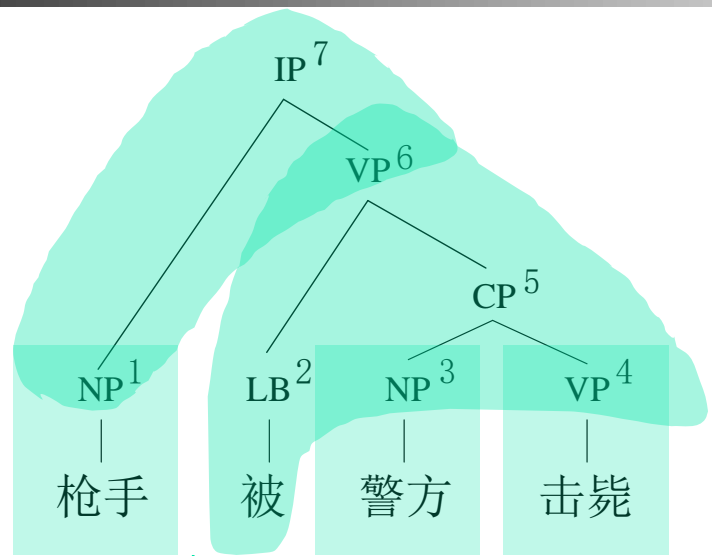
11.2.10 树翻译模型



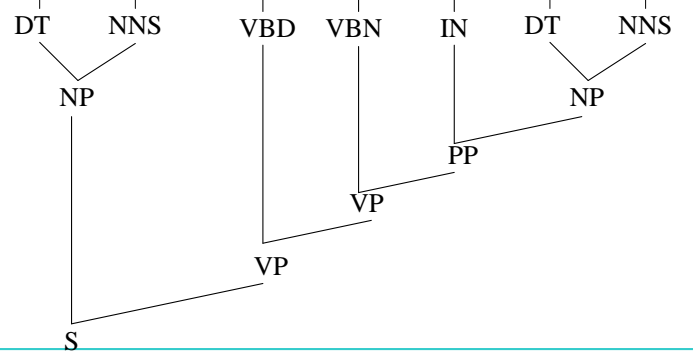
11.2.10 树翻译模型



11.2.10 树翻译模型



the gunmen were killed by the police





11.2.10 树翻译模型

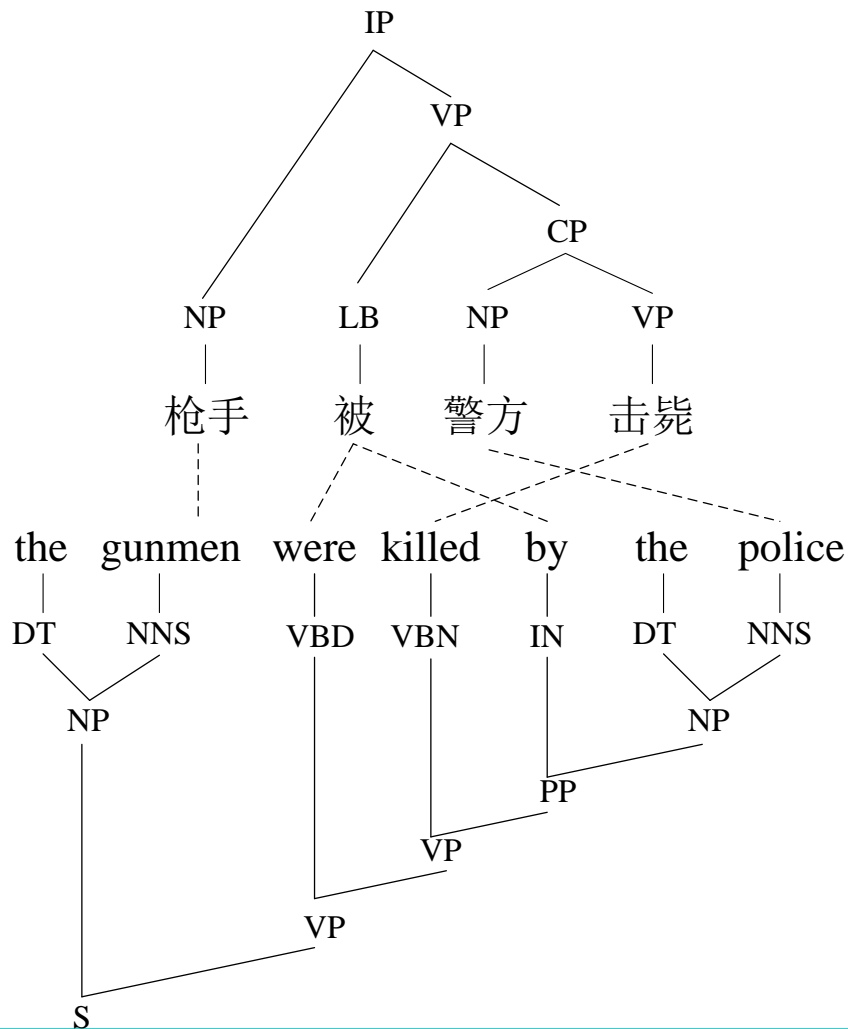
◆ 树到树翻译模型的规则学习

- **树到树翻译规则学习算法：** 给定源语言和目标语言的双语平行句对（经过词语对齐、源语言和目标语言端经过句法分析），抽取满足词语对齐的树到树翻译规则

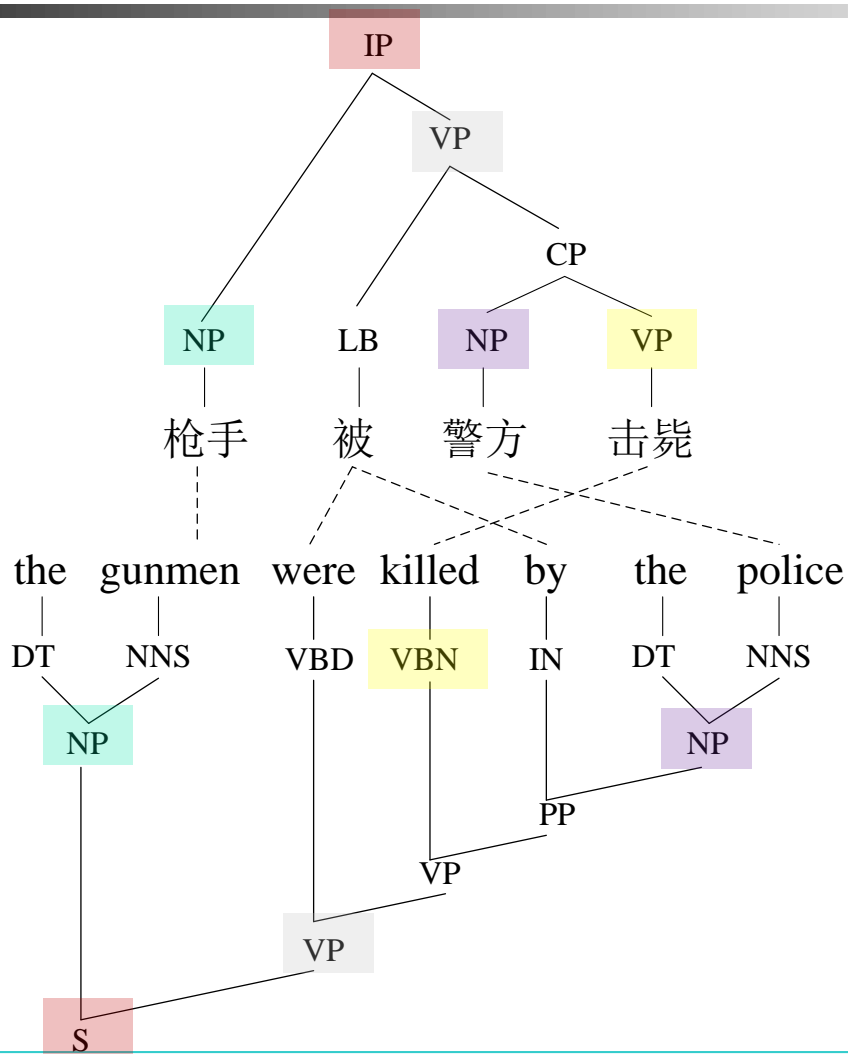
11.2.10 树翻译模型

- **确定满足词语对齐的树节点：** 源语言句法树节点所覆盖的源语言子串与目标语言句法树节点所覆盖的目标语言子串满足词语对齐约束。为每对节点抽取树到树翻译规则。

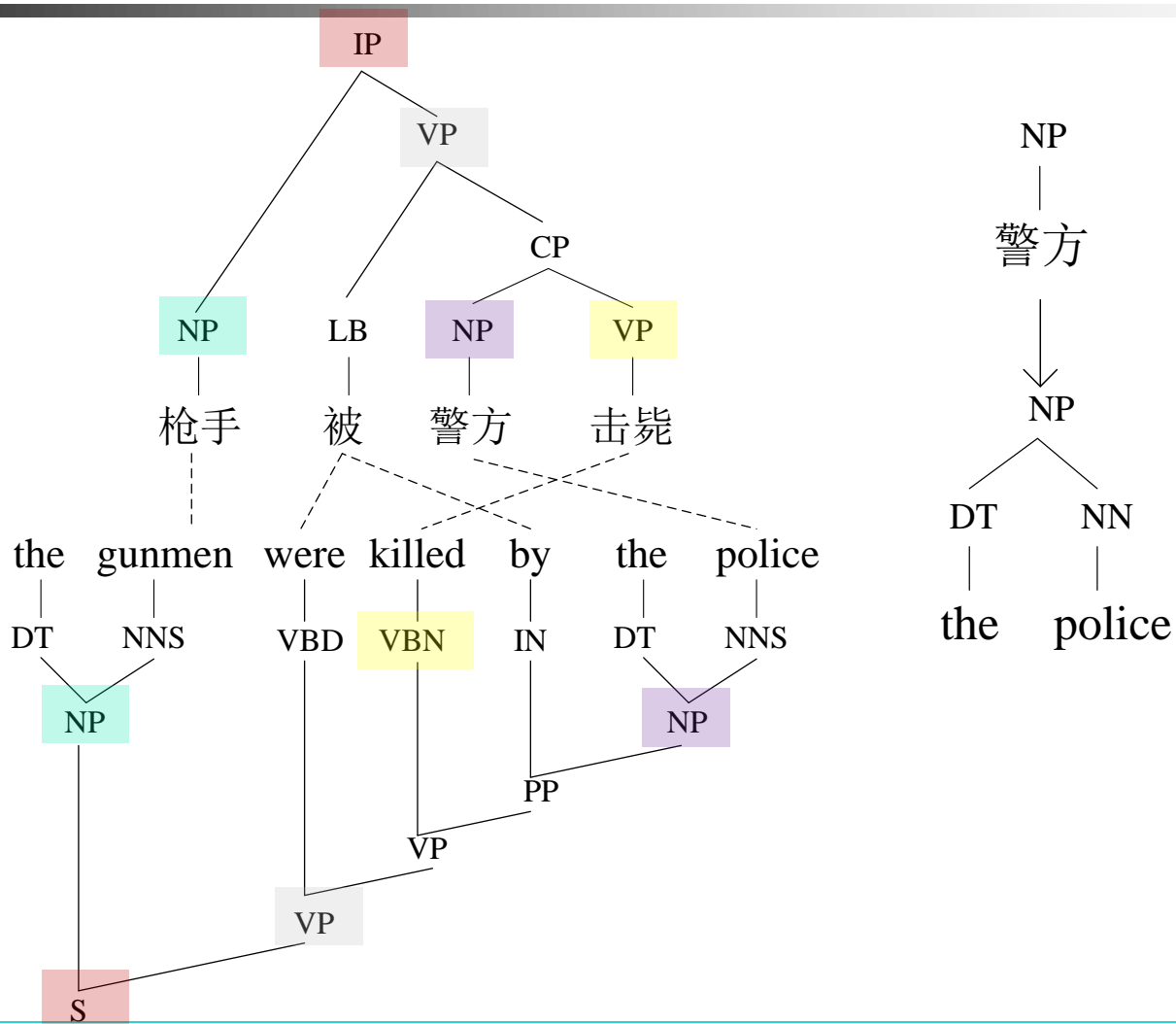
11.2.10 树翻译模型



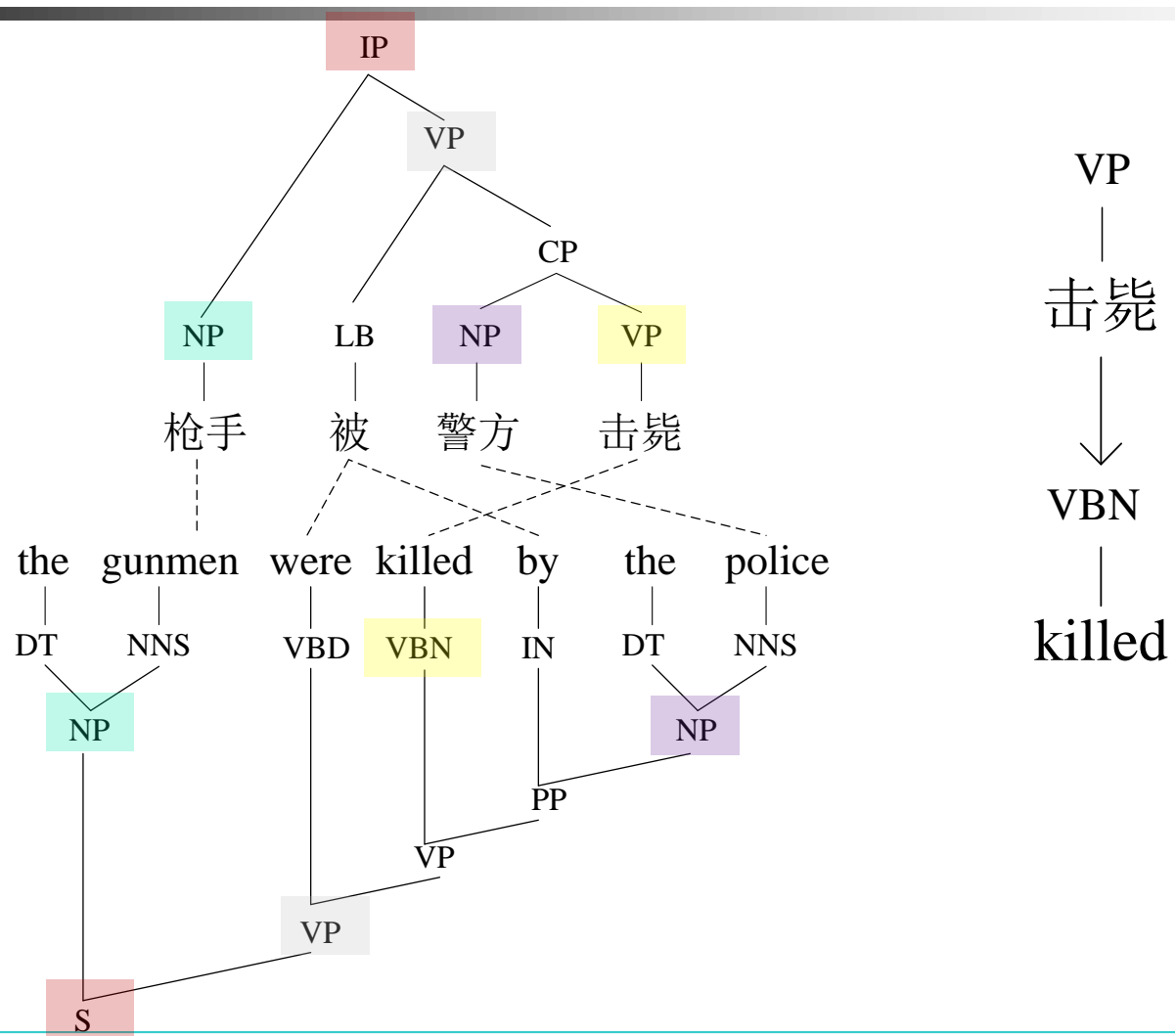
11.2.10 树翻译模型



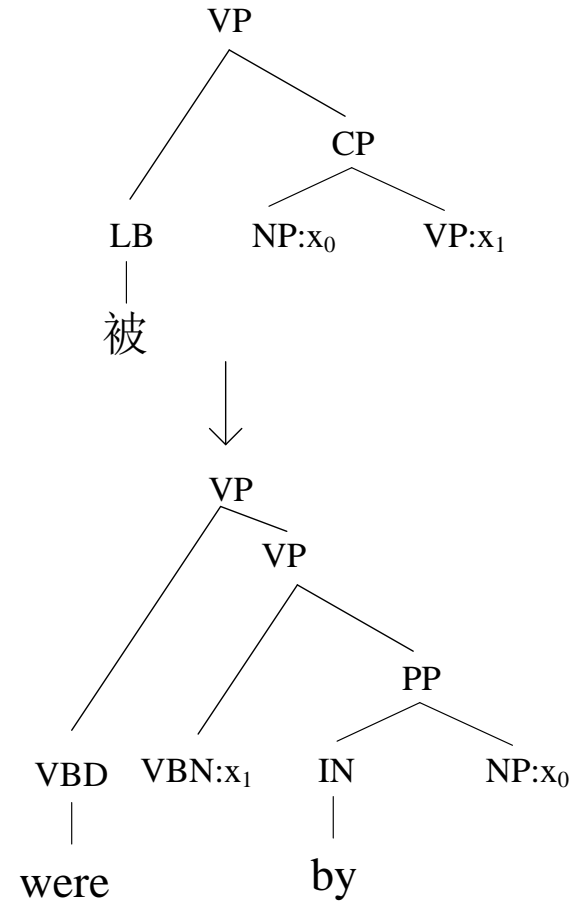
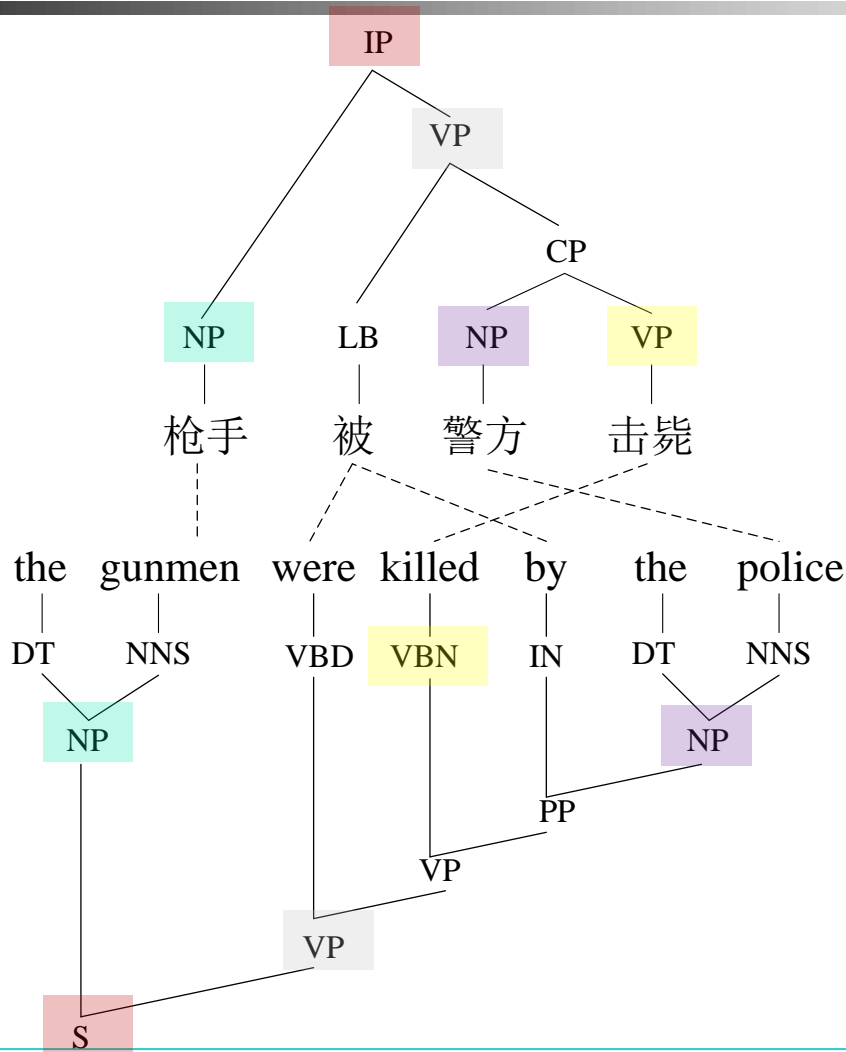
11.2.10 树翻译模型



11.2.10 树翻译模型

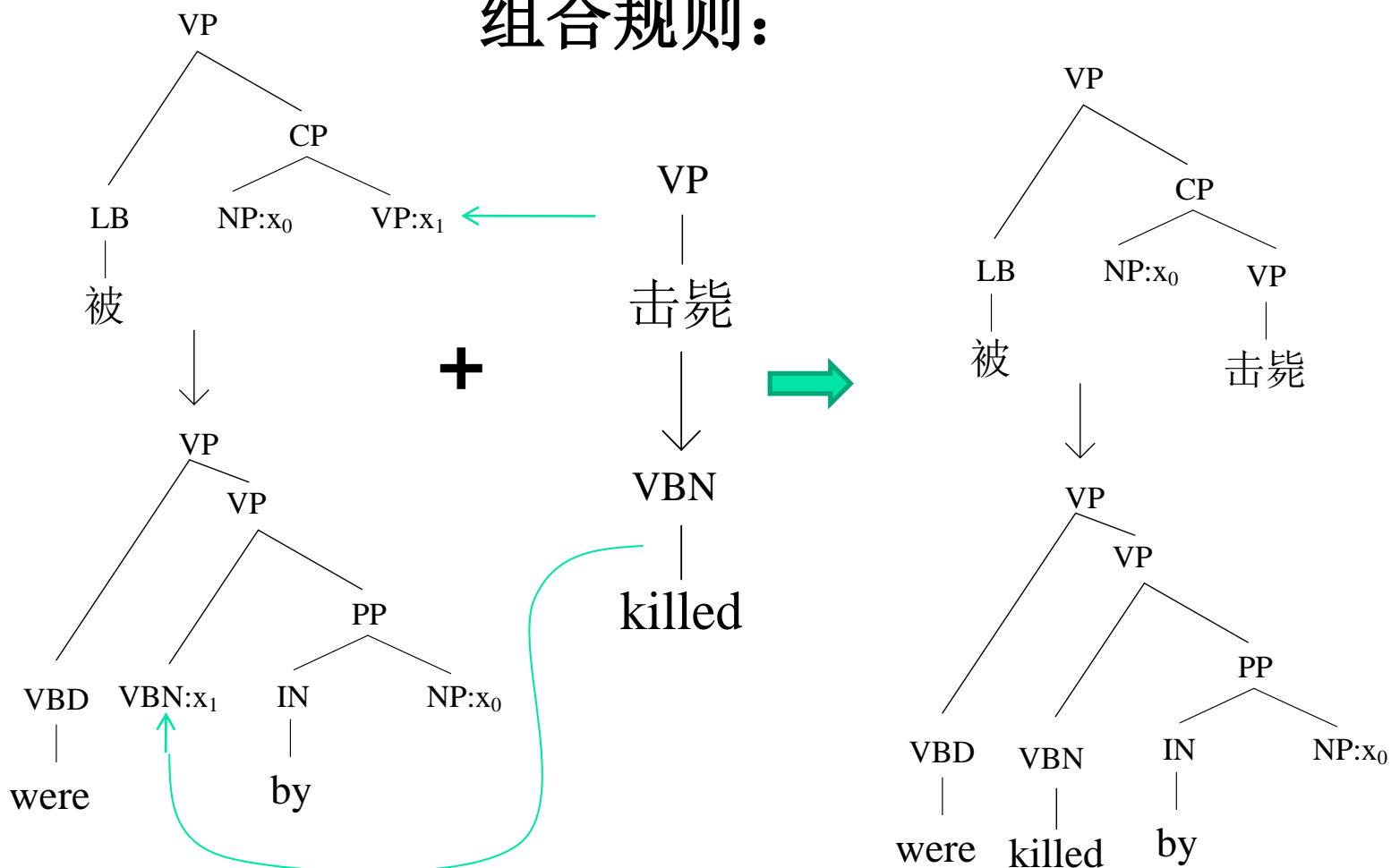


11.2.10 树翻译模型



11.2.10 树翻译模型

组合规则:





11.2.10 树翻译模型

■ 树到树模型的优势

- 搜索空间小、解码效率高

■ 树到树模型的不足

- 强烈依赖于源语言和目标语言句法分析的质量
- 利用两端句法结构精确匹配，数据稀疏非常严重
- 翻译质量差



11.2.10 树翻译模型

◆ 串到树的翻译模型

Galley et al. (2004, 2006) , Marcu et al. (2006) 提出了串到树的翻译模型。

- **串到树的转换：** 利用串到树转换规则，将源语言句子分析为一棵目标语言句法结构树，拼接叶结点得到译文。

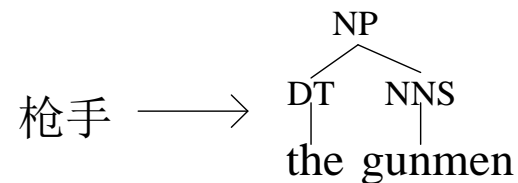
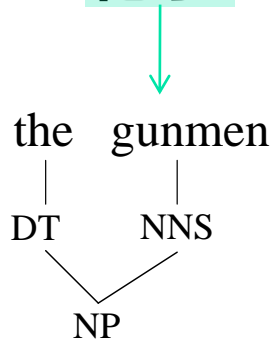


11.2.10 树翻译模型

枪手 被 警方 击毙

11.2.10 树翻译模型

枪手 被 警方 击毙



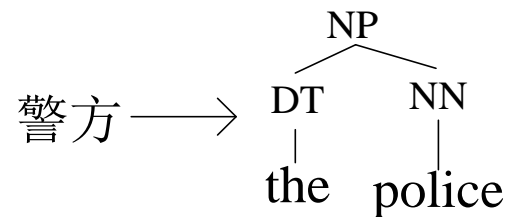
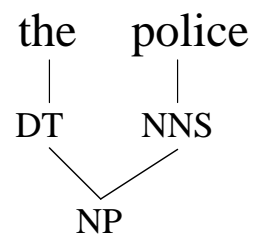
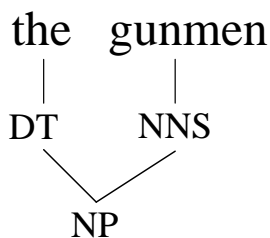
11.2.10 树翻译模型

枪手

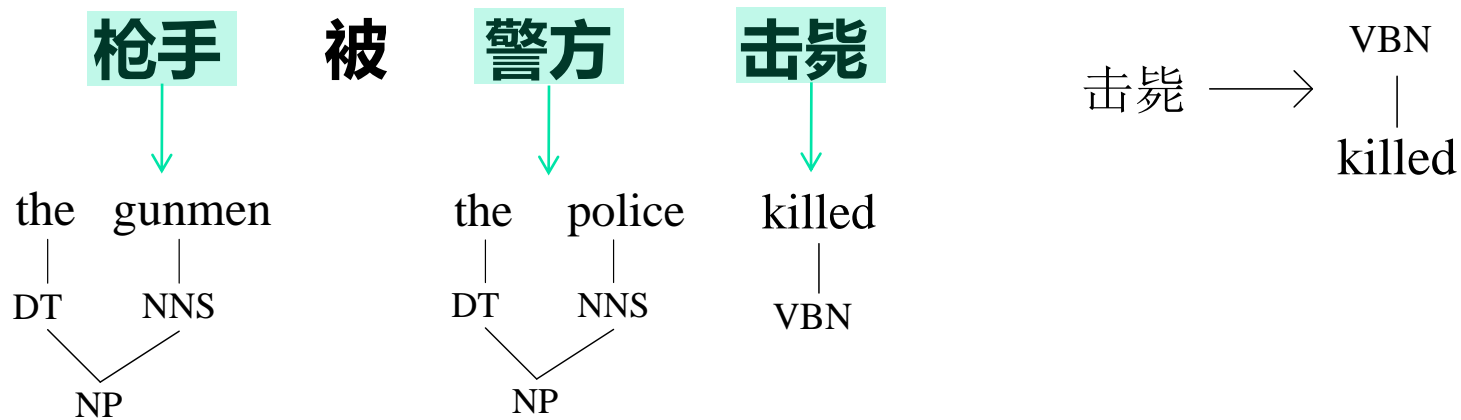
被

警方

击毙

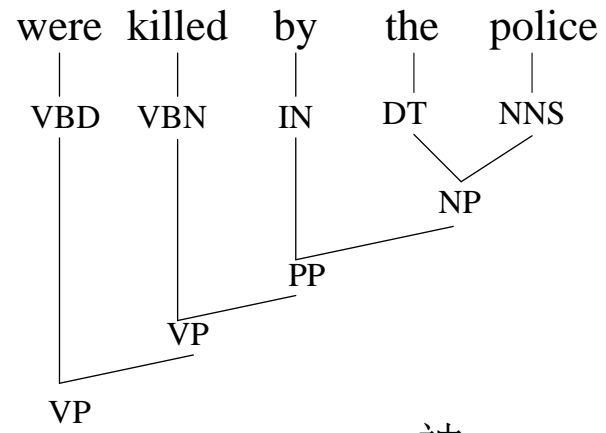
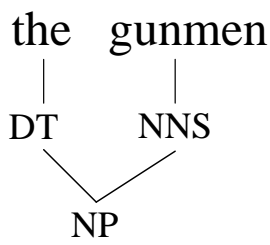


11.2.10 树翻译模型

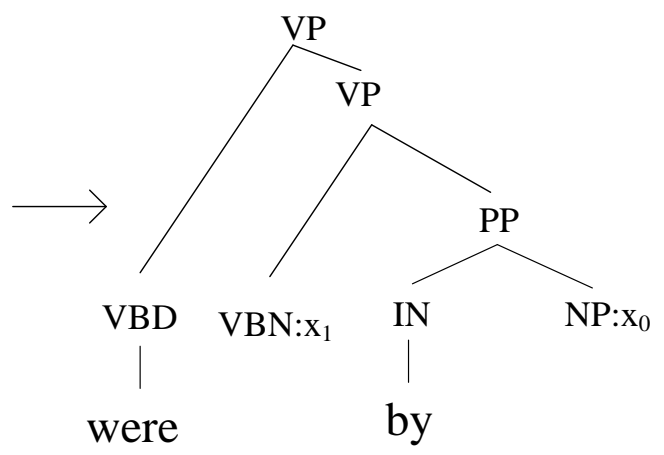


11.2.10 树翻译模型

枪手 被 警方 击毙

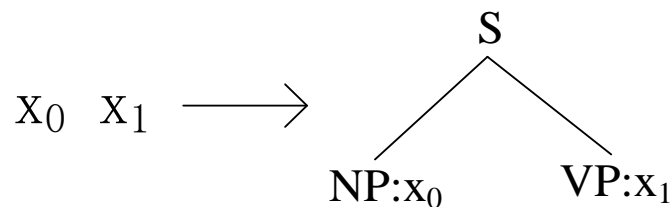
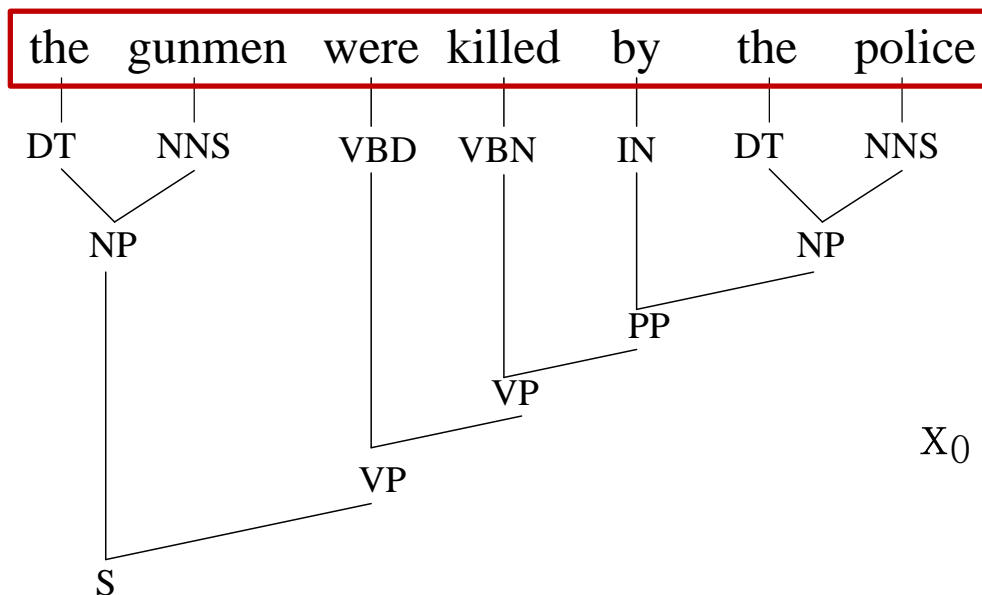


被 x_0 x_1



11.2.10 树翻译模型

枪手 被 警方 击毙



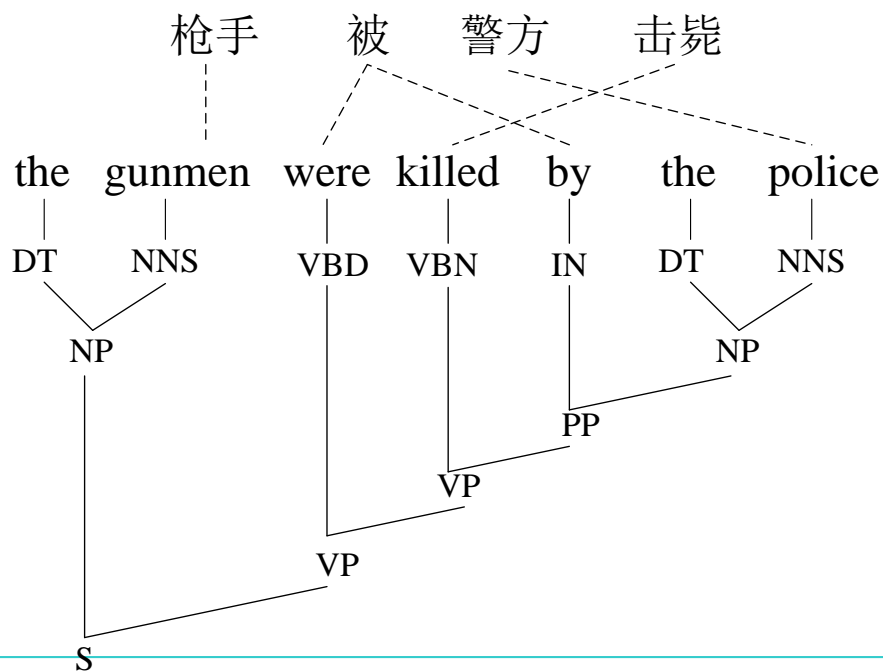


11.2.10 树翻译模型

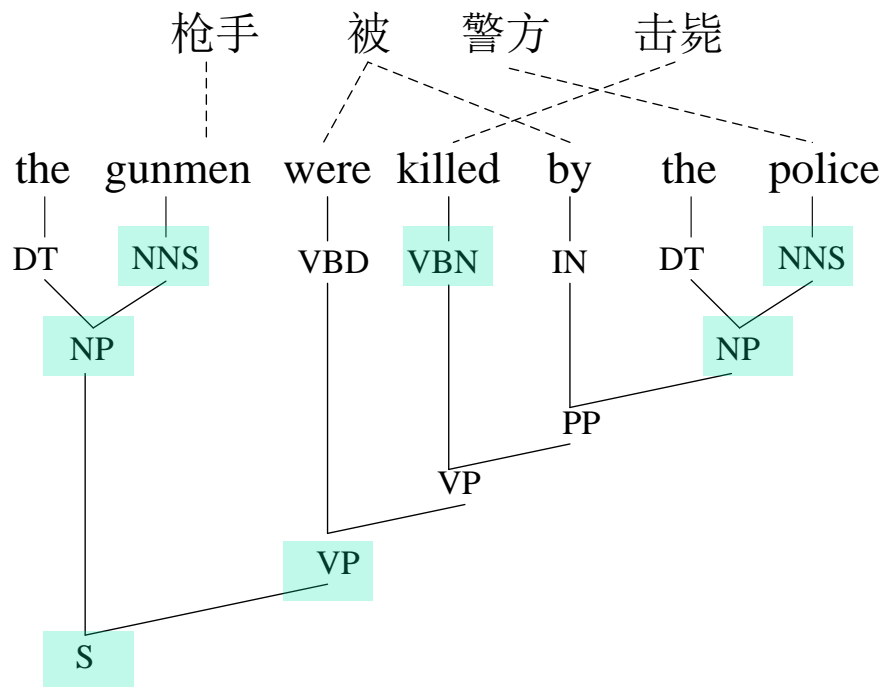
- **串到树翻译规则抽取：** 给定源语言和目标语言的双语平行句对（经过词语对齐、目标语言端句法分析），抽取满足词语对齐的串到树翻译规则

11.2.10 树翻译模型

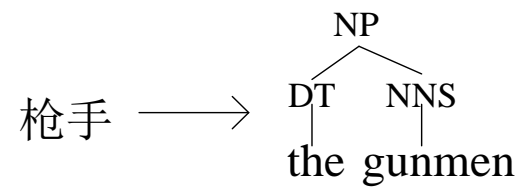
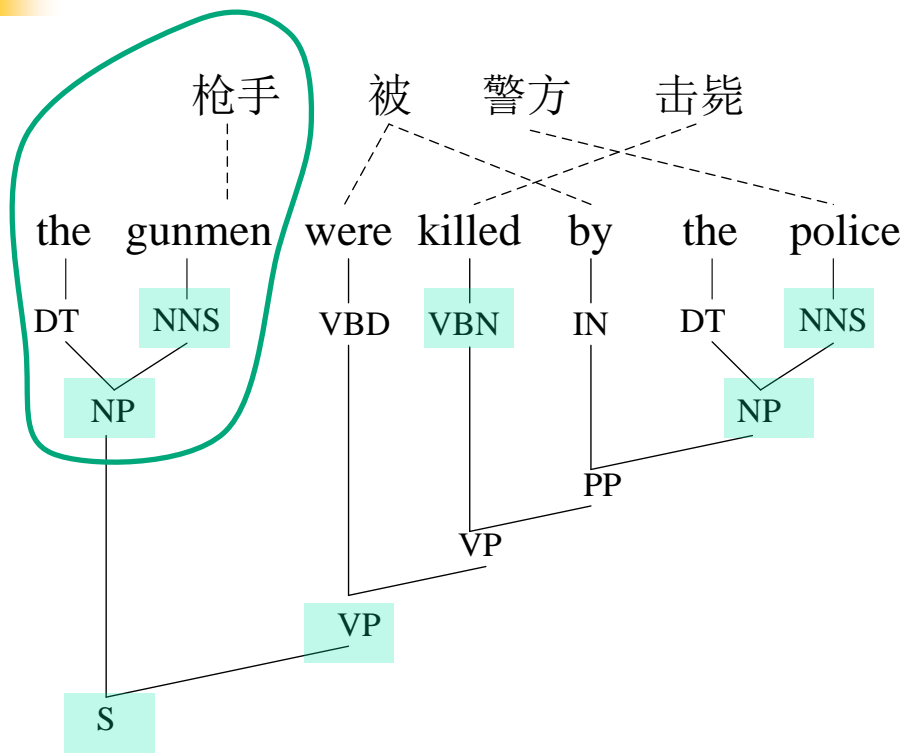
- **确定满足词语对齐的树节点：** 目标语言句法树节点所能到达的源语言子串与该树节点覆盖的目标语言子串满足词语对齐约束



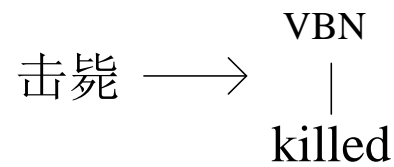
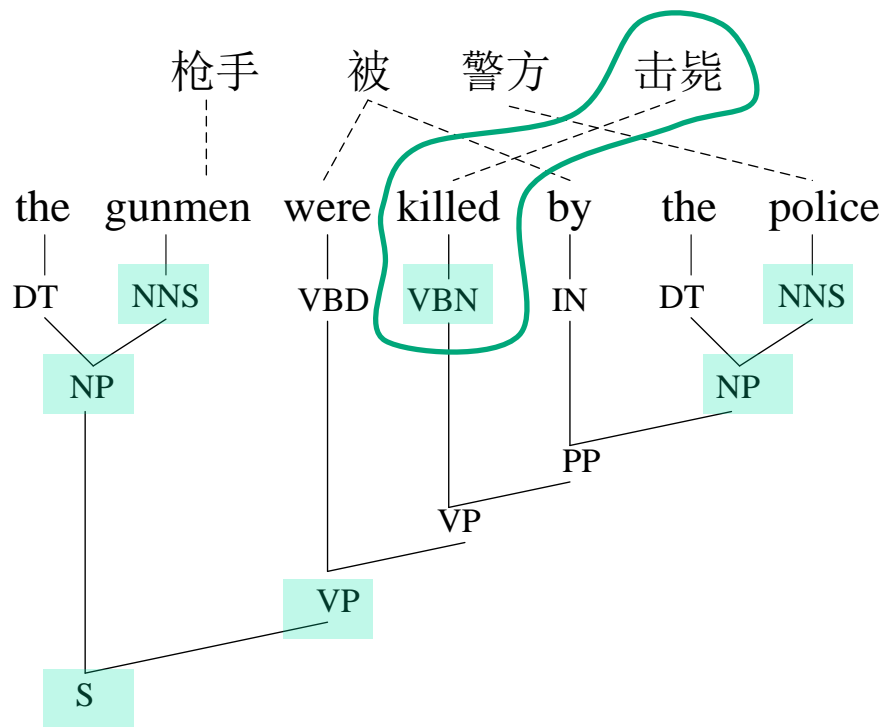
11.2.10 树翻译模型



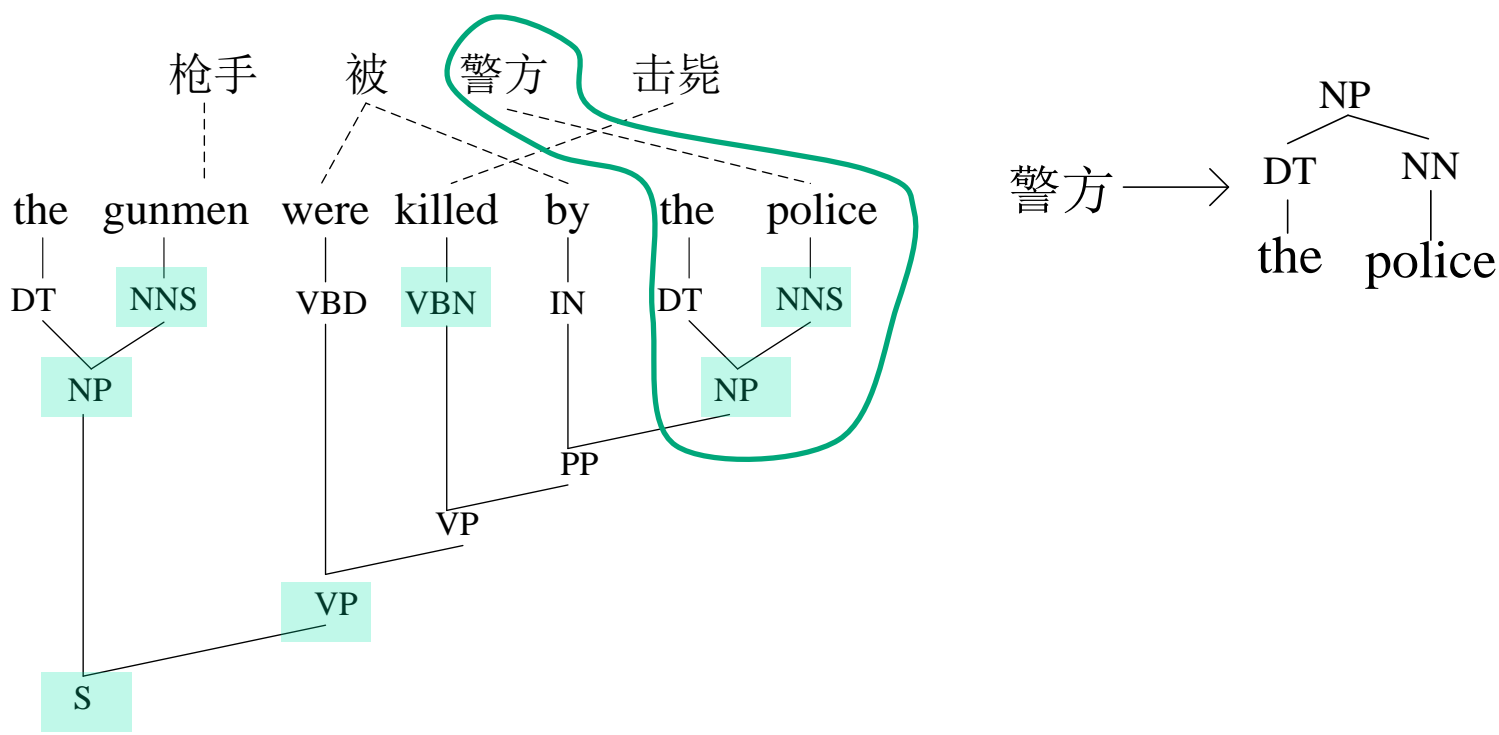
11.2.10 树翻译模型



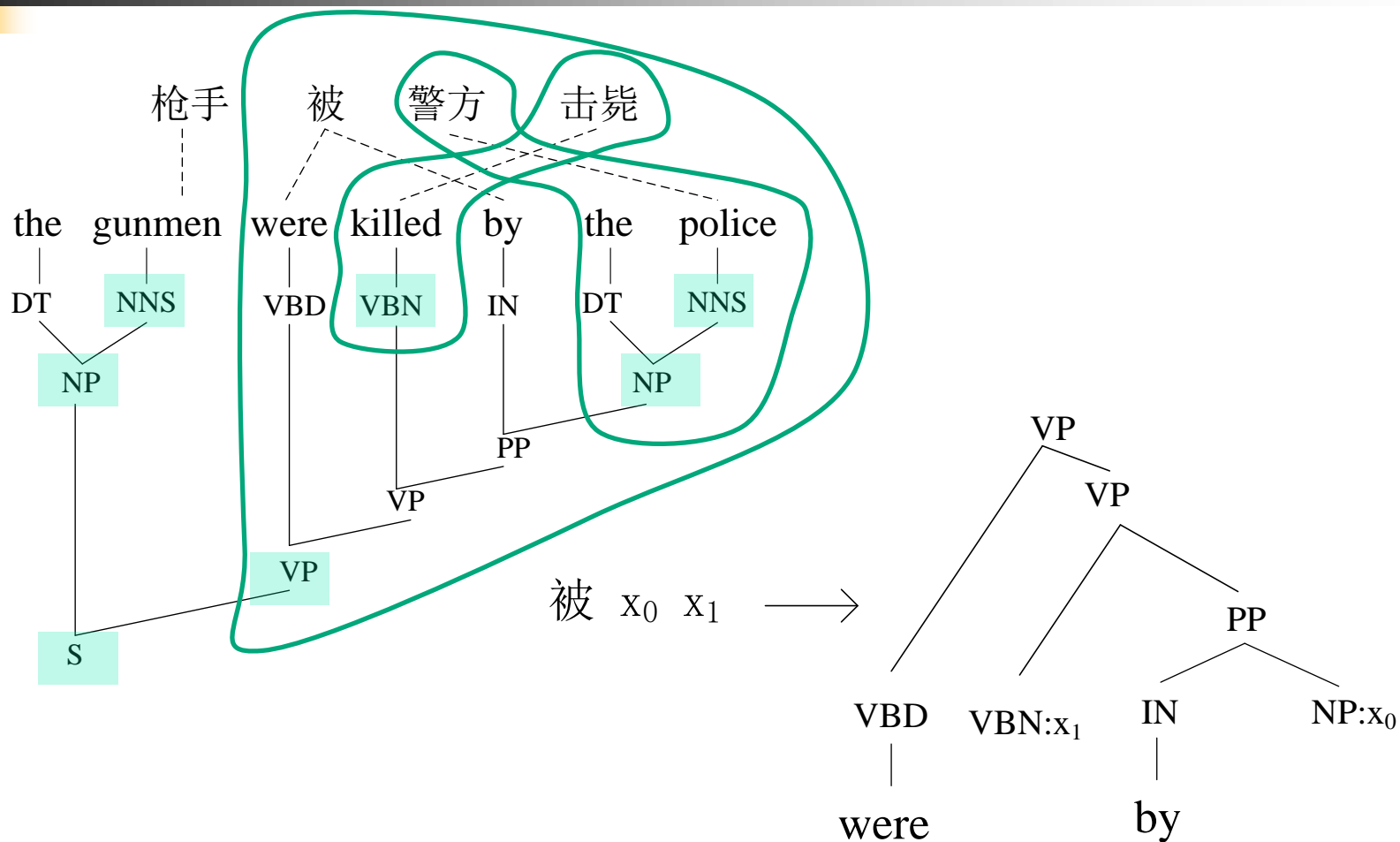
11.2.10 树翻译模型



11.2.10 树翻译模型

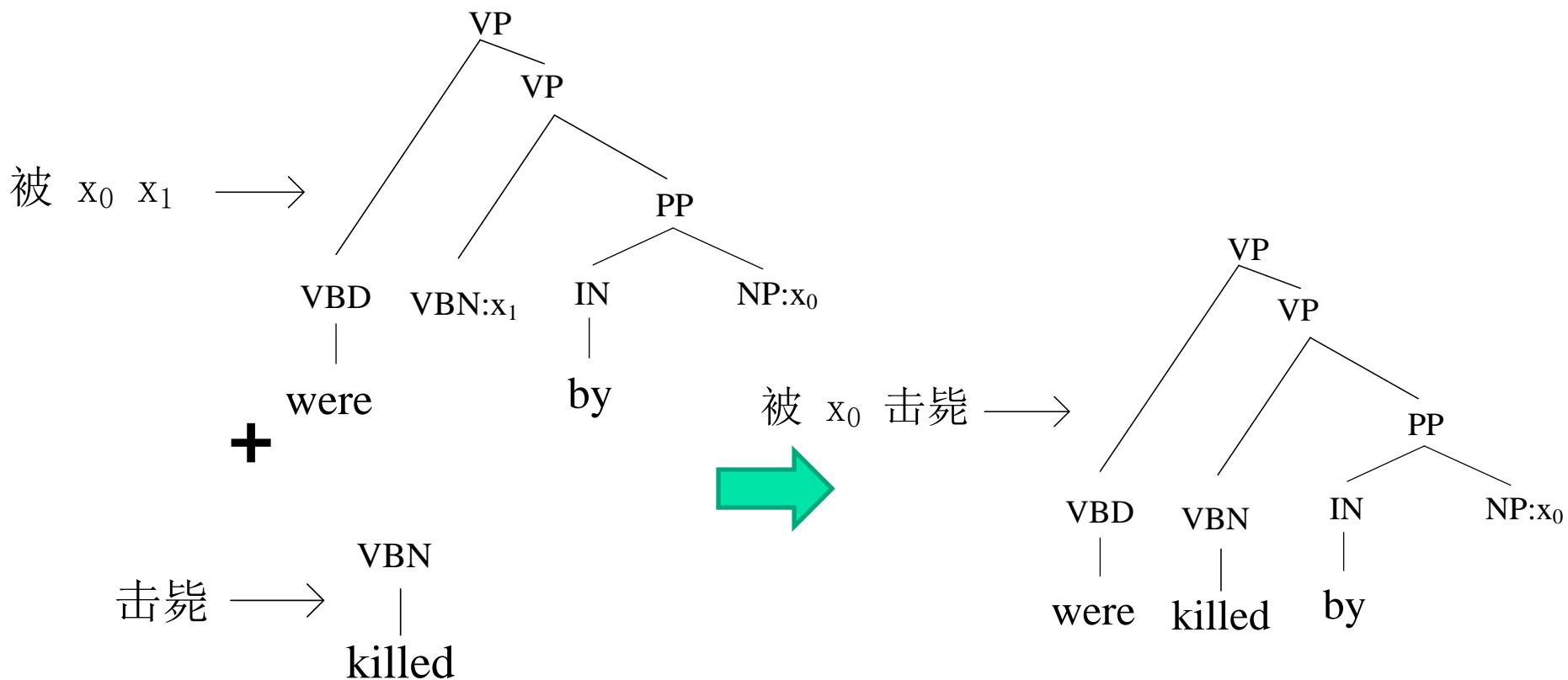


11.2.10 树翻译模型



11.2.10 树翻译模型

组合规则:





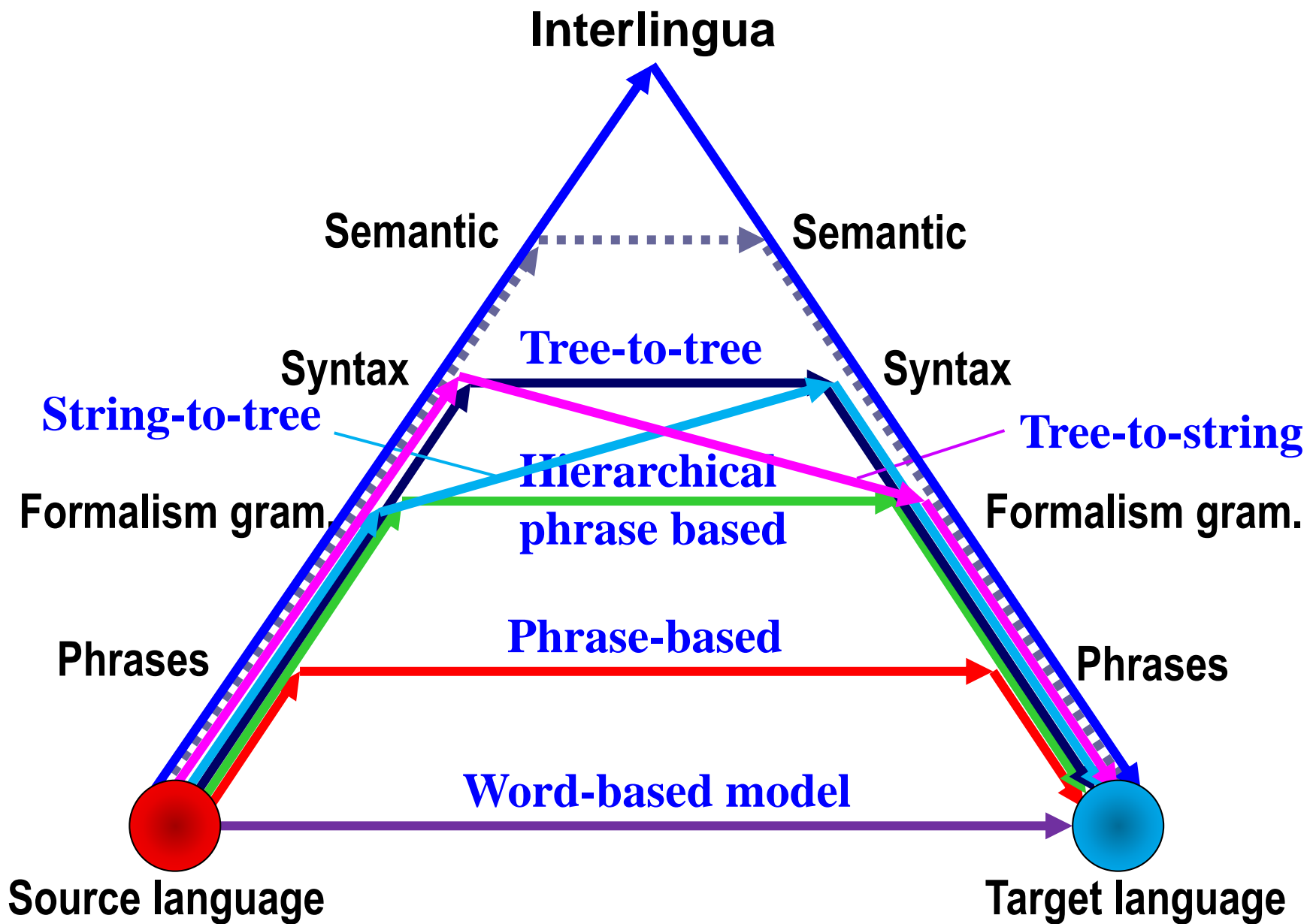
11.2.10 树翻译模型

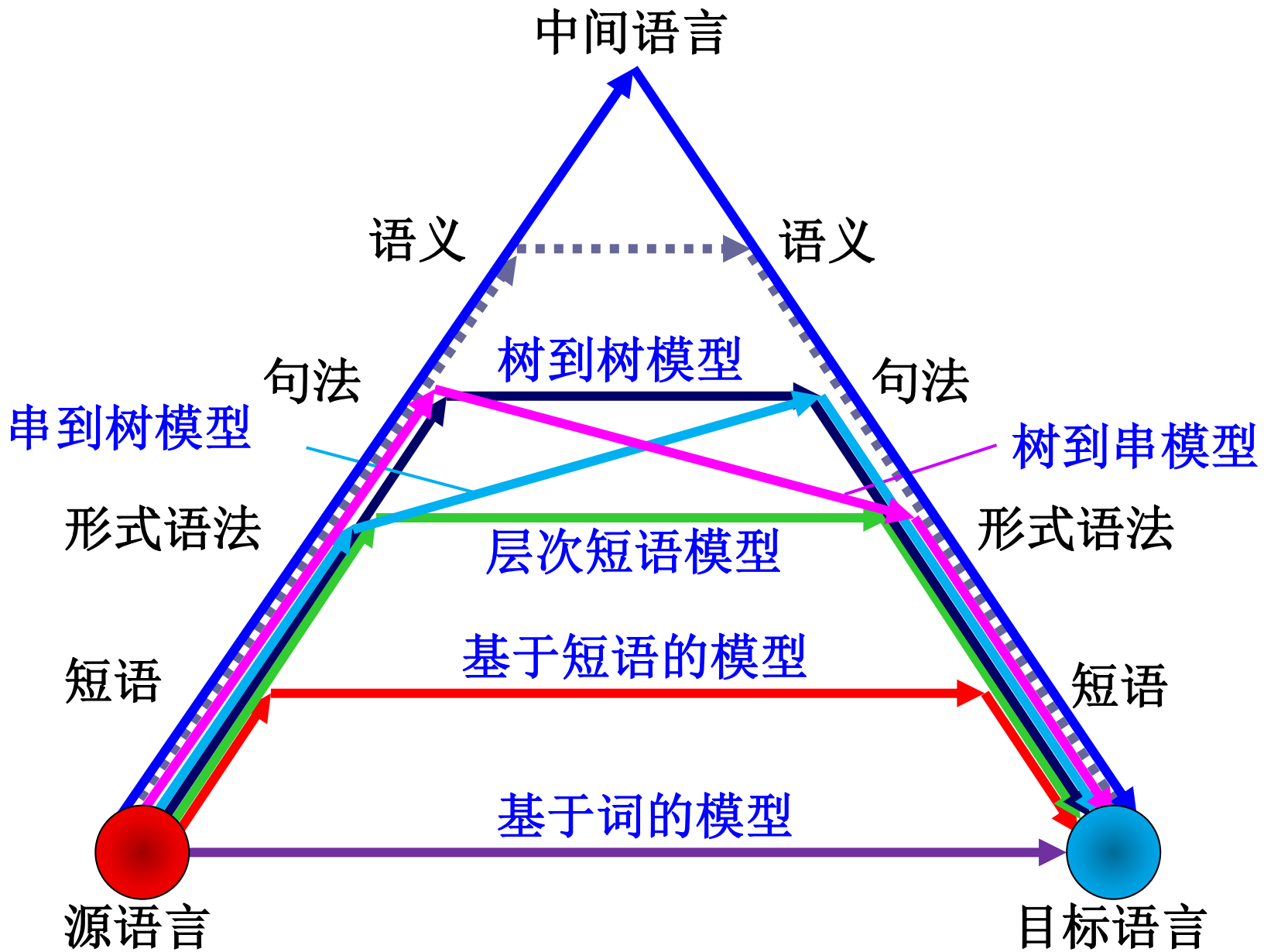
■ 串到树模型的优势

- 搜索空间大，保证译文符合文法，翻译质量高

■ 串到树模型的不足

- 解码速度受限
- 未使用源语言端句法知识，存在词义消歧问题



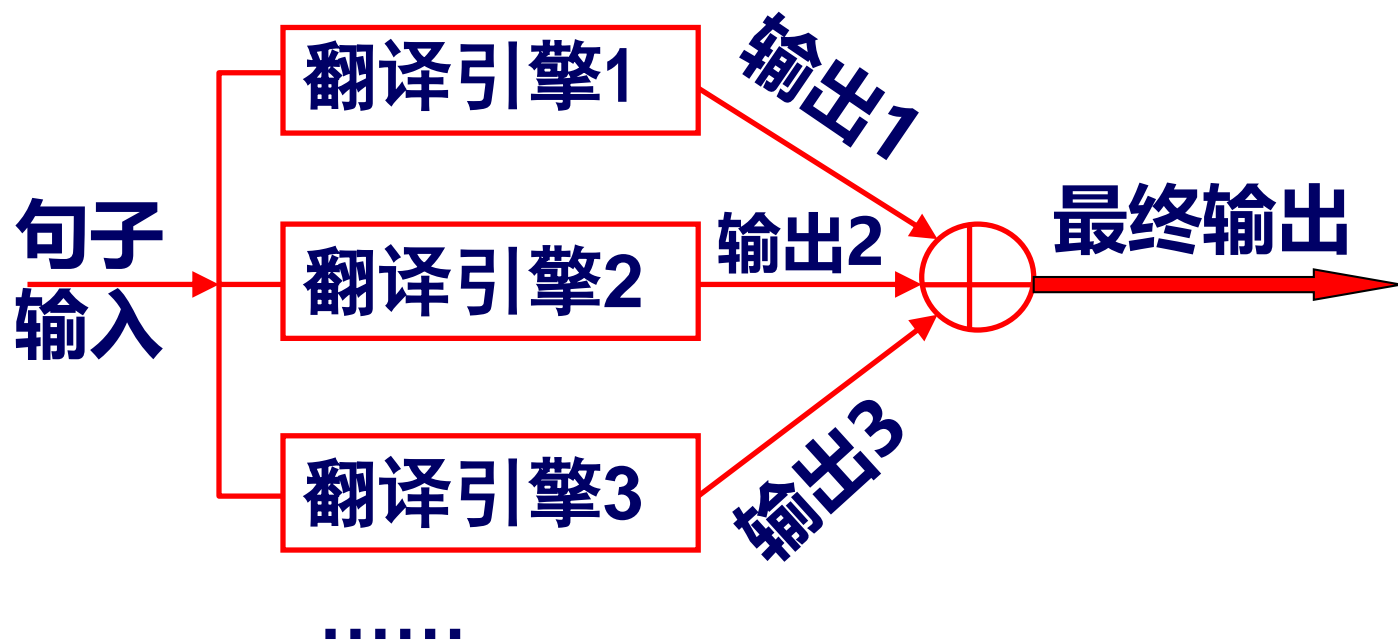




11.2.11 系统融合

11.2.11 系统融合

◆多翻译引擎集成策略



11.2.11 系统融合

◆ 系统融合方法：

(1) 句子级系统融合

针对同一个源语言句子，利用最小贝叶斯风险解码或重打分方法比较多个机器翻译系统的译文输出，将最优的翻译结果作为最终的一致翻译结果。

(2) 短语级系统融合

(3) 词语级系统融合



11.2.11 系统融合

澳洲是与北韩有邦交的少数国家之一

- 1. Australia is having diplomatic relations with North Korea one of countries**
- 2. Australia is one of the few countries have diplomatic relations with North Korea**
- 3. Australia was one of the few countries with diplomatic relations with North Korea**
- 4. Australia is one country that have diplomatic relations with North Korea**

... ..



11.2.11 系统融合

◆ 系统融合方法：

- (1) 句子级系统融合
- (2) 短语级系统融合

利用多个翻译系统的输出结果，重新抽取短语翻译规则集合，并利用新的短语翻译规则进行重新解码。

- (3) 词语级系统融合



11.2.11 系统融合

澳洲是与北韩有邦交的少数国家之一

1. **Australia is** having diplomatic relations with North Korea one of countries
2. Australia is **one of the few countries** have diplomatic relations with North Korea
3. Australia was one of the few countries with diplomatic relations with North Korea
4. Australia is one country **that have diplomatic relations with North Korea**

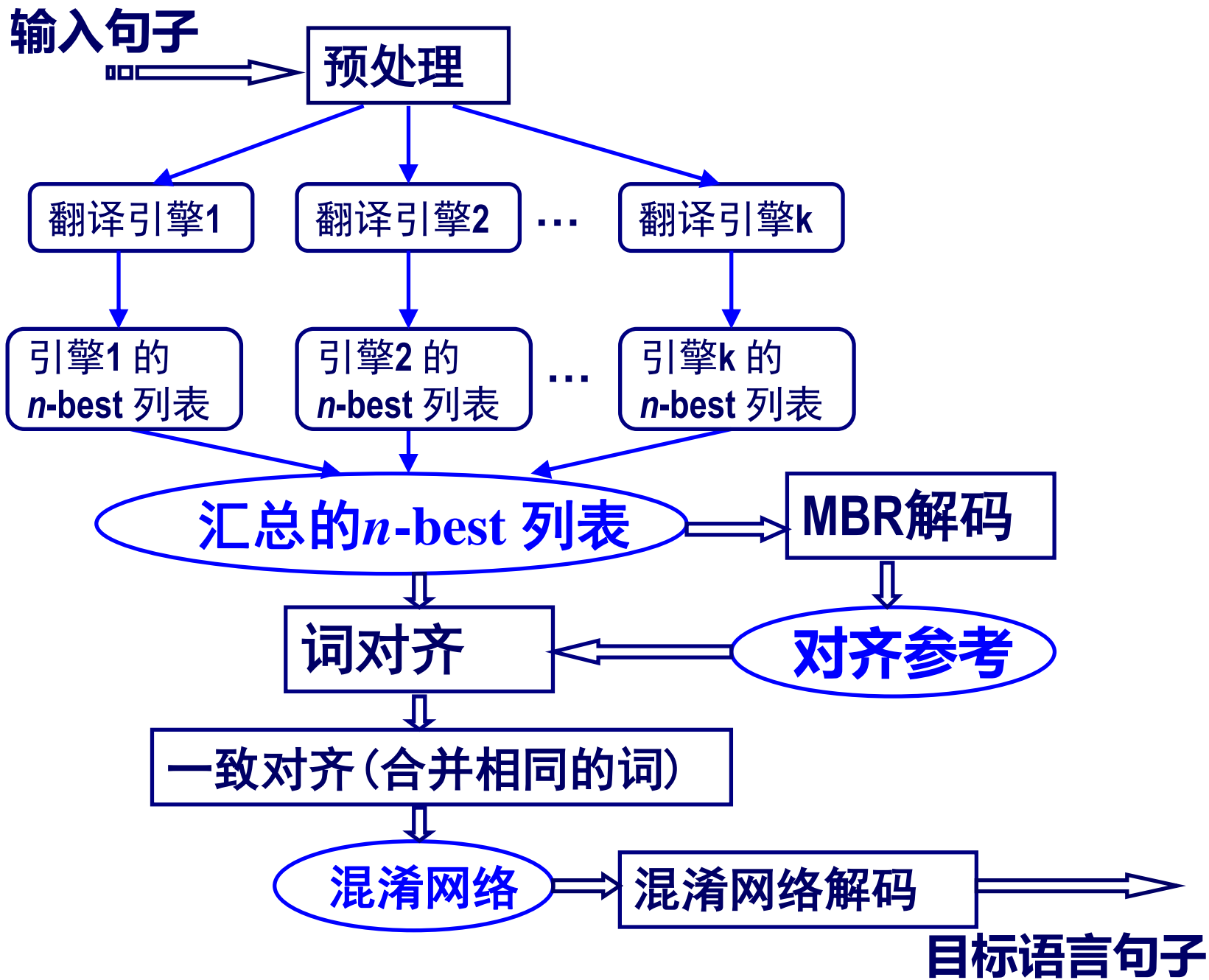
... ..

11.2.11 系统融合

◆ 系统融合方法

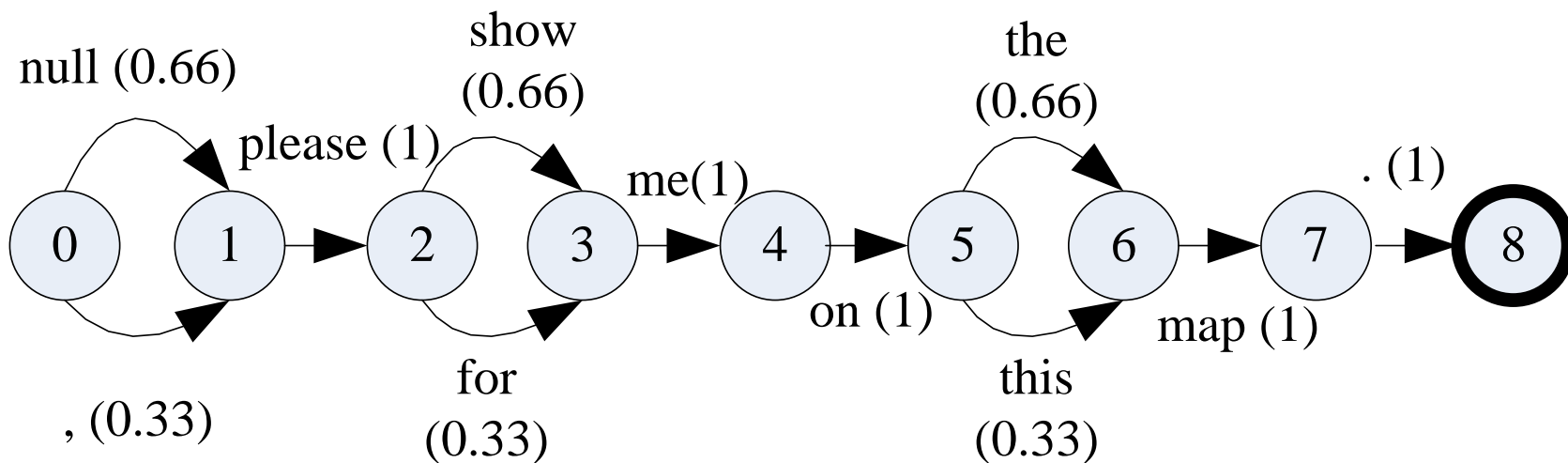
- (1) 句子级系统融合
- (2) 短语级系统融合
- (3) 词语级系统融合

首先将多个翻译系统的译文输出进行词语对齐，构建一个混淆网络，对混淆网络中的每个位置的候选词进行置信度估计，最后进行混淆网络解码。



11.2.11 系统融合

例如：
null please show me on this map .
null please for me on the map .
, please show me on the map .





11.2.12 译文评估方法

11.2.12 译文评估方法

◆ 常用的评测指标

- 主观评测：(1)流畅度；(2)充分性；
(3) 语义保持性。

流畅性	
5	完美的英语表达 (Flawless English)
4	较好的英语表达 (Good English)
3	非母语的英语表达 (Non-native English)
2	不流畅的英语表达 (Disfluent English)
1	无法理解的英语表达 (Incomprehensible)

11.2.12 译文评估方法

充分性

5	全部信息都已充分表达了出来 (All information)
4	绝大部分信息已经表达了出来 (Most information)
3	很多信息被表达了出来 (Much information)
2	表达了少量信息 (Little information)
1	没有表达任何信息 (None)

* 2004年日本ATR组织口语翻译评测(IWSLT)评测时使用了“流畅性”和“充分性”这两个标准。

11.2.12 译文评估方法

语义保持性 (meaning maintenance)

0	意思完全相反 (Total different meaning)
1	部分语义相同，但引入了误导信息 (Partially the same meaning but misleading information is introduced)
2	部分语义相同，没有引入新的信息 (Partially the same meaning and no new information)
3	意思几乎相同 (Almost the same meaning)
4	意思完全相同 (Exactly the same meaning)

* 2005年CMU组织IWSLT评测时除了使用“流畅性”和“充分性”以外，还使用了这一标准。



11.2.12 译文评估方法

源语言句子: This boy is very lovely.

译文1: 这个小孩很可爱。

译文2: 这个桌子很可爱。

译文3: 这个小孩不可爱。

11.2.12 译文评估方法

◆ 客观评测

(1) **句子错误率**：译文与参考答案不完全相同的句子为错误句子。错误句子占全部译文的比率。

(2) **单词错误率**(Multiple Word Error Rate on Multiple Reference, 记作 mWER): 分别计算译文与每个参考译文的编辑距离, 以最短的为评分依据, 进行归一化处理



11.2.12 译文评估方法

(3) 与位置无关的单词错误率 (Position independent mWER, 记作mPER) : 不考虑单词在句子中的顺序

(4) METEOR 评测方法

对候选译文与参考译文进行词对齐, 计算词汇完全匹配、词干匹配、同义词匹配等各种情况的准确率 (P)、召回率(R)和 F 平均值

11.2.12 译文评估方法

$$F = \frac{10PR}{R + 9P}$$

$$Score = F \times (1 - Penalty)$$

$$Penalty = 0.5 \times \left(\frac{\#chunks}{\#unigrams_matched} \right)^3$$

#chunks 指系统译文中所有被映射到参考译文中一元文法可能构成的语块的个数；

#unigrams_matched 为所有匹配的一元文法的个数。不区分大小写。

分数取值为0~1，0为译文最差，1为最好。

11.2.12 译文评估方法

(5) BLEU评价方法 [Papineni, 2002]

— **Bi**Lingual **E**valuation **U**nderstudy, IBM

➤ 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 **n 元词**数目，得到评测结果。



11.2.12 译文评估方法

例如：

- 系统译文：**the the the the the the the.**
- 参考译文1：**The cat is on the mat.**
- 参考译文2：**There is a cat on the mat.**

按照上述计算方法，如果 n 取1的话，该候选译文可以得到 7/7 的打分，但显然这种翻译结果几乎没有任何意义。

11.2.12 译文评估方法

修正的计算一元语法精确度的方法：针对某个待评测的系统译文句子，首先统计每个单词在所有参考译文中出现次数的最大值 Max_Ref_Count ，然后，统计该单词在系统译文中出现的总次数 $Count$ ，取 $Count$ 和 Max_Ref_Count 两者中小的一个，即

$$Count_{clip} = \min(Count, Max_Ref_Count)$$

这样保证了每个系统译文中的单词计数不会超过该词在某个参考译文中出现次数的最大值。

11.2.12 译文评估方法

把系统译文中所有单词的 $Count_{clip}$ 值累加起来，得到 $Total_Count_{clip}$ ，即待评测的系统译文中出现在参考译文中的单词个数，

最后，用 $Total_Count_{clip}$ 除以系统译文中全部单词的个数。

在上面的例子中，系统译文中的单词 **the** 在参考译文1中出现的次数最多， $Max_Ref_Count=2$ ，而 **the** 在系统译文中出现的次数为7，即 $Count=7$ ，因此， $Count_{clip}=\min(7, 2)=2$ 。候选译文中全部单词的个数等于7，因此，该例中修正后的一元语法精确度为 $2/7$ 。

11.2.12 译文评估方法

候选译文: It₁ is₂ a₃ guide₄ to₅ action₆ which₇ ensures₈ that₉ the₁₀ military₁₁ always₁₂ obeys₁₃ the₁₄ commands₁₅ of₁₆ the₁₇ party₁₈.

参考译文1: It is a guide to action that ensures that the military will forever heed Party commands.

参考译文2: It is the guiding principle which guarantees the military forces always being under the command of the Party.

参考译文3: It is the practical guide for the army always to heed the directions of the party.

单词1~6, 8~11, 14~15, 17~18 均出现在参考译文1中, 单词7出现在参考译文2中, 单词12、16出现在译文2和3中。只有第13个词没有出现在任何译文中。

因此, 一元语法精确度为: 17/18。

11.2.12 译文评估方法

如果考虑 bi-gram {(it is), (is to), (to insure) ...}, 那么, 修正的2元语法的精度为: 10/17。

对于含多个句子的测试文本 *Candidates*, 以句子 *C* 为单位分别计算 *n*-gram 的匹配情况, 然后, 累计所有翻译句子修正后的 *n*-gram 计数, 及所有句子的所有 *n*-gram 计数, 二者相除, 得到修正后的精确度记分:

$$P_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count_{clip}(n\text{-gram})}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-gram}' \in C'} Count(n\text{-gram}')}$$

11.2.12 译文评估方法

考虑到在修正的 n 元语法精度计算中，随着 n 值的增大精度值几乎成指数级下降，因此，BLEU方法中采用了修正的 n 元语法精度的对数加权平均值，相当于对修正的精度值进行几何平均， n 值最大为4。

另外，考虑到句子的长度对上述BLEU评分也有一定的影响，例如，如果一个机器翻译系统只翻译最可靠的词汇，译文句子就可能比较短，按上述方法计算出的精度值就会较高。因此，需要进一步考虑候选译文的句子长度对计算评分的影响。

11.2.12 译文评估方法

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

长度过短句子的
惩罚因子

$$w_n = 1/N$$

最大语法的阶
数，实际取4。

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 为候选译文中单词的个数， 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。

11.2.12 译文评估方法

(6) NIST 评测方法

National Institute of Standards and Technology

- ▶ BLEU评分公式中采用的 n 元语法同现概率的几何平均方法使评分值对于各种 n 元语法同现的比例具有相同的敏感性，但实际上，这种做法存在着潜在的矛盾，因为 n 值较大的统计单元出现的概率较低。
- ▶ 基本思想：因此，NIST的研究人员提出了另外一种处理方法，就是用 n -gram同现概率的算术平均值取代几何平均值。另外，如果一个 n 元词在参考译文中出现的次数越少，表明它所包含的信息量越大，那么，它对于该 n 元词就赋予更高的权重。

11.2.12 译文评估方法

根据上述观点，NIST的研究人员对参考译文中的 n -gram单元给出了一种信息权重的计算公式：

$$Info(w_1 \cdots w_n) = \log_2 \left(\frac{Num(w_1 \cdots w_{n-1})}{Num(w_1 \cdots w_n)} \right)$$

其中，分子和分母分别表示 $w_1 \cdots w_{n-1}$ 和 $w_1 \cdots w_n$ 出现的次数。

在NIST方法中还使用了F-比 (F-ratio) 的概念。所谓的F-ratio就是指不同系统之间的得分偏差除以某一系统本身的得分偏差。不同系统之间的得分偏差是指不同系统得分平均值的变化，而某一系统本身的得分偏差是指一个给定的系统在不同测试集上得分的差异。因此，F-ratio越大，评测方法越好。

11.2.12 译文评估方法

基于信息权重和 F-比，NIST 评分的计算公式如下：

$$Score = \sum_{n=1}^N \left\{ \frac{\sum_{\substack{\text{all } w_1 \cdots w_n \\ \text{that co-occur}}} Info(w_1 \cdots w_n)}{\sum_{\substack{\text{all } w_1 \cdots w_n \\ \text{in sys output}}} (\mathbf{1})} \right\} \times \exp \left\{ \beta \log^2 \left[\min \left(\frac{L_{sys}}{\bar{L}_{ref}}, 1 \right) \right] \right\}$$

其中， \bar{L}_{ref} 是参考译文句子的平均长度； L_{sys} 是评测系统的译文长度； β 为经验值，当 L_{sys} 为 \bar{L}_{ref} 的 2/3 时， β 的取值使长度惩罚因子等于 0.5；一般取 $N=5$ 。

NIST 评分值为不小于 0 的实数，0 分表示译文质量最差。



11.2.12 译文评估方法

- ◆ 其他评测方法：TER, GTM 等
- ◆ 关于评测方法的评测



11.2.12 译文评估方法

◆ 统计翻译中的译文错误

- (1) 模型错误：概率最高的译文不是正确的。
- (2) 搜索错误：概率最高的译文是正确的，但搜索算法找不到。这类错误大约占5%。

11.2.12 译文评估方法

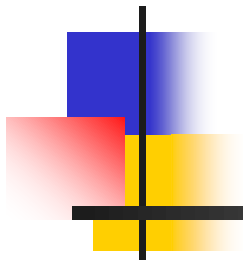
◆ 机器翻译评测活动

- **IWSLT: International Workshop on Spoken Language Translation**
 - 2004 年起，2004~2017 每年一次，针对口语翻译
- **DARPA 资助的机器翻译评测 NIST**
 - NIST 组织，2002 年 6 月首次评估，几乎每年一次
- **欧盟资助的机器翻译评测 WMT**
 - 2006 年起，2006~2018 每年一次，针对欧洲语言之间的翻译
- **全国机器翻译评测 (CWMT: China Workshop on MT)**
 - 2007，2008，2009，2011，2013，2015，2017，2018



本讲小结

- ◆ 统计机器翻译的基本原理
- ◆ IBM 1~5 翻译模型
- ◆ 翻译模型的改进
- ◆ 解码算法
- ◆ SMT 系统实现方法
- ◆ 译文评估方法



Thanks

谢谢!