

机器翻译原理与方法

第二讲 机器翻译方法概述

刘群

中国科学院计算技术研究所

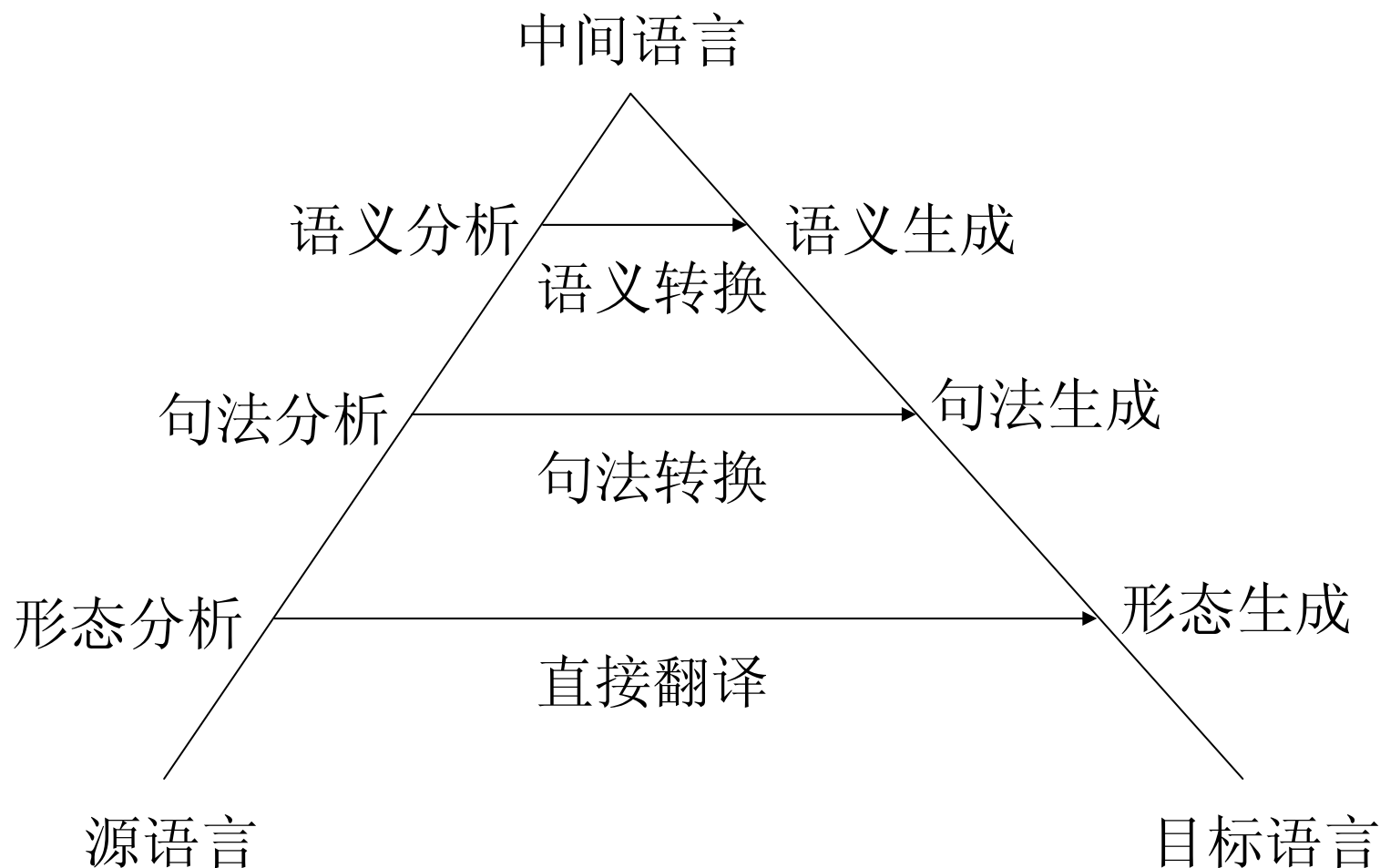
liuqun@ict.ac.cn

中国科学院计算技术研究所2008-2009年度秋季课程

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

按转换层面划分的机器翻译方法



内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

直接翻译方法

- 通过词语翻译、插入、删除和局部的词序调整来实现翻译，不进行深层次的句法和语义的分析，但可以采用一些统计方法对词语和词类序列进行分析
- 早期机器翻译系统常用的方法，近期**IBM**提出的统计机器翻译模型也可以认为是采用了这一范式
- 著名的机器翻译系统**Systran**早期也是采用这种方法，后来逐步引入了一些句法和语义分析

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

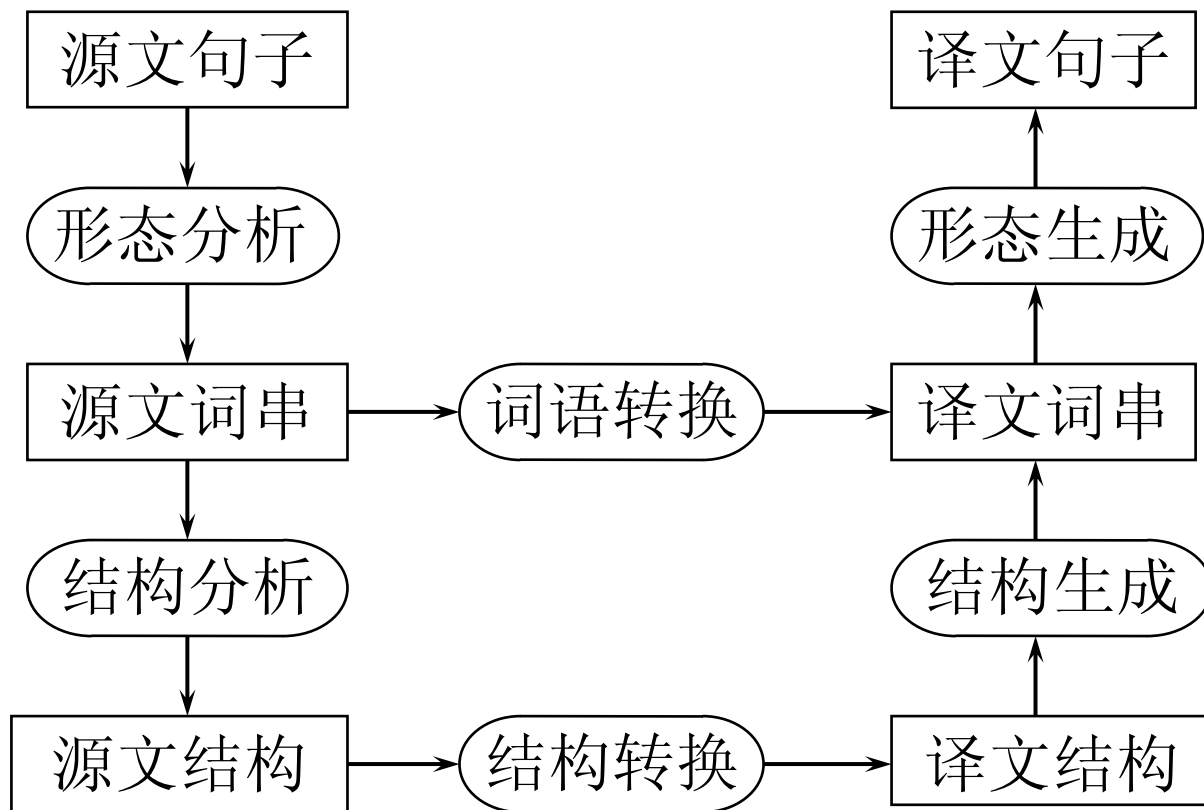
转换方法 (1)

- 整个翻译过程分为“分析”、“转换”、“生成”三个阶段；
- 分析：源语言句子 → 源语言深层结构
 - 相关分析：分析时考虑目标语言的特点
 - 独立分析：分析过程与目标语言无关
- 转换：源语言深层结构 → 目标语言深层结构
- 生成：目标语言深层结构 → 目标语言句子
 - 相关生成：生成时考虑源语言的特点
 - 独立生成：生成过程与源语言无关

转换方法 (2)

- 理想的转换方法应该做到独立分析和独立生成，这样在进行多语言机器翻译的时候可以大大减少分析和生成的工作量；
- 转换方法根据深层结构所处的层面可分为：
 - 句法层转换：深层结构主要是句法信息
 - 语义层转换：深层结构主要是语义信息
- 分析深度的权衡
 - 分析的层次越深，歧义排除就越充分
 - 分析的层次越深，错误率也越高

转换方法 (3)

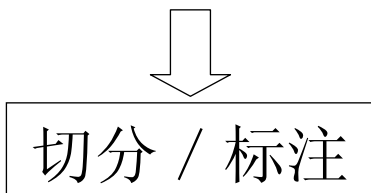


基于转换方法的翻译流程

机器翻译原理与方法讲义(02)机器翻译方法概述

句法层面的转换方法 (1)

她把一束花放在桌上。 \implies She put a bunch of flowers on the table.



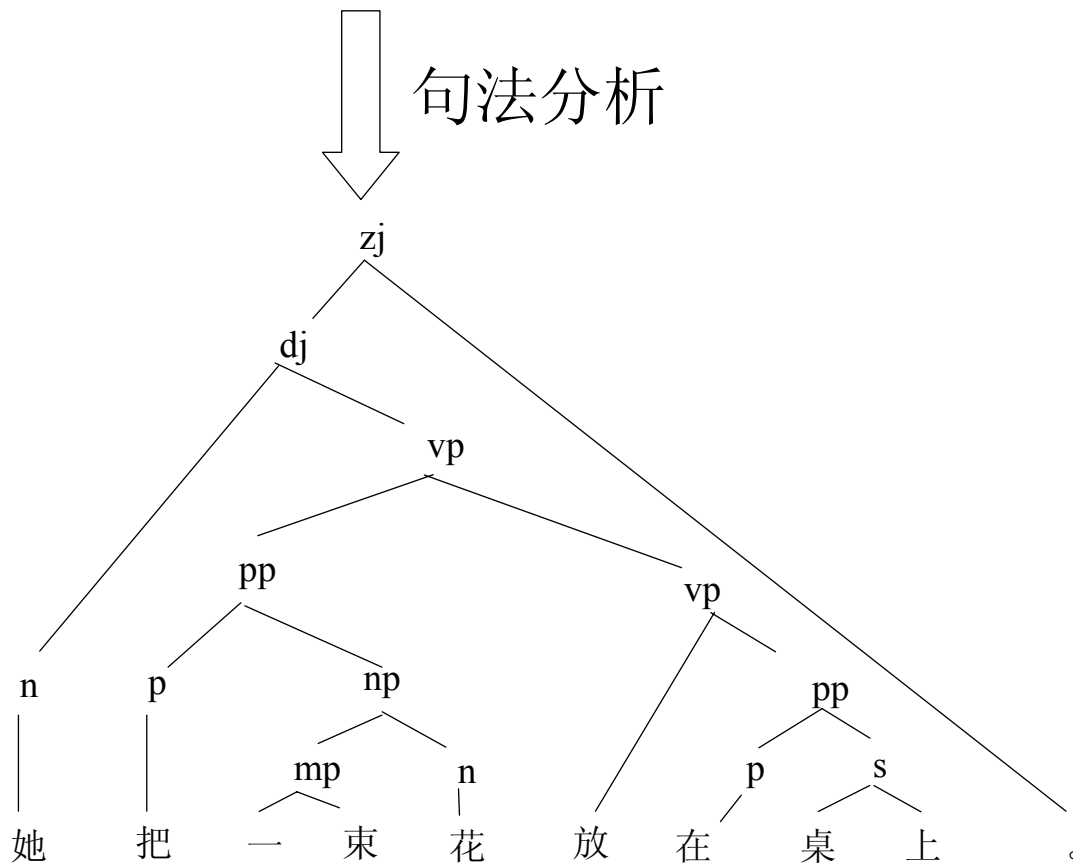
她/r 把/p-q-v-n 一/m-d 束/q 花/n-v-a 放/v 在/p-d-v 桌/n 上/f-v 。 /w



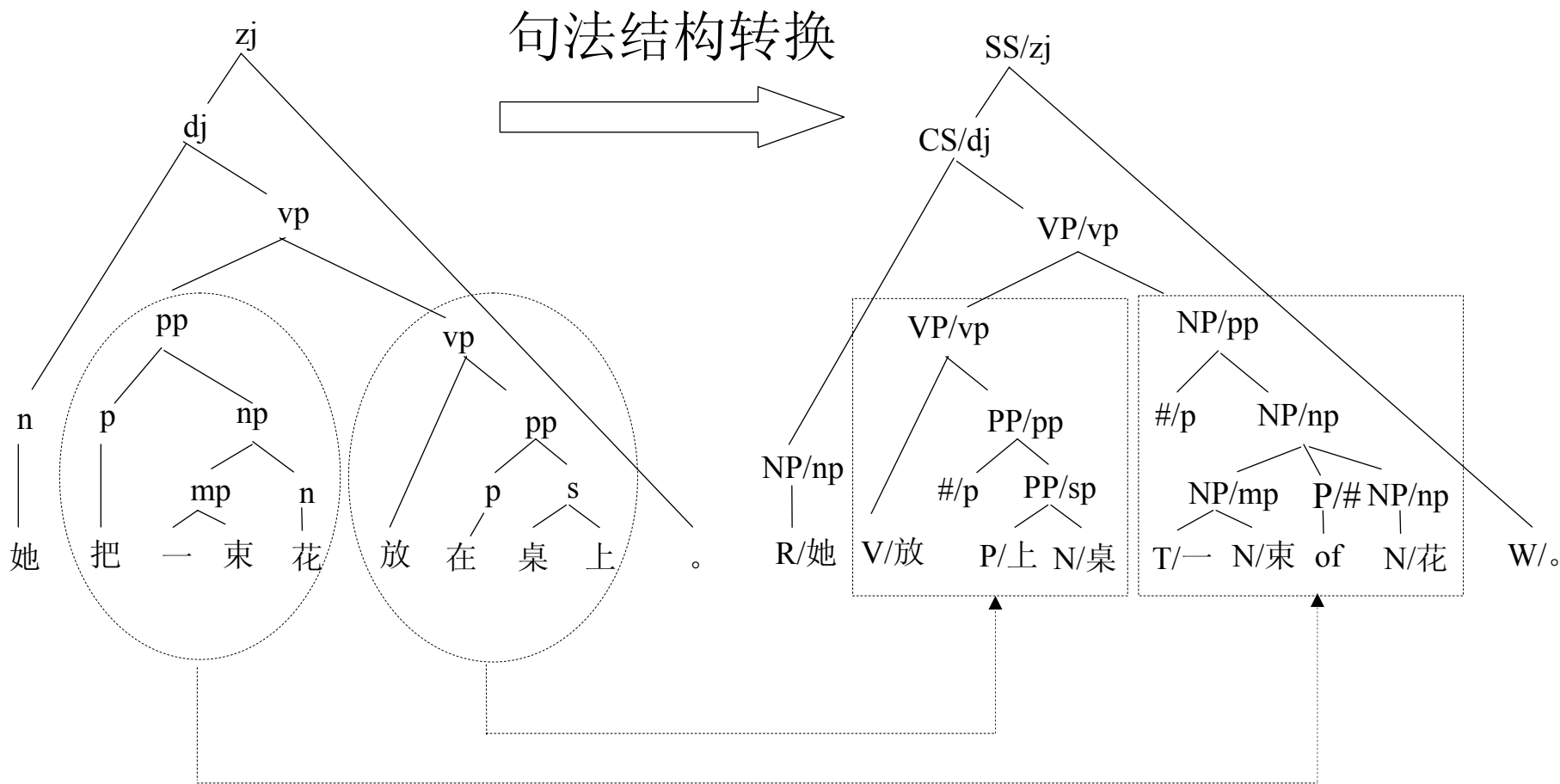
她/r 把/p 一/m-d 束/q 花/n 放/v 在/p-v 桌/n 上/f-v 。 /w

句法层面的转换方法 (2)

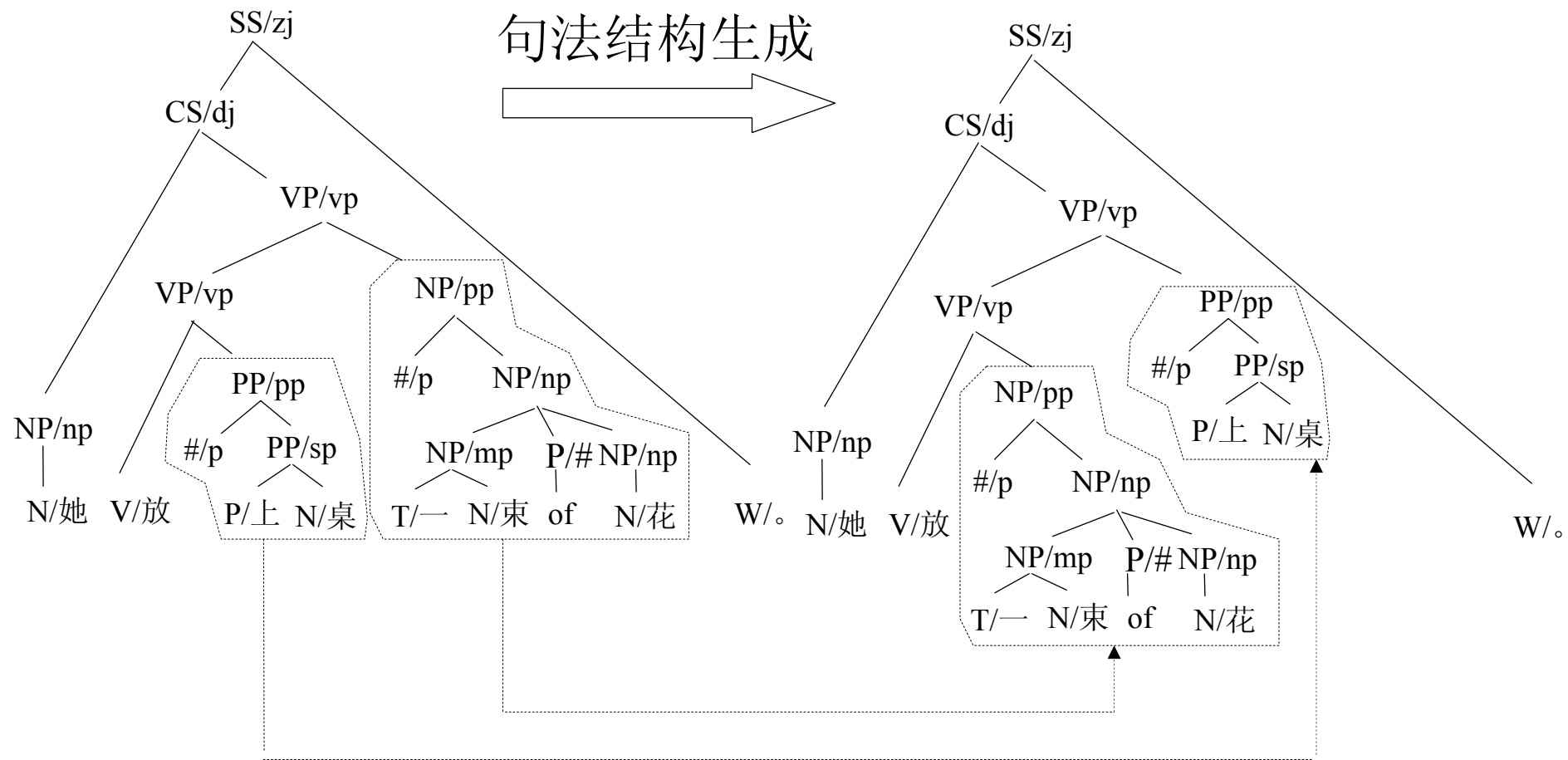
她/r 把/p 一/m-d 束/q 花/n 放/v 在/p-v 桌/n 上/f-v 。 /w



句法层面的转换方法 (3)

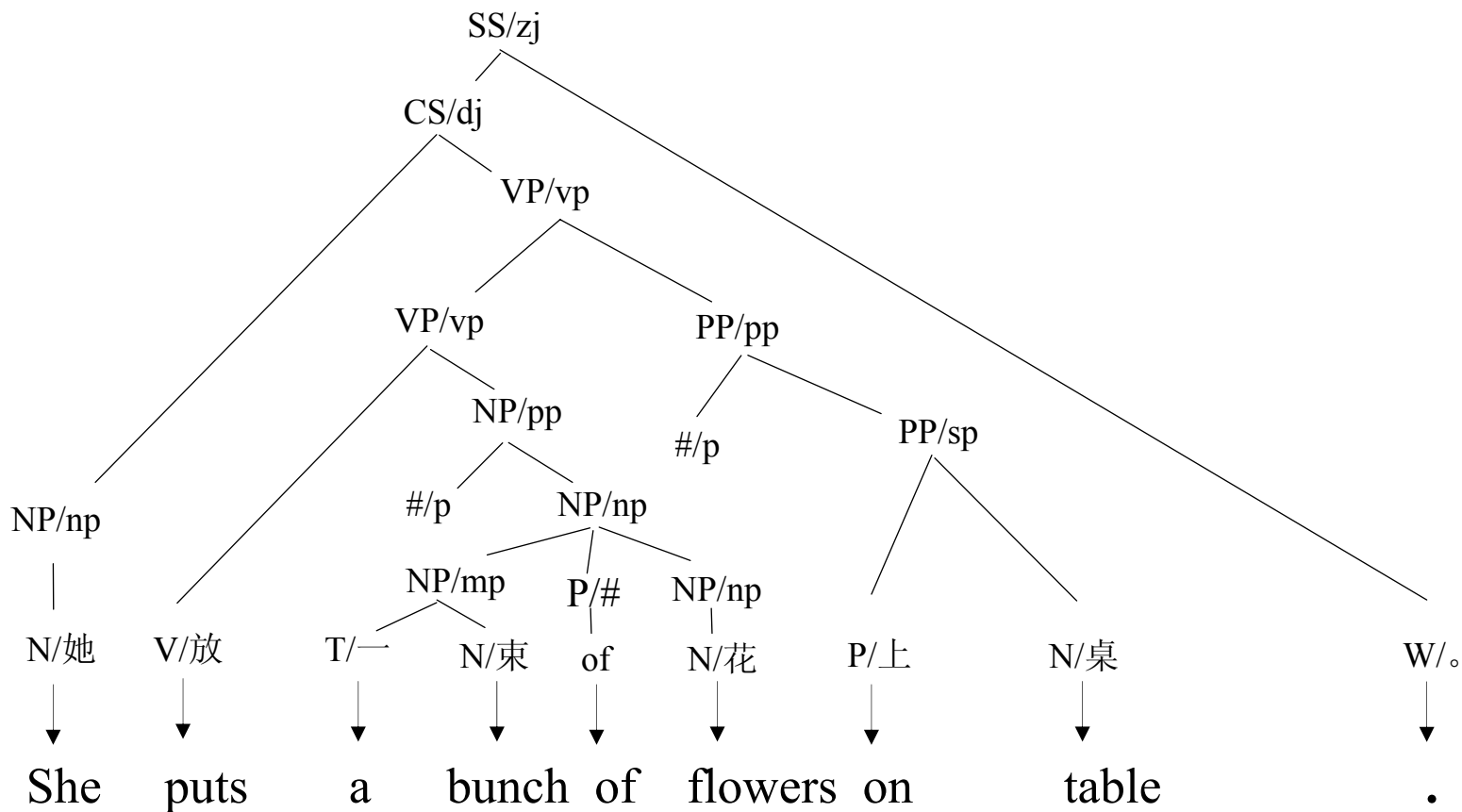


句法层面的转换方法 (4)



句法层面的转换方法 (5)

词语转换
与
词语生成



内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

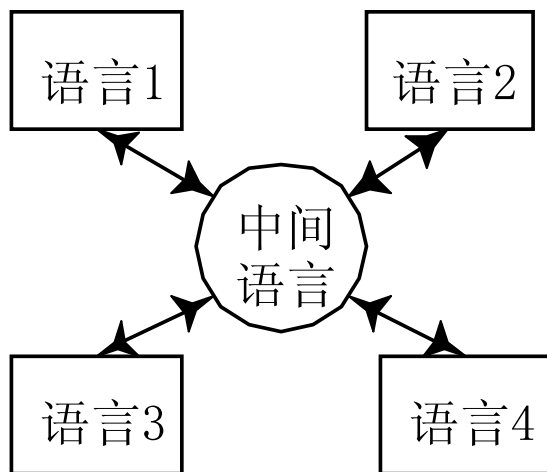
中间语言方法 (1)

- 利用一种中间语言（interlingua）作为翻译的中介表示形式；
- 整个翻译的过程分为“分析”和“生成”两个阶段
- 分析：源语言 → 中间语言
- 生成：中间语言 → 目标语言
- 分析过程只与源语言有关，与目标语言无关
- 生成过程只与目标语言有关，与源语言无关

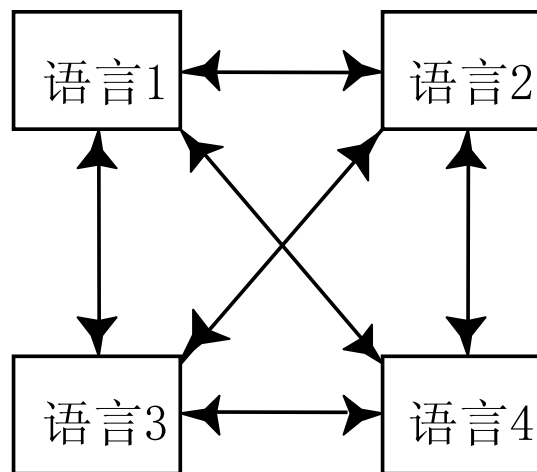
中间语言方法 (2)

- 中间语言方法的优点在于进行多语种翻译的时候，只需要对每种语言分别开发一个分析模块和一个生成模块，模块总数为 $2*n$ ，相比之下，如果采用转换方法就需要对每两种语言之间都开发一个转换模块，模块总数为 $n*(n-1)$

中间语言方法 (3)



中间语言方法



转换方法

中间语言方法 (4)

- 中间语言的类型
 - 自然语言：如英语、汉语
 - 人工语言：如世界语
 - 某种知识表示形式：如语义网络
- 以某种知识表示形式作为中间语言的机器翻译方法有时也称为基于知识的机器翻译方法

中间语言方法 (5)

- Makoto Nagao (Kyoto University) said: “.. when the pivot language [i.e. interlingua] is used, the results of the analytic stage must be in a form which can be utilized by all of the different languages into which translation is to take place. This level of subtlety is a practical impossibility.” (Machine Translation, Oxford, 1989)
- Patel-Schneider (METAL system) said: “METAL employs a modified transfer approach rather than an interlingua. If a meta-language [an interlingua] were to be used for translation purposes, it would need to incorporate all possible features of many languages. That would not only be an endless task but probably a fruitless one as well. Such a system would soon become unmanageable and perhaps collapse under its own weight.” (A four-valued semantics for terminological reasoning, Artificial Intelligence, 38, 1989)

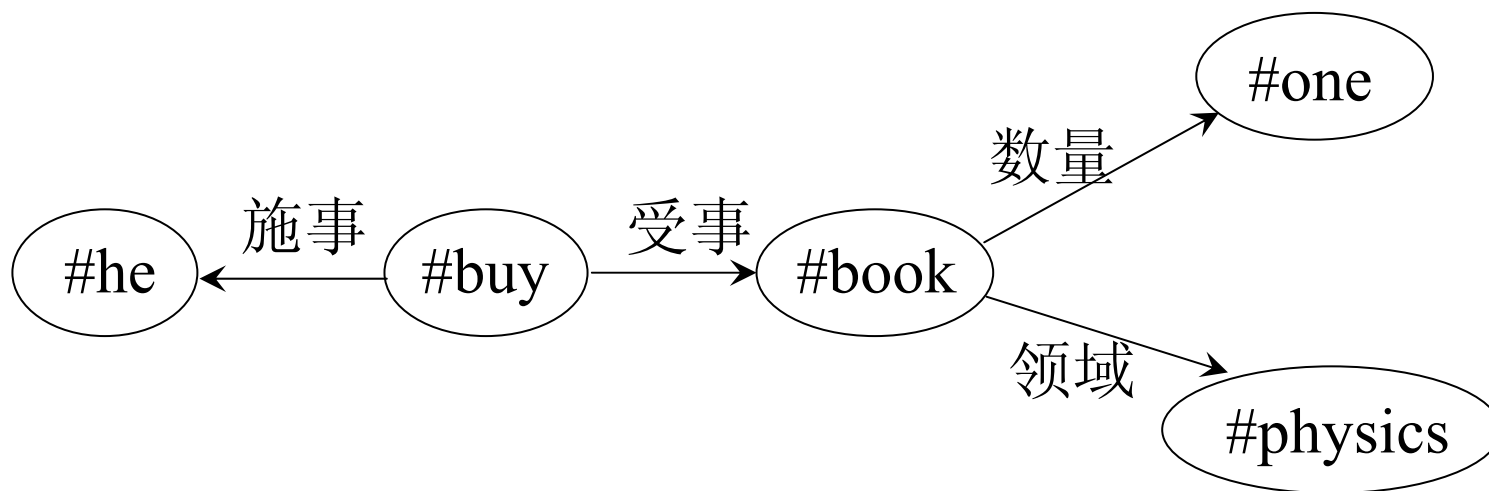
中间语言方法 (6)

- 基于中间语言方法一般都用于多语言的机器翻译系统中；
- 从实践看，基于中间语言的机器翻译系统还没有比较成功的先例，如日本主持的亚洲五国语言机器翻译系统，总体上是失败的；
- 在**CSTAR**多国语音机器翻译系统中，曾经采用了一种中间语言方法，其中间语言是一种语义表示形式，由于语音翻译都限制在非常狭窄的领域中（如机票预定），语义描述可以做到非常精确，因此采用中间语言方法有一定的合理性。

中间语言示例一语义网络

英语：He bought a book on physics.

汉语：他买了一本关于物理学的书。



说明：这里#后面表示的是概念，而不是英语词。

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

按知识表示划分的机器翻译方法

- 基于规则的机器翻译方法
- 基于实例的机器翻译方法
 - 基于翻译记忆的机器翻译方法
 - 基于模板（模式）的机器翻译方法
- 基于统计的机器翻译方法

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

基于规则的方法 (1)

- 采用规则作为知识表示形式
 - 重叠词规则
 - 切分规则
 - 标注规则
 - 句法分析规则
 - 语义分析规则
 - 结构转换规则 (产生译文句法语义结构)
 - 词语转换规则 (译词选择)
 - 结构生成规则 (译文结构调整)
 - 词语生成规则 (译文词形生成)

基于规则的方法 (2)

- 优点
 - 直观，能够直接表达语言学家的知识
 - 规则的颗粒度具有很大的可伸缩性
 - 大颗粒度的规则具有很强的概括能力
 - 小颗粒度的规则具有精细的描述能力
 - 便于处理复杂的结构和进行深层次的理解，如解决长距离依赖问题
 - 系统适应性强，不依赖于具体的训练语料

基于规则的方法 (3)

- 缺点

- 规则主观因素重，有时与客观事实有一定差距
- 规则的覆盖性差，特别是细颗粒度的规则很难总结得比较全面
- 规则之间的冲突没有好的解决办法（翘翘板现象）
- 规则一般只局限于某一个具体的系统，规则库开发成本太高
- 规则库的调试极其枯燥乏味

基于规则的方法—译词选择

\$\$ 开

**{v} v \$=[...]

|| \$.主体=是,\$.主体.语义类=植物

→ V<bloom> \$=[...]

|| \$.客体=是,\$.客体.汉字=灯|机|器

→ V(!V<turn> D<on>) \$=[...]

|| \$.客体=是,\$.客体.语义类=交通工具

=> V<drive> \$=[...]

|| OTHERWISE

=> V<open> \$=[...]

基于规则的方法—结构转换

&& {mp7} mp->r !mp :: \$.内部结构=组合定中,...

|| %mp.定语.内部结构=单词, %mp.定语.yx=一,%mp.量词子类=集体|种类|容量|时量|度量|成形

=> NP(T/r !NP/mp) %T.TNNUM=%NP.NNUM /*这一年*/

|| %mp.定语.内部结构=单词, %mp.定语.yx=一,%mp.量词子类=个体

=> T(T/r M<one>) /*这一个 哪一个*/

|| %r.yx=这|那, IF %mp.定语.内部结构=单词,%mp.定语.yx=一 FALSE

=> NP(T/r !M/mp) %T.TNNUM=PLUR,\$.NNUM=PLUR /*这两张*/

=> NP(T/r !NP/mp) %T.TNNUM=PLUR,\$.NNUM=PLUR

|| %r.yx=~这~那,IF %mp.定语.内部结构=单词,%mp.定语.yx=一 FALSE

=> NP(T/r !M/mp) \$.NNUM=%M.NNUM

=> NP(T/r !NP/mp) %T.TNSUB=%NP.NSUBC,...

基于规则的方法—结构生成

{ NPMP1 } NP(T !NP(T !N))

=> NP(T/T !NP/NP(!N/N))

/* this a kind => this kind */

{ NPATN1 } NP(AP(!A) !NP(T !N))

=> P(T/T !NP/NP(AP/AP(!A/A) !N/N))

/* red this book => this red book */

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

基于语料库的机器翻译方法

- 机器翻译的实例方法和统计方法都是基于语料库的机器翻译方法
- 优点
 - 使用语料库作为翻译知识来源，无需人工编写规则，系统开发成本低，速度快
 - 从语料库中学习到的知识比较客观
 - 从语料库中学习到的知识覆盖性比较好
- 缺点
 - 系统性能依赖于语料库
 - 数据稀疏问题严重
 - 语料库中不容易获得大颗粒度的高概括性知识

基于实例的机器翻译 (1)

- 长尾真(Makoto Nagao)在1984年发表了《采用类比原则进行日-英机器翻译的一个框架》一文，探讨日本人初学英语时翻译句子的基本过程，长尾真认为，初学英语的日本人总是记住一些最基本的英语句子以及一些相对应的日语句子，他们要对比不同的英语句子和相对应的日语句子，并由此推论出句子的结构。参照这个学习过程，在机器翻译中，如果我们给出一些英语句子的实例以及相对应的日语句子，机器翻译系统来识别和比较这些实例及其译文的相似之处和相差之处，从而挑选出正确的译文。
- 长尾真指出，人类并不通过做深层的语言学分析来进行翻译，人类的翻译过程是：首先把输入的句子正确地分解为一些短语碎片，接着把这些短语碎片翻译成其它语言的短语碎片，最后再把这些短语碎片构成完整的句子，每个短语碎片的翻译是通过类比的原则来实现的。
- 因此，我们应该在计算机中存储一些实例，并建立由给定的句子找寻类似例句的机制，这是一种由实例引导推理的机器翻译方法，也就是基于实例的机器翻译。

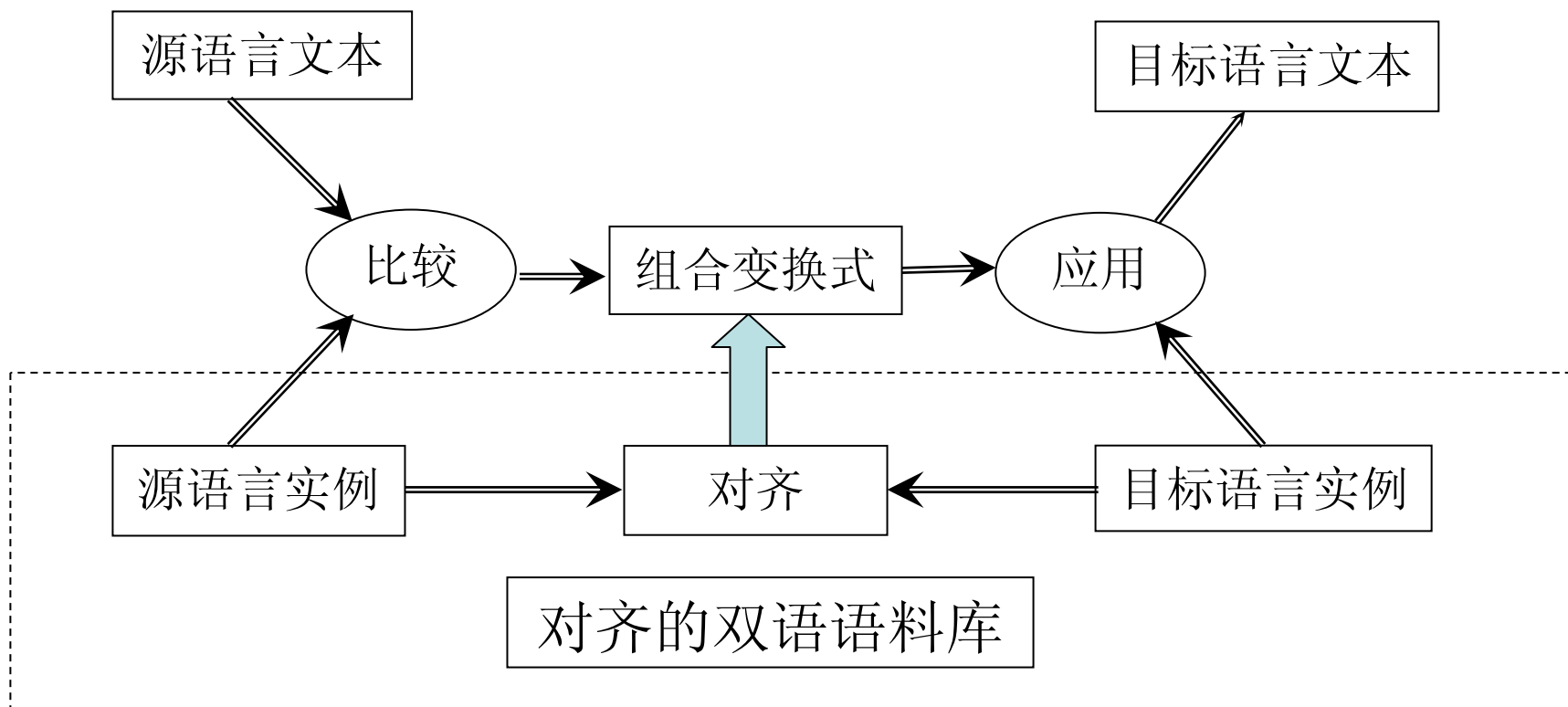
基于实例的机器翻译 (2)

- 在基于实例的机器翻译系统中，系统的主要知识源是双语对照的翻译实例库，实例库主要有两个字段，一个字段保存源语言句子，另一个字段保存与之对应的译文，每输入一个源语言的句子时，系统把这个句子同实例库中的源语言句子字段进行比较，找出与这个句子最为相似的句子，并模拟与这个句子相对应的译文，最后输出译文。
- 基于实例的机器翻译系统中，翻译知识以实例和义类词典的形式来表示，易于增加或删除，系统的维护简单易行，如果利用了较大的翻译实例库并进行精确的对比，有可能产生高质量译文，而且避免了基于规则的那些传统的机器翻译方法必须进行深层语言学分析的难点。在翻译策略上是很有吸引力的。

基于实例的机器翻译 (3)

- 优点
 - 直接使用对齐的语料库作为知识表示形式，知识库的扩充非常简单
 - 不需要进行深层次的语言分析，也可以产生高质量的译文
- 缺点
 - 覆盖率低，实用的系统需要的语料库规模极大（百万句对以上）

基于实例的机器翻译系统结构



基于实例的机器翻译一举例

要翻译句子:

(E1) He bought a book on physics.

在语料库中查到相似英语句子及其汉语译文是:

(E2) He wrote a book on history.

(C2) 他写了一本关于历史的书。

比较(E1)和(E2)两个句子, 我们得到变换式:

(T1) replace(wrote, bought) and replace(history, physics)

将这个变换式中的单词都换成汉语就变成:

(T2) replace(写,买) and replace(历史,物理)

将(T2)作用于(C2)

(C1)他买了一本关于物理学的书。

基于实例的机器翻译 需要研究的问题

- 正确地进行双语自动对齐(**alignment**): 在实例库中要能准确地由源语言例句找到相应的目标语言例句, 在基于实例的机器翻译系统的具体实现中, 不仅要求句子一级的对齐, 而且还要求词汇一级甚至短语一级的对齐。
- 建立有效的实例匹配检索机制: 很多研究者认为, 基于实例的机器翻译的潜力在于充分利用短语一级的实例碎片, 也就是在短语一级进行对齐, 但是, 利用的实例碎片越小, 碎片的边界越难于确定, 歧义情况越多, 从而导致翻译质量的下降, 为此, 要建立一套相似度准则(**similarity metric**), 以便确定两个句子或者短语碎片是否相似。
- 根据检索到的实例生成与源语言句子相对应的译文: 由于基于实例的机器翻译对源语言的分析比较粗, 生成译文时往往缺乏必要的信息, 为了提高译文生成的质量, 可以考虑把基于实例的机器翻译与传统的基于规则的机器翻译方法结合起来, 对源语言也进行一定深度的分析。
- 开展浅层句法分析(**shallow parsing**)的研究: 浅层句法分析以建立语段(**chunk**)之间的依附关系为目标, 进行语段的识别, 分析语段之间的依附关系。由于分析的语言单位的颗粒度比较大, 歧义就比较少, 有利于提高双语对齐的准确度。

实例库的匹配 (1)

- 实例匹配的目的在于将输入句子分解成语料库中实例片断的组合，这是基于实例的机器翻译的关键问题之一，实例匹配的各种方法有很大的差异，还没有那种做法显示出明显的优势；
- 实例库匹配的效率问题：由于实例库规模较大，通常需要建立倒排索引；
- 实例库匹配的其他问题：
 - 实例片断的分解：
 - 实例片断的组合：

实例库的匹配 (2)

- 实例片断的分解
 - 实例库中的句子往往太长，直接匹配成功率太低，为了提高实例的重用性，需要将实例库中的句子分解为片断
 - 几种通常的做法：
 - 按标点符号分解
 - 任意分解
 - 通过组块分析进行分解

实例库的匹配 (3)

- 实例片断的组合
 - 一个被翻译的句子，往往可以通过各种不同的实例片断进行组合，如何选择一个最好的组合？
 - 简单的做法：
 - 最大匹配
 - 最大概率法：选择概率乘积最大的片断组合
 - 有点像汉语词语切分问题

片断译文的选择

- 由于语料库中一个片断可能有多种翻译方法，因此存在片断译文的选择问题；
- 常用的方法：
 - 根据片断上下文进行排歧；
 - 根据译文的语言模型选择概率最大的译文片断组合

基于实例的机器翻译系统

- MBT1和MBT2系统：由日本京都大学长尾真和佐藤研制。该系统的翻译过程分为分解(decomposition)、转换(transfer)、合成(composition)三步。在分解阶段，系统根据提交的源语言词汇依存树检索实例库，并利用检索到的实例碎片来表示该源语言句子的依存树，形成源匹配表达式；在转换阶段，系统利用实例库中的对齐信息将源匹配表达式转换成目标匹配表达式；在合成阶段，将目标匹配表达式展开成为目标语言词汇依存树，输出译文。
- PANGLOSS系统：由美国卡内基-梅隆大学研制，这是一个多引擎机器翻译系统(Multi-engine Machine Translation)。这个系统的主要引擎是基于知识的机器翻译系统，基于实例的机器翻译系统只是它的一个引擎，为整个多引擎机器系统提供候选结果。
- ETOC和EBMT系统：由日本口语翻译通信研究实验室 ATR研制。ETOC系统能够检索出与给定的源语言句子相似的实例，EBMT系统能够利用实例库来消解歧义，这两个基于实例的机器翻译系统还不完整。
- 我国清华大学计算机系的基于实例的日汉机器翻译系统。

翻译记忆方法 (1)

- 翻译记忆方法（Translation Memory）是基于实例方法的特例；
- 也可以把基于实例的方法理解为广义的翻译记忆方法；
- 翻译记忆的基本思想：
 - 把已经翻译过的句子保存起来
 - 翻译新句子时，直接到语料库中去查找
 - 如果发现相同的句子，直接输出译文
 - 否则交给人去翻译，但可以提供相似的句子的参考译文

翻译记忆方法 (2)

- 翻译记忆方法主要被应用于计算机辅助翻译（CAT）软件中
- 翻译记忆方法的优缺点
 - 翻译质量有保证
 - 随着使用时间的增加匹配成功率逐步提高
 - 特别适用于重复率高的文本翻译，例如公司的产品说明书的新版本翻译
 - 与语言无关，适用于各种语言对
 - 缺点是匹配成功率不高，特别是刚开始使用时

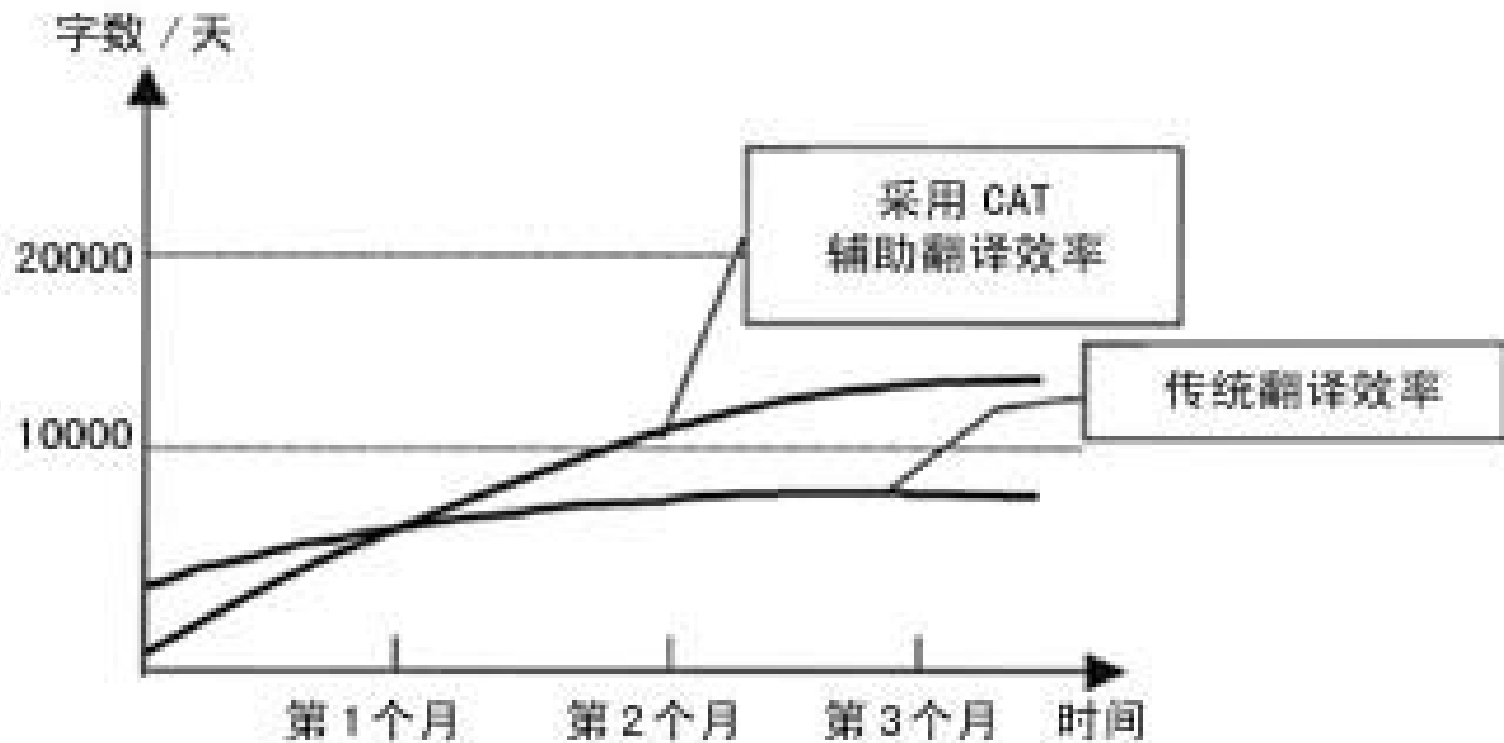
翻译记忆方法 (3)

- 计算机辅助翻译（CAT）软件已经形成了比较成熟的产业
 - TRADOS
 - 号称占有国际CAT市场的70%
 - Microsoft、Siemens、SAP等国际大公司和一些著名的国际组织都是其用户
 - 雅信CAT
 - 适合中国人的习惯
 - 产品已比较成熟
 - 国际组织：LISA（Localisation Industry Standards Association）
- 面向用户：专业翻译人员
- 数据交换：LISA制定了TMX（Translation Memory eXchange）标准。

翻译记忆方法 (4)

- 完整的计算机辅助翻译软件除了包括翻译记忆功能以外，还应该包括以下功能
 - 多种文件格式的分解与合成
 - 术语库管理功能
 - 语料库的句子对齐（历史资料的重复利用）
 - 项目管理：
 - 翻译任务的分解与合并
 - 翻译工作量的估计
 - 数据共享和数据交换

翻译记忆方法 (5)



基于模板(模式)的机器翻译方法(1)

- 基于模板（**Template**）或者模式（**Pattern**）的机器翻译方法通常也被看做基于实例的机器翻译方法的一种延伸
- 所谓“翻译模板”或者“翻译模式”可以认为是一种颗粒度介于“翻译规则”和“翻译实例”之间的翻译知识表示形式
 - 翻译规则：颗粒度大，匹配可能性大，但过于抽象，容易出错
 - 翻译实例：颗粒度小，不易出错，但过于具体，匹配可能性小
 - 翻译模板（模式）：介于二者之间，是一种比较合适的知识表示形式
- 一般而言，单语模板（或模式）是一个常量和变量组成的字符串，翻译模板（或模式）是两个对应的单语模板（或模式），两个模板之间的变量存在意义对应关系

基于模板(模式)的机器翻译方法(2)

- 模板举例：
 - 这个 X 比 Y 更 Z。
 - The X is more Z than Y.
- 模板方法的主要问题
 - 对模板中变量的约束
 - 模板抽取
 - 模板的冲突消解

模板的自动提取

- 利用一对实例进行泛化
 - Jaime G. Carbonell, Ralf D. Brown,
Generalized Example-Based Machine Translation
<http://www.lti.cs.cmu.edu/Research/GEBMT/>
- 利用两对实例进行比较
 - H. Altay Guvenir, Ilyas Cicekli, Learning Translation Templates from Examples
Information Systems, 1998
 - 张健, 基于实例的机器翻译的泛化方法研究, 中科院计算所硕士论文, 2001

通过泛化实例得到翻译模板

- 已有实例：
 - Karl Marx was born in Trier, Germany in May 5, 1818.
 - 卡尔·马克思于1818年5月5日出生在德国特里尔城。
- 泛化：
 - <Person> was born in <City> in <Date>
 - <Person>于<Date>出生在<City>
- 对齐
 - <Person> ⇔ <Person>
 - <City> ⇔ <City>
 - <Date> ⇔ <City>

通过比较实例得到翻译模板

- 已有两对翻译实例：
 - 我给玛丽一支笔 \Leftrightarrow I gave Mary a pen.
 - 我给汤姆一本书 \Leftrightarrow I gave Tom a book.
- 双侧单语句子分别比较，得到：
 - 我给 #X 一 #Y #Z \Leftrightarrow I give #W a #U.
- 查找变量的对应关系：
 - #X \Leftrightarrow #W
 - #Y \Leftrightarrow ϕ
 - #Z \Leftrightarrow #U

内容提要

- 机器翻译方法（按转换层面划分）
 - 直接翻译方法
 - 句法转换方法
 - 语义转换方法
 - 中间语言方法
- 机器翻译方法（按知识表示形式划分）
 - 基于规则的方法
 - 基于实例的方法（含模板方法、翻译记忆方法）
 - 统计方法

统计机器翻译

- 统计机器翻译概述
- 历史回顾：一些重要事件回放
- 一种新的研究范式
- 统计机器翻译论文发表数量的增长
- 近年来国际机器翻译评测的最好成绩
- 统计机器翻译目前的水平
- 统计机器翻译的优缺点

统计机器翻译概述

- 统计机器翻译也是基于语料库的机器翻译方法，不需要人工撰写规则，而是从语料库中获取翻译知识，这一点与基于实例的方法相同
- 为翻译建立统计模型，把翻译理解为搜索问题，即从所有可能的译文中选择概率最大的译文。基于实例的机器翻译无需建立统计模型
- 与基于实例的方法的区别在于，基于实例的机器翻译中，语言知识表现为实例本身，而统计机器翻译中，翻译知识表现为模型参数

历史回顾：一些重要事件回放 (1)

- 1990年代初IBM首次开展统计机器翻译研究
- 1999年JHU夏季研讨班重复了IBM的工作并推出了开放源代码的工具
- 2001年IBM提出了机器翻译自动评测方法BLEU
- 2002年NIST开始举行每年一度的机器翻译评测
- 2002年第一个采用统计机器翻译方法的商业公司Language Weaver成立
- 2002年Franz Josef Och提出统计机器翻译的对数线性模型

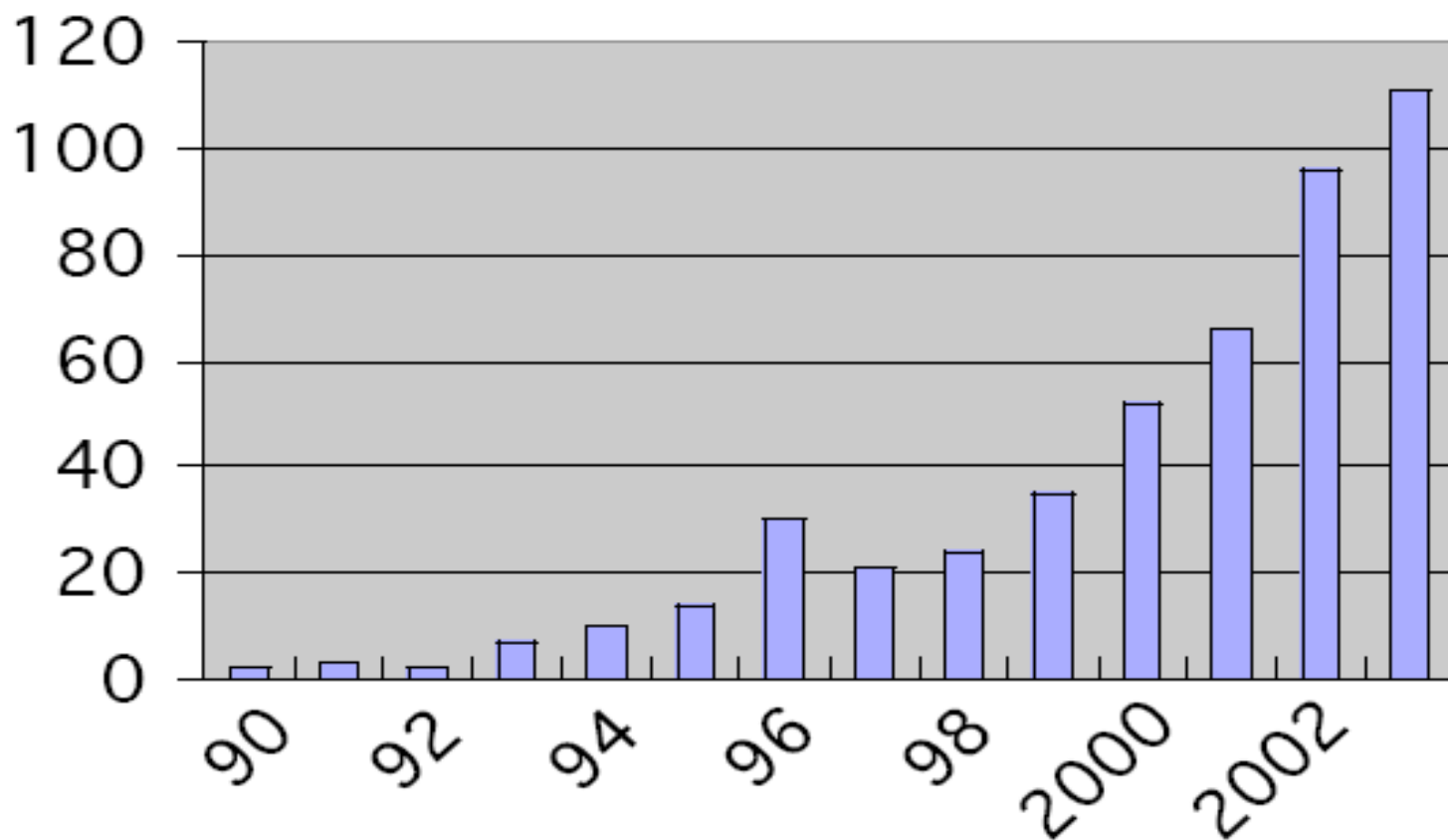
历史回顾：一些重要事件回放 (2)

- 2003年Franz Josef Och提出对数线性模型的最小错误率训练方法
- 2004年Philipp Koehn推出Pharaoh（法老）标志着基于短语的统计翻译方法趋于成熟
- 2005年David Chiang提出层次短语模型并代表UMD在NIST评测中取得好成绩
- 2005年Google在NIST评测中大获全胜，随后Google推出基于统计方法的在线翻译工具，其阿拉伯语-英语的翻译达到了用户完全可接受的水平
- 2006年NIST评测中USC-ISI的树到串句法模型第一次超过Google（仅在汉英受限翻译项目中）
- 2007年Google推出采用统计机器翻译技术的跨语言检索网站

统计机器翻译：一种新的研究范式

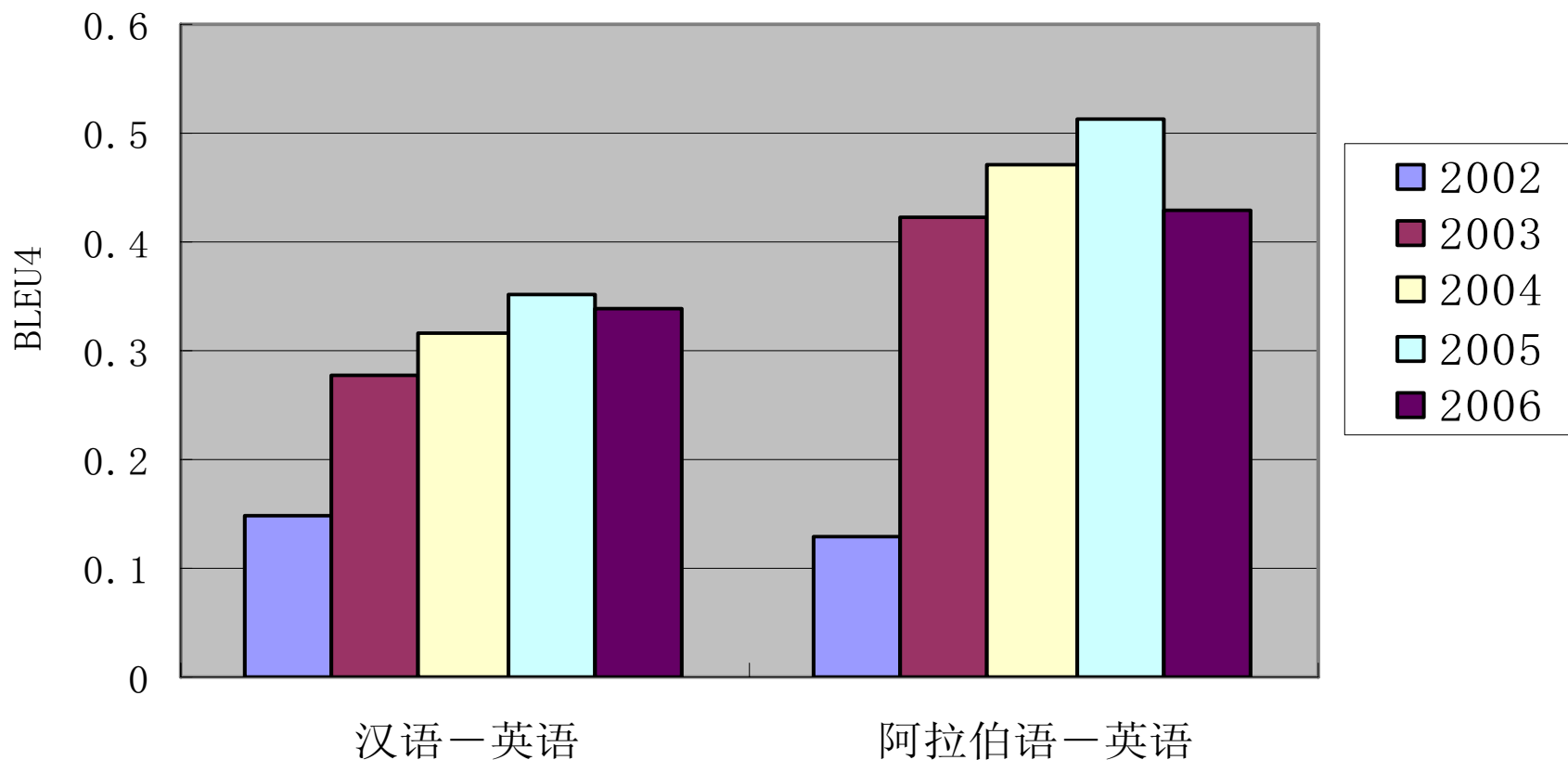
- 统计机器翻译的成功在于采用了一种新的研究范式（paradigm）
- 这种研究范式已在语音识别等领域中被证明是一种成功的翻译，但在机器翻译中是首次使用
- 这种范式的特点：
 - 公开的大规模的训练数据
 - 周期性的公开评测和研讨
 - 开放源码的工具

近年来统计机器翻译论文发表数量

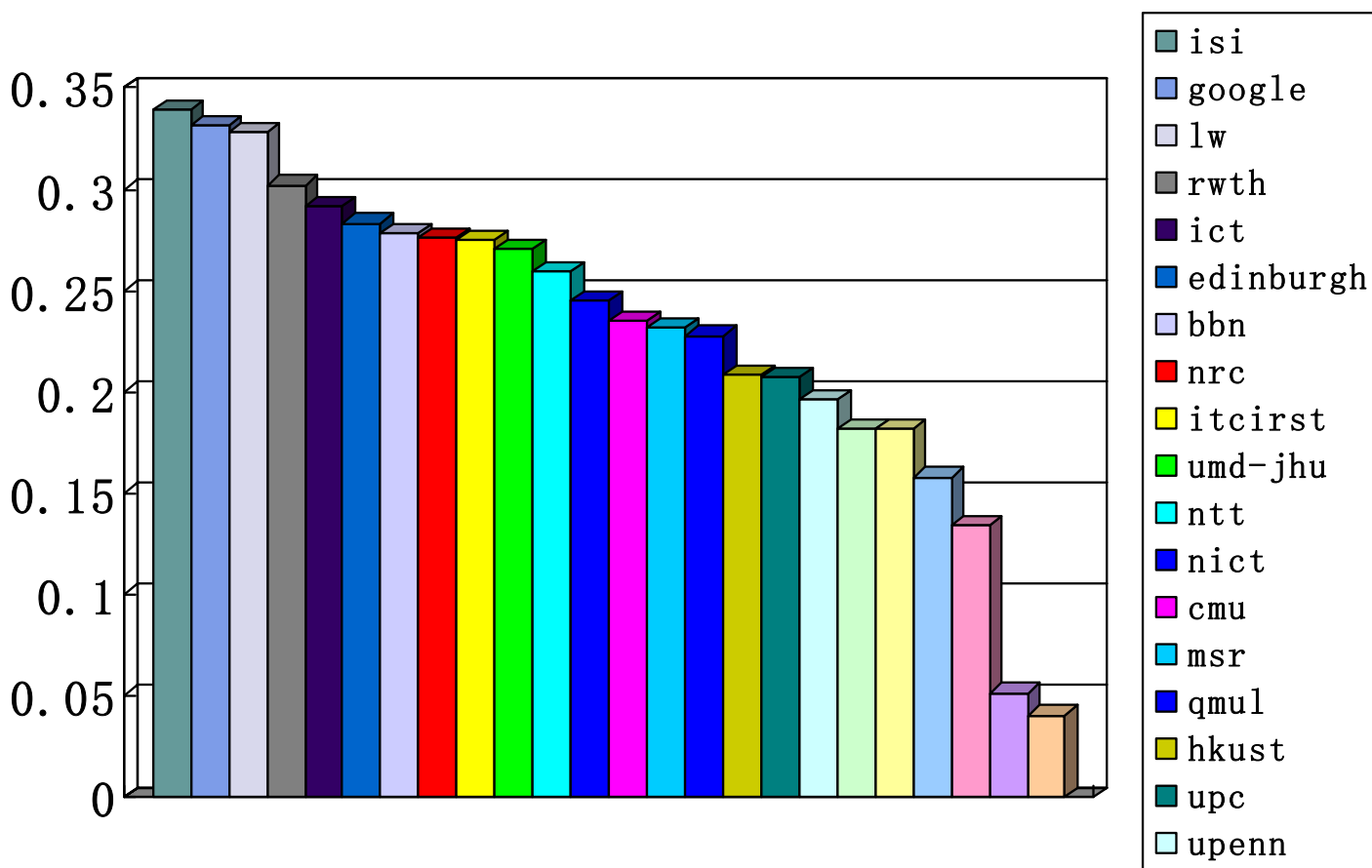


引自 Franz Josef Och, Statistical Machine Translation: Foundations and Recent Advances, Tutorials on MT Summit X, September 13-15, 2005, Phuket, Thailand
机器翻译原理与方法讲义(02)机器翻译方法概述

近年来国际NIST评测最好成绩



Results on NIST 2006 Evaluation: Large Data Track, NIST Subset



统计机器翻译目前的水平

- 以**Google Translator**为例，实地考察一下统计机器翻译的水平
 - 阿拉伯语—英语
 - 汉语—英语
 - 英语—汉语

Google Translator 阿拉伯语-英语

半岛电视台
网站新闻

阿拉伯语原文

الجمعة 1428/9/30 هـ - الموافق 2007/10/12 م (آخر تحديث) الساعة 7:14 (مكة المكرمة)، 4:14 (غرينتش)

الصفحة الرئيسية: دولي

متملقات

متعلقات من الأرشيف

- آلاف يحتشدون بنيويورك للاستماع للدلاي لاما
- مسؤولون أميركيون كبار يلتقون الدلاي لاما
- يكن تحتج على زيارة الدلاي لاما لواشنطن
- الصين تحتج على زيارة الدلاي لاما للبريتان
- يكن تنتقد استضافة واشنطن للرئيس التايواني والدلاي لاما

أهم أخبار الصفحة الرئيسية

- الخرطوم تتهم جنوبيين بالتآمر وسلفاكير يدعو لتدخل دولي
- ملايين المسلمين يؤنون شعائر عيد الفطر اليوم
- العنف يتواصل والمارينز يضغطون لترك العراق
- عباس يلتقي ولش ويشترط تسلم غزة لحوار حماس
- أردوغان يهتد بإجراءات إضافية ضد واشنطن بشأن إبادة الأرمن

بوش يستقبل الدلاي لاما قبيل توشيح الكونغرس له



الدلاي لاما

يستقبل الرئيس الأميركي جورج بوش في البيت الأبيض الأسبوع المقبل الزعيم الروحي للبوذيين في التبت الدلاي لاما في خطوة من المرجح أن تزعج الصين.

وسيلتقي بوش الدلاي لاما، في إطار خاص بجيدا عن وسائل الإعلام، كما قال المتحدث باسم البيت الأبيض غوردون جوندرو، على غرار ما فعل في السابق.

وسيحضر بوش في اليوم التالي في واشنطن حفلا رسميا يقلد خلاله الكونغرس الدلاي لاما ميدالية الكونغرس الذهبية، وهي أعلى وسام يمكن للكونغرس أن يمنحه.

وحفل منح الوسام سيكون المرة الأولى التي يظهر فيها بوش علانية مع الدلاي لاما الذي سبق له أن زار البيت الأبيض لكن دائما في اجتماعات غير رسمية.

وردت الصين بغضب عندما قرر الكونغرس الأميركي منح الدلاي لاما الوسام وتسجبت القرار قائلة إنه تتخل في شؤونها الداخلية.

وتعتبر الصين الدلاي لاما -الذي فر من التبت عام 1959 بعد انتفاضة فاشلة على السلطات الصينية- انفصاليا.

وتتهم الصين "بعض البلدان أو الأشخاص" باستغلال الدلاي لاما كما قال المتحدث باسم وزارة الخارجية الصينية ليو جيانشاو، قبل الإعلان عن اللقاء بين بوش والدلاي لاما.

وتؤكد الصين أنها حررت لتثبيت من الظلم الإقطاعي لدى سيطرتها عليها عام 1949، قبل أن تقم فيها منظمة تتمتع بالحكم الذاتي في 1965.

المصدر: وكالات

Google Translator 阿拉伯语-英语

半岛电视台
网站新闻

Google 的
英文译文

(Friday 30/9/1428 12/10/2007 e -approved m (Updated) at 7:14 (Mecca), 4:14 (GMT)

متعلقات International:Home

متعلقات من الأرشيف

- Thousands gathered in New York to listen to Dalai Lama
- Senior American officials meet Dalai Lama
- Beijing protests the visit of the Dalai Lama to Washington
- China protests Dalai Lama's visit to Portugal
- Beijing to host criticize Washington for Taiwanese President and the Dalai Lama

أهم أخبار الصفحة الرئيسية

- Khartoum accuses of conspiring southerners and Sgakhir calls for international intervention
- Millions of Muslims perform rites Eid Al-Fitr today

Bush received the Dalai Lama as a Touchih Congress



Receives American President George Bush at the White House next week, the spiritual leader of Buddhists in Tibet the Dalai Lama in a step likely to upset China

And Bush will meet with the Dalai Lama, in the framework of special away from the media, as White House spokesman Gordon Jondro, as it did in the past

The Dalai Lama

Bush will attend the next day in Washington, a ceremony which officially imitate Congress Dalai Lama Congressional Gold Medal, the highest decoration that can be granted by Congress

The granting of the medal ceremony will be the first time that Bush is displayed publicly with the Dalai Lama who has previously visited the White House, but always in informal meetings

China replied angrily when the American Congress decided to grant the Dalai Lama medal and denounced the resolution, saying that interference in their internal affairs

China considers the Dalai Lama - who fled from Tibet in 1959 after a failed uprising on the Chinese authorities separatist

China accuses "certain countries or persons" to exploit the Dalai Lama as the spokesman of the Chinese Foreign Ministry Liu Jianchao, before the announcement of the meeting between Bush and the Dalai Lama

The China Tibet liberated it from feudal injustice to the control by the year 1949, prior to evaluate the autonomous regions in 1965

Agencies :Source

Google Translator 阿拉伯语-英语

- الروحي الزعيم المقبل الأسبوع الأبيض البيت في بوش جورج الأميركي الرئيس يستقبل الصينين تزعم أن المرجح من خطوة في لاما الدلاي التبت في اللبوذيين.
Receives American President George Bush at the White House next week, the spiritual leader of Buddhists in Tibet the Dalai Lama in a step likely to upset China. 😊
- البيت باسم المتحدث قال كما، الإعلام وسائل عن بعيدا خاص إطار في، لاما الدلاي بوش وسيلتقي السابق في فعل ما غرار على، جوندرو غوردون الأبيض.
And Bush will meet with the Dalai Lama, in the framework of special away from the media, as White House spokesman Gordon Jondro, as it did in the past. 😊
- ميدالية لاما الدلاي الكونغرس خلاله يقد رسميا حفلا واشنطن في التالي اليوم في بوش وسيحضر يمنحه أن للكونغرس يمكن وسام أعلى وهي، الذهبية الكونغرس.
Bush will attend the next day in Washington, a ceremony which officially imitate Congress Dalai Lama Congressional Gold Medal, the highest decoration that can be granted by Congress. 😊
- زار أن له سبق الذي لاما الدلاي مع علانية بوش فيها يظهر التي الأولى المرة سيكون الوسام منح وحفل رسمية غير اجتماعات في دائما لكن الأبيض البيت.
The granting of the medal ceremony will be the first time that Bush is displayed publicly with the Dalai Lama who has previously visited the White House, but always in informal meetings. 😊
- إنه قائلة القرار وشجبت الوسام لاما الدلاي منح الأميركي الكونغرس قرر عندما بغضب الصين وردت الداخلية شؤونها في تدخل.
China replied angrily when the American Congress decided to grant the Dalai Lama medal and denounced the resolution, saying that interference in their internal affairs. 😊

Google Translator 汉语-英语

新浪新闻

中文原文

土耳其抗议美国会有有关亚美尼亚大屠杀议案

<http://www.sina.com.cn> 2007年10月12日01:23 新京报



10月11日，土耳其，伊斯坦布尔，土耳其反对者举着旗帜和标语反对美“亚美尼亚”大屠杀议案。

[点击观看本新闻视频](#)

据亚美尼亚方面的史料记载，1915年至1923年期间，土耳其奥斯曼帝国对其统治的亚美尼亚人实施种族灭绝，导致150万亚美尼亚人死亡。

土耳其历届政府均对此予以否认，认为这是奥斯曼帝国崩溃过程中出现的非正常死亡。土耳其认为，那些人死于当时的内战和社会动荡，而且这一数字被夸大了。

美国国会众议院外交事务委员会定于10日表决通过一项关于“亚美尼亚大屠杀”的议案，并准备提交众院全体会议表决。

由于这项议案可能损害美国与重要盟友土耳其的关系，总统乔治·W·布什当天呼吁众院拒绝表决。土耳其已就这项议案向白宫提出抗议。

Google Translator 汉语-英语

新浪新闻


Google 的
英文译文

Turkey, the United States will protest the Armenian massacre motion

[Http://www.sina.com.cn](http://www.sina.com.cn) 2007, 12 October 01:23 Xin Jing Bao



October 11, Turkey, Istanbul, Turkey, opponents holding banners and placards opposed to the "Armenian" massacre motion.

 [Click on the video to watch the news](#)

According to the Armenian side of historical records, in 1915 to 1923, [its rule of the Turkish Ottoman Empire implementation of the Armenian genocide](#), leading to the death of 1.5 million Armenians.

Turkey, successive governments have denied this, believing that this is the collapse of the Ottoman Empire appeared in the process of unnatural deaths. Turkey believes that those who died at that time of social unrest and civil war, but that figure has been exaggerated.

The U.S. House of Representatives Foreign Affairs Committee is scheduled to vote on the adoption of a on the 10th on the "Armenian Massacre," the motion and to be submitted to the House plenary vote.

As a result of this motion could damage the United States and Turkey, an important ally of the relationship between President George W. Bush appealed to the House of Representatives refused to vote on the same day. Turkey has been on the motion to the White House protest.

Google Translator 汉语-英语

- 土耳其历届政府均对此予以否认，认为这是奥斯曼帝国崩溃过程中出现的非正常死亡。
Turkey, **successive** governments have denied this, believing that **this is the collapse of the Ottoman Empire appeared in the process of unnatural deaths.** (语序混乱) 😞 😞
- 土耳其认为，那些人死于当时的内战和社会动荡，而且这一数字被夸大了。
Turkey believes that **those who** died at that time of social unrest and civil war, **but** that figure has been exaggerated. 😞 😊
- 美国国会众议院外交事务委员会定于10日表决通过一项关于“亚美尼亚大屠杀”的议案，并准备提交众院全体会议表决。
The U.S. House of Representatives Foreign Affairs Committee is scheduled to vote on the adoption of a on the 10th on the "Armenian Massacre," the motion **and to** be submitted to the House plenary vote. 😊 😊

Google Translator 英语-汉语

CNN新闻

英文原文

updated 3 hours, 34 minutes ago EMAIL SAVE PRINT

Turkey recalls ambassador over genocide resolution

STORY HIGHLIGHTS

- **NEW:** Top Dem on defense says resolution could hinder redeployment from Iraq
- Turkish Ambassador Nabi Sensoy will head home after U.S. House panel vote
- Vote calls killing of Armenians during World War I genocide
- Condoleezza Rice set to call Turkish leaders to express "deep disappointment"

[Next Article in Politics >](#)

READ VIDEO MAP

WASHINGTON (CNN) -- Turkey on Thursday recalled its ambassador to the United States and warned of repercussions in a growing dispute over congressional efforts to label the World War I era killings of Armenians by Ottoman Turkish forces "genocide."



Members of the Workers Party protest the U.S. House resolution Thursday in Istanbul, Turkey.

The U.S. House Committee on Foreign Affairs passed the measure 27-21 Wednesday. President Bush and key administration figures lobbied hard against the measure, saying it would create unnecessary headaches for U.S. relations with Turkey.

Turkey -- now a NATO member and a key U.S. ally in the war on terror -- accepts Armenians were killed but call it a massacre during a chaotic time, not an organized campaign of genocide.

The full House could vote on the genocide resolution as early as Friday. A top Turkish official warned Thursday that consequences "won't be pleasant" if the full House approves the resolution.

"Yesterday some in Congress wanted to play hardball," said Egemen Bagis, foreign policy adviser to Turkish Prime Minister Recep Tayyip Erdogan. "I can assure you Turkey knows how to play hardball."

Asked about Ambassador Nabi Sensoy's recall after the news broke, a State Department spokesman said he could not confirm it. "People are sometimes called back for consultation; sometimes they're called back for other reasons," said spokesman Tom Casey.

"If they wanted to bring their ambassador back for consultations or do something else, that is their decision. I certainly think that it will not do anything to limit our efforts to continue to reach out to Turkish officials, to explain our views, to engage them on this issue and again to make clear that we intend to work on this with Congress."

被过滤广告

Most Popular

STORIES

Most Viewed	Most Emailed	Top Searches
1	Paparazzi 'snapped Diana in crash'	
2	Turkey recalls ambassador	
3	Raid finds U.S. soldiers' weapons	
4	Writer suspect in dismembering...	
5	Madonna to sign \$120M record deal	
6	Plane fire after emergency landing	
7	Raid finds U.S. soldiers' weapons	
8	Report: Myanmar prisoners abused	
9	Iraqi families sue Blackwater	
10	Spears wins expanded visitation	

VIDEOS TOPICS

We Recommend

Stories you may be interested in based on past browsing

Google Translator 英语-汉语

CNN新闻

Google的
中文译文

updated 3 hours, 34 minutes ago EMAIL SAVE PRINT

土耳其驻华大使回忆种族灭绝决议

战争聚焦

- **新：顶层**数字高程模型对国防部说，决议案可能会阻碍调配从伊拉克撤军
- 土耳其驻华大使纳比申索伊会回家后，美国众议院小组投票
- 投票呼吁杀害亚美尼亚人在第一次世界大战种族灭绝
- 赖斯定致电土耳其领导人表示“极度失望”

[明年第从政»](#)

阅读

TEXT SIZE

华盛顿（美国） 土耳其对周四召回其驻美国大使，并警告反响，在越来越多的争议，美国国会努力标签第一次世界大战时期屠杀亚美尼亚人，由奥斯曼土耳其军队的“种族灭绝”



美国众议院外交委员会通过这项措施队医周三。布什总统及主要政府人物努力游说反对这项措施，认为这将有助于制造不必要的烦恼，为美国与土耳其的关系。

土耳其现在是北约成员国和美国的重要盟友，在反恐战争-接纳亚美尼亚人被杀害，但它称作是屠杀，在一个混乱的时候，没有一个有组织的运动，灭绝种族罪的人。

充分众议院可以投票就种族灭绝决议案早在周五。顶级土耳其官员警告说，周四的后果“不会愉快”，如果充分众院通过这项决议案。

“昨天，在国会要扮演强硬，说：“埃盖曼巴厄什，外交政策顾问的土耳其总理埃尔多安。”我可以向你保证，土耳其也知道如何发挥强硬”

问大使彩蝶申索伊召回的消息传出后，打破了，美国国务院发言人表示，他无法证实这一消息。“的

人有时被称为回谘询公众;有时，他们正在召回其他原因，说：“发言人汤姆凯西。

“如果他们希望把他们的大使回国磋商或去做别的事，那是他们自己的决定。当然，我认为它不会做任何限制，我们的努力，以继续接触，以土耳其官员，以解释我们的看法，从事他们对这个问题，并再次明确表示，我们打算在这方面的工作与国会合作。”

Most Popular

▼ STORIES

最受欢迎	大多数电子邮件	搜索排名
一	狗仔队‘室戴安娜在坠机’	召回大使
三	空袭认定美军士兵的武器	作家嫌疑人在肢解... ..
四	麦当娜签署\$ 1.2记录处理	架飞机火警后紧急降落
六	空袭认定美军士兵的武器	：缅甸囚犯受虐待
九	伊拉克家屋控告黑水	
十	布兰妮赢得扩大探视	

▶ VIDEOS

▶ TOPICS

Google Translator 英语-汉语

- Turkey recalls ambassador over genocide resolution
土耳其驻华大使回忆种族灭绝决议 😞 😞
- Members of the Workers Party protest the U.S. House resolution Thursday in Istanbul, Turkey.
委员对工人党为抗议美国众议院决议案周四在土耳其的伊斯坦布尔。 😞 😊
- The U.S. House Committee on Foreign Affairs passed the measure 27-21 Wednesday.
美国众议院外交委员会通过这项措施队医周三。 😊 😊
- President Bush and key administration figures lobbied hard against the measure, saying it would create unnecessary headaches for U.S. relations with Turkey.
布什总统及主要政府人物努力游说反对这项措施，认为这将有助于制造不必要的烦恼，为美国与土耳其的关系。 😊 😊
- Turkey -- now a NATO member and a key U.S. ally in the war on terror -- accepts Armenians were killed but call it a massacre during a chaotic time, not an organized campaign of genocide. 😞 😞
土耳其-现在是北约成员国和美国的重要盟友，在反恐战争-接纳亚美尼亚人被杀害，但它称作是屠杀，在一个混乱的时候，没有一个有组织的运动，灭绝种族罪的人。（语序混乱）

统计机器翻译的优缺点

- 优点：
 - 无需人工编写规则，利用语料库直接训练得到机器翻译系统；（但可以使用语言资源）
 - 系统开发周期短；
 - 鲁棒性好；
 - 只要有语料库，很容易适应新的领域或者语种。
- 缺点：
 - 时空开销大；
 - 数据稀疏问题严重；
 - 对语料库依赖性强；
 - 引入复杂的语言知识比较困难。

思考题

- 采用基于句法层面的转换式机器翻译系统，尝试写规则将完成以下句子翻译：
 - 原文：他对语言学很有兴趣。
 - 译文：He is very interested in linguistics.
 - 源文：买这本书花多少钱？
 - 译文：How much does it take to buy this book?