

# 英汉机器翻译的转换生成策略

傅爱平

中国社会科学院语言研究所

**摘要** 由于源语和目标语对比差异的原因,我国研制的英汉机器翻译系统采用了比较特殊的转换生成策略。这就是把转换生成的起点分散在源语分析的各个层次,用共性和个性两种规则描述英汉语之间的转换对应关系。本文在分析英语和汉语句法结构、语义关系对比差异的基础上,从词汇、句法和语义等方面叙述了目前我国英汉机器翻译广泛采用的转换生成方法。

**关键词** 英汉机器翻译 转换和目标语生成

## A Transfer-generation Strategy in English-Chinese Machine Translation

Fu Aiping

Institute of Linguistics

Chinese Academy of social Sciences

**Abstract** For the sake of translation divergences between language pair, most English-Chinese MT systems developed in China have adopted a rather different scheme in building their transfer & generation models with two distinctions from the conventional approaches: In one the transfer has been performed at all possible levels in translation process: lexical, syntactic and semantic. In the other, two kinds of transfer rules, general and specific, have been employed to define the transfer mapping from source to target language. After the discussion of translation divergences in question, the paper describes such method of transfer & generation which is widely used in English-Chinese MT systems in China.

**Key words** English-Chinese machine translation, Transfer & generation

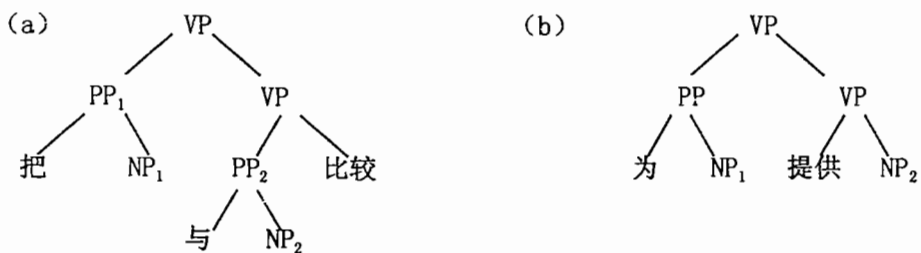
中国的机器翻译研究是从五十年代末开始的。在四十年的研制经历中,涉及最多的语言对是英译汉,最常采用的策略是转换法。

转换法是每一个机器翻译研究人员都熟悉的(图1),包括基于词汇的转换(Lexical Transfer)、基于结构的转换(Structural Transfer)和基于语义(Semantic Transfer)的转换。一个系统的转换通常在某一个层次上进行,譬如句法结构层。这时转换模型描写源语和目标语在句法结构上的对应关系,系统根据这些对应关系把源语的句子结构转换成目标语的句子结构,再经过形态处理生成句子的表层词语串。

然而,我国研制的多数英汉翻译系统却没有采用这种经典的转换模式。我们往往把转换更多地和译文生成联系在一起,与图1相比,这种转换生成策略主要有以下两点不同之处:

第一点是把转换的起点分散在源语分析的各个层次,在词汇、句法、语义分析之后,都可以开始转换生成处理。第二点是转换规则有共性和个性两种。共性指以句法关系或语义关系为基础定义的转换,譬如名词及其修饰成分的词序,在汉语中是修饰语在前,中心语在后,而多个修饰语当中,还要根据各自与中心语的语义关系确定词序;个性的转换与特定的词有关,例如动词短语:

(a) compare NP<sub>1</sub> with NP<sub>2</sub>      →      把 NP<sub>1</sub> 与 NP<sub>2</sub> 比较  
 (b) provide NP<sub>1</sub> with NP<sub>2</sub>      →      为 NP<sub>1</sub> 提供 NP<sub>2</sub>



(a) 和 (b) 在英语中结构相同，转换成汉语在句法结构上却有很大的差异(如上图)。这种差异适合用个性规则在词汇项的基础上描述。

英语和汉语分属不同的语系，从词汇系统到造句方式都有很大的差别。使用不同类型的规则在词汇、句法和语义各个层次分别作转换，目的是用不同的方法、在适当的时机解决不同的翻译差异问题。

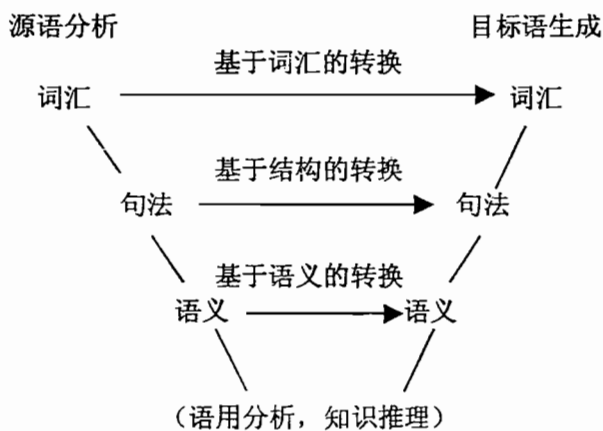


图 1

## 一、词法分析阶段的转换生成

多数英语的习惯表达 (Idiomatic Expressions) 和固定词组 (Word Group) 在源语的形态分析阶段 (查词典、词法分析) 被直接转换成汉语的表层词语。转换以后，习惯表达和固定词组都收缩为一个结点，保留在整个英语句子的句法结构中，以便参加后面的句法和语义分析。例如：

How do you do      →      你好      由多个结点合为一个结点 S  
 the quantity of heat      →      热量      由多个结点合为一个结点 N

显然，词义的选择和结点的合成都与特定的词有关，所以形态分析阶段的转换是个性的转换。对整个句子来说，这是在源文分析的较浅层次上的局部转换，与局部表层译语的生

成直接有关。

## 二、句法分析阶段的转换生成

句法层的转换规则有个性的和共性的两种。前者主要处理短语，后者主要处理句子成分。

转换在M T系统中的作用是建立语言对之间的对应关系。基于结构的转换应该在源语和目标语的句法结构之间建立对应规律。如果能系统地实现这样的对应，确实是M T最理想的情形。然而英语和汉语的结构很难对应，至少很难找到一一对应。比如英语的介词短语可以翻译成汉语的时间词短语、方位词短语、动词短语和介词短语，而汉语的动词在英语中可以翻译成名词、形容词、介词和动词。翻译要译的是意义，句法结构的相互对应是以表达相同的意义为基础的。把句法结构作为转换的平面实际上是在借用句法作语义的事情。因此，英汉两种语言的句子能否在句法结构平面上找到对应规律，最终还要看它们各自的句法结构和语义结构之间有没有对应规律。

Comrie(1981)认为，与俄语相比，英语句子的语法结构和语义角色之间的联系相对松散，也就是说，英语句子的结构关系和语义关系之间的对应相对复杂。然而如果用英语和汉语比较，我们会发现，汉语的这种对应更为复杂，复杂得我们很难找到一种系统的方法，来建立汉语的句子结构和语义角色之间的对应关系。所以在这里，我们只需看看汉语句法关系和语义角色相互作用的情形，就可以说明在英汉语的语法结构之间系统地建立对应关系有多麼困难。

汉语句法结构和语义结构之间的联系相当松散，或者说句法成分和语义成分的配位很灵活，一个句法结构经常表示多种意义，一种语义内容也经常可以用多个句法结构来表示。譬如，在一本现代汉语教材中（钱乃荣，1990），列有十六种既可以用主谓结构实现，也可以用动宾结构实现的语义关系类型。这种多对多的结构—语义对应现象在汉语中十分普遍。我们至今还没有找到辨别它们的有效办法。有的汉语语法学家认为，在汉语语法结构中，有两个值得注意的事实：语义蕴含上的兼容性和形式选用上的趋简性（邢福义，1997）。这就是说，同一个语法结构可以包容多种意义，而结构形式则通过缩减趋于简化。这两个事实对M T都非常不利。多年来汉语语法一直注意研究结构和意义的关系，已经揭示的参数除了词汇—语义信息以外，还有句法、话语、语用、认知等信息，但是这些研究还不够系统。M T无法在这些研究的基础上，系统地建立汉语表层句法结构和深层语义表示之间的对应关系，因此系统地描写英语和汉语句法结构之间的对应关系，也就非常困难。

困难恐怕主要来自我们过于理想的目标：抽象、而且概括能力强的数学模型、参数和规则。曾经有人在很小范围的语料内作过类似的尝试，杨平（1986）在翻译了38个句子之后，遗憾地指出：我们还缺乏描写汉语的数学模型，从参数到结构，都不够清楚。

问题的可解与否常常取决于这个问题所处的参照系，句法平面上的不确定性在词汇—语义平面常常可以得到消解。对机器翻译的转换生成，更多的人采用了比较灵活的策略：容易找到的对应规律用抽象的共性规则去描写，那些不好找抽象规律的现象则以词汇项为基础用个性规则去描写。以词汇项为基础，英语的许多短语可以确切地转换成汉语的短语，例如前面的动词短语(a)和(b)以及下面名词短语的结构转换：

the effect of NP<sub>1</sub> on NP<sub>2</sub> ----> NP<sub>1</sub> 对 NP<sub>2</sub> 的影响

为了提高效率，在多数系统中，这样的个性转换往往与源语分析连在一起。当源语分

析识别出这些英语短语时，短语内部的句法关系和语义关系就已经清楚了，转换规则的作用是指出与它对应的汉语短语的表层句法结构和词汇选择。由于是个性规则，这些目标语表层的信息很具体，包括句法结构、语义关系、词义选择、增加或删除结点的位置和词义，甚至还有表层词序的信息。实际上，转换和译语表层生成的某些过程合在了一起，这样可以充分利用源语分析得到英汉对比差异，避免在源语和目标语之间设置过多的接口参数。当然，这样作也会使转换生成在某种程度上受源语分析的制约，另外转换的两端在形式上也往往不对等，即源语的句法—语义结构，有时可能会转换成译语较浅层次的表层结构。

我们曾经注意到关于“翻译差异”的研究。Dorr (1995) 借助词汇—语义差异，在深层语义表示和表层句法结构之间建立对应关系，也是以词汇项为基础的，这一点与我们个性转换的思路一致。而对于下面要说的共性转换，处理的方法则有所不同。

在句法层次，共性的转换是以句法关系为基础的转换。如果不考虑修辞和语用效果，英语的一部分句子结构可以在句法平面上转换成汉语的句子结构。譬如下面的结构转换规则：

N + V + O	—————>	N + V + O	或	O + 被 + N + V
N + V + O <sub>1</sub> + O <sub>2</sub>	—————>	N + V + O <sub>1</sub> + O <sub>2</sub>	或	N + 把 + O <sub>2</sub> + V + O <sub>1</sub>
N + V + O + P	—————>	N + V + O + P	或	N + 把 + O + V + 为 + P
			或	O + 被 + N + V + P

再如名词及其修饰成分：

N <sub>1</sub> + of + N <sub>2</sub>	—————>	N <sub>2</sub> + 的 + N <sub>1</sub>
N + Ving-clause	—————>	Ving-clause + 的 + N
N + Ved-clause	—————>	Ved-clause + 的 + N
N + which/that-clause	—————>	which/that-clause + 的 + N

对于整个句子的句法结构树，个性的转换发生在树叶或靠近树叶的分支上，共性的转换则发生在子树或整个树上。

### 三、语义分析阶段的转换生成

由于英汉之间的差异，句法结构的转换只靠句法层次的信息往往是不够的。例如，修饰名词的成分常常有好几个，在汉语中都需要放在名词前面，它们之间的位置怎么确定呢？有的汉语语法学者认为，形容词在偏正结构中的排列顺序，与它们和名词的疏密程度有关，离名词近的与名词的关系紧密，离名词远的与名词的关系疏远（马庆株，1995）。这种疏密程度可以通过修饰成分与名词的语义关系来描写：

领有性成分 < 状态或描写性成分 < 属性类成分 < 形状类成分 < 名词。  
(方、圆、斜、歪)

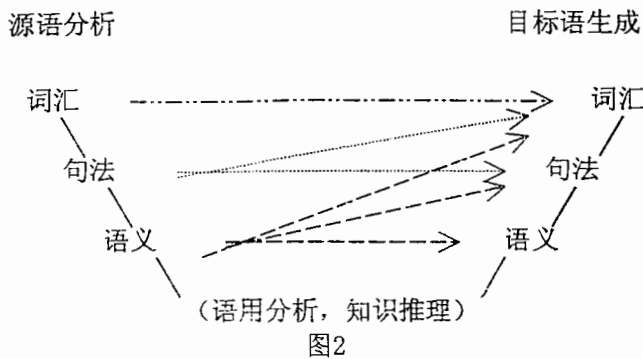
其中属性类成分又可以按语义特征分为：

性质属性 < 感官属性 < 几何属性 < 颜色属性。  
(新、旧、美、丑) (香、臭、热、凉) (长、短、厚、薄) (黑、白、青)

源语分析确定了名词与其各个修饰成分的语义关系以后，根据上述规则，就可以经转换得到目标语的表层词序。动词及其修饰成分之间也有类似的转换。这些都是在结构转换的前提下从语义关系到表层结构的对应，属于共性的转换。

还有一种基于语义的转换不受句法结构制约，这就是按句子中各个概念之间的逻辑语义关系进行的转换。转换的起点是经过源语分析得到的“逻辑语义结构+各结点上必要的转换参数”，终点是与逻辑语义结构对应的目标语的表层结构（董振东，1986）。这是一种在深层结构和表层结构之间寻求对应的共性转换，属于我们在前面说的那种理想的方法。

这种分层的、共性和个性并存的转换生成方式（图2）在印欧语言对的机器翻译中也许还不多见，但目前我国研制的英汉机器翻译系统大部分采用了这种方式。很多其他外汉系统也这样处理转换生成问题，譬如德汉系统（邸海燕等，1995），俄汉系统（李向东等，1999）。经过多年的研究和开发，这种转换生成的技术已经比较成熟。



#### 四、这种方式的局限性

多层次转换的目的是用各种办法找到源语和目标语表层结构的对应关系。而它的主要缺点也是常常不能找到最恰当的那种对应关系。

英语翻译成汉语时，如果不考虑修辞、语用功能和交际效果，一个句子可以译成多个汉语的表层形式。例如（1）和（2）、（3）、（4）：

- （1）He cried so much that his eyes became blood-shot.
- （2）他哭红了眼睛。
- （3）他哭得眼睛红了。
- （4）他哭得如此厉害以致于他的眼睛变红了。

使用本族语的人都知道，（2）是地道的汉语译文。可是它和（1）的句法结构完全不一样。（2）是动补格句式，其中主语是动作的施事，补语是动作的结果，这个结果使宾语和补语又形成了主谓关系，宾语是补语的感事（experiencer），宾语和动词之间没有语义关系。这种动补格句式结构紧凑，寓意丰富，可以说明汉语语义上的兼容性和形式上的趋简性。吕叔湘先生也认为动补格句式形式和意义之间的关系变化多端，表现了汉语句法的灵活性与节约性（吕叔湘，1986）。

用我们的转换生成方法，MT很难生成这样的译文。倒是（4）虽然听起来不够地道，可意思却没错。最值得注意的是，它与（1）的句法结构接近得多。目前我们转换生成的结

果，常常是在意思大致相同的几种译文里边和源语句子结构最为接近的那种表层结构。汉语具有“意合”的造句方式，句子结构不象英语那样，受由一致关系控制的“主—谓”格局的限制，同时语序也比较自由。借助英语的结构格局，只要源语分析到位，多数译文能够作到易懂或可读。汉语译文的可读性或懂性，是一个相当宽泛、容错程度较高的标准，因为汉语译文的读者绝大多数以汉语为母语，具有“意合”的能力。

这样作实际上是在汉语语法理论还无法支持精确的转换模型的情况下，利用汉语的表义方式在外汉机器翻译中有时能够帮助“藏拙”这一事实，尽量回避汉语生成的困难，以一部分译文带“机器味儿”的代价，让英汉机器翻译系统尽快达到实用，走向市场。从下面三组句子中，我们可以再次体会到那种和原文相比结构形式差不多，意思不走样，带“机器味儿”的译文（每组的最后一句）：

The noises made me awake.

喧闹声吵醒了我。

喧闹使我醒了。

Excitement deprived me of all power of utterance.

我兴奋得什麼话也说不出。

兴奋使我丧失了所有的说话的能力。

The thick carpet killed the sound of my footsteps.

我走在厚厚的地毯上，一点儿脚步声也没有。

我的脚步声在厚地毯上一点儿也听不见。

厚地毯使我的脚步的声音消失了。

当然，也还有一些结构形式差异太大的句子用这样的转换生成方法翻译不好，例如：He said nothing about it. 就很难转换成：他什麼也没说。

## 参考文献

- Comrie, Bernard, 1981, *Language Universals and Linguistic Typology*.  
中译本：语言共性和语言类型 沈家煊译，1989，北京：华夏出版社。
- Dorr, Bonnie J., 1995, A lexical-semantic solution to the divergence problem in machine translation. *Computational Lexical Semantics*, ed. by P. Saint-Dizier and E. Viegas, 367-395. New York: Cambridge University Press.
- 邱海燕 柴佩琪 许玉祥 基于双语语料库和规则库的德汉复合句的转换生成，计算语言学进展与应用，清华大学出版社，1995 264-270
- 董振东 关于应用型机译系统的几点考虑，语言和计算机，No. 3 中国社会科学出版社，1986，23-27
- 李向东 周清波等 智能型俄汉机器翻译系统的句法规则库的设计原则，1999，中文信息学报，Vol. 13, No. 1 16-19
- 吕叔湘 动补结构的歧义性，中国语文，1986，No. 1
- 马庆株 多重定名结构中形容词的类别和次序，中国语文，1995，No. 5，357-366
- 钱乃荣 现代汉语，高等教育出版社，1990，233-251
- 邢福义 汉语语法结构的兼容性和趋简性，世界汉语教学，1997，No. 3，3-8
- 杨平 汉语生成问题，语言和计算机，No. 3，中国社会科学出版社，1986，28-46