

工艺语言描述中的自动翻译模型的建立

周继如, 陈实

(南京师范大学电气与电子工程学院, 210042, 南京)

[摘要] 根据机械制造工艺语言的特点, 在机械制造工艺汉英自动翻译系统中, 对汉语工艺语言的词语进行了合理地分类, 并建立了词类判别的数学模型. 同时, 简单介绍了工艺语言识别和处理的过程.

[关键词] 自动翻译, 工艺语言, 数学模型

[中图分类号] TH16; H085; [文献标识码] A; [文章编号] 1672-1292(2002)03-0034-04

1 机械制造工艺语言中词的分类

汉英自动翻译中, 为了更好地实现翻译自动化, 原文(即中文)句子中词的合理分类是关键. 在汉语语言学上, 通常把词分为实词、虚词两大类. 实词即意义比较具体的词, 包括名词、动词、形容词、数词、量词和代词; 虚词即不能单独成句, 意义比较抽象, 有助于造句的词, 包括副词、介词、连词、助词、叹词、拟声词. 根据机械制造工艺语言的特点, 在机械制造工艺汉英自动翻译系统中我们将词分为以下几类:

名词 代表抽象的或具体的事物或概念的词, 包括代词;

动词 陈述人或事物的动作、情况或变化的词;

形容词 修饰或限定名词的形状、性质或状态的词;

数量词 表明人、事物或动作的数目和单位的词;

副词 所有修饰或限定动词或形容词的范围、程度、时间、频率、语气、情貌等的词. 可以有实际意义, 已不全属于虚词;

介词 引出修饰或限定名词或短语的标志词;

搭配词 标志介词的引出部分和结束的词;

时态助词 帮助标识动词的时态和句子的语气词;

结构助词 帮助造句的词, 包括的、得等;

连词 连接词与短语的词, 包括和、或等.

如此定义的原因有以下6点:

(1) 由于工艺文件中较少出现代词, 加上名词和代词在汉英翻译过程中所遵从的语法规则基本上是一致的, 因此将名词和代词合并为一类.

(2) 表示单数概念的量词对可数名词而言是汉语所特有的, 英文只表现为数词. 对不可数名词, 英文虽也有相当量词的概念, 但实际上与汉语的联系甚微. 如: 一张图纸和一件夹具中的一张和一件都对应为 a piece of. 表示复数概念的量词, 如: 打、堆、捆等, 一般可用于限定多种名词. 此类量词也要与数词复合后才能完成词的转换. 由此可见, 量词和数词的分离对于翻译帮助不大, 反而会给词的分类带来不必要的麻烦, 因此将量词与数词合称为数量词.

(3) 由于我们仅研究工艺文件中简单陈述句的词的自动定类, 叹词和拟声词一般不会出现, 所以在

收稿日期: 2002-06-17

作者简介: 周继如, 1971-, 女, 硕士, 南京师范大学电气与电气工程学院讲师, 主要从事机电一体化方面的研究.

分类中略去。

(4) 介宾结构在工艺文件中经常出现, 它常由介词带上动词短语或含动词的名词短语组成。这个短语被独立出来后常可成为一个完整的句子。如: 在加工零件的时候 中 加工零件 这个短语可以当作子句拿出来独立分析。因此, 定出介词 在 和其搭配词 时候 有助于对工艺文件的分析。我们将搭配词单列一类添加在词的分类中, 搭配词就是介宾短语的典型的结尾用词, 通常是名词中的方位词, 然而不是每个介宾短语中都有搭配词。

(5) 在工艺文件的翻译过程中, 对时态助词、结构助词的处理都有较大差别, 时态助词在翻译时直接与其后的动词结合, 如: 可能完成 may finish。结构助词 的 常用于取消句子的独立性, 在翻译时往往使其前面的动词被当作分词或是句子来处理。因为, 工艺规程中的句子大多为陈述句, 所以可以不考虑语气助词。因此, 将各种助词分为时态助词和结构助词两类。

(6) 在上述分类中对保留下来的词类名称的外延也做了若干变动, 以便于转换。其中, 需要特别提出的是:

在汉语中, 副词的个数是有限的。有的形容词加 地 后虽可用来修饰和限定动词, 但对认定词类不便。在英语中, 没有 地 字, 而通常要采用不很规则的形态变化。所以, 我们在分类时, 就把这些形容词的词类直接归为副词, 以利转换。

在翻译复句时, 当将连接句子的连词译成英文后, 句中的其它部分即被拆为两个独立的单句来处理。因此我们将连词定为连接词与短语的词。

总之, 我们在分类时, 既根据汉语语言学的理论, 又充分考虑到工艺语言的特点。每种类型的词还可以取值, 如:

名词次类: 一般名词、专用名词、人称代词、物主代词、关系代词。

动词次类: 情态动词、连系动词、助动词、一般动词。

词组类型: 名词词组、形容词词组、前置词词组、后置词词组、动词句、从句。

数: 单数、复数。

时态: 现在时、过去时、将来时。

在机械制造工艺汉英自动翻译中, 对工艺语言的词语合理分类之后, 就可以在对汉语的语法分析中, 依靠词的分类而不是具体的词建立分析规则, 这样既能保证规则的有限性和抽象性, 也能满足排他性的要求。从语言学的角度建立一套科学、合理的词的分类法, 可以为原文分析和译文综合提供尽可能多的有用信息, 使翻译更准确、可靠。

2 工艺语言描述的数学模型

由于在自动翻译中都是以每个单词的词类而不是具体的单词作为语法分析对象的, 所以工艺语言自动翻译系统中描述工艺语言的数学模型主要指词类的模型, 包括三个方面:

2.1 代码体系

1 名词, 2 动词, 3 时态助词, 4 数量词, 5 副词, 6 形容词, 7 介词,
8 搭配词, 9 结构助词, 10 连词

2.2 集合的定义

左词性集 对一给定的词性其左邻的所有可能的词性的代码的集合。

右词性集 对一给定的词性其右邻的所有可能的词性的代码的集合。

作者根据语法规则制定了如表 1 所示的左词性集和表 2 所示的右词性集:

表 1 左词性集	
词性代码	左词性集元素
1	{ 1, 2, 3, 4, 6, 7, 8, 9, 10 }
2	{ 1, 2, 3, 4, 5, 6, 7, 8, 10 }
3	{ 2, 5 }
4	{ 2, 3, 6, 7, 8, 9, 10 }
5	{ 1, 2, 3, 5, 6, 7, 8, 9, 10 }
6	{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 }
7	{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 }
8	{ 1, 2, 3, 4, 5, 9 }
9	{ 1, 2, 4, 5, 6, 8 }
10	{ 1, 2, 3, 4, 6, 7, 8, 9 }

表 2 右词性集	
词性代码	右词性集元素
1	{ 1, 2, 5, 6, 7, 8, 9, 10 }
2	{ 1, 2, 3, 4, 5, 7, 8, 9, 10 }
3	{ 1, 2, 4, 5, 6, 7, 8, 10 }
4	{ 1, 2, 6, 7, 8, 9, 10 }
5	{ 2, 3, 5, 6, 7, 8 }
6	{ 1, 2, 4, 5, 6, 7, 9, 10 }
7	{ 1, 2, 4, 5, 6, 7, 10 }
8	{ 1, 2, 4, 5, 6, 7, 9, 10 }
9	{ 1, 4, 5, 6, 7, 8, 10 }
10	{ 1, 2, 4, 5, 6, 7 }

2 3 词类的判别

我们在机制工艺汉英自动翻译中, 对句子中每个词的词类的判别方法是按下述原则指定的: 句子中各个词在该句子中的词性是合理的充分必要条件是该句中任意一组相邻的三个词满足, 中间词的词性应在右边词的左词性集和左边词的右词性集的交集中. 该规则用于消除兼类.

每次取 3 个词的数学模型如下, 设 m 为句长, 对 $i = 1, 2, \dots m$ 计算

$$C[i + 1] = C_p[i] \cap C_q[i + 2]$$

其中设: $W[i]$ 代表一个词(注: $[]$ 内数字表示正整数, 下同); $C[i]$ 一个词性; $C_q[i]$ 一个词的左词性集; $C_p[i]$ 一个词的右词性集.

若对每一个 $i = 1, 2, \dots m$ 有 $C[i + 1]$ 则有解, 否则无解(可能为非线性结构) 该文法可用下图表示:



在作了上述数学抽象后, 就可以对此数学模型进行与汉语语法规则等价的运算, 从而得到各词正确的词性. 试举机制工艺语言中一例说明.

在句子 外圆的加工参照图纸 中, 加工 一词的词性可以为动词(如 按简图加工零件), 也可以为名词(如本例), 我们可以根据上述文法消除兼类. 具体如下:

加工 为 $W[3]$, 词性未知; 其左邻词 的 为 $W[2]$, 词性 $C[2] =$ 结构助词; 其右邻词 参照 为 $W[4]$, 词性 $C[4] =$ 动词. 由表 1 和表 2 可知, $C[3] = C_p[2] \cap C_q[4] =$ 名词. 这样就确定例句中 加工 一词的词性是名词, 消除了兼类.

3 工艺语言的识别和处理

建立工艺语言描述模型的目的, 是为了在自动翻译中更好地识别和处理工艺语言. 机制工艺汉英自动翻译系统对汉语工艺语言的处理采用的分析文法包括两方面内容, 一是前面所述的用于消除词性兼类的词性判断文法, 二是基于汉语完全语法树的文句分析文法. 文句分析文法认为由于工艺文件中都是

陈述句, 而陈述句的语法结构都是: (定) + 主 + (状) + 谓 + (补) + (定) + 宾. 所以, 对工艺语言文句的分析采用了能够涵盖所有汉语陈述句结构的汉语完全语法树(Chinese integrated syntax tree), 从而使汉语工艺语言分析纳入了逻辑推理的轨道, 而不再是个例描述.

在对工艺语言的处理中, 利用词类判断文法确定句子中各词的词性之后, 根据文句分析文法将句子中的各词装入汉语完全语法树的六大成分块中, 从而得到该句的深层结构, 然后在深层结构上进行汉英转换, 最终得到正确的英译文. 其中的过程是复杂而繁琐的, 但工艺语言的描述, 即词的合理分类以及词性的确定是整个自动翻译的基础和关键.

[参考文献]

[1] 吴蔚天, 罗建林. 汉语计算语言学[M]. 北京: 电子工业出版社, 1994, 100~ 106.

Establishment of Automatic Translation Model to Describe the Language of Technology

Zhou Jiru, Chen Shi

(College of Electrical and Electronic Engineering, Nanjing Normal University, 210042, Nanjing, PRC)

Abstract: According to the characteristics of the language of technology, the words of the language of technology are rationally classified, with the mathematics model for differentiating the parts of speech established. Besides, the process to distinguish and deal with the language of technology is described.

Key words: automatic translation, language of technology, mathematics model

[责任编辑: 刘健]

(上接第 29 页)

Preparation of Cu_xS ($x = 1, 2$) thin films by a Asynchronous pulse spray pyrolysis method

Wang Wei¹, Wang Shengyue²

(1. College of Physics science and Technology, Nanjing Normal University, 210097, Nanjing, RPC;

2. National Laboratory of Molecular and Biomolecular Electronics, Southeast University, 210096, Nanjing, PRC)

Abstract: Asynchronous pulse ultrasonic spray pyrolysis technique was employed to deposit Cu_xS ($x = 1, 2$) thin films on glass. The deposited films chemically close to CuS were found to be polycrystalline phases, the Cu_2S films were a mixture of polycrystalline as well. The crystalline phase of particles was highly depended on the molar ratio of thiourea to CuCl_2 and the pyrolysis temperature. XRD and XPS analysis showed pure stoichiometric Cu_xS (Covellite CuS , Chalcocite Cu_2S). Raman shifts of the films were measured at 472 cm^{-1} (Cu_2S) and 474 cm^{-1} (CuS).

Key words: Copper sulfide, Thin films, Spray pyrolysis

[责任编辑: 刘健]