

# 对‘被’字句进行汉语句法分析的研究

聂昱 傅其芳 王斌 戴大为

(武汉大学计算机科学系 430072)

汉英机器翻译一直是计算机科学中一个极具吸引力而又充满困难的课题。我们经过几年的努力,终于设计,编制成功了一个基于规则的,应用分段循环分析策略的汉英机器翻译分析器及与之相应的电子词典,规则库。在此基础上,对一些汉语句子的结构特点及基于规则对它们进行分析进行了一些研究,本文将介绍对一种比较典型的汉语句式——被字句的分析的实现。

首先介绍一下整个系统工作的概况。一个汉语句子首先经过分词,然后分析器开始工作。在分析器中,规则放在16个规则段中。每一个规则段解决一定的语法问题,分析采用自底向上的策略,若输入的词串最终能合一成为一个语法范畴,则分析成功,否则分析失败。在分析的过程中同时生成了英语语法树,在分析结束后经过调整生成英语句子。在分析过程中,要从词典和规则库中取得信息,规则库分为公共规则库和专用规则库。公共规则是指使用比较频繁,覆盖面比较广的规则;专用规则则是指使用不太频繁,仅适用于一个或几个词的规则。

规则的一般格式如下:

$C1+C2 \rightarrow [C3, COND, A, E]$

HEAD=X, PASSAGE=Y

C1, C2, C3 为语法范畴,各带有几十个特征项。我们用其词性特征项 cat 来对其进行标识。COND 为条件集合,说明规则在什么条件下适用。A 为对中文语法树的操作的集合。E 为对英文语法树的操作的集合。HEAD 若为 1,则新生成的语法范畴 C3 将主要继承 C1 中的语法信息,HEAD 若为 2,则 C3 将主要继承 C2 中的语法信息。PASSAGE 则说明这条规则是属于哪个规则段。

被字句是指含有介词‘被’的句子,‘被’的句法功能是构成被动句,其前一般是动词的受事格。本文通过列举出经过实际检验的基本规则,介绍了对下面七种‘被’字句的分析,并且还举出了相应的例句分析过程。

- 1 名(受事)+‘被’+[名(施事)]+[助]+动
- 2 名(受事)+‘被’+[名(施事)]+动介+名(处所)  
名(受事)+‘被’+[名(施事)]+动趋+名(处所)
- 3 名(受事)+‘被’+[名(施事)]+动 1+动 2
- 4 名(受事)+‘被’+[名(施事)]+动+名(与事)  
名(与事)+‘被’+[名(施事)]+动+名(受事)
- 5 名(受事)+‘被’+名(施事)+动 1+动 2  
名(受事)+动 1+‘被’+名(施事)+动 2
- 6 名(施事)+动 1+名(受事)+‘被’+名(施事)+动 2
- 7 名(受事)+‘被’+名(施事)+‘把’+名(受事)+动

除了对被字句的分析,我们对把字句,兼语句,连动句,双宾语句等也作了一些研究,所有的规则都在我们的句法分析器上测试通过,相互之间互不干扰。这说明基于规则进行汉英翻译是有一定前景的。