

智能型俄汉机器翻译系统中的源语言分析

李向东 周清波 周天荣
(北京外国语大学俄语学院, 100081)

Source language analysis in Russian-Chinese Intelligent Machine Translation System

Li Xiangdong, Zhou Qingbo, Zhou Tianrong
(Beijing Foreign Studies University, 100081)

智能型俄汉机译系统是一套面向实用的机译系统。系统设计以SC文法为基础。由于俄语形态变化丰富、词序灵活多变等特点,源语言的分析、处理难度很大。如何针对俄语特点进行有效的结构分析与语义分析是我们的主要攻关方向。

1. 动词是结构分析的关键。俄语的句子成分格局表现为以动词为中心的语言组织形式。谓语动词是句子的结构联系和语义联系的中心,起框架作用。我们把谓语动词作为分析句法结构的首要标志,调整句子词序的主轴,通过它搜寻、确认句中各成分之间的相互关系,逐层递归,生成目标语句型。

基于SC文法体系的特点和俄语的句法结构特点,我们综合分析了动词的语法特征、语义特征、结构特征、功能特征、俄汉两种语言的对应关系等特点将俄语动词分为四大类:V100, V200, V300, V400。V100和V200类动词做谓语的句法结构特点是主谓补形式, V100类的补语结构可以是直接补语和间接补语两种形式, V200类只能是间接补语; V300类动词做谓语的句法结构特点是主谓形式,以V400类动词做主要成分的句法结构特点是无主语结构。

我们的动词分类体系既描述了动词与句子其他成分之间的表层结构关系,又在一定程度上反映了它们之间的深层语义关系;既体现了俄语动词的性质、功能,又兼顾了汉语对应体的特点,在动词性谓语句型的结构分析中起着关键作用。

2. 自然语言中多义现象的分析与处理一直是困扰机器翻译界的一大难题。这一问题解决得好坏将直接影响机器翻译的质量。文中指出,上下文综合分析是排除语言多义现象的主要前提和依据。

我们利用SC文法中的Search函数的方法,通过在词典库和规则库中设置语境条件即采用外部限定,进行上下文相关处理。与此同时对结构内每一成分所应具有的词法形式、句法特性及语义特征进行严格的限定(内部限定)。内、外部限定相结合的处理方法,使机译系统在一定程度上具有排除语言多义性的功能。文中就机器翻译中常遇到的语法、语义、结构、语用四个方面的多义现象进行了具体分析,提出了解决办法。

最后,作者指出,要建立有健全知识体系的高智能的自动翻译系统,语言学家与计算机专家的有机结合是必不可少的先决条件。语言学家要对语义学、语法学、语用学、篇章分析做出更加深入的研究,不断丰富知识库,而计算机专家应着力解决知识的获取、表示、应用等问题,提高系统处理自然语言的能力。