

# 一种新的适合中文计算机处理的文法理论

## ——核心推导语法

宋继平 陈肇雄 黄河燕

王能忠

(中国科学院计算技术研究所机译中心, 100080)

(西南师范大学计算机科学系, 630715)

自然语言处理的效果受到文法理论及其分析技术的直接制约和影响。现有文法理论都因自身固有的缺陷影响了实际系统处理能力的提高, 迫切需要建立适合中文计算机处理的新的文法理论体系。陆丙甫先生提出的核心推导语法较为有效地把握了汉语的实际特征, 有利于实现语言分析的突破。

在一个结构体中确定一个核心, 对于结构关系分析的简化和明确有关键性的作用。核心推导语法从人类信息处理能力的角度进行分析, 以核心为出发点对各种语法现象进行解释和研究。陆丙甫认为存在两种核心: 词汇核心和功能核心。其中, 词汇核心根据语义标准和独立性来确定, 有利于对汉语语句的合理切分与分析。

核心推导语法有效体现了语言构成的层次性和实际语言使用的无限递归性。该语法认为句子和短语都是各自核心及其直接联系的附属成分组成, 这些附属成分也可由局部核心及其附属成分构成。因此, 在核心推导语法中, 与核心动词无直接语义联系的“定语”不是句子的直接构成成分。

核心推导语法以核心为出发点, 通过对附属成分同核心词语义联系强弱和相对距离关系的研究, 发现了语序和语义之间内在的规律性联系。根据与核心语义联系的直接与否则形成的以核心动词为中心的轨层结构, 总结出了汉语省略主语后的优势语序, 并对汉语复杂的语序变化进行了合理有效地解释, 即变化的范围一般仅限于轨层内部或属于“大块外移”和“小块内移”情况。

将各组成部分线性(而不是通常的树状)排列, 则轨层结构表示的便是语言成分的实际表达顺序。核心推导语法以句子核心为出发点, 得到了汉语句子成分出现的相对稳定区域, 研究了不同区域成分的属性 and 功能, 得出了“职有定位”的结论, 并赋予了相应部分一个较为稳定的序号。这实际限定了各类成分出现的相对顺序, 可根据对序号的确定与分析实现对语法、语义的处理, 简化了语言分析处理的难度。

现有的某些系统进行了语义甚至语用处理, 但重心在于语法分析, 语义、语用分析由于知识表示和处理等困难只是用来处理少量非常简单的针对性强的问题。由于汉语本身的复杂以及当前对其认识的局限, 语法规则难以有效完整地表达实际的语法现象, 严重影响系统的语言分析处理能力。核心推导语法在一定程度上避免了这个近期难有实质性进展的问题, 有效地将语法、语义要求统一于对语序的处理。既符合汉语形态不发达, 主要通过语序和虚词表示的特点, 又采用了全新途径进行语言分析, 有助于解决目前存在的困难, 具有很强的可操作性。

核心推导语法为中文信息处理通过了一条新的语言分析的途径, 具有一定的优势, 但作为一种新的文法理论, 其本身尚不够成熟, 有待进一步完善。