

面向词义自动标注的多义动词选择限制研究

Research on Selectional Restriction of Polysemous Verb for Word Sense Tagging

作者:蒋媛 导师:苏新春
毕业院校:厦门大学 答辩日期:2015年5月

DOI:10.16499/j.cnki.1003-5397.2016.02.020

多义动词的处理是词义自动标注的重点、难点,在组合关系中基于选择限制产生的搭配特征是词义自动标注的主要依据。论文对903个多义动词在真实语境中的分布、组合、搭配进行分析,讨论了多义动词在计算机环境中的识别条件与方法。

第一章对选题的基础理论“义项形式论”和核心概念“区别性形式特征”进行了阐释。机器处理语言具有重形式的特点,义项形式特征指义项在具体语境使用中呈现出来的有规律,能够加以归类、概括,并能被计算机识别的标志物(以语义搭配为主),是语言单位间选择限制的结果。义项形式论讨论的是义项的意义内涵差别与义项形式特征的关系。论文立足于多义动词义项的“对立”与“差异”,集中关注形式特征的“区别价值”。

第二章讨论了多义动词选择限制知识的获取与形式化。文中建立了描写框架,在语料库、义项库、语义分类库、语法库等资源库的结合使用中对知识的获取问题进行了分析。形式化指把属于自然语言范畴的区别性形式特征转化为计算机可以识别、处理的人工语言。通过“形式化”得到词义自动标注所需的“规则”。规则的形成与质量是词义自动标注的关键,论文结合实例,从可靠性、简约性、拓展性三个方面讨论了知识的形式化问题,并通过验证分析了规则的有效性。

第三章建立了选择限制的分析模式,对多义动词选择限制在语法、语义层面的表现进行了分析。总结了多义动词选择限制的特点,探索了选择限制的规律与机制。多义动词的选择限制具有不平衡性、系统性的特点,选择限制的力度有强弱之分。选择限制力的强弱与动词动作性的强弱、词义的限定元素等有关。

第四章结合多义动词规则的验证,探索了多义动词机用义项库的建设问题。机用义项库的建设是词义自动标注的基础与关键,直接影响着规则的提取与效用。义项的设置应具备独立性、排他性、概括性,要在语料统计的基础上,以语境中的区别性形式特征为线索,以核心语义特征为依据进行处理。义项库的调整应形成科学、统一的标准。义项的粒度可以因词而异。

论文针对词义自动标注工程中最为核心的知识源问题,对语言特征的识别与标示进行了探讨,具有全面、典型的特点。讨论的结果对词义自动标注知识源问题、标注策略的改进、机用义项库建设等具有一定参考价值。