

智能式机器翻译中指代关系的 FAST 处理

Anaphora Interpretation through FAST for AI-Machine Translation

孙 介 铭

Sun Jie Ming

上海交通大学

Shanghai Jiao-Tong University

摘 要

全文机器翻译中的指代关系实际上涉及到省略、隐含、借喻以及前后呼应等自然语言处理和机器翻译中最难处理的一些问题。指代关系也不限于先行词和代词之间的照应关系,也可能是一个先行词组或一个句子,甚至一个语段和另一个词,例如副词或词组之间的指代关系。

本文讨论用函子—变元结构转换器(FAST)对篇章层次中指代关系的归结,特别考虑到在英、德语作为源语,汉语作为目标语的智能式机器翻译中的应用,并对形态、句法、语义以及带有知识库的各层次处理的结果作了分析比较。

指代关系三个层次的处理为:(1)形态语法层;(2)形态语法加语义层;(3)形态语法语义加知识库层。在处理的过程中特别考虑到候选先行词和指代关系的下列诸因素:(1)角色一致因素;(2)符合一致因素,这里主要指的是形态符合;(3)结合原则,例如紧靠旁边的变元不能考虑作为先行词;(4)靠近因素,但应把结合原则作为“禁区”排除在外;(5)主题语义优先因素;(6)句子主题优先因素;(7)概念一致因素等。然后用计算积分高低优先原则选取候选先行词/组。

全文机器翻译指代关系通过 FAST 和逐渐完善的知识库必定能在可见的将来获得高质量的译文结果。同时能较好解决隐含、省略、比喻等公认较难处理的问题。应用 FAST 还有一个优点,即可用 AI-语言 Prolog 极方便地实现,最新发表的由 PDC Prolog 发展而来的 Visual Prolog for Windows 是作者所设计的全自动德汉全文翻译系统所乐于采用的方便工具。