
基于统计的汉英句子对齐研究*

吕学强 李清隐 任飞亮 姚天顺

(东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

E-mail: studystrong@sohu.com

摘要: 翻译比较规范的汉英语料适合用统计方法实现句子对齐,但评价函数中的参数计算不能采用处理印欧语种的方法,而要针对汉英语种的特点进行调整。在已有两种评价函数的基础上,文中又提出五种评价函数,并进行了对比研究。

关键词: 双语语料; 句子对齐; 评价函数; 参数计算

作者简介: 吕学强(1970—),男,辽宁抚顺人,博士生,讲师,主要研究领域为自然语言处理、机器翻译;李清隐(1980—),男,北京人,硕士生,主要研究领域为自然语言处理、机器翻译;任飞亮(1977—),男,山西大同人,硕士生,主要研究领域为自然语言处理、机器翻译;姚天顺(1934—),男,上海人,教授,博士生导师,主要研究领域为自然语言处理、机器翻译。

Towards Chinese-English Sentence Alignment Based on Statistical Method*

LV Xue-qiang Li Qing-yin Ren Fei-liang Yao Tian-shun

(School of Information Science and Engineering, Northeastern University Shenyang 110004, China.)

E-mail: studystrong@sohu.com

Abstract: The Chinese-English corpora that is translated literally is suitable to be aligned on sentence level with statistical method. But the method used to compute the parameters in processing Indo-European languages must be adjusted before applying to Chinese-English corpora. In this paper five evaluation functions are presented based on the former two evaluation functions, and the seven functions are researched comparably.

Keywords: bilingual corpora; sentence alignment; evaluation function; parameter computation

*本项目受到国家自然科学基金资助(60083006)。

*Supported by the National Natural Science Foundation of China under Grant No.60083006