

<<汉英平行语料库的平行处理>>

图书基本信息

书名：<<汉英平行语料库的平行处理>>

13位ISBN编号：9787510041105

10位ISBN编号：7510041104

出版时间：2011-11

出版时间：世界图书出版公司

作者：冯敏萱 著

页数：224

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<汉英平行语料库的平行处理>>

### 内容概要

汉英语料的平行处理技术，旨在利用双语信息以解决歧义现象。  
本课题研究了汉英平行语料中的未登录词识别、词性标注、词义标注及句法分析等层面的平行处理技术及其有效性，实现了汉—英双向平行处理。  
在词汇未对齐平行语料中，采用了基于个性规则的词性、词义消歧方法。  
精加工1000句对的汉英平行语料，可作为加工大规模平行语料的资源。  
平行处理技术可以有效解决单语处理时的一系列困难，有助于汉英机器翻译知识的自动获取。

## <<汉英平行语料库的平行处理>>

### 作者简介

冯敏萱，江苏南京人。  
语言学及应用语言学专业文学博士。  
现任南京师范大学文学院语言科技系讲师，获校第八届青年教师教学大赛“教学十佳”称号。  
主要研究方向为中文信息处理、语料库语言学。  
发表论文20篇，代表作有《带后缀“者”的派生词识别》、《英语人名的汉译名平行识别》和Parallel Processing Strategy for Segmentation Ambiguity。  
现主持江苏省社会科学基金项目1项，参与国家社会科学基金项目等4项。

## <<汉英平行语料库的平行处理>>

### 书籍目录

#### 第一章 引言

第一节 平行语料库的现状

第二节 本课题的研究意义及价值

第三节 本课题的研究内容及方法

#### 第二章 平行语料库概述

第一节 平行语料库的建设与发展

第二节 平行语料库的加工与利用

第三节 平行语料库与其他相关研究

第四节 本章小结

#### 第三章 词汇分析的平行处理

第一节 词汇分析研究的现状及难点

第二节 平行处理实验的设计

第三节 平行处理实验结果的分析

第四节 本章小结

#### 第四章 词性标注的平行处理

第一节 词性标注研究的现状及难点

第二节 平行处理实验的设计

第三节 平行处理实验结果的分析

第四节 本章小结

#### 第五章 词义标注的平行处理

第一节 词义标注研究的现状及难点

第二节 平行处理实验的设计

第三节 平行处理实验结果的分析

第四节 本章小结

#### 第六章 句法结构的平行处理

第一节 句法结构研究的现状及难点

第二节 平行处理实验的设计

第三节 平行处理实验结果的分析

第四节 本章小结

结语

参考文献

附录

附录一 PCCE1000观察语料示例

附录二 PCCE1000统计数据总表

附录三 术语索引

附录四 图表索引

后记

## <<汉英平行语料库的平行处理>>

### 章节摘录

机器翻译的实质在于输入的源语与目标语通过后台平行语料库建立对应关系，包括词汇和结构的对应。

从这个角度看，决定机器翻译质量最重要的两个因素是：一、后台语料库内容的丰富与否。

简单地说，就是语料库里有没有源语与目标语的对应，这将决定机器能否顺利翻译。

二、在后台的若干对应中，能不能选择出合适的对应，这个因素决定了翻译是否正确、到位。

因此，从原理上来讲，应该建立大型的双语平行语料库，使得需要翻译的句子在语料库中能够检索到相同或相近的句子并自动生成翻译产品或翻译的粗产品（余国良，2009）。

在这方面，已经有学者开始考虑同时利用两种语言的信息来进行研究，穗志方等（1998）提出，基于实例的汉英机器翻译系统，应该在比较两句整体结构相似的基础上进行语句相似度计算，除了利用骨架依存分析法，识别汉语谓词中心词，还根据汉英例句集中英语例句的谓语中心词来识别相应的汉语例句的谓语中心词，实验对3000个汉语单句进行谓语中心词的自动识别，正确率达到87.3%。

但是，随着研究的深入，人们发现基于平行语料库的翻译信息获取，对于机器翻译等有以下的不  
足：（1）获取的翻译信息的质量一定程度上依赖于平行语料的质量，即译文的质量；（2）新出现的词往往超前于平行语料库，从平行语料库中难以获取新词及其译词，而对于新词（未登录词）的处理能力对机器翻译却十分重要，尤其在专业领域，单词术语和多词术语的获取对机器翻译尤其重要。

.....

<<汉英平行语料库的平行处理>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>