

语言工程

LANGUAGE ENGINEERING

陈力为 袁琦 主编

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是全国第四届计算语言学联合学术会议(1997年8月15—17日,北京)的论文集。书中选录的62篇论文全文和12篇论文详摘是从大会征集到的113篇论文中精选出来的,内容包括下列七类:(1)汉语的句法和语义;(2)语料库建设与语料加工技术;(3)基于语料库的语言分析方法;(4)文本分析与生成;(5)机器翻译;(6)文本检索、自动文摘、文本校对;(7)智能型汉字输入方法和人机接口。

本书展现了近两年来我们计算语言学研究与应用的最新进展,对今后中文信息处理的基础研究和产品开发具有重要的参考价值。本书可供计算机、语言学等专业的科研人员、工程技术人员、大学教师和研究生选读。

书 名: 语言工程
作 者: 陈力为等
出 版: 清华大学出版社(北京清华大学校内, 邮编 100084)
印 刷: 清华大学印刷厂
发 行: 新华书店总店北京科技发行所
开 本: 787×1092 1/16 印张: 26 字数: 666 千字
版 次: 1997年8月第1版 1997年8月第1次印刷
书 号: ISBN 7-302-02618-1/TP·1342
印 数: 001~600
定 价: 60.00 元

全国第四届计算语言学联合学术会议组织情况

日期：1997年8月15日～17日

地点：北京清华大学

发起单位：中国中文信息学会
中国计算机学会
中国人工智能学会
北京市语言学会

赞助单位：电子工业部计算机与微电子发展研究中心
美国 IBM 公司中国研究中心
微软（中国）有限公司
台湾英业达集团电子技术有限公司
中国科学院计算机语言信息工程研究中心
清华大学计算机科学与技术系

组织单位：北京清华大学

大会名誉主席：陈力为
大会主席：姚天顺
大会副主席：刘开瑛 张 普

程序委员会主席：黄昌宁
副 主 席：俞士汶 董振东
委 员：（以姓氏拼音为序）
蔡莲红 曹右琦 陈群秀 傅爱平 嘎日迪 黄河燕 姬东鸿
林杏光 刘开瑛 鲁 川 茅于杭 孙茂松 许寿椿 姚天顺
易绵竹 苑春法 张 普 周 强

组织委员会主席：陈群秀
副 主 席：曹右琦 苑春法
委 员：张凤中 王建章 来春红 赵 军 李涓子 张 磊

语言工程

前言

由中国中文信息学会、中国计算机学会、中国人工智能学会和北京市语言学会共同发起的“全国第四届计算语言学联合学术会议”(JSCL-97)即将在北京清华大学举行。我们仅代表上述四个发起单位和全体与会代表向主办本届大会的东道主——北京清华大学的同仁表示由衷的感谢；并预祝大会取得圆满成功。

当今世界信息化浪潮席卷全球。Internet 的发展非常迅速，目前全世界 Internet 的用户数已达到 6 千万左右，并且每年仍在以极高的速度增加。中国的 Internet 用户已在两、三年中迅速发展到了 20 万户。在面向 21 世纪的网络通信时代，以文本信息的智能化处理为主要对象的语言工程，已成为国际上关注的热点。1996 年由联合国大学高等研究所 (UNU/IAS) 发起的通用网络语言 (UNL) 工程，将为实现网络时代的自然语言处理应用开辟广阔的领域。从另一方面看，当前计算机国际互联网是一个以英语为主导语言的网路，因此极大的限制了我国在网络上进行信息交换与实现资源共享的能力。对于我国计算语言学工作者来说，以汉语的句法语义分析为基础的应用研究是一项任重道远的任务。目前，根据市场需求牵引的指导原则，我们认为具有网络环境功能、基于自然语言处理技术的中文文本自动分类、信息检索、信息提取与生成，以及机器翻译等智能化技术将是我国“九五”期间语言工程的重点研究开发任务。

本届大会在全国各地和海外一共征集到 113 篇论文。大会的程序委员会本着严肃认真的态度对全部应征论文进行了评审，最后精选出 62 篇论文在本届大会上宣读交流。评选出的论文充分展示了国内计算语言学研究与应用的最新进展。注重实用性的研究与开发依然是它们的一个重要特点。本书中选录的论文大体上可归纳为如下 7 类：

- (1) 汉语的句法和语义
- (2) 语料库建设与语料加工技术
- (3) 基于语料库的语言分析方法
- (4) 文本分析与生成
- (5) 机器翻译
- (6) 文本检索、自动文摘、文本校对
- (7) 智能型汉字输入方法与人机接口

根据程序委员会的决定，本书除选录上述 62 篇论文的全文之外，还择优录用了另外 12 篇论文的详细摘要。

最后，衷心的感谢中国中文信息学会计算语言学专委会主任、本届大会程序委员会主席黄昌宁教授与陈群秀副教授在大会论文的组织 and 评审中所做出的贡献。

编者 1997年6月18日 于
电子部计算机与微电子发展研究中心(CCID)

语 言 工 程

目 录

论 文 全 文

I. 汉语的句法和语义

“请”的句法语义特征和兼语句的理解策略	刘永耕…………… 1
“甚”的语法功能的聚类分析与词性演变模型的建立	王秋月…………… 7
隐含型语句函子的求解方法	陆汝占 靳光瑾…………… 13
现代汉语槽关系研究	林杏光 张庆旭…………… 19
名词空位的控制性同指、照应性同指与词汇性同指	沈 阳…………… 25
自然语言理解的人本中心论	易绵竹 汤庆国…………… 31
对汉语中包含结果补语的句子分析策略	胡正微 胡继红…………… 37
汉语句子的信息安排及其句法后果 — 以“周遍句”为例	温锁林…………… 43
现代汉语自动词句模研究	苗传江…………… 49
现代汉语插入语研究	邢红兵…………… 55
语言的多媒体性与多媒体语言知识的作用	张 普…………… 61

II. 语料库建设与语料加工技术

实现 500 万词级《现代蒙古语文数据库》的主要措施	华沙宝…………… 69
《现代蒙古语词频统计系统》的设计与实现	
吉日木图 嘎日迪 赛音 白晓玲 达·巴特尔 阿拉腾敖道…………… 76	
几种常用汉语词典收词的统计分析	张永奎 齐新战…………… 82
大规模语料库中词语接续对的统计与分析	邱超捷 宋 柔 欧阳龙根…………… 88
并行语料库中的双语知识获取	周莉娜 廖雅国 俞士汶…………… 95
汉语句子的主题 — 主语标注	陈小荷 石定栩…………… 102
基于概率模型的语言的聚类方法：根据多语种语料库进行语言系统树的再构造	北研二…………… 109
万维网上的双语文本的获取及对齐	许玉祥…………… 115
基于英汉双语语料库的词汇自动对齐实验系统	李 竹…………… 121

III. 基于语料库的语言分析方法

局部统计在汉语未登录词辨识中应用和实现方法	沈达阳 孙茂松 黄昌宁	127
基于语料库统计方法在汉字文本识别应用中的若干问题		
	夏 莹 马少平 朱小燕 金奕江 姜 哲 常新功	133
基于特征关联度的汉语文本自动分类系统的设计与实现	张玥杰 姚天顺	139
汉语语义关联网的研究	苑春法 许 伟 黄昌宁	145
现代汉语紧缩结构定量分析	梁蕴华	151
从标注语料库中归纳语法规则：“V + N”序列实验分析	孙宏林	157
基于组合实例的双向优化聚类	李涓子 姬东鸿 黄昌宁	164

IV. 文本分析与生成

基于词典的汉语名词语义信息的自动分析与获取		
	翟高寿 张永奎 杨尔弘 查建中	170
汉语天气预报文本内容规划器的设计与实现	王 纤 姚天昉	176
歧义字段的最大概率切分算法	刘 挺 王开铸 姜兴海 邵艳秋	182
汉语句子分块研究	王厚峰 戴大为	188
汉语匹配分析算法的实现	周 强	194
中文文本歧义切分技术研究	郑家恒 刘开瑛	201
A Lexicalized Tree Adjoining Grammar for the Generation of Chinese Weather Forecasts	Yu-Fang Wang	207
基于维吾尔中心语驱动句法分析器的结构及其实现		
	玉素甫·艾白都拉 吾守尔·斯拉木	214

V. 机器翻译

面向受限领域的汉英口语翻译	王海峰 高 文 李 生	219
汉语方言机器翻译	张小衡 石定栩	225
实用型日语自动分词系统的算法及初步实现	陈群秀 孙勇 陈利人 莫林峰	232
汉语句子规范化处理策略的研究	宗成庆 章 森 陈肇雄 黄河燕	239
基于译后启发式信息的知识辅助获取系统：HIKAS	汪 冰 黄河燕 陈肇雄	245
朝语情态问题研究	毕玉德	251
基于依存关系分析的日语句法分析器	简幼良 唱红涛 王秀坤 黄德根	258
日汉机器翻译系统中的断段分析法研究与实现	曹睿妮 陈群秀	266
用概率方法处理汉英机器翻译系统中的歧义	刘 颖 常宝宝 刘 群 王 斌	274
基于软件构件的机器翻译研究方法	刘 群 张 祥	280
词的语义分类在汉英机器翻译中所起的作用以及难以处理的问题		
	詹卫东 刘 群	286
机器翻译基于从属关系的语义分析	傅爱平	292
带有对译信息的惯用表示的收集	田中康仁	298

VI. 文本检索、自动文摘、文本校对

基于选择生成文摘法的自动文摘系统研究与实现	杨晓兰 宋 帆 钟义信..... 313
汉语文本校对字词级查错处理的研究	孙 才 罗振声..... 319
中文校对系统中的修改建议提供算法	郭志立 裘照明..... 325
汉语文摘系统中文本结构的自动分析	薛翠芳 李晓黎 郭炳炎..... 332
中文校对软件标准评测系统的构造	郑 佩 杨力平 裘照明..... 338

VII. 智能型汉字输入方法与人机接口

语句拼音 — 汉字转换的智能处理	章 森 宗成庆 陈肇雄 黄河燕..... 345
以杂讯通道 — 讯息重建模型自动评测语言处理系统	张照煌..... 351
语音在人机接口中的作用和特点	方志刚 吴晓波..... 357
汉语查询可移植接口的设计与分析	邰晓英 童 颖 徐洁磐..... 363
容错型拼音汉字转换系统 — 一些调查及基本方案	(日) 叶 斌 应江黔 川上裕忠 松本忠博 後藤宗弘..... 369
计算机汉字语句输入的最短码长	王 轩 王晓龙 377
基于扩充 LR 分析的数据库汉语接口的句法分析	王力红 383
基于短语分析和动态语义搭配的拼音汉字智能转换	张小衡 陆发兴 389

论 文 详 摘

基于类比的计算语义处理机制	刘海涛..... 396
有限词汇、非特定人关键词识别系统	温艳红 李 维 张子范 张吉庆 王 粤..... 397
会议公告信息抽取系统	付宝宇 李侯运 吕克林..... 398
无词典德语自动分析	孙介铭..... 399
专用机器翻译系统 SCDOS 的技术研究与实现	张琳琳 王能忠..... 400
意义段划分问题研究	刘 挺 吴 岩 王开铸 邵艳秋..... 401
一种应用句法分析技术的汉语自动分词模型	单玉秋..... 402
基本词典与专业词典的关系	刘志杰 刘 倬..... 403
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..... 404
汉英机器翻译中的简单语境处理	刘海军 陈肇雄 黄河燕..... 405
维吾尔语词法分析器的探讨	
玉素甫·艾白都拉 吾守尔·斯拉木 赛依提·阿不都拉..... 406	
实用型汉英机器翻译系统的研究与实现	孙广范..... 407