

计算语言学进展与应用

ADVANCES AND APPLICATIONS ON COMPUTATIONAL LINGUISTICS

陈力为 袁琦 主编

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是全国第三届计算语言学联合学术会议(1995年11月5—7日,上海)的论文集。书中收入的56篇论文全文和15篇论文详摘是从大会征集到的137篇论文中精选出来的。内容包括:汉语的句法和语义、语料库建设与语料加工技术、基于语料库的语言分析方法、句法分析与篇章理解、机器词典、机器翻译系统及其评测方法、机器翻译方法研究、文本检索与自动文摘、智能型汉字输入方法等九类。

本书展现了近两年来我国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。

计算语言学进展与应用
ADVANCES AND APPLICATIONS ON
COMPUTATIONAL LINGUISTICS

陈力为 袁琦 主编

☆

清华大学出版社出版

北京 清华园

清华大学印刷厂印刷

☆

开本: 787×1092 1/16 印张: 23.5 字数: 602千字

1995年10月第1版 1995年10月第1次印刷

印数: 001—600

ISBN 7-302-01998-3/TP·922

定价: 56.00元

全国第三届计算语言学联合学术会议组织情况

时间:1995年11月5日—7日 •

地点:上海交通大学

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

美国微软公司北京代表处

台湾英叶达集团电子有限公司

科智集团公司

清华大学计算机系

组织单位:上海交通大学

大会名誉主席:陈力为

大会主席:刘 倬

副主席:盛焕烨 刘开瑛

程序委员会主席:黄昌宁

副主席:张普 林杏光

委员:(以姓氏拼音为序)

曹右琦 陈群秀 陈肇雄 蔡莲红 冯志伟 傅爱平

刘 倬 陆汝占 姚天昉 俞士汶 苑春法

组织委员会主席:盛焕烨

副主席:曹右琦 陈群秀

秘书长:陈群秀 姚天昉

委员:何厚存 黄人杰 王永成 陈梅菊

计算语言学进展与应用

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“全国第三届计算语言学联合学术会议”(JSCL-95)即将在上海交通大学举行,我谨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—上海交通大学的同仁表示由衷的感谢,并预祝大会取得圆满成功。

本届大会在全国各地和海外一共征集到 137 篇论文,它们分别来自 63 个单位,其中包括国内的 41 所高等院校和 9 个研究院所。来稿最多的前 5 个单位分别是:清华大学(21 篇),上海交通大学(14 篇),哈尔滨工业大学(8 篇),东北大学(7 篇)和山西大学(7 篇)。来稿数量之多和来源之广泛都是前两届大会所未见的,这说明计算语言学的研究与应用已在我国各地深深扎根,国内的研究队伍正在日趋壮大。

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。每篇论文都事先经过两位委员的书面评审,然后将综合评审结果提交 8 月 19 日召开的程序委员会全体会议复议,最后精选出 56 篇论文(占应征论文总数的 41%)在本届大会上宣读交流。通过评选的论文大体上可归为如下 9 类:

- (1)汉语的句法和语义,8 篇;
- (2)语料库建设与语料加工技术,7 篇;
- (3)基于语料库的语言分析方法,7 篇;
- (4)句法分析与篇章理解,8 篇;
- (5)机器词典,6 篇;
- (6)机器翻译系统及其评测方法,4 篇;
- (7)机器翻译方法研究,11 篇;
- (8)文本检索与自动文摘,3 篇;
- (9)智能型汉字输入方法,2 篇。

根据程序委员会的决定,本书除收入上述 56 篇论文的全文以外,还择优录用了另外 15 篇论文的详细摘要。

在本届大会上交流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。例如,汉语自动分词、词性自动标注、短语边界辨识和双语平行文本的自动对齐等大规模语料的加工技术与标注工具,两年来已取得了长足的进步,为基于语料库的语言知识获取和语言分析中的语料库方法奠定了良好的基础。又如,象机器词典、机器翻译、汉语分析、汉语文本自动校对、全文检索和自动文摘等一批具有广阔应用前景的实用化系统,也取得了令人瞩目的成果。相比之下,国内基础理论的研究仍显薄弱,同国外同行相比我们还有较大的差距,这种状况应当引起我们的重视。

在1994年8月日本京都召开的国际计算语言学大会(COLING-94)上,国内有四篇学术论文入选。与我们自己相比,这已是很大的进步;但同周边国家或地区相比,我们的成绩仍不能令人满意。我殷切希望在明年召开的COLING-96上,国内能有更多的优秀论文入选。

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在大会论文的组织 and 评审中所做出的贡献。

编者 1995年9月16日

计算语言学进展与应用

目 录

论 文 全 文

I. 汉语的句法和语义

- “VP+的+N”结构 N 省略的条件描述 陆汝占 靳光瑾……1
汉语的二元语义关系模型 万建成……7
一种汉语句子语义表达模型及其实现策略 黄海 姚天昉…… 12
基于格关系的现代汉语述语动词分类系统 林杏光…… 18
现代汉语述语动词框架研究 张庆旭…… 23
动宾短语中的格关系及其理解 胡正微…… 30
汉语名词的内在知识及其表示 姬东鸿 黄昌宁 赵军…… 37
动词套合与动词重置 沈阳…… 43

II. 语料库建设与语料加工技术

- 一个人机互助的汉语语料库多级加工处理系统 CCMP 周强 俞士汶…… 50
汉语词切分及词性自动标注一体化方法 白拴虎…… 56
基于长度算法的中—英双语文本对齐的试验 刘昕 周明 黄昌宁…… 62
中国地名的自动辨识 沈达阳 孙茂松 黄昌宁…… 68
现代汉语语料库分词中的若干问题 孙宏林…… 75
汉语自动分词中的歧义问题 侯敏 孙建军 陈肇雄…… 81
单词 bi-gram 统计分析的一种自适应算法 应江黔 兵藤安昭 池田尚志…… 88

III. 基于语料库的语言分析方法

基于语料库的汉语短语边界划分的研究	张国焯 郁梅 王小华……	94
汉语文本自动查错与确认纠错系统的研究	慕勇 孙才 罗振声……	100
中文字字同现概率统计及应用		
	夏莹 常新功 马少平 朱小燕 金奕江……	106
双语语料库研究中的若干统计结果		
	赵铁军 李生 王海峰 杨彦 毛永明 吴学伟……	112
基于语料库的中文最长名词短语的自动抽取		
	李文捷 周明 潘海华 林耀燊 黄锦辉……	119
从大规模语料库看双语词典释义的适用性和客观性原则		
	解建和 谭家祥 肖建平……	125
现代汉语中二字复合词的构词格式研究——基于汉语		
语素数据库的研究之一	苑春法 黄昌宁 周思宇……	133

IV. 句法分析与篇章理解

现代汉语分析系统 UCAS 的研究与实现	朱靖波 王宝库 姚天顺……	139
歧义消解策略初探		冯志伟……146
一种汉语句子间相似度的度量算法及其实现		
	张民 李生 赵铁军 周明……	152
汉语句型成分自动分析中谓语识别策略的研究		
	罗振声 孙长健 孙才……	159
自然语言理解及在数据查询中的应用	徐九韵 叶延滨 王新民……	165
分析汉语句子的解释学习方法	李晓黎 郭炳炎……	171
TBIRS 中的信息抽取方法	张永奎 李红涛……	177
清华联机手写汉字识别系统——中文人机接口的智能化		
	杨泽红 夏莹……	183

V. 机器词典

汉语机用同义词词典的建造技术	杨尔弘 刘开瑛……	189
利用基于语义信息的名词识别方法来建造现代汉语名词机器词典		
	翟高寿 张永奎 杨尔弘……	195
机器词典的动态管理技术	孙健 胡海文 王启祥……	201
现代汉语述语动词机器词典的研究(二探)		陈群秀……207

配价制导的德汉机译系统中词素词典的微结构	许玉祥	柴佩琪	……213		
汉语语音识别用电子词典的自动建立方法研究	张树武	黄泰翼	徐波	马斌	……219
VI. 机器翻译系统及其评测方法					
通用德汉机器翻译系统	谢金宝	何星	梁桂森	……225	
基于动词配价模式日汉机器翻译系统的设计与实现		孙勇	陈群秀	……232	
基于语料库的机器翻译研究环境的设计	王挺	陈火旺	史晓东	李志伟	……239
面向用户的 MT 系统评测方法			傅爱平	……245	
VII. 机器翻译方法研究					
基于 PG 文法的俄语动词句型描述	易绵竹	薛思奎	……251		
基于实例的日汉机器翻译方法中的句子相似度计算研究	陈利人	陈群秀	……257		
基于双语语料库和规则库的德汉复合句的转换生成	邸海燕	柴佩琪	许玉祥	……264	
代换法——基本句型和短语的翻译方法		刘孝叔	……271		
英语语言中的兼类问题及英汉机器翻译系统的对策	刘志杰	刘倬	……277		
清华英汉机译系统中的歧义处理	刘月荣	苏玉宏	陈圣信	……283	
机器翻译中的消歧策略	朱靖波	王宝库	姚天顺	……289	
经验知识在机器翻译系统中的应用			任福继	……296	
KDD 和基于语料库的机译系统	范建华	李德毅	张亚非	……302	
机器翻译中的智能规则系统	程学旗	唐泓英	姚天顺	……308	
THECMT 系统中规则的处理机制	苏玉宏	刘月荣	……315		
VIII. 文本检索与自动文摘					
非受限域中自动文摘系统的研究与实例	王开铸	李俊杰	吴岩	……321	
汉语全文检索中的义项标注技术研究			刘开瑛	……327	
一个新的中文信息分类检索模型与系统实现	彭甫阳	何新贵	……333		

IX. 智能型汉字输入方法

- 一个拼音汉字自动转换系统的设计与实现 成华 尹宝林……340
统计分析方法相结合的小型拼音——汉字智能转换系统 PCTS
郑嵘 梁晋清 何厚存……346

论 文 详 摘

- 计划语言和计算语言学 刘海涛……352
对现代汉语语义格的认识与划分 贾彦德……353
仿人汉英口语机译系统 王永成……354
智能型俄汉机器翻译系统中的源语言分析
李向东 周清波 周天荣……355
汉英机器翻译 THCEMT Version 1.0 的设计
严浩 顾宇杰 陈圣信……356
智能式机器翻译中指代关系的 FAST 处理 孙介铭……357
中文自动校对系统的研究与实践 刘挺 王开铸……358
无词典自动分词的研究 王开铸 李俊杰 吴岩……359
词典规则设计中的分段和继承思想 傅奇芳 聂昱 王斌 戴大为……360
对“被”字句进行汉语句法分析的研究
聂昱 傅奇芳 王斌 戴大为……361
《拼音—汉字变换输入法》的性能测试 何厚存……362
基于德语语料库词性标注和统计方法的研究 姚天昉 林莉……363
自动词细分类 苗传江……364
机译评估方法评述 罗爱荣 段慧明……365
用 n 元语法的统计确定短语的边界 郭志立 苑春法 黄昌宁……366