

<<汉语虚词知识库的建设>>

图书基本信息

书名：<<汉语虚词知识库的建设>>

13位ISBN编号：9787562238850

10位ISBN编号：7562238855

出版时间：2009-5

出版时间：华中师范大学出版社

作者：刘云

页数：324

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>

<<汉语虚词知识库的建设>>

前言

随着历史的发展,社会的进步,科技的发达,语言学在整个世界范围内越来越展示出强大的活力和能量。

中国语言学是世界语言学的重要组成部分。

为了对中国语言学事业有所推动,我们组织撰写“华中语学论库”。

作为专用名称,这里的“语学”主要指汉语语言学,近期的15年时间里以现代汉语语法专题研究为重点。

“语学论库”,这是汉语语言学的一个系统工程,如果将来主客观条件具备,在研究范围上可以不断扩大,在研究时间上可以无限延展,在研究队伍上可以辈辈交接,代代传承。

“华中”一词,既跟研究队伍的华中群体相关,又跟华中师范大学出版社的名称相关。

汉语语言学源远流长。

千百年来,特别是《马氏文通》出版以来,尤其是20世纪70年代之后,由于一代代学者的不懈努力,汉语语言学沿着“创业——拓新——发展”的轨道不断推进。

目前,汉语语言学所统括的汉语语法学、汉语语音学、汉语方言学、汉语词汇学、汉语语用学等等学科,都已出现了初步繁荣的喜人局面。

<<汉语虚词知识库的建设>>

内容概要

《汉语虚词知识库的建设》共分五章。

第一章“面向信息处理的语言研究”主要介绍了中文信息处理的发展与成就、汉语自动分析的难点、计算机背景下的汉语语法研究以及“句管控”与汉语信息处理等内容。

第二章“虚词知识库的建设”主要介绍了虚词与虚词研究概貌、汉语虚词知识库的重要性、虚词知识库建设的难点以及虚词知识库建设的构想等内容。

第三章“虚词词典的内容”主要介绍了副词机器词典、连词机器词典和介词机器词典等内容。

第四章“复句层次和关系的自动分析”主要介绍了复句自动分析的目标和意义、难点和对策、二重复句的自动分析、复句关系词语的离析度以及关系词语驱动的复句关系和层次自动判定。

第五章是全书的结语，主要是对已有研究工作的总结和进一步的研究计划。

<<汉语虚词知识库的建设>>

书籍目录

序前言第一章 面向信息处理的语言研究第一节 中文信息处理的发展与成就一、引言二、中文信息处理的发展三、中文信息处理的成就第二节 汉语自动分析的难点一、引言二、词法分析的难点三、句法分析的难点四、转换过程的难点五、小结第三节 计算机背景下的汉语语法研究一、研究对象的转换二、研究目的的转换三、研究手段的转换四、研究视角的转换五、小结第四节 “句管控”与汉语信息处理一、引言二、词语切分与词性标注三、句法与语义四、理论与实践五、小结第二章 虚词知识库的建设第一节 虚词与虚词研究一、虚词概貌及其作用二、虚词在通用语料库中的使用情况三、虚词研究第二节 汉语虚词知识库的重要性一、词法分析中的作用二、句法分析中的作用三、具体应用中的作用第三节 虚词知识库建设的难点一、虚词语法意义的概括二、虚词用法差异的揭示第四节 虚词知识库的建设一、引言二、“三位一体”的虚词知识库建设三、小结第三章 虚词词典的内容第一节 副词机器词典一、副词及其分类二、副词字段的设立三、副词目录四、副词机器词典摘录第二节 连词机器词典一、连词及其研究二、连词字段的设立三、连词目录四、连词机器词典摘录第三节 介词机器词典一、介词及其分类二、介词字段的设立三、介词目录四、介词机器词典摘录第四章 复句层次和关系的自动分析第一节 复句关系和层次自动分析的目标和意义一、汉语复句自动分析的目标二、复句自动分析的意义第二节 汉语复句自动分析的难点与对策一、复句自动分析的难点二、复句自动分析的策略三、单复句的确认对策第三节 二重复句自动划分研究一、引言二、复句关系词语的包孕机制三、小结第四节 复句关系词语离析度一、引言二、复句关系词语的离析度三、复句关系词语离析能力差异分析第五节 关系词语驱动的复句关系和层次自动判定一、引言二、复句关系词语的作用三、复句关系词语的表现四、复句自动分析的流程第五章 结语第一节 研究工作回眸第二节 研究工作设想附录1 复句层次和关系标注语料库样例及说明附录2 复句层次和关系自动分析系统后记

<<汉语虚词知识库的建设>>

章节摘录

1991年,北京语言学院汉语水平考试中心组成研制小组,以国内影响较大的七个动态性频度统计词典、词表和字表为主要依据,编制了《汉语水平词汇与汉字等级大纲》。

其编制总原则为:(1)多学科定量统计与群体性定性分析相结合;(2)语言学科学原财与对外汉语教学需要相结合。

共得常用词8822个,分成甲、乙、丙、丁四级,写进《大纲》(修订稿)。

这是一次较为典型的把主、客观标准有机结合起来选择汉语常用词的实践。

由于20世纪90年代初期有多个常用字词统计结果的推出,随后常用字词的统计相对沉寂了几年。

随着网络等媒体的兴盛,大规模字词统计又开始复兴。

进入21世纪后,先后有两次大规模的词汇统计,分别是《中国语言生活状况报告(2005)》和《中国语言生活状况报告(2006)》。

这两次调查统计有三个重要的特点:一是规模大。

《中国语言生活状况报告(2005)》调查语料分为平面媒体、有声媒体、网络媒体三种,共计892034个文本文件,909429700字符次,其中汉字出现732143010字次。

《中国语言生活状况报告(2006)》调查的语料分为平面媒体、有声媒体、网络媒体三种,共计1311749个文本文件,1170367879字符次,其中汉字出现978994406字次,规模达10亿次,这是迄今为止最大规模的汉语字词统计。

二是统计项目全面。

两次调查的主要对象是汉字和词语,调查项目有“频次、频率、累加频率、出现文本数”等。

三是成果丰硕。

<<汉语虚词知识库的建设>>

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>