

<<大规模真实文本汉语字母词语考察研究>>

图书基本信息

书名：<<大规模真实文本汉语字母词语考察研究>>

13位ISBN编号：9787561534892

10位ISBN编号：7561534892

出版时间：2010-4

出版时间：厦门大学出版社

作者：郑泽芝

页数：295

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>

## 前言

郑泽芝博士的专著《大规模真实文本汉语字母词语考察研究》是她在博士论文的基础上修改增订而成的，她于2005年毕业获博士学位，而本书在她毕业四年之后才提上出版日程，2005年后她又在《语言文字应用》、《中文信息学报》、《语言科学》、《厦门大学学报》等学术期刊和“中国语言学会”、“中国社会语言学国际会议”、“汉语和汉字关系国际会议”等学术会议上陆续报告了多篇相关论文。

从2004年到2009年，她共发表了近20篇有关字母词语的研究论文，在她读博期间，字母词语研究就得到国家语言资源监测与研究中心项目“中国主流报纸大规模真实文本字母词语考察”的支持，毕业后她到厦门大学工作，又继续申请了福建省社科规划项目“基于大规模真实文本的汉语字母词语分类考察研究”以及厦门大学科研启动基金的资助。

确定研究目标之后，“咬定青山不放松”，持之以恒。

她不急于出书，而是先发表阶段性成果，既有先发优势，也得以在深入观察研究分析之后再系统全面论述，这是泽芝的研究特点之一。

她抢占先机抓紧不放的特点，我还有一个小故事作佐证：2003年10月初，我去国家语委参加关于“字母词”的讨论会，泽芝的字母词研究还处在收集材料和预研阶段（所以她在《后记》中说她对字母词语的研究“始于2004年”）。

离开会还有一两天时间，我问她：“你能不能给我一些最基本的统计数据？”

比如到底在我们的报纸上字母词语的实际使用情况是多少？

《人民日报》上是多少？

《北京青年报》又是多少？

## <<大规模真实文本汉语字母词语考察研究>>

### 内容概要

汉语中字母词语的使用情况已经无法仅仅用外来语来涵盖，比如：HSK、GB、3C、阿Q等。而且这种语言现象在汉语中的使用，呈泛滥的态势，有的字母串已经产生了很强的构词语能力，字母参与构词也已屡见不鲜，它们已经是汉语中不容忽视的一种新语言现象。它们的影响还不仅限于此，对于汉语信息处理而言，字母词语不仅是汉语中的一类新词语，而且其中含有大量的各领域术语，其中许多术语正在进入汉语语言社会的各个层面，它们已经成了情报翻译、自动分词、信息检索、机器翻译中不可低估的影响因素。

语言学界见仁见智，态度各异，有人主张保卫祖国语言的纯洁性，有人为字母词语叫好。本书认为字母词语在汉语中的使用不是孤立的，只有经过大规模真实文本的考察，才可以用事实说话，并据实给予规范和制定相关的政策。

本书采用语料库技术与内省相结合、定量和定性相结合的研究方法，以大规模真实文本为研究基础，历时5年，跟踪考察了近两亿真实文本语料。

界定了文字字母；对字母词语所用字符进行了考察和归纳，界定了字母词语；在共时截面上对字母词语在主流报纸媒体的使用状况进行了全面客观的描述和分析；对字母词语的主题领域分布、用法分布状况进行了量化分析研究；从不同角度对字母词语分类、构成成分进行了研究；从历时发展的角度对字母词语演化状况进行了跟踪研究；从语言信息处理的角度对字母词语的统计特征和字母词语的自动识别进行了研究；对通用字母词语的量化提取进行了探索；对字母词语将给汉语文字系统带来怎样的影响进行了探索等等。

客观真实地描写和分析了字母词语的各种情况，取得了大量第一手的资料。

全书共分15章，书后附有作者整理的字母词语研究文献索引录，双语对释字母词语表，5年历时字母词语表，《人民日报》、《北京青年报》、《羊城晚报》共有字母词语表等附录。

本书对社会语言学、汉语词汇学、新词语、外来语的研究有重要的参考价值，是中文信息处理领域的人员和相关单位关注的对象，同时也可作为政府有关部门制订相关政策提供参考。

## 书籍目录

序 / 张普第一章 绪论1 字母词语的界定2 字母词语与真实文本3 本书的结构第二章 主流报纸媒体字母词语的使用状况1 语料的选择2 文本中使用的非汉字符3 外文字母在汉语文本中的使用情况4 媒体字母词语的量化概况5 各报纸媒体字母词语概况5.1 《人民日报》字母词语概况5.2 《北京青年报》字母词语概况5.3 《羊城晚报》字母词语概况6 媒体字母词语的一次性使用情况7 不同性质媒体中字母词语的使用情况8 媒体共有字母词语情况9 小结第三章 字母词语中的标点1 字母词语中的标点符号2 字母词语中标点符号出现情况及问题2.1 字母词语中标点的出现情况2.2 与标点相关问题的处理3 小结第四章 字母词语中的数字1 数字在字母前的情形2 数字在字母后的情形3 含数字字母词语之管见第五章 单字母、多字母、光杆字母词语与字母词语并列结构1 单字母字母词语2 多字母字母词语考察3 光杆字母词语与外语原文缩略语3.1 光杆字母词语使用状况3.2 字母词语原文缩略和汉语拼音成词问题4 字母词语的并列结构第六章 双语对释字母词语1 双语对释字母词语在文本中的使用情况概述2 对释汉字串的分布考察3 双语对释字母词语语形历时演化分析4 双语对释字母词语使用方式演化4.1 共有双语对释字母词语I型、型分布4.2 双语对释全称简称形式的演化4.3 字母词语中的新词语5 小结第七章 字母词语领域分布特点1 媒体主题领域聚类2 字母词语在各主题领域的量化分布规律2.1 一般字母词语领域分布情况2.2 双语对释字母词语领域分布情况3 领域共有字母词语的情况分析4 各主题域独有字母词语情况分析5 小结第八章 字母词语有多少种用法1 字母词语的用法分类2 字母词语的用法分布情况3 字母词语领域和用法交叉分布情况4 小结第九章 字母词语的演化态势1 字母词语的使用数量的历时演化2 字母词语一次性使用的历时变化3 字母词语变化程度及变化趋势4 一般字母词语的演化情况分析4.1 数量变化4.2 字母词语搭配汉字的情况4.3 共有—般词语的分类考察5 字母词语稳定性第十章 字母词语的使用与规范1 字母词语的层次划分1.1 字母词语三圈划分1.2 外语原词、计算机相关词语在文本中的使用情况2 字母词语来源分析3 外文字母使用例析4 字母词语不规范情况分类5 字母词语的词性问题6 字母词语的归化7 关于字母词语的几点建议第十一章 字母词语对汉语文字系统的影响1 语言与文字2 从外来语到字母词语2.1 传统外来语的影响2.2 字母词语与汉语言系统3 汉语文字系统4 汉字系统可否发生变化5 小结第十二章 通用字母词语提取研究1 通用词语的各种量化属性2 字母词语量化属性观察分析3 通用字母词语提取算法3.1 绝对指标提取方法3.2 通用度方法3.3 散布均匀度方法……第十三章 字母词语的自动识别第十四章 字母词语的社会接受程度考察第十五章 结束语附录后记

## 章节摘录

面的、口头的、日常生活中街市门面、各种标牌的……语言的使用涉及生活的方方面面，边边角角，各个层面，各个年龄段。

我们只是从一个侧面考察了字母词语的使用情况，虽说一滴水可以折射大海，但毕竟客观实际中还存在大量应该但却没有被反映的现实。

如果可以，我们应该从更多的途径考察更多的媒体中外文字母的使用情况，以便更全面地反映外文字母在汉语中的使用状况。

自动提取和人工校对的结果不可谓不丰富，十几万条次的工程字母词语，足以产生一些很有价值的结论，但因投入的时间、人力、物力等原因，我们目前不可能对它们一一标记语义、领域、结构、语法等属性以供后人更好地研究它们，或为国家语言生活提供更好的参考。

面对如此大量的条目进行人工归类分析，工程非常巨大，我们认为必须对计算机辅助字母词语自动聚类进行研究，以减少人工投入；工程中还存在语言学需要解决的问题、各领域术语的映射问题，如：语言学对专名描述就不够精细，拿来说事时往往捉襟见肘；语言学的专名与语言信息处理学界的命名实体之间是一种什么样的关系，我们弄不清楚或者就没有想去弄清楚，这一点也反映出语言学和信息处理研究上缺乏沟通和互补。

因而，我们对《人民日报》的提取结果进行的分类分析，也只是粗浅的、不太成熟的，但已经可以初见字母词语在汉语中使用状况的端倪。

考察中我们发现字母词语一次性出现的比率达到甚至超过了70%，这种现象引起了我们的注意，我们发现这种现象除了与字母词语是新词语、出现的次数低有关系外，也和它们多为产品名、公司名、各种术语协定名有关。

这类对象只在各种专业或行业文本中高频出现。

还有一个更重要的原因就是字母词语缺乏使用规范，造成字母词语使用的混乱状态，同一个对象有多个不同的指称，如“全国男篮甲A联赛（CBA）”这个比较普通的字母词语在文本中还有“全国男子篮球甲A联赛（CBA）”这一孪生兄弟，所以对字母词语结果进行。

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>