

<<史记字频研究>>

图书基本信息

书名：<<史记字频研究>>

13位ISBN编号：9787100045827

10位ISBN编号：7100045827

出版时间：2006-3

出版单位：商务印书馆

作者：李波

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>

<<史记字频研究>>

内容概要

本书利用《史记》文献语料库和由语料库得到的文献数字化信息，全面描述并分析《史记》用字的量和位。

早在上个世纪80年代初期，我就着手筹划编纂《史记索引》。

传统的索引是依据选定的文献底本，手工断句，剪贴编排而成。

燕京哈佛学社引得编纂处的那些引得，叶圣陶先生的《十三经索引》，都是手工编纂索引的先例。

前辈编纂索引的经验告诫我，人工为20万字以上的文献做详尽的索引是异常困难的，手工操作必然产生的差错率甚至会断送千辛万苦的努力。

计算机的使用给编纂大型文献索引带来转机。

当时在黑龙江计算机领域已经崭露头角的李晓光告诉我，使用计算机可以代替人工编纂详尽的文献索引。

工作程序是先做计算机机读本，制成数据库，再进行一系列运算，而后即可完成编纂索引的工作。

要给《史记》这样一部大部头的文献编纂索引，就非得使用计算机不可了。

那时，所能见到的只是相当原始的微型计算机。

即使是那样的微型机，轮到个人使用也很罕见。

为了工作，我必须求得一台计算机，后来好友王新生竟然从一个小公司里借出来一台PC机为我专用。

接着，我们寻求到了一块汉卡，那是如今已经很少有人知道的“仓颉输入法”专用卡。

最初的机读本《史记》就是借助仓颉汉卡实现的。

在制作《史记》机读本和对《史记》语言进行分析的时候，问题总是不断地提出来，用东北话说是“零揪”，这对编程工作来说，无疑是反反复复的折磨。

合作伙伴李晓光却总是能在艰苦的条件下，对我提出的各种要求给出完满的解决方案。

当时微型计算机的处理速度和容量，与现在普遍使用的“奔4”相比可能相差千倍有余，处理《史记》这样一个大部头文献，仅在计算机处理能力上的难度就可想而知了。

但是，对计算机的新鲜感和用计算机处理大型古籍的憧憬，驱使我们做成这件事。

不懈的努力，终于有了回报，我们的机读本《史记》和《史记》语料库都达到了令人满意的地步。

1987年5月27日举行的“使用IBMPC—XT微型电子计算机处理大型古代文献——《史记》鉴定会”上，我们的工作得到专家们肯定。

鉴定会在哈尔滨师范大学召开，邀请的专家有：郭锡良（北京大学教授，鉴定会主任委员）、吴几康（中科院计算机研究所研究员）、刘涌泉（中国中文信息协会研究员）、吕冀平（黑龙江大学教授）、花栅（哈尔滨船舶工程学院教授）、郭福顺（哈尔滨工业大学教授）、曹先擢（国家语言文字工作委员会研究员）、曹乃木（商务印书馆编审）、傅永和（国家语言文字工作委员会副研究员）、叶长阴（哈尔滨师范大学教授）、刘博文（黑龙江电子计算机服务公司工程师）、张晋梗（哈尔滨师范大学副教授）、李传靖（黑龙江电子计算机服务公司工程师）、郑继志（黑龙江电子研究所工程师）。

鉴定会的成果报告、技术审查报告、用户意见和鉴定证书都说明了这个成果的可靠性和先进性。

专家们认为，在实现中文文献检索的现代化和计算机的功能开发方面，这项工作处于全国领先地位。鉴定会以后，《史记索引》顺利出版了。

这些成果，后来获得了黑龙江省科委的计算机软件奖、社会科学编著奖等奖项。

郭锡良先生在鉴定会座谈时曾讲，《史记》数据库的实现，就像是挖到了一座金山，吃不完用不尽，对语言研究来说应该继续往深里挖掘开发。

我觉得可以把文献语料库比作冰山，当这座大冰山浮出水面以后，在语言学的研究领域中出现生机，语言各个层面的研究差不多都可以在语料库的环境里实现。

就《史记》研究而言，借助《史记》语料库提供的大量的文献数字化信息，以汉字为计量单位，可以极方便地从定量和穷尽两个方面对《史记》作全面分析，进而对《史记》的语言开展多方面的研究，并获取那些单凭抽样调查不可能得出的结论。

然而，这只是冰山的一角。

<<史记字频研究>>

文献语料库的出现，使语言研究领域需要做的和能够做的工作一下子增加了许多倍，包括综合语料库的建设，数据信息的提取，计算方法的探求，等等。

使用文献语料库作语言研究，要有与之相适应的方法。

依据语料库作研究总是要着眼全局，所研究的内容又具有很强的相关性，为了说明一个问题，必须给出多种量的证明，否则，就不可能发挥语料库研究的优势，不能称其为语料库语言研究。

这样，一部文献的语料库常显得孤立无援，必须有多部文献的语料库才能互相辅证。

所以这些年来我和几个同志坐下来，试着先搞一些文献语料库，同时也在学习研究汉语文献的运算方法，试图在开发语料库功能的同时解决老问题，发现新问题。

随着工作的向前推进和计算机功能的急剧拓展，我们陆续研制出了多部文献语料库，包括十三经、前四史、《国语》、《战国策》及《汉语大词典》辞目语料库等。

利用文献语料库进行研究的成果之一，就是编纂文献索引，多部索引已经由中国 / “ 播电视出版社出版了。

在出版这些索引的时候，我们总是从语言研究的角度考虑，不断更新这些索引的格式和内容，读者如果手头有这些索引，按时间顺序将其排列一下，加以比较，就会看到其中改进的情况。

本书的写作，源于戴昭铭先生的建议。

我本来想写一篇文章来表述对《史记》字频表的分析，介绍我们通过计算《史记》语料库分析《史记》字表的方法和结论。

戴昭铭先生认为，语料库是有利于大家的東西，字表又是借助语料库研究文献语言的大纲，与其只介绍研究的方法，不如用一本小书把研究方法和研究的结果都公布出来，与研究《史记》的同仁共享。

戴先生说得对。

我把这个想法跟一些同志谈过，他们都很赞同，还认为这项工作对一些大型文献语料库的研制和使用都会有意义。

这部书讲述了文献语料库的研制和在此基础上所做的工作，公布了《史记》字表的各个字频区的用字，并对其进行分析，阐述了字频表的功能。

这些工作是使用《史记》语料库研究《史记》语言的一部分基础工作。

从长远看，使用《史记》语料库研究《史记》语言，还是使用文献语料库研究文献语言的组成部分，是一项探索性很强的工作。

对《史记》语言的深入研究，以及对文献语言的研究工作，都需要更多的文献语料库；大型综合文献语料库的研制和使用，应该从现在开始开展起来。

<<史记字频研究>>

书籍目录

前言第1章 文献语料库和文献的数字化信息 1.1“例不十，法不立”的原则和穷尽性研究 1.2文献语料库的品格和研制思路 1.3汉字的计量优势 1.4《史记》语料库和文献的数字化信息 1.5文献用字的坐标系第2章 《史记》语料库的制作 2.1文献版本的选定 2.2机读本《史记》的制作 2.3语言研究的配套程序 2.4参比文献语料库第3章 《史记》分卷用字渊查 3.1分卷研究的意义 3.2《史记》130卷的分卷数据 3.3《史记》130卷用字数据分析第4章 《史记》用字概况和字频区的划分 4.1《史记》用字的概况 4.2《史记》语料库的数据提取 4.3《史记》字区的划分标准 4.4《史记》的字区 4.5《史记》与其他典籍数据的比较 4.6《史记》与《十三经》、现代汉语字频的比较 4.7《史记》用字分类研究的方法第5章 《史记》核心字区的分析 5.1《史记》的核心字区 5.2核心字的7种数据 5.3数据和内容的讨论 5.4数据和内容的验证 5.5结论第6章 《史记》高频字区的分析 6.1《史记》的高频字区 6.2高频字的7种数据 6.3数据和内容的讨论 6.4数据和内容的验证 6.5结论第7章 《史记》中频字区的分析 7.1《史记》的中频字区 7.2中频字的7种数据 7.3数据和内容的讨论 7.4数据和内容的验证 7.5结论第8章 《史记》低频字区的分析 8.1《史记》的低频字区 8.2低频字的7种数据 8.3数据和内容的讨论 8.4数据和内容的验证 8.5结论第9章 《史记》罕用字区的分析 9.1《史记》的罕用字区 9.2一用字的7种数据 9.3数据和内容的讨论 9.4结论第10章 《史记》的句长 10.1《史记》断句的依据 10.2《史记》分篇的句长统计表 10.3句长的有关数据结语参考文献

<<史记字频研究>>

编辑推荐

这部书讲述了文献语料库的研制和在此基础上所做的工作，公布了《史记》字表的各个字频区的用字，并对其进行分析，阐述了字频表的功能。

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>