

语言与数学



[语言与数学 下载链接1](#)

著者:冯志伟

出版者:世界图书出版公司

出版时间:2011-1

装帧:

isbn:9787510028069

如何实现语言的机器翻译？如何从海量的语言文字中抽取有用信息？如何利用自然语言

进行人机对话？自然语言与计算机程序语言是否一致？如何自动合成语音……信息时代对自然语言的处理使语言学与数学紧密结合在一起。

作者从事计算语言学研究几十年，潜心探索出语言符号的七种新特性。本书详细论述了这七种特性与数学的关系，反映了当前国内外语言与数学关系研究方面的最新成果。

语言符号的随机性与统计数学

语言符号的冗余性与随机过程

语言符号的离散性与集合论

语言符号的递归性与公理化方法

语言符号的层次性与图论

语言符号的非单元性与复杂特征的运算

语言符号的模糊性与模糊数学

作者介绍:

目录: 信息时代语言学研究的基础——读《语言与数学》有感
再版前言

绪言——语言学是数学和人文科学之间的桥梁

第一章 语言符号的随机性与统计数学

第1节 语言符号的随机性

第2节 字频和词频的统计

第3节 语音统计研究

第4节 方言研究中的统计方法

第5节 计算风格学

第6节 古代语言研究中的统计方法

第二章 随机过程与语言符号的冗余性

第1节 语言的使用与马尔可夫链

第2节 语言的熵和语言符号的冗余性

第三章 语言符号的离散性与集合论

第1节 语言符号的离散性

第2节 语言的集合论模型

第四章 语言符号的递归性与公理化方法

第1节 语言符号的递归性

第2节 生成语法的公理化方法

第五章 语言符号的层次性

第1节 语言符号的层次性

第2节 树形图

第六章 语言符号的非单元性与复杂特征的运算

第1节 语言符号的非单元性

第2节 复杂特征的运算

第七章 语言符号的模糊性与模糊数学

第1节 语言符号的模糊性

第2节 模糊数学在语言研究中的应用

附录：胡耀邦同志鼓励我研究数理语言学

1989年，当时的北京大学校长丁石孙教授决定出版一套《数学·我们·数学》的丛书，内容包括“数学与经济”、“数学与军事”、“数学与教育”、“数学与语言”等许多方面，丁石孙校长委托北京大学数学系马希文教授找我，希望我写一本《数学与语言》。我是一个语言学工作者，当时正从事机器翻译和信息检索等应用系统的开发研究，接触到不少语言学中的数学方法问题，对于数学与语言之间关系的问题作过一些思考，因此，我欣然接受了丁石孙教授的这个任务，历时两年，写成了这本《数学与语言》专著，于1991年出版。

在这本专著中，我从数学的角度，对于自然语言的性质进行了深入的思考，明确地指出，除了索绪尔过去提出的语言符号的任意性之外，语言符号还具有另外7个明显的特性，它们是：语言符号的随机性、语言符号的冗余性、语言符号的离散性、语言符号的递归性、语言符号的层次性、语言符号的非单元性、语言符号的模糊性。我提出的语言符号的这7个新的特性，显然补充了索绪尔关于语言符号任意性的思想，使我们对于语言符号的特性有了更加深刻的认识。

我在本书中提出语言符号的这7个新的特性之后，引起了许多语言学家的关注，有的语言学家指出，语言符号的这7个特性，反映了“信息时代的语言观”。语言学家刘海涛在本书的书评中指出，语言符号的这7个特性应当成为“信息时代语言学研究的基础”，并指出，本书“值得每一位语言工作者研读”。19年过去了，我提出的语言符号这7个特性的“语言观”仍然显得很有生命力，越来越多的事实将会继续证明这种“语言观”的正确性。

本书在1991年出版时只印了1400册，早已销售一空。不少读者尽管听说过此书，但是，踏破铁鞋无觅处，根本买不到此书。

总之，电子计算机的出现和广泛使用，就像催化剂一样促进了数学和语言学的结合。数学渗透到了形态学、句法学、词汇学、语音学、文字学、语义学等语言学的各个分支部门，促进了语言学的数学化。而语言学的数学化则是语言学现代化的一个重要内容，这些内容被概括在“数理语言学”（mathematical linguistics）这个新兴学科中，并得到了迅速的发展。

1955年，美国哈佛大学首先创办了数理语言学讨论班，1957年正式开设了数理语言学课程。接着，麻省理工学院、密歇根大学、宾夕法尼亚大学、印第安纳大学、加利福尼亚大学都相继开设了数理语言学课程。同年，日本成立了计量语言学会，创办了数理语言学杂志《计量国语学》，德国的波恩大学也开设了数理语言学课程，苏联在莫斯科大学、列宁格勒大学及莫斯科国立第一外国语师范学院也进行了数理语言学的研究工作。1958年，莫斯科大学、高尔基大学、萨拉托夫大学、托姆斯克大学，分别给数学系及语文系的学生开设了数理语言学的选修课，并在列宁格勒大学设置了数理语言学专业。此外，罗马尼亚、匈牙利、捷克、英国、法国、挪威、波兰、瑞典等国，都先后开展了数理语言学的研究工作，有的国家还创办了专门的刊物，成立了专门的研究机构。

我国从20世纪50年代起便开展了数理语言学的研究工作。1982年，北京大学中文系给汉语专业的学生开设了《语言学中的数学问题》的选修课，首先在我国开设了数理语言学方面的课程。①1985年，上海知识出版社出版了我国的第一本数理语言学专著——《数理语言学》。数学的定量的研究方法已逐渐受到我国传统的语言学家们的注意，并开始用到他们的研究工作中去，在用数学方法研究汉语的句子结构、汉字频率统计、汉语单词频率统计、频率词典的编制、方言定量分析、汉字熵值测定等方面，都取得一定的成绩。

数理语言学的研究常常要用电子计算机作为工具，因此，它与计算语言学的研究是联系在一起的。当前，数理语言学与计算语言学有合流的倾向。这清楚地说明，语言学、数学、计算机科学有着不解之缘。

• • • • • [\(收起\)](#)

[语言与数学 下载链接1](#)

标签

数学

语言学

计算语言学

NLP

语言

自然语言处理

冯志伟

语言与数学

评论

讲语言冗余度的时候，举的例子是俄语印刷体挡住上面一半还能认出来，你怎么不说手写体呢哈哈哈哈哈。总算知道当年他是怎么手工统计汉语熵的了。统计和代数都要好好学。

言简意赅的综述让人看得心旷神怡，语音分析和语法，句法，词法，乔姆斯基

可惜只是简介...

老冯是真心厉害。书可以当科普用，不过有些部分估计得学点知识才能读。

冯老太强了。。。

难道是时代局限？我觉得相当一部分内容和观点都陈旧，且有的理论简直就是胡扯，比如根据Zipf定律估计求汉字熵的样本量——完全是毫无意义的凑数字游戏嘛！

interesting also complicated. Prof.Feng gave a big and clear image. Chomsky is freakly awesome.

Definitely an eye-opening piece..
对习以为常的事物更要多加思考才有可能领悟其本质。看到熵和马尔可夫链以为是巅峰了，但其实模糊性才是最精彩的部分啊啊啊。最如果学了CS/Math一定要选ling，太扎劲了。。有点想去做AI，草。+interdisciplinary才是未来啊+生成语法和GEB的递归很是相似哦，快去把GEB捡起来..

冯志伟老师最早研究的是俄语，所以书中很多来自前苏联和俄罗斯的理论都用的是俄语，不好理解，如果可以换成英语就好了。整本书超赞。

[语言与数学 下载链接1](#)

书评

[语言与数学 下载链接1](#)