

## 多语言信息网络时代的语言学家：冯志伟



2008年8月6日在北京大正语言知识处理科技有限公司作学术报告

冯志伟，教育部语言文字应用研究所研究员，博士生导师，所学术委员会委员，中国当代著名语言学家，主要研究方向为计算语言学和应用语言学及语言规划研究。冯志伟先生是一个多语者（polyglot person），通晓汉、英、德、法、俄、日等多种语言，一直从事语言学和计算机科学的跨学科研究。针对多语言信息网络时代的特点，他在机器翻译、自然语言的计算机处理、术语数据库的研究等方面，进行了长期的坚持不懈的探索，得到了国内外学术界的高度评价，1991年10月1日获国务院颁发的政府特殊津贴证书。

冯志伟先生1939年4月15日出生在云南省昆明市，1957年从云南昆明第五中学考入北京大学学习地球化学专业。他在一本英文《信息论》杂志上看到了国外用数学方法研究语言的文章，对语言的数学模型产生兴趣，便立志研究数理语言学。1959年转入中文系语言专业学习语言学，师从王力、岑麒祥、朱德熙等教授，受过严格的文字、音韵、训诂等传统语言学和西方现代语言学的训练。大学期间，因学习成绩优异，曾经被错误地认为是“不关心政治”而受到非议，但是，当时的团中央书记胡耀邦在1961年11月11日亲自在北京富

强胡同的家中接见了冯志伟，给他热情的鼓励，使得他能够在当时复杂的政治气氛下继续研究数理语言学问题。1964年考取北京大学理论语言学专业的研究生，师从岑麒祥教授学习理论语言学和法语，不久发生文化大革命，研究生学习中断。1967年离开北京大学，回到家乡昆明担任中学物理教员，利用业余时间，模仿信息论之父 Shannon 估算英语字母的熵的方法，在世界上首次估算出汉字的熵为 9.65 比特。

1978年文化大革命结束，高考制度恢复，但根据毛泽东生前的指示，文科暂时不招生，冯志伟只好报考理工科，以高分考入中国科学技术大学研究生院。冯志伟入学不久就在《计算机科学》创刊号上发表论文，由于学习成绩突出，学校公派他到法国格勒诺布尔理科医科大学应用数学研究所(IMAG)自动翻译中心(CETA)，师从法国著名数学家、国际计算语言学委员会主席沃古瓦(B. Vauquois)教授，专门研究数理语言学和机器翻译问题。在法国学习期间，他提出了“多义多标记树形图模型”(MMT 模型)和静态信息与动态信息分开处理的“双态原则”(DSP 原则)，研制了世界上第一个汉语到法、英、日、俄、德等多种外语的机器翻译系统 FAJRA。1981年回国，在中国科学技术信息研究所计算中心担任机器翻译研究组组长，研制了德汉机器翻译系统 GCAT 和法汉机器翻译系统 FCAT。1985年调入国家语言文字工作委员会语言文字应用研究所，担任计算语言学研究室主任，同时兼任中国科学院软件研究所研究员。1986年，受中国科学院软件研究所的派遣，到德国夫琅禾费研究院(FhG)新信息技术与通信系统研究所担任客座研究员，研制了世界上第一个中文术语数据库 GLOT-C，并提出了“潜在歧义论”(PA 论)。1990-1993年在德国特里尔大学语言文学院(Fachbereich-II)担任客座教授，用德文出版了《汉字的历史和现状》(德国特里尔科学出版社，1994年版)一书。1996年在德国康斯坦茨高等技术学院国际术语学和应用语言学中心(CiTaL)担任技术顾问。1999年10月到2000年8月，再次在德国特里尔大学语言文学院(Fachbereich-II)担任客座教授。2000年8月到2001年8月，在桑夏自然语言处理研究院担任高级研究员。2001年9月到2002年8月，在韩国科学技术院(KAIST)电子工程与计算机科学系(EECS)担任访问教授，用英语讲授“自然语言处理技术”(NLP-II)的博士课程。2003年2月到5月，在英国伯明翰大学担任高级研究员。2003年10月到2004年10月，再次在韩国科学技术院(KAIST)电子工程与计算机科学系(EECS)担任访问教授。2007年和2008年先后两次在香港城市大学担任高级研究员。现为国家教育部语言文字应用研究所研究员，博士生导师，所学术委员会委员。

在国内，冯志伟先生是中国语文现代化学会副会长、中国人工智能学会理事、中国社会科学基金语言学评审组成员、中国自然科学基金信息科学部计算机学科评审组成员、北京市自然科学基金计算机学科评审委员、国家语言文字工作委员会 21 世纪语言文字规范(标准)审定委员会委员、全国科学技术名词审定委员会委员、全国术语标准化技术委员会委员，《中国语文》、《语言文字应用》、《中国科技术语》、《语言研究》等期刊编委，北京大学、浙江大学、黑龙江大学、中国传媒大学、北京师范大学、北京外国语大学兼职教授，清华大学、大连海事大学讲座教授，中国科学院自动化研究所国家模式识别重点实验室学术委员会委员。

在国际上，冯志伟先生是 TELRI (Trans-European Language Resources Infrastructure) (跨欧洲语言资源基础建设工程学会) 的顾问委员会委员，第一、二、三届 LREC (Language Resources and Evaluation Committee) (语言资源与评测会议) 的国际顾问委员会委员，IJCL (International Journal of Corpus Linguistics) (语料库语言学国际杂志) 的编委，IJCC(International Journal of Chinese and Computing)(中文与计算国际杂志)的编委，英国 Continuum 出版公司系列丛书《语料库与话语研究》(Research in Corpus and Discourse) 的编委。

冯志伟先生是国内外知名的机器翻译、计算语言学和语料库研究方面的专家，20世纪80年代以来受聘为多所国际知名大学和研究院的兼职教授和客座研究员。他在中国的学术活动中做了很多创新性的研究：他首次测定汉字熵值，研究了生成语法的公理化方法，提出了多叉多标记树模型，提出了语言特征信息的双态原则；他开发了世界上第一个汉外多语言机器翻译系统，建成了世界上第一个中文术语数据库；他最早在北京大学开设数理语言学课程，出版了中国的第一本《数理语言学》、第一本《现代术语学引论》和第一本《现代语言学流派》的专著。他是中国最早一批从事多语言术语数据库开发的学者，在长期的研究实践中，他提出了“术语形成的经济律”、“生词增幅递减律”、“潜在歧义论”等理论和汉字结构的括号式表示法，并建立了汉语文本自动切分的形式词理论。

冯志伟先生主要研究计算语言学中的各种理论和技术问题，也关心语言生活中的现实问题和语言规划的研究。他担任中国语文现代化学会副会长，主张语文现代化。他积极支持中国政府推行《汉语拼音方案》和汉语拼音正词法，主张维护汉字的法定地位，主张汉字是中国的法定文字。他认为，汉语拼音对于中国的语文现代化具有重要的作用，主要应当用在汉字不能使用和不便使用的场合，不主张把拼音作为一种正式的法定文字，更不主张全盘的拼音化。他在北京大学中文系求学期间，曾经接受过中华文化的熏陶，比较注意中华优秀传统文化的传承，他曾经用德文写过论述汉字的专著，向国外介绍汉字的专门知识，他对于汉字的本质和特性有深刻的认识，所以，他始终坚持地认为：汉字和拼音应当是休戚与共的朋友，而不应当是势不两立的敌人。

冯志伟先生重视术语学的研究，认为术语是人类科学技术知识在自然语言中的结晶。在术语学研究中，他除了研究术语与概念的关系之外，还重视基于知识本体的术语研究；除了研究术语的规范化和标准化之外，还重视术语的描写研究；除了研究术语的静态的共时特性之外，还重视术语的动态的历时特性研究。冯志伟对于术语学基本理论的这些原创性探讨，丰富了传统的术语学，引起了国际学术界的注意。2006年6月30日，联合国教科文组织奥地利委员会（Austrian Commission for UNESCO）、维也纳市（City of Vienna）和国际术语信息中心（INFOTERM）联合给冯志伟颁发了Wüster（维斯特）奖（Wüster Special Prize），以表彰他在术语学理论和术语学方法研究方面做出的突出贡献。

冯志伟先生重视翻译外文著作，他翻译过长达588页的大部头英文专著《自然语言处理综论》（电子工业出版社，2005年版）。为了帮助年轻学子阅读外文语言学原著，他还为一些外文语言学名著写导读。他写的导读有：《应用语言学中的语料库》（世界图书出版公司&剑桥大学出版社，2006年版），《译者的电子工具》（外语教育与研究出版社，2006年版），《人工智能在第二语言教学中的应用》（世界图书出版公司，2007年版），《语言学中的数学方法》（世界图书出版公司，2009年版）等等。

冯志伟主持过多个国家社会科学基金项目、国家自然科学基金项目和国际合作项目；出版了大量学术专著，并多次荣获国家图书奖和中国出版集团优秀图书奖。

冯志伟最近在香港《语文建设通讯》（第91期）发表的《机器翻译，任重道远》一文中，对于他在机器翻译方面的研究做了如下的总结：“我研究机器翻译已经将近五十年的时间了，五十年前，我还是一个不谙世事的十九岁的小青年，现在，我已经是白发苍苍的老人了，我们这一代人正在一天天地变老；然而，我们如痴如醉地钟爱着的机器翻译事业却是一个新兴的学科，她还非常年青，充满了青春的活力，尽管她还很不成熟，但是她无疑地有着光辉的发展前景。我们个人的生命是有限的，而科学知识的探讨和研究却是无限的。我们个人渺小的生命与科学事业这棵常青的参天大树相比较，显得多么地微不足道，有如沧海之一粟。想到这些，怎不令我们感慨万千！”他这段语重心长的话，也是他在自然语言的计算机处理、术语数据库和语言规划等方面所做的长期坚持不懈的探索和研究的总结，反映了他“烈士暮

年，壮心不已”的学者情怀。

#### 附录：冯志伟先生的主要著作

1. 《数理语言学》，上海知识出版社，1985 年版。
2. 《自动翻译》，上海知识出版社，1987 年版。
3. 《现代语言学流派》，陕西人民出版社，1987 年版。
4. 《现代汉字和计算机》，北京大学出版社，1989 年版。
5. 《中文信息处理与汉语研究》，商务印书馆，1992 年版。
6. 《数学与语言》，湖南教育出版社，1991 年版。
7. 《自然语言机器翻译新论》，语文出版社，1995 年版。
8. 《自然语言的计算机处理》，上海外语教育出版社，1997 年版。
9. 《Chinesische Schriftzeichen - Vergangenheit und Gegenwart》, Wissenschaft Verlag, Trier, 1994  
(《汉字的历史和现状》，德文版，德国特里尔科学出版社，1994 年版)。
10. 《现代术语学引论》，语文出版社，1996 年版。
11. 《应用语言学综论》，广东教育出版社，1998 年版。
12. 《计算语言学基础》，商务印书馆，2001 年版。
13. 《计算语言学探索》，黑龙江教育出版社，2002 年版。
14. 《机器翻译研究》，中国对外翻译出版公司，2004 年版。
15. 《汉语教学与汉语拼音正词法》(与新加坡许福吉博士合著), Marshall Cavendish Academic  
出版社(新加坡), 2005 年版。
16. 《术语浅说》，语文出版社，2005 年版。
17. 《机器翻译今昔谈》，语文出版社，2007 年版。
18. 《自然语言处理的形式模型》，中国科学技术大学出版社，2009 年版。



冯志伟部分著作