

载《新华书目报》·《科技新书目》第 919 期，A05 版，《科技人物》
专栏，2010 年 12 月 9 日出版



乔姆斯基与冯志伟合影（2010 年）

冯志伟 与数理语言学共舞

冯志伟，1939 年 4 月生，云南省昆明市人，计算语言学家，专门从事语言学和计算机科学的跨学科研究。他先后在北京大学和中国科学技术大学研究生院两次研究生毕业，获双硕士学位。先后到法国格勒诺布尔理科医科大学、德国夫琅禾费研究院 (FhG) 新信息技术与通信系统研究所、德国特里尔大学语言文学院、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)、桑夏自然语言处理研究院、韩国科学技术院 (KAIST) 电子工程与计算机科学系 (EECS) 学习、工作。现为国家教育部语言文字应用研究所研究员、博士生导师、学术委员会委员。

跨学科奇才

语言学家冯志伟教授是极为罕见的跨学科奇才。过去的著名语言学家有的只懂社会科学，不懂自然科学；许多人只懂古代汉语、现代汉语或普通语言学，一般只着重研究汉语的语音、词汇、语法或文字等某一个方面的问题，研究的问题和领域比较单一；他们中的一些佼佼者，至多也只懂得两三门外语，视野不够开阔，语言的纵横向对比研究都不够，有一定的局限性。而冯志伟先生却懂得理科中的数学、物理、化学和计算机科学，又懂得语言学中的古代汉语、现代汉语、文字学、音韵学和普通语言学，深研过汉、英、法、德、俄、日等语言的语音、词汇和语法的自动处理，并把各方面的知识紧密地结合起来综合应用，在计算机上加以实现，成为文理兼通的语言学专家。冯志伟教授不但在国外著名大学的电子工程与计算机科学系讲授理科的“机器翻译的方法和技术研究”、“自然语言处理的算法研究”、“计算语言学专题研究”等艰深的博士课程，而且还在国内外大学的中文系讲授“汉魏六朝散文”、“唐诗”、“宋词”、“古代汉语”、“现代汉语”、“汉字的历史与结构”等饶有风趣的课程，他还能给学生们辅导英、德、法、俄、日等外语课程，提高学生们的外语口语表达能力和书面写作能力。他的散文和诗歌也写得很好，他为哀悼好友马希文教授（著名数学家）英年早逝而写的诗歌《长歌当哭》，凄楚感人，读后令人潸然泪下。他还是一位翻译专家，出版过翻译著作。这样的人才确实是很罕见的。

冯志伟自幼聪慧沉静，18岁就以优异成绩考入北京大学地球化学

专业。他在北大图书馆偶然看到了美国语言学家乔姆斯基 (N. Chomsky) 的论文《语言描写的三个模型》(Three models for the description of language), 被乔姆斯基在语言研究中的新思想深深地吸引, 继而从理科转到中文系语言学专业从事语言学的学习。在中文系读书期间, 冯志伟一面学好传统语言学的各门课程和外语, 一面利用课余时间, 继续研究数理语言学的问题, 他充分地利用北京大学图书馆丰富藏书和最新的杂志, 跟踪着国际上数理语言学发展的足迹, 他成为了班上名列前茅的学生。1964 年考上北京大学语言学理论的研究生后, 经导师岑麒祥教授同意, 他的研究生毕业论文的题目定为《数学方法在语言学中的应用》, 在我国语言学研究, 首次系统地、全面地来研究数理语言学这个新兴学科。

这样, 我国的数理语言学研究便首先在北京大学正式地开展起来。北京大学中文系的著名语言学家王力先生和朱德熙先生都支持冯志伟的数理语言学研究, 王力先生曾对冯志伟说: “语言学不是简单的学问, 我们应该像赵元任先生那样, 首先做一个数学家、物理学家、文学家、音乐家, 然后再做一个合格的语言学家。”朱德熙先生曾对冯志伟说: “数学和语言学的研究都需要有逻辑抽象的能力, 在这一方面, 数学和语言学有共同性。”北京大学的这些第一流的学者, 总是站在科学的最前沿来看待学术的发展, 他们的鼓励给了冯志伟以巨大的力量。

1978 年, 之前毕业于北京大学中文系研究生的文科学生冯志伟, 以优异的理科成绩考上了中国科学技术大学研究生院信息科学系的

研究生，又开始了理科的学习。1979年，入学不到一年，他就在《计算机科学》杂志创刊号上发表了《形式语言理论》的长篇论文，用严格的数学表达方式向计算机科学界说明数理语言学中的形式化方法如何推动了当代计算机科学的发展，并且指出：在数理语言学研究发展中发展起来的形式语言理论，事实上已经成为了当代计算机科学不可缺少的一块重要的理论基石，计算机科学绝不可忽视形式语言理论。许多人认为这篇文章一定是资深的计算机科学家写的，后来，当计算机界的一些专家了解到，这篇论文的作者竟然是文革前北京大学中文系的一个文科研究生的时候，感到非常惊讶。

1990年~1993年，冯志伟应邀在德国特里尔大学用德语讲授一系列的汉语语言学课程，并且用德语出版了语言学的专著，显示了他的外语才能。在我国中文系出身的语言学家当中，像冯志伟这样通晓多门外语的人还不多。他不仅是一位善于深思的语言学的理论家，而且还是一位勇于实践的能操多种外语的多面手。

梅花香自苦寒来

在多学科间游走是需要深厚的知识积淀的，知识积淀则来自于勤奋。冯志伟在上世纪六十年代初期就已经学会了4门外语，而且能够使用这4种外语阅读数理语言学的外文文献。他取得这样的学习成绩，不仅是天赋，更多的是因为勤力。为了学习英语，他就买一本中型的英汉词典来，一页一页地记忆和背诵，背完一页就撕去一页。几年来，冯志伟先后撕完了英汉、俄汉、法汉、德汉、日汉等多部词典，

他就用这样的笨方法，学会了多门外语。

又比如，“文革”期间，冯志伟在朋友们的帮助下，用了将近 10 年的时间，对数百万字的现代汉语文本（占 70%）和古代汉语文本（占 30%）进行手工查频，从小到大地逐步扩大统计的规模，建立了 6 个不同容量的汉字频度表，最后根据这些不同的汉字频度表，逐步地扩大汉字的容量，终于计算出了汉字的熵。在当时的政治环境和技术环境下，能实现如此繁重的工作，难度可想而知。

冯志伟不仅自己的学习时候勤奋有加，对待教学也是一丝不苟。在特里尔大学文学院任教期间，冯志伟用德语给德国学生讲授《汉魏六朝散文选》、《唐诗宋词选》、《中国现代散文选》、《汉字的发展与结构》、《汉语拼音正词法》、《汉语词汇史》、《机器翻译的理论和方法》等课程。为了讲好课，他苦练德语口语，认真用德语备课，在上每一节课之前，他都要先用德语把讲课的内容自己对自己叙述一遍或多遍，直到能够熟练地背诵为止，他把“备课”当作了“背课”。由于冯志伟的备课特别认真，课堂教学效果很好，他的讲课受到德国学生们的一致好评。

冯志伟在韩国科学技术院任教期间，利用全部的业余时间来翻译《语音和语言处理 - 自然语言处理，计算语言学和语音识别导论》（原文为英语）一书，晚上加班到深夜，连续工作了 11 个月，当翻译完 14 章（全书的三分之二）的时候，他不幸患了黄斑前膜的眼病，视力出现障碍，难于继续翻译工作。剩下的 7 章由中国科学院软件研究所的一位年轻的副研究员协助他完成。这位副研究员把剩下的 7 章

逐一翻译成中文，通过计算机网络一章一章地传给在韩国的冯志伟，冯志伟使用语音合成装置，让计算机把书面的文本读出来，通过读出来的语音进行译文的校正。2004年，全书的翻译大功告成，由电子工业出版社以《自然语言处理综论》的书名出版。这本书的出版受到广大读者的欢迎，而冯志伟为此却损害了自己的视力，他不得不借助于语音合成装置来阅读了。

硕果累累著作丰

天赋和勤勉成就了跨学科奇才，冯志伟在数理语言学领域不断前行，解决一个又一个难度很高的问题，硕果累累。

在法国留学期间，经过在计算机上编写程序进行潜心的钻研和反复的试验，冯志伟提出了“多叉多标记树模型”（Multiple-labeled and Multiple-branched Tree Model，简称 MMT 模型）。这个模型提出后，立即引起了国际计算语言学界的高度重视。冯志伟根据他提出的 MMT 模型，于 1981 年完成了汉-法/英/日/俄/德多语言机器翻译试验，建立了 FAJRA 系统（FAJRA 是“法语-英语-日语-俄语-德语”的法文首字母缩写），在 IBM-4341 大型计算机上，把二十多篇汉语的文章自动地翻译成英文、法文、日文、俄文、德文，当时在实验室工作的外国朋友亲自目睹他们视为天书的一篇篇汉语文章被计算机翻译成他们懂得的五种外国语之后，无不拍手叫绝。这是世界上第一个汉语到多种外语的机器翻译系统，开创了多语言机器翻译系统之先河。

从法国回国之后，冯志伟在中国科技信息研究所计算中心担任机

器翻译研究组的组长，在王力先生的鼓励之下，他利用当时北京遥感技术研究所的 IBM-4361 计算机，于 1985 年进行了德—汉机器翻译试验和法—汉机器翻译试验，建立了 GCAT 德—汉机器翻译系统和 FCAT 法—汉机器翻译系统，进一步检验了 MMT 模型分析汉语和生成汉语的能力，试验结果良好。

在北京大学“语言学中的数学问题”选修课讲稿的基础之上，冯志伟写出了我国第一部数理语言学的专著，书名就叫做《数理语言学》，于 1985 年 8 月由上海知识出版社出版。接着，他又出版了《自动翻译》的专著，深入地探讨自然语言机器翻译的理论和实践问题。这两本专著的出版，受到了我国计算语言学界的欢迎。当时不少出国学习计算语言学的留学生，出国时都带着这两本书，作为入门的向导，在本书的引导下，他们很快就接触到了国外计算语言学研究中的前沿问题。

冯志伟上世纪 80 年代在德国夫琅禾费研究院新信息技术与通讯系统研究所(FhG)担任客座研究员期间，克服了重重困难，使用 UNIX 操作系统和 INGRES 软件，建立了数据处理领域的中文术语数据库 GLOT-C，并且把这个数据库与 FhG 的其他语言的术语数据库相连接，可以快速地进行多语言术语的查询和检索，并且能够处理汉字（当时计算机上还没有成熟的中文操作系统）。这是世界上第一个使用汉字的中文术语数据库，具有开创作用。

2006 年，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)给冯志伟教授颁发了维斯特奖(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。维斯特(Eugen Wüster, 1898-1977)是奥地利著名科学家,是术语学和术语标准化工作的奠基人。维斯特奖是专门为那些对于术语学和术语标准化工作有出色成就的科学家而设置的。

今年七十一岁的冯志伟著述颇丰,其内容涉及不少领域,引起国内外不少同行专家的广泛关注和高度评价。他现在已在多家著名出版社出版了二十余部专著,翻译国外重要论著数十篇(部),并用汉、英、法、德文等撰写了有代表性的重要学术论文百余篇,在国内外许多著名刊物上发表。最主要的代表作如:专著《数理语言学》、《自动翻译》、《现代语言学流派》、《现代汉字和计算机》、《中文信息处理与汉语研究》、《数学与语言》(新版更名为《语言与数学》)、《自然语言机器翻译新论》、《应用语言学综论》、《应用语言学新论》、《计算语言学基础》、《计算语言学探索》、《机器翻译研究》、《现代术语学引论》、《自然语言的计算机处理》、《汉语教学与汉语拼音正词法》、《自然语言处理的形式模型》、《汉字的历史与现状》(德文版),译著《自然语言处理综论》,等等。他还主持和参与了若干个国家标准的制定,为我国标准化做出了贡献。他又是《中国大百科全书》语言文字卷编辑组成员、《数学辞海》总编委会委员、《计算机百科全书》和《中国少年百科全书》等大型工具书的撰稿人,为这些权威性工具书写了不少重要条目。他还为一些当代语言学名著的外文版写导读,如《应用语言学中的语料库》、《语言学中的数学方法》、《译者的电子工具》、《人

工智能在第二语言教学中的应用》、《牛津计算语言学手册》、《自然语言生成系统的建造》等，帮助国内读者阅读外文原著。

冯志伟对于他的这些成就却看得很平淡，他很少对别人谈起他的成就。除了平时喜欢喝白开水和游泳之外，他几乎没有什么特殊的嗜好，他每日粗茶淡饭，过着非常清贫的生活。他从来不以为自己是什么“专家”，没有任何的架子，总是谦和地对待他的学生和周围的同志。在北京大学中文系 59 级同学纪念册上，他写下了这样的人生感言：“先天不足，后天失调；岁月蹉跎，艰辛备尝；老当益壮，穷且越坚；平生无悔，褒贬由之。”他始终认为自己是一个很平常的普通人，对于别人的褒贬，他是看得很平淡的。他对于生活的信条是：“在科学探索的过程中，我所知道的东西终究是有限的，而我不知道的东西始终是无限的，只要平生无悔就很好了！”