

以下访谈载冯志伟《机器翻译研究》，中国对外翻译出版公司，2004 版。

关于机器翻译和计算语言学的对话

(冯志伟教授口述访谈记录)

受 访 者：冯志伟教授

访 谈 者：张 宜（北京外国语大学语言研究所，博士研究生）

整理 / 注释：张 宜

地 点：冯志伟教授家

时 间：2003 年 10 月 16 日，上午 9: 00—12: 30

张 宜：今天是 2003 年 10 月 16 日，我现在是在冯志伟教授的家里，我今天请他讲的主题是计算语言学、机器翻译。冯老师，您好！您是怎样走上语言学道路的？您为什么要研究语言学？

冯志伟教授：我原来是学理科的，我是 57 年通过理科高考考上北大。刚开始的专业是地球化学，主要是研究稀有元素跟一些放射性元素在地球上的分布。

张 宜：是地球物理系吗？

冯志伟教授：嗯，应该是地质系的地球化学专业，叫做 Geochemistry。

张 宜：那是本科了。您是从昆明考过来的？

冯志伟教授：是从昆明考过来的。我学习地球化学的目的是由于当时国家建设需要弄清中国的资源嘛，所以我认为地球化学很重要，因为从元素的角度来看，它要研究各种元素的分布。而地质学是从矿物的角度来看，但从元素角度来看是更基本的层次。

张 宜：当时是您自己选择的这个专业，还是老师建议的？

冯志伟教授：是我自己报的。我的兴趣主要是在稀有元素上，钨钼钽钨一族，在元素周期表上是排在比较后的元素，因为这些国家很需要。我当时也没有任何想从事其他学科的想法，这个学科确实也很有意思。地球化学在 57 年也是属于国家尖端学科之一，国家重点发展，在地球科学里面也是属于最先进的学科。

张 宜：您当时是在昆明的哪所中学？

冯志伟教授：昆明市第一中学。我们昆明一中是很好的学校，它的前身是昆华中学。它是云南省最好的学校。我们这所学校出过一些名人，比如说艾思奇，写《大众哲学》的哲学家。另外还出过更重要的人物，诺贝尔奖的获得者杨振宁，他曾在昆华中学（昆明一中的前身）从高一读到高三。他从昆华中学直接考上西南联大，然后就从事物理学了。所以我们学校是昆明很好的学校，我当时是昆明一中唯一考上北大的学生。云南属于边疆地带，我们上北大很不容易。当年云南省大概有两个人考上北大。我认为我选择的这个专业还是很有意思的，因为在云南，我可以接触到各种各样的矿物，对自然界的東西非常感兴趣。我对我的学科很有兴趣。

北大是所开放性的大学。57 年时办学思想还不像现在这样开放。我们北大学生相对于其他院校的学生还是比较自由的。北大学生可以进外文图书室看杂志，在其他院校几乎是不行的，所以可以看到很多东西。当时我的第一外语是俄语，当然英文我也懂，我也能看英文书。我当时在一个偶然的的机会里看了一本杂志 *Transactions on Information Theory*，叫做《信息论学报》，这个杂志是美国一个比较有名的关于信息论、计算机科学基础理论的杂志。我在上面看到了 Chomsky 的文章，那是 57 年。我想他当时应该是很年轻了，还不太有名，他

是在 57 年出了一本书¹后才逐渐有名的。所以他的文章发表在《信息论学报》这本自然科学的杂志上，题目叫做《语言描写的三种模型》（“Three Models for the Description of Language”²）。我是无意当中看到的。他在这篇文章里谈到几点。他认为形式语言，包括早期的程序语言、自然语言以及描写程序的机器语言都有着共同的基础，数学的基础是一样的。他提出有限状态模型是基于一种有限状态的 Markov³过程来描述的。我看到他描写英语的情况以后觉得很新奇。我也学过一点语言学，但是我不是语言学家，我知道一些语言学知识，知道我们传统的语言学怎么做。一般来说要决定主语和谓语，画个图说明主语在哪儿，谓语在哪儿。我们管这个过程叫 parse⁴，就像任何一个学英文、俄文的学生都要经过一个句子的 parse 了以后才能理解这个句子。上课老师提问时也说把这个句子 parse 一下，说明哪个是主语，哪个是谓语，哪个是宾语。如果你说错了，你就是没有懂。用 Chomsky 的有限状态模型方法来说明自然语言不行，它是个很简单的数学模型。另外一个模型叫做短语结构模型，就是后来他所谓的上下文无关语法，他说短语结构模型描述性比有限状态模型更强。他还提出一种叫做 0 型语法，可以用图灵机器来描述的语法，这种语法很宽阔。Chomsky 在文章里讲，用图灵机来描述的语言太宽，可以生成的句子太多，有绝大部分是自然状态里没有的句子。如果用有限状态机器来描述的语法又太窄，有好多现象又描述不了，特别是一些递归的（recursive）语言现象，嵌套的句子，比如：“if... then...”，if 和 then 离开很远；“the man who... said, ...”，主语（man）离谓语（said）很远，主语和谓语间夹了一个很长的定语，这种句子用如果通过几个状态是很难描写的，这种叫嵌套型句子。如果改用上下文无关短语结构语法，叫做“上下文无关语法”context-free grammar，当时 Chomsky 把它叫做“可推导语法”，用这种语法来描写就可以了，因为这种语法具有递归的特性。我觉得他说的东西很新奇，因为当时中国基本上还没有人知道这个事情。我当时数学比较好，所以我对数学有兴趣，对于他的研究中的数学实质我看得比较清楚，我甚至预感到有一件很伟大的事情就要发生了，我有这种感觉。他的这篇文章虽然很普通，但是我认为对我们人类的智能研究会有很大的作用，因为他已经找到了一种比较实质性的东西，这可能就是在后面支配着我们人类语言行为的东西，它是一种描述人类语言行为的形式化机制。

张 宜：这时是 大一 还是 大二？

冯志伟教授：大一的下学期，57 年年底到 58 年上半年吧。这时我自己就开始考虑能不能从事这个专业，我有这个想法。但是当时是不大可能的，因为我也问过一些中文系的老师，问他们我们国家有没有搞语言学的，他们说我们北大中文系就有搞语言学的，但我们不知道什么 Chomsky。对于用数学方法来搞语言，他们认为是一种古怪的事情，觉得很奇怪。在这种情况下，我就有一点想要转到中文系去学语言学的念头。当时，那个老师跟我说如果你要转过来的话，你得跟着我们学，就是说我教你什么，你学什么，你这个想法可以说是一种异端邪说吧。第一，我们没有老师教你；第二，你也得收敛一点，你不要想得太多，你得跟着我们学。这个事情快谈成时，大二快结束了。这就意味着到了该下决心的时候。因为我原来这个专业挺好的。当时中文系说，你是理科学生，你的文科程度我们还不知道，那么你能不能来学，我们还得研究研究，而且尽管我们让你过来，可能你也不能从大三读起，你得放弃你两年学理的学历。因为我在地球化学专业也就是读了两年的书而已，有一些基本的数理化知识，但是还没有学会我本专业的好多东西。我要从中文系的一年级读起，所以这些事我得考虑。当时还有一个最后促使我下决心的因素是我的眼睛有些问题，我发现我的眼睛的辨色

¹ 指 *Syntactic Structure*, Mouton & Co., S-Gravenhage, 1957. 中文译名:《句法结构》，邢公畹、庞秉钧、黄长著、林书武译，中国社会科学出版社 1979 年版。

² *Three Models for the Description of Language*, I. R. E. Transactions on Information Theory, vol. IT-2, Proceedings of the Symposium on Information Theory, 1956.

³ 马尔可夫，俄罗斯著名数学家。

⁴ 英语，意为：对（句子）进行语法分析。汉语可翻译为“剖析”。

能力有些不好，本来我高考时不是色盲，但自从我学了地球化学后，因为我要利用光谱来搞稀有元素分析，分辨稀有元素在光谱区的位置，在光谱区不同的元素有不同颜色的谱线，我能够从数学上算出谱线的位置，但是我不准，因为我辨别细微谱线的能力不如其他同学，所以这些因素就促使我下决心转系。如果我继续下去，将来我只能从事理论工作，因为我实际眼力不行，而我的专业需要用眼，我只能算出它，估计出它的位置和特性，但是我不能够亲自去实践它，我的生理能力决定了我不能像其他眼力正常的同学去辨别化学元素。这意味着我将来能否把我的专业学得很优秀，能不能搞出成绩来，将是很成问题的。这样就促使我下决心。我就到北大中文系初步表态我还是想来，但是还要由学校同意。学校说，原则上不允许你转系，因为你是大改行，从理科系转到文科系，如果你接受条件从一年级读起，另外还需要考试，我们要考你文科基础，如果考不及格，没有北大中文系学生的水平，你就不能进北大中文系的行列。我有些害怕。老实说，我虽然喜欢学语言学，但行不行我也不知道啊。后来中文系让我写一个理由，说明我为什么要改行，要转系，我为什么想到要搞语言学，为什么我不想搞地球化学了。我就动了一番脑筋写，好像写得还比较好。后来北大中文系说，我们从你写信的水平可以知道你的文科水平还可以。好像我写得比较有感情，但我已经记不清这位老师了。这样就不用考试了，我不用准备考试了，我也不必担心我的水平是否低于北大中文系的学生，我就可以保险进入北大中文系，我就改过来了。这时是 1959 年的 9 月份，我等于在理科读了两年。

张 宜：您去读中文系的三年级还是一年级？

冯志伟教授：一年级。这是我答应中文系的条件，他们说你不能插班，因为你没有学过原来的课，而且你理科的课没有一门跟我们的一样，所以你必须从头学。等于我浪费了两年的时间，这对于我来说是一个牺牲，不过这是我兴趣的驱使。当时我脑子里没有功利的观点，我没有想搞语言学将来能不能赚钱，那时的学生很少去考虑这个，基本上都想搞点学问，就是一种兴趣吸引着我过来。我到了中文系以后出现了一些问题。首先说，中文系的学生全都是文科基础的人，他们好像从小就喜欢文学、语言，对自然科学毫无兴趣，而我对数学很有兴趣。当时我每天还做数学题，还继续读北大数学系的书，像《数学分析》啊，《高等代数》什么的，另外还做吉米多维奇的《数学分析习题集》的 4 000 多个题吧，我一道题一道题地做，所以说我跟中文系的学生不怎么合拍。我一边上中文系的课，一边研究数学，我想把数学搞好一点，说不定将来可以像 Chomsky 那样，用他这种（数学）的方法来研究中文，我有这个想法，所以当时对于我来说比较苦。第一，我不想彻底放弃我的理科学习，我还继续学些数学。第二，我还要把中文系的课学好。我到了中文系，我不能做很差的学生，我还得保持在班上的前几名。不然你将来什么都干不成，等于我是肩负两重任务。另外，当时的气氛不是读书的气氛，57 年以后就逐渐不读书了。反右以后，政治气氛越来越浓，政治压力也很强，不太主张学生读书，所以，我就显得有些怪：平时做些数学题，经常还读点外国书，这在当时都是很不合拍的。

张宜：特别是在北大这种政治空气很浓的校园里。

冯志伟教授：特别浓。北大是中国时代的晴雨表。国家的政治斗争，中央的政治斗争马上就能反映出来，所以我在当时的日子也不好过。周围的同学看我很奇怪：你怎么从理科跑到了文科，而且你跑到了文科你还没有放弃理科。对我来说，为了证实我自己，我必须把功课学好。我当时在班上的成绩还不错，包括古代汉语、汉语史、文字学、音韵学和训诂学，我都学得不错，因为我要使他们认为我不是一个外来者，我们还是一样的，你懂我也懂。另外一方面，我在业务上也还不能太放纵。为什么呢？当时的形势是不让大家读很多书，书读得太多，可能别人会对你有看法。北大当时还抓白专学生，如果戴上这顶帽子，将来找工作都不好找，所以在当时为人很难为。一方面，你要学习好，但你又不能太拔尖，因为你太拔尖，就会给你看成白专分子。当时我对学习很有兴趣，一方面要好好学习，一方面要表现出成绩

不是特别的好。这些都是麻烦。考试的时候，要故意做错一、两道题，显得自己不是走白专道路的人。你不这么做，就会很麻烦。而且还要给人一种印象，你是关心政治的，你是一个投入社会生活的人，所以当时做人比较难做，和现在的学生完全不一样。现在的学生把书读好了，大家就说你好。在当时你书读好了，大家不一定说你好，甚至还会对你产生好多想法：你这人是不是想成名成家，走白专道路等等。到了64年我在中文系快毕业时，我在班上的成绩算是比较好的，因为在大学里不可能用数学的方法来做语言工作，我只能学习数学，所以我在大学主要的任务是把中文系的各门课程学好，但我的兴趣是在语言学史方面。我的导师是岑麒祥教授。今年是岑麒祥教授的百年诞辰，他是搞西方语言学史的。那么我跟岑麒祥先生学了法文，他是从法国回来的，我比较系统地读了一些语言学的书，1964年我考上了他的（硕士）研究生跟他学理论语言学。实际上我是搞理论语言学的，因为当时还没有计算语言学这门学科。岑老对我的帮助很大，他教我学了法语，为我增加了一门外语。另外，他有一个重要的观点，要治语言学史，最好不要看译作，因为译本有很多译错的地方，治史须读原著。这一点也促使我这时读了一些语言学的原著，像索绪尔的《普通语言学教程》和布龙姆菲尔德的《语言论》(Language)等。

张 宜：这些您读的是原文吗？

冯志伟教授：是的。

张 宜：那就是说索绪尔的《普通语言学教程》您读的是法文版本？

冯志伟教授：对，我读的是法文版本。这些东西帮助我后来写了《现代语言学流派》⁵一书。这些都是我向岑老学习的结果。我看了一些原文的语言学著作，所以书中的引文基本是从外文直接引用的，基本上不是第二手材料，所有的引文我都几乎核对过原文，所以是比较准的。另外，我也不太相信译本，所以对国内的一些说法我也不太相信。我是按照自己的认识去写。岑老主张不要随便篡改人家原来的思想，也不要用自己的主观思想去强加于人，应该首先把原著读懂以后，然后再进一步去研究。如果不读原著，仅用一些二手材料，道听途说，或是看了别人的评论再去评论，这样经过几手之后，可能你加进去的东西和原来的已经不一样。这些对我都很有启发。但是同时我也想把我用数学方法研究语言学的想法跟岑麒祥教授说一说，我希望能得到他的支持，但是这个过程比较艰难。因为系里的老师都知道我到中文系的目的，我想用数学方法搞语言学。但是读本科的阶段不可能啊，必须要上很多课，在负担重的情况下，我只能搞搞数学，我没法搞计算语言学，因为当时没有计算语言学这门学科，在语言学里不算一门学科。现在读研究生，岑老是我的导师，我必须把我的想法跟他说。我虽然学语言学，但我的目的还不是这个，我还有其他的目的，我还想在数学方面做一点。岑老是一位典型的人文学者，他跟我说，你这样做的话，你是不是就不打算搞语言学了。我说我正是为了搞语言学我才来学语言学的。岑老说，你要搞的已经是理科了啊。他不能接受，于是我就跟他说了我的各种各样的理由。后来岑麒祥教授基本上不太反对，但是他不是很支持，他说你可以做这个。我的毕业论文题目《数学方法在语言学中的应用》是个很大的题目，岑老原来是要我搞语言学史的，我没有搞，我写了这个题目。跟他讨论几次之后，他基本上不反对，但他也没有特别的支持。这个题目已经不是语言学史了，我想用数学方法来研究语言学。为此，我已经做了很多准备，岑麒祥教授也同意了，他也算承认我搞的是语言学。这在当时已很不容易了。

除了他以外，当时北大中文系还有些老师是支持我的，一个是朱德熙先生，因为他是学物理的，他是从物理转学语言学的，所以他很知道理科思维在语言学中非常重要，所以他看到我到中文系很高兴，他跟我很谈得来。朱先生他对我的研究很支持；另外一个就是王力教授，他认为语言学家应该知道逻辑，逻辑不好没法从事语言研究。他认为我能从数学角度看

⁵ 《现代语言学流派》，陕西人民出版社，1987（第一版），1999年（第二版）。

语言学是对的。当时朱德熙教授还很年轻，他还是副教授，但王力先生已经是北大中文系的泰斗了，那么他一说话，岑老基本上就不反对。岑老是位人文学者，他认为这么搞怪怪的，不过他看到朱先生和王力先生都支持，他也不太反对了。这样我的论文题目就定了。于是我就开始做论文。做到66年的上半年，我的论文就快写完了。因为我原来就有些基础，我有好多想法，我当时想用公理化系统的学说来建立一套自然语言系统，想做这个工作。那么架子已经有了，这些领导同意以后我就好办了，所以我用了不到两年的时间，论文就基本成形了。当时还包括我们去湖北的江陵搞“四清”的时间。所以说我的论文其实就用了半年的时间就写完了。

张宜：您原来学习的那些东西都已成为了你的基础。

冯志伟教授：对。到了1966年的四月份就定稿了。当时我跟岑老讲，我可不可以提前答辩。我家里很困难，我有三个弟弟和三个妹妹，他们都要上大学，那么我是家里第一个大学毕业生。我父母都是很普通的人。我母亲是商店的营业员，我父亲是个菜市场里的会计。我们属于在城市生活比较贫困的人。我弟弟写信来说，他考上云南大学物理系了，他问我读研究生可不可以给他点钱，我当时一个月42块，还得给他20块。我也希望赶快工作，工作以后就可以挣到56块。

张宜：那时候研究生读几年，三年还是两年？

冯志伟教授：读三年。

张宜：你想读到两年就答辩。

冯志伟教授：我想答辩。就在准备过程当中，岑老也觉得我差不多了，文章他也看了。虽然他看不太懂，他想既然我已写出来，大概已经成熟了吧，所以他基本上也不反对，他还是比较开明的。但是这时候发生了一件事情，就是1966年的5月25日，第一张马列主义的大字报贴到了北大饭厅的门口。我记得很清楚那一天是5月25日，因为那一天我要去买一本法文词典，因为当时的《法汉词典》很不好，很简单，单词太少了。我学过日文，我的日文还可以，岑老说你去买本《仏和词典》吧，《仏和词典》是《法日词典》的日文写法），我就到外文书店买了一本《仏和词典》。大概是中午吧，我刚刚在城里买词典吃完饭回来，回来一看北大饭厅前人头攒动。我伸头一看大字报贴着呢。上写“陆平、彭佩云你们要走往何方？”，言词很激烈，陆平是北大的校长，彭佩云是北大的党委书记，彭佩云现在是全国妇联的领导，他们俩当时被认为是北京市委的黑线人物，当时彭真已被揪出来了。我一看到大字报，我就知道我的论文泡汤了，一场很大的革命就要来了。

张宜：您做了第二次的预见。

冯志伟教授：在这种情况下北大大乱了。过了几天《人民日报》就发表了社论说，“这是一张马列主义的大字报”，一下把火点起来了。北大就进入“文化大革命”状态，保守派和造反派就打起来了。这时候我的提前毕业、减轻我家家庭负担等想法就不行了。这样我就有一个很重要的改变。“文化大革命”要来了，我就得放弃所有的业务，投入“文化大革命”。虽然我跟导师谈我家比较困难，我能不能快点把论文写完，快点答辩，然后找个个工作做。这个事情比起当时的“文化大革命”实在是很小的事情了，“文化大革命”是中国的大事！

张宜：您也就不敢再提了吧？

冯志伟教授：不提了。从此，就进入“文化大革命”，这时是1966年的5月份，5月以后，所有的业务就停下来了。我的论文写好了就放下了，等着参加运动了。我一直在北大呆到68年，我应该67年毕业，但不能毕业，不但不能毕业，还要坚持到68年，因为乱套了。68年分配工作了。我们等于是半巴拉子的学生，因为我们只读了两年的研究生，所以当时也没有算我们是研究生，就不明不白的毕业了。我们算什么呢？大学毕业又在北大呆了两年，又搞了两年的“文化大革命”，就属于这么一种情况，所以也没有说我们是不是研究生。好像工宣队发给我一张文凭，说我在北大研究生读了三年，因为后来多数的时间都在参加“文

化大革命”，不算数，大概勉强算个研究生吧。当时不敢说自己是研究生毕业，这是 1968 年。当时我在北大参加“文化大革命”不是特别的积极，因为我对“文化大革命”不太懂。我奇怪人们为什么要互相打、互相杀，原来关系很好的人为什么一下子成了仇人。我想不明白，也许是我的政治敏感性不太强。

张宜：也许是您的整个学生经历比较单纯。

冯志伟教授：所以老实说，学校里斗“反动权威”的时候，我跟这些“反动权威”的关系还都不错，他们被斗了以后，我见到他们还照样跟他们点头，他们都很奇怪。我记得王力先生后来碰到我时跟我说：我被斗以后，人人见到我愤怒得不得了，回家时见到冯志伟，冯志伟还给我鞠个躬，我很感动。他当时心里想怎么还会有这样的学生，其他人对我那么仇恨，怎么还会有人尊重我，所以我当时对运动不是特别积极。

张宜：当时王力先生和岑麒祥先生都挨斗了吗？

冯志伟教授：岑麒祥先生受到的冲击比较小，因为他的权威性不像王老师那么高。王力先生的权威性早就知名了，他自然就成了众矢之的，他被斗得很厉害。我记得当时还有人打他，有人拿棒球棒打他，我看在眼里觉得他们这些权威们挺惨的。当时王力先生的年纪跟我现在差不多，60 多岁，头发都白了。他们排好队，一个一个等着挨打，我很同情他们，所以一般我都不参与，我不明白这些人为什么坏，坏在哪里，我很同情他们。到了 1968 年的夏天，我分配到天津。离开了北大以后，我有了很大的改变，对语言学的看法有很大的改变。在北大时，我从理科转到文科来，而且又学了那么多文科的课，同时还学数学，花了很多的功夫，自己认为很有学问，但是离开了北大以后，跨出了语言学的圈子以后就有很多问题，别人不知道你搞的这是什么玩艺儿。你说什么主语、谓语，人家说你是疯子，你搞这个东西干吗？

张宜：就是说现实生活使您不得不作重新思考？

冯志伟教授：对呀。语言学有什么用啊。我在北大的研究生中属于温和派，当时我是北大的井冈山派，分配我们的是聂元梓。我是温和派，也没有得罪过他们，我们这一派有好多人给分到山西插队去了，而我能到天津大城市算是不错，因为我比较温和，没有仇人，就是这种情况。天津市教育局把我分到唐口三中教书。去的第二天就碰到了专业问题。唐口三中地处天津经济比较落后的地区，属于河东区，是搬运工人聚居的地方，这里的学生基本是工人子弟，工人也不是很正规的产业工人，而是搬运工人等。当时这个学校的“文化大革命”搞得很好，因为贫苦人的子女多。我本来想当语文老师，但是我去了以后发现社会上对于我们这种人的理解完全是两回事。在北大，王先生、朱先生、岑先生认为我是好学生，脑子比较好，外语比较好，文笔也不错；另外我还有点不同，我的数学比较好，所以在老师们的心中我是比较好的学生。到了社会上，人家不这么看，他们看你有什么真本事。我在唐口三中报到以后，学校领导问我你是学什么的，我说我是学语言学的。他说语言学是什么，我就说我念的是北大中文系，我希望教语文。他说语文课我们都能教，中国人不会讲中国话还行吗？我们都认识字，我们不需要语文老师，你能不能干其他的。后来我就想我还会什么呢？他问，那你懂不懂外语呢？我说我懂，于是我就在唐口三中教外语。教外语就教一句话：毛主席万岁。教英文：Long Live Chairman Mao！这句话要翻来覆去教一年，所以首先我让学生把每一个字母背了，然后把 Long Live Chairman Mao！写出来。

张宜：您教的是几年级的学生？一年里就教这一句话？

冯志伟教授：对啊。我教初三吧。好像初一、初二也都教这一句话，所以我当时得有些办法，不然这课我没法上啊！因为不让教其他的东西，所以我得想些招儿。一个是我让学生去跑步，就是把外语课变成体育课，一面跑步，一面喊 1、2、3、4，然后同学们一起喊：Long Live Chairman Mao！学生好像还有点兴趣。慢慢学生烦了，我就又想“击鼓传馨”的办法。大家坐在一起，然后拿个东西传，一边喊：Long Live Chairman Mao！如果一个同学说到“Mao”的时候，那个东西刚传到他手里，我一喊停，他就要受罚唱首歌。只有这样搞些花招才能把

这节课进行下去，学生也觉得腻，但是课堂上不敢教其他内容，因为当时提倡反对“地道英语论”：中国人讲的英语不能跟美国人讲的一样，因为我们是无产阶级，是最革命的人，只能讲 Long Live Chairman Mao！后来我觉得没办法，就去请示能不能增加点儿内容。领导说你要增加什么？我说再增加两句话。当时不是战备吗？我说增加“提高警惕，保卫祖国，要准备打仗”，这样说的东西能多一点儿。于是我就增加了两句话：Heighten the Vigilance, Defend Our Motherland。天天教这个，这样混了一年多，我实在呆得太腻味了，也没什么意思。后来我就想回到昆明去，因为我当时已经结婚了，我爱人在昆明，我是 70 年结婚的，在天津工作了两年后回昆明结婚的。我就想到昆明去。当时正好林彪有个“战备令”，动员大城市的人到边疆去，到“三线”去。云南属于边疆，我就在学校报名了。领导说很好，他们认为他们的任务可以完成，我又可以回老家，调动上没有什么问题，我就到昆明教育局报到。

我在昆明也碰到了同样的问题，就是对语言学的看法。我到了昆明教育局，分配我到昆明市第五中学教书，但是教什么还没定，反正就是当个老师吧。首先工宣队的人问我，你原来是教什么的，我说我是教英文的。他说，冯老师，我们这儿的英文老师已经够了。我说我是学语言学的，我可以教中文。他说我们不需要中文老师，教中文的我们拍拍脑袋就是一个！谁都懂中文，随便拍拍脑袋就是个教中文的，不需要什么技术。你是北大毕业的，能不能教点儿复杂的课。我说你是什么意思。他说你能不能教点儿数理化，我们教不了的。语文、政治没有问题，我们都能教。外语呢，我们这儿已经有老师，不需要了。我就说那你需要我教什么呢？他说教物理行不行？我说那就教物理吧。因为物理我也懂啊，我原来是学理科的嘛。于是我就在昆明五中当物理老师，从 1970 年开始一直当到 1978 年，当了 8 年的物理老师。这 8 年当中也在思考人们对语言学的看法，为什么我会教物理呢？因为大家对语言学没有兴趣，我们原来学习的时候自己觉得很有理论色彩，什么 Saussure, Bloomfield……等等，一到下面脱离了圈子，人家不管你 Saussure 的“语言”（langue）跟“言语”（parole）的差别，谁管你这些，人家认为你是疯子。因为人家根本不研究这些，而且研究这个也没有什么经济效益。比如说我教数学和物理，大家认为这是个本领。所以我认为我们搞语言学的人一旦离开了我们语言学这个圈子确实是一事无成，而且别人会认为我们是一群很古怪的人，虽然在圈里人家可能会说这位先生成绩很好啊，有很多贡献啊，但圈子很小。如果你把自己放进整个世界整个社会里面一看，你这个成就很渺小，而且老百姓根本不承认，认为你是废物。我就有这种想法，所以还搞不搞语言学我当时很怀疑。至于说用数学的方法研究语言学，我还有些兴趣，我还经常找些书来看。

张宜：这八年里您还经常找书来看吗？

冯志伟教授：我主要靠云南省图书馆，他们还进点儿杂志。还有云南科技情报所，他们进些外文杂志。我同这些单位关系搞得很好，因为我的外文还可以，我的英文、法文、俄文、日文都懂，所以对它们还有些帮助。比如它们出现一些问题，我能帮助它们解决，我还帮助云南科技情报所翻译了一部法文的关于海底开采石油的科技电影短片，所以，我和他们的关系搞得很好。他们有新的杂志，我可以看到，跟他们很熟，所以我通过这些还知道一些情况。比如我读到了 ALPAC 报告，就是美国的机器翻译报告，一般叫做黑皮书报告，也就是语言自动处理咨询报告，英文全称是 Automatic Language Processing Adversary Committee ——语言自动处理咨询委员会。我在昆明看到了这份报告。这个报告是美国 66 年写的，直到 70 年才传到中国。看了这个报告，我知道美国的机器翻译这时进入了低潮。因为机器翻译产生后，美国好多公司都投资搞机器翻译，政府也给了很多钱，投资了几百万美元，但最后搞的名声很差，因为机器翻译的效果不好，所以美国的科学院就组织了一个调查，成立了这个 ALPAC 委员会，专门来调查语言自动处理这个问题，得出的结论是在目前来说在美国的基础条件下，机器翻译不合算。首先是因为美国人基本不用学外语，科技文献都是用英文发表的。他们只要学点法语和俄语就够了，而且是少数人，没有必要大家都学，没有好多工作量，

所以花这个钱不值得。ALPAC 报告之后，美国就把所有关于机器翻译的课题都停了，所有的投资全部停掉，美国的机器翻译进入了低潮。我当时是在昆明了解到这一情况，可能国内还没有人知道。70 年代初，我在云南省图书馆和云南省科技情报所得知了这些信息。过了五年后在中国知道这些信息就算是新的了，因为当时中国把门都关起来了。唯一的渠道是偶尔看到这些资料知道一些消息。另外，大的一些理论研究进展我也知道一些，而且当时我知道 Chomsky 已经成了名人了。在 50 年代我念书时他还是普通人，到了 70 年代 Chomsky 已经很有名了。有好多夸赞他的文章，评他的形式语法，这些我都看到了，了解到了。这个人已经起来了，跟我原来的预见差不多——伟大的事情果然发生了。当时语言学发展的动态我还知道，而我自己将来能不能搞语言学，我非常怀疑。因为我是个中学老师，我的本职工作又是教物理。那时中国也没什么人搞语言学。当时北大老师跟我基本上没有来往，他们都到江西插队去了，包括王力先生都去了。岑麒祥先生当时七十几岁，也去五七干校插队去了。我在昆明的业余时间里还了解到一些计算语言学方面的情况，了解到一些消息。但总体来说，将来能不能做，我很怀疑。另外，我对语言学的用处也非常非常怀疑，对于原来所谓的权威啊、水平啊我都已看得很淡了，没有人承认你这个。比如说我到了云南省图书馆和云南省科技情报所，并不是因为我学过一些文字、音韵和训诂，大家才对我关照，而是因为我懂的外文多，外文有用，所以当时我想一个语言学家应该是“多语者”，polyglot，能够说不同的语言，说明你是个语言学家。如果一个语言学家只会说一种语言，而且仅知道某个词的内部变化，人家不需要你。人家老百姓根本不管你这些，人家需要的是雪中送炭，听不懂的时候问你什么意思，你能说出来，这才算是个语言学家。因为语言的功能是交际嘛，你能在他交际出现障碍的时候帮他解决困难，这才是语言学家。这样我对语言学的看法发生了很大的变化，原来只知道语言学的理论很强，要搞个系统，此时这个想法都没有了。

张宜：原来您是想搞得很深很专吧。

冯志伟教授：社会实践以后，你发现社会不需要你的东西，在圈里大家可以互相吹捧，说你很有水平啊，很有成就啊，但是到底有没有成就最后用社会来衡量。我对语言学的认识有很大的改变，包括现在我对语言学研究也是如此，我不愿搞那些离开现实特别特别远的题目。当然我并不反对别人搞，因为他们没有我这样的经历。我只好学点具体的本事。1978 年之前，我一直搞物理，我还参加了编写云南省物理教材。我在云南省还算是比较有名的物理老师，但是大家都不知道我是北大中文系毕业的，因为在物理课上也用不着我中文的东西。另外，在这段时间里，我还利用教物理的业余时间测出了汉字的熵(entropy)是 9.65 比特，一个汉字的熵也就是一个汉字包含的信息量的大小，这是汉字一个非常重要的自然常数，是汉字信息处理的基础数据。它的重要性正如测定圆周率对于数学的重要性一样。我认为这是我在 20 多年前做的最有意义的研究工作，可惜很多人长期不能理解这项研究的重要性。最近，由于统计方法在计算语言学中得到越来越广泛的应用，为了估计自然语言处理的效果，需要了解汉字的熵，熵的重要性越来越明显了。我想，随着科学的发展和时代的进步，我对于汉字熵的基础性研究还会引起学术界更多的重视。除此之外，由于条件的限制，在这段时间里，在计算语言学方面，我几乎不可能做其他的研究。“四人帮”倒台后，1978 年招生制度恢复了。毛泽东在世的时候讲过大学还需要办，我理解他的话是理工科大学还要办。所以根据毛泽东的指示，招生制度恢复以后，首先要恢复理工科大学，文科大学要不要恢复毛泽东没有指示，所以 1978 年首先招生的是中国科技大学研究生院，就是位于北京八宝山现在叫中国科学院的研究生院。当时我也写信问过北大，“四人帮”倒台了，你们还要不要搞语言学，北大没有很快给我回音，因为北大到底招不招生还不知道呢，但是科技大学肯定招生。当时我在昆明也很难考，我在昆明教物理已经有名了，而且高考制度恢复以后第一届 77 年的高考已经考了，我的学生已经有考上大学的了。我这个人比较会做题，我培养一些学生是做题高手，他们高考的成绩都很好，家长一下子对我特别尊重。我所在的昆明五中在大观楼附近，

有很多解放军的子女在我们那儿读书，他们的子女考上大学以后就说冯老师很厉害，经他一点拨学生考试的成绩都很好，所以我还很难走。

后来北大恢复了文科，但（研究生）招生没有恢复，北大中文系准备要搞语言学。当时有几位老师推荐我说，这个人读书还可以，是不是他若想回来就让他回来。北大军宣队的领导跟云南省委宣传部长是老战友，当时北大想要我就想通过关系“走后门”。北大就通过这位军宣队的代表给云南省委宣传部长写了一封信，他看了这封信说，我们云南省还有北大的人，那不错嘛，现在“四人帮”垮台了，要搞“四化”建设，需要人才啊。于是他就批文：此人北大要，说明他很有用，云南省不能放。这个批示就被送到云南省教育局去了。我活动了半天，上级领导把我叫去了说：你是不是叫冯志伟？省里有批示说你不能走。结果我就回不了北大了，我就没有办法了。当时研究生招生工作已经开始了。科技大学开始招生，这种情况下怎么办？当时我在云南呆得也不错，但我也想得到机会再发展，所以我就决定干脆考理科吧，一考就考上了。

张宜：你考的是物理吗？

冯志伟教授：不是，叫信息科学系。这样我就到北京来了。因为当时还没有新校舍，现在的北京林业大学，当时叫林业学院，文革时期下放了，所以房子是空着的，暂时拨给中国科技大学作为研究生院。我们研究生都住在那儿。当时我已经 39 岁了。我估计可能在科技大学研究生里算比较大的了。进去以后就等于我改行了。为什么我选择了信息科学专业呢？因为我有一个想法，信息科学跟数学的联系比较多，而信息科学往往牵涉到语言，这个系当时设有信息检索专业和机器翻译专业，都与语言学有关系，所以我也不能跑得太远，比如搞物理就意味着彻底离开语言学了。来了以后，不久就可以出国。

张宜：您 1978 年去了法国吧？

冯志伟教授：对。我是 1978 年考上研究生的，到了北京以后知道还有出国机会，学校让我们考，一考我就考过了，我就作为公派生到了法国。当时送我到法国，可能也是因为我年纪大了，因为一般年轻一些的都往美国送。人家可能认为我没有多少培养前途，我已经 39 岁了嘛，这样我就到了法国。

张宜：是法国的格勒诺布尔大学？

冯志伟教授：对。法国的格勒诺布尔大学应用数学研究所。这个所有“机器翻译中心”（CETA），也就是法文的“Centre de Traduction Automatique”。因为我是应用数学所的人，所以我还要学一些数学课，因为自动翻译也是从数学角度来研究机器翻译的。当时有个老师叫 B. Vauquois（沃古瓦），是个数学家，实际上是我的导师，是法国概率论专家，他是从事信息科学的概率论专家，是机器翻译中心的主任，同时，他又是国际计算语言学学会的主席，算是当时世界上第一流的计算语言学家了。他用数学方法搞机器翻译，而我对数学有兴趣，所以我们自然是一拍即合，我就由他来指导，成了他的学生。

张宜：您留学是去读学位吧？

冯志伟教授：当时是这样的，走的时候科技大学跟我说你的年龄大了，所以我们对你的政策是“国外开花，国内结果”，什么意思呢？你去法国留学把学分修完，把论文写完，但是你不能在法国取得学位，回到科技大学我们给你学位，但你要把你的学习成果拿回来。就是说不让我跑，要把我拴住。科技大学已经和法方达成了协议了，我是属于“国外开花，国内结果”的人，所以我必须回来。我在法国把论文写完了，把试验做了，然后我就回来了。在法国期间，因为我在应用数学研究所，我原来又是搞语言的，我就可以很好地把这些结合起来，这是一个很好的机会。所以我在法国我主要做了一个机器翻译系统，汉译多种语言的，汉语译成五种语言的，因为原来这几种语言我基本都会了，汉语译成法语、德语、英语、日语、俄语。我做了一个五种语言的翻译系统。

张宜：后来您是不是在这方面有一些课题？

冯志伟教授：当时我做这个系统的考虑是，第一，因为当时我有一种想法，就是国际上对于 Chomsky 的理论有很多改进。我对改进的情况基本上清楚，基本上有两派。一派是 Chomsky 本人，意识到自己的语法有问题，他的短语结构语法到了解决歧义问题时出现了好多问题。我们知道该短语结构语法比一般的线性语法要好，因为短语结构语法强调层次，层次比线性看法要深。所谓线性语法是说如果两个语言成分、两个语言结构如果它们的词序相同，词性也相同，那么它们就应该是两个相同的成分，Chomsky 或者一部分的结构主义者认为这还不行。词序相同，词性也相同不能说它们是相同的结构，还要求在层次上也相同。Chomsky 最有名的一个例子是：They are flying planes。关于 are flying，他有两种不同的解释：如果 flying 修饰 planes，这句的意思就是：它们是正在飞行中的飞机。如果 flying 是 are 后面的现在分词，这句的意思就成了：他们正在驾驶飞机。Chomsky 认为虽然 flying planes 词序相同，词性也相同，但层次不一样，于是就有两个不同的意义，这个理论对过去的语言学是个很大的冲击，是很新的思想，而且它是对的，就是说层次很重要。后来 Chomsky 发现这个还不行，因为有好多种现象还是不能解释。比如说层次相同就是完全一样的结构嘛？不一定。比如他举例说：the shooting of hunters。一个意思是“射击猎人”，一个是“猎人射击”。到底是打猎人还是猎人被打？层次一样，词性一样，词序一样，但是它们是不同的结构，怎么解释？这在语言学上是很难的题目。Chomsky 后来说我的短语结构语法已经进入了困境，已经解决不了了。他提出这个问题是因为转换造成的，这些东西虽然形状一样，层次一样，但因为是不同句子转换过来的。如一个是 Hunters shoot，猎人射击。一个是 Someone shoots hunters，有人向猎人射击。变成名词短语后都成了 the shooting of hunters。所以说 Chomsky 的观点是用转换的方法解释歧义，这种解释显然比靠层次来解释更好，更进一步了。那么 Chomsky 自己解决自己的问题，他自圆其说，后来进一步发展，产生了他的中期理论，就是“标准理论”。中期理论基本上是用转换的概念，就把转换进一步地发展了。这是 Chomsky 的想法。另外，还有一派，特别是计算语言学的学者认为转换不方便。转换要找句子间的关系，计算语言学上很难找，计算机很难判断不同句子之间的关系，所以他们认为应该就句子本身来解释句子的问题，也就因此提出各种各样的方法。当时在法国，我就处在语言学上这么一个重要的时期。大家都在想办法解决 Chomsky 解决不了的问题。两个语言成分，它们词性相同，词序相同，层次也相同的时候，你怎么能辨别它们是不同的，而且又不走 Chomsky 的转换的道路。我在法国时也同样产生了这种思想。我当时在法国跟我的导师商量做中文的 translation，把中文翻成五种外语。这个目的是很大的，关键是能不能把中文分析好，把中文的 parser（分析程序）做好，做中文的 parser 首先碰到的一个问题，就是 Chomsky 解决不了的问题到处存在，比如说“点心吃了”，实际上是“点心被吃了”，但中文一般不用“被”字；“张三吃了”，实际上是“张三把点心吃了”。“张三”是个 NP，“点心”也是个 NP，“吃了”是个 VP，就是 NP+VP= 一个 Sentence（句子），那么层次相同，词序相同，词性也相同，但他们却是不同的东西，怎么来解释？你用机器做出来肯定是一样的树形图，不同的只是在结点上的词不一样，上面是一样的 S=NP+VP 嘛，很简单，但是为什么是两个含义不同的句子？怎么解释？中文里到处都是这样的句子，因为中文里的被动关系不清楚。当时我跟导师 B. Vauquois 讲了，他说中文是种 terrible language（糟糕的语言），他说哪种语言能够不分主动和被动，人吃了和被人吃了怎么能是一样？怎么这么乱？我说，一点儿也不乱，我们中国人分得很清楚，因为我们中国人知道在一般情况下，人是不能被吃的。所以“小王吃了”，不能是“小王被吃了”，而点心它不吃东西，所以它必定是被吃的。汉语是靠词汇来解决语法问题的，但是对于法国人来讲，他们认为没有被动语态的语言就 terrible，我认为汉语一点儿也不 terrible。所以我们不能按照法语来解决问题，英语没有这个问题，因为 Chomsky 已经解决这个问题了。汉语中最常见的主谓句子“王冕死了父亲”，到底谁死了，王冕死了，还是父亲死了，在思想上必须要思考一种新的语法理论来解决这个问题，这也是我在法国解

决的一个最主要的问题。当时国际上的语言学家都在想这个问题，就是如何要在计算上不很麻烦，同时又不走 Chomsky 的转换道路，自己走一条简捷的道路。当时我提出了两点：一点是我们要采用多标记的办法，就是做出一个树形图以后，在结点上不能只打出 NP，我要说明它是个什么 NP。如果我能假定“小王吃了”，我能说小王是个 agent，是个施事者，绝对是小王吃东西了；如果“点心吃了”，NP 就是 patient，受事者，“点心”是受事，绝对是点心被吃了。如果我能通过汉语语法把 agent 和 patient 分离出来，而且能反映在我的计算语言学的树形图的节点上，这回就不能出错了，这就是我的“多标记”思想。而 Chomsky 是“单标记”的思想：S=NP+VP。多标记要在程序上下功夫，如何让树形图表示多标记。另外，我觉得在处理中文当中不能用“二叉树”binary-tree，因为语言成分具有“二分性”。在 Chomsky 理论中也叫 Chomsky 的范式。Chomsky 范式是二分的，Chomsky 认为“二叉树”是自然语言描写的最好形式，Chomsky Norm Form 在语言学上叫做 Chomsky 的范式。本来 Chomsky 的短语结构语法是多义的。但他认为“二叉”是最好的。中国的语法学家模仿美国人，比如说丁声树写过一本《现代汉语语法讲话》⁶，就说过“二叉”。他提出汉语有 6 种结构：主谓结构、述宾结构、偏正结构、述补结构……，都是二义的：主+谓，述+宾……等等，彻底的“二叉化”。我想如果做汉语的话，“二叉化”不行。我考虑到，一个兼语式，比如说：“请小王吃饭”。“小王”做“请”的宾语，他又做“吃饭”的主语。我用树来表示的话，“小王”上面就会有两个“父亲结点”，因为“小王”既是主语又是宾语嘛，而一个结点上上面有两个“父亲结点”就不是树，树只允许一个“父亲结点”和多个“儿子结点”，因为“儿子结点”必从一个父亲结点“生”出来，不能从两个父亲结点上“生”出来，两个父亲结点就不能是树的形状了，就违反了树的定义。如果坚持二叉树或者二分，那么到底“请 / 小王吃饭”，还是“请小王 / 吃饭”，怎么做呢？这在机器上没法做，所以兼语式只能采取三分，就是“请 / 小王 / 吃饭”。“小王”是兼语，“吃饭”是第二个动词。这样可以避免树的交叉。另外状语修饰述宾结构时，状语到底是修饰动词还是整个述宾结构呢？比如说：“努力学习法语”。“努力”是修饰“学习”的，还是“努力”修饰“学习法语”？语法学家有很多争论了，有人认为是修饰“学习”的，有人认为是修饰“学习法语”，其实我认为这样的争论没有什么意义，因为语感不一样，都有一定的道理。如果采用二分的话，就又有困难了，到底分成“努力 / 学习法语”还是“努力学习 / 法语”，这就发生矛盾了。我在机器上就没办法做。所以遇到这种情况，我们就采取三分法，“努力 / 学习 / 法语”，这样就不用争论了。基于这种原因，我主张用“多分法”来处理兼语结构和状动结构。连动结构，我也是主张采用多分的方法。比如说，“他骑车上街去买菜”最好不用二分，用多分最好，至于是什么意思，再去辨别。我当时就提出这么两点：第一是多标记方法，第二是多义的观点，把“二叉”变成“多义”，把“单标记”变成“多标记”，这么两点是对 Chomsky 的一种改革。Chomsky 的理论是单标记的，我的办法则是多标记，可以解决中文问题。其次，Chomsky 主张二叉而我主张多义的树分析法，这就是我在法国完成的研究，叫 MMT Model，即多义多标记树模型，Multiple-labeled and Multiple-branched Tree Model。我提出了多标记多义的树分析法，让我在机器上造出一种多标记多义的树，这是我在语言结构方面的观点，用它来做机器翻译。另外，当时我还接受到了“依存语法”，叫 Dependency Grammar，主要是由法国的 Lucien Tesnière 提出来的，因为当时我必须找到一种理论支持。我在 Chomsky 的语法中发现了上述的两个毛病，我提出了我自己的观点来改进它。另外我发现动词的分叉很重要，所以我必须要把动词研究好。要研究动词的话，就要找到一种理论，所以我就找到了 Dependency Grammar，现在多译成“依存语法”，我看了“依存语法”的书，叫做《结构句法基础》，是用法文写的，英文叫 *Element of Structure Syntax*⁷。这本书给我的

⁶ 《现代汉语语法讲话》，丁声树等编，商务印书馆 1961 年版。

⁷ L. Tesnière, *Element de Syntaxe Structurele*, Paris, 1956.

最大启发是我认识了“价”的概念，知道可以用动词的“价”来把动词的特性区分开来，用动词的价来搭起整个句子的结构，那么这两点是我在法国学习的最主要的成就。一个是搞了一个 MMT，一个是用 Dependency Grammar 来做我整个研究的基础。我在机器翻译中，首先用多叉树把汉语的句子做出来，另外中心的东西是根据动词的配价来做的。我要把一个动词前面带什么了，后面带什么了，都做出来。我们管这些叫 argument，叫“论元”。我们把“论元”弄清楚，以这个来做整个的汉语分析。事实证明这样做是成功的，因为我整个的实验做得很快，我大概花了一年的时间就把整个中文的分析基本上做出来了。原来我用 Chomsky 的方法做，遇到像“小王吃了”、“点心吃了”这样的句子，一下子就完蛋了，做不下去。现在我就豁然开朗了，全都能做了。然后我就分析了中文的大概几十篇文章，基本上是《光明日报》上的一些短文。比如说有中国特色的文章，如“张衡的候风地动仪”，“发雕（在头发上雕字）”，“中国猿人”等文章作为我的语料。当然我的语料也很小，我只想通过分析这几十篇文章来检验我的理论。把中文做了以后，我就开始把它翻成外语，我原来只是想做一个 prototype（原型系统），因为在程序上我想采用模块化的方法，如果语种越多就越难，因为要转换；这样，我就故意把它做成六种语言的系统，汉语是源语言，然后自动翻译成五种目标语言。当时我才 40 岁，我还年轻，我就拼命的干活，所以做了一个很大的程序，做了 6 个模块，每一模块都自己写规则，我做英语，我就看英语的语法书，然后用我的模块把英语的语法写完。然后把上面提到的几十篇文章一一译成五种目标语，这样我大概用了两年多的时间做完了。做汉译外很有好处，因为我当时身处外国，外国人都不懂中文。我把中文输进机器，一旦出来法文，法国人很兴奋，因为他们终于知道输入的那些乱七八糟的东西是什么意思了。包括后来的英国人，日本人等整个实验室中的工作人员看到我做出来的东西一片欢腾。这件事后来在巴黎还报道过，算是我一个不错的成果吧。这是在 1981 年吧。这时我的工作已经做完了，等于我已经在法国做了一个机器翻译的模型吧，是多语言的，名字叫做 FAJRA（汉—法/英/日/俄/德系统）。这个系统是世界上第一个把汉语译成多种外语的机器翻译系统，也是我们中国人做的第一个汉译外系统，也可以说是我们中国人第一次在机器上真正玩汉语，把复杂的汉语玩通了，所以这个工作还是挺有意思的。到了 1981 年，我的工作完了，很快我就回国了，因为我是“境外开花，境内结果”，我就回国通过我的论文（答辩）。

张宜：这回您是正儿八经的硕士毕业生了吧？

冯志伟教授：对，科技大学硕士毕业，1981 年毕业。回国后，我是一个理科学生，我已经不是文科学生了，所以回国以后，我没有想到要回到语言学的队伍中去，因为我认为我在语言学的队伍里摔了一个很大的筋斗，这一摔就是十几年，摔得很厉害，我要从理的话，可能就不会有这事。我去天津，去边疆转了这么一大圈，然后又文科跑到理科，国外转了一圈又回到北京。后来科技大学把我分配到中国科技情报所计算中心，现在叫中国科技信息研究所计算中心。我在计算中心当机器翻译组的组长，率领一帮人做机器翻译。回国后，我又接着做了一些机器翻译系统，主要是倒过来做：做了法—汉，德—汉，用的方法还是我的 MMT 模型，多叉多标记的方法，还能做得通。我的工作是在科技信息所完成的。同时我们还做了一个实用型的系统。翻译英文的 title，翻译科技文章的题目，这是一个英译汉的系统。我在科技情报所待了好几年。从 1981 年到 1985 年，我在那儿待了四年。

张宜：我注意到 1986 年您又一次出国，这次您去的是德国吧？

冯志伟教授：我后来就来到现在这个单位（国家语言文字工作委员会语言文字应用研究所）。

张宜：那是什么时候呢？

冯志伟教授：1985 年的下半年。回国后，我跟语言学界还有些联系。这个联系主要是跟北大。我曾是北大的学生。1982 年，应北京大学的邀请，我在北京大学中文系汉语专业开设了“语言学中的数学问题”的选修课。这是国内首次在高等学校全面地、系统地讲述数理语

言学的课程，受到学生们的欢迎。北京大学校长、著名数学家丁石孙教授很支持这门新课程，他是数学家，他认为数学对语言学有用。在他的专著《数学与教育》一书中，对我的这门课程作了如下的评价：“1982年，北京大学中文系开设了“语言学中的数学问题”，这是给汉语专业学生开的选修课程，许多同学对这门学科产生了很大的兴趣，经过一个学期的学习，同学们初步认识了现代数学的发展给语言学注入了生机，觉得获益匪浅，对语言学这门古老的学科分支的发展充满了信心，而且，这一举动冲击了相当多的人的旧概念，使闭塞的中国学术界认识到，即使在人文科学教育中，数学也在逐渐起作用。”在北京大学讲稿的基础上，我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。但是我的绝大部分时间是在科技情报所做机器翻译工作。我们当时还没有汉字。回国后的一个大问题是没有汉字，因为做外译汉，你要有汉语的 transcript（转写文本）。没有汉字就做不了。翻译时机器里出的代码我们看不懂。1981年还没有汉字软件系统。当时如何输出汉字是我们的大问题。在科技信息研究所，我的工作带有实用价值。我们为国家把外文用计算机译成中文，然后让领导和科学家们看到汉字。译得好不好先不说，至少要出来汉字。但是你如果说，对不起，现在只能给你看“穿孔纸带”，这个“眼儿”代表的是什么什么，那还不如原来的外文好读呢！所以这是个很大的问题。对于我们来说，出现这个“眼儿”，我们已经成功了。但是别人不知道这些“眼儿”代表的是什么意思，所以汉字的问题是个大问题。当时中国的微机还没有普及。我们在 IBM-4381 机器上做机器翻译，在当时这是很大型的机器，我们还在日本高千穗公司制造的 TK-70 计算机上做机器翻译，这也是当时很好的计算机了，结果出来的是一些代码，不出来汉字，这是一个很大的问题。80年我们只做过一个汉字编码，仅是国标而已，就是 GB2312-80。国标码就是个标准，当时还没有出现汉字，所以我们自己还解决不了汉字的输入和输出问题。当时我们查看各种文献，知道日本有这样的东西。我们把日本的 T4100 汉字处理机买来，花了很多钱。当时情报所比较有钱，这样把我们的代码变成了日本汉字，这是在1982年。首先把汉字代码输到 T4100 机器里，变成日本汉字输出来，就可读了。当时我很兴奋，我估计这是在中国土地上第一次由计算机输出汉字，可惜这个技术是日本技术。到了1983，1984年以后，就逐渐有了汉字系统了。286个人计算机（IBM PC）出现了，可以逐步输出汉字了。1985年国家语言文字工作委员会成立。我们这个大院原来是文字改革委员会，它是中华人民共和国成立以后政府为了改革汉字简化汉字搞汉语拼音而成立的，直属于国务院，委员会的主任是吴玉章，他是毛泽东的老师，所以这个委员会的级别在当时是相当高的。文字改革到了1985年的时候，就不能再改了，因为不能天天改革文字，文字要有相对的稳定，所以政府也觉得文字改革委员会这个名称不太好。有外国专家来访看了这个名字也说，你们的文字天天都在改革吗？老百姓怎么办呢？这是个误解，我们不是天天在改，文字不能天天变。每天改一个字的话，那就乱了，所以政府要把文字改革委员会的名称改了。1985年改为国家语言文字工作委员会。这个名称比较好，比较稳定，当时的语言学界认识到了信息处理的问题。当然当时的认识还很有局限，以为语言文字的信息处理仅仅是把汉字输到机器中去，所以出现了各种各样的输入法专家。语言学家认为这是个技术领域，他们不大愿意介入这个领域。他们需要一些懂技术的人，所以国家语言文字工作委员会成立后就要找这样的人。当然，我就会被他们“瞄准了”，因为我有北大中文系那一段历史，知道我是学语言学的，现在此人又在理工科，就要把我弄回来，跟我也谈过话。我当时确实不大想过来，我的想法是我就在理工科待着，然后搞点儿语言学算了。这样子，第一，我经济上比较好过，搞理工科工资多，出国机会也多。到文科有各种各样的障碍和限制。另外，我跟文科的人交流可能有各种各样的障碍，他们的想法跟我的不大一样。我是自然科学的头脑，那么文科人讲究各种各样的论证，我不太懂，我已经脱离文科多年了。当物理老师就当了八年，完全是一道题怎么做，怎么解，怎么做实验，所以文科的方法我已经不是很熟悉了。当时我就有些犹豫，后来几位领导陆续找

我谈话非让我过来，1985年我就重新回到语言学队伍中来。

张宜：当时语委的领导是谁？

冯志伟教授：陈章太。他现在已经退休了。他当时是语委的副主任，主抓业务的，主要由他跟我联系。他动员了我好多次，我们也是老朋友了。后来我也就不好意思了，我就听他的，我就过来了。

张宜：您原来的单位放您走吗？

冯志伟教授：我原来是中国科技信息研究所计算中心机器翻译的头儿，我也都安排好了，我在那儿还培养了几个学生，也都被派到法国去学机器翻译。我说不久他们也要回来。当时刚刚从法国回来一个博士，我走正好给他安排一个位置，接替我做机器翻译组的组长，所以在阻力不太大的情况下，就把我放了，因为我的走对他们的工作没有太大影响，我都已经安排好了，他们没有表现出太大的不情愿，我也就安全的过来了。这样我就开始搞语言学了。过来以后，我又去了一次德国，那时我已在这儿工作了一年多。我是1986年去的，因为当时我还是理科的人，中国科学院有一个协作项目，这个项目需要一个懂德语和法语的专家，又要有信息科学基础，科学院没有这样的人，他们大多数只懂英语，而我会说法语和德语。所以，科学院软件研究所的所长许孔时教授就极力推荐我去。

张宜：您是什么时候学的法语？

冯志伟教授：我原来不是和岑麒祥教授学过一点儿法语吗？后来到了法国以后，我的法文还不够好，因为我要听课，要考试，虽然年纪大些，但我是学生啊，所以我就苦读法文，差不多有一年的时间。因为我在法国已经做过德语的机器翻译，我能阅读德文，但当时我还不会讲德语。我在法国做德语机器翻译我得学德文呀，我得把德语语法书看完又要能看懂德语的句子，我才能 parse 它。

那时我还是软件所的兼职研究员，所以我并没有完全脱离理科，后来科学院就把任务分到软件所了，软件所的所长许孔时就说这个任务给冯志伟是最好的，可他现在到语委去了。于是，许孔时他们就亲自来语委找语委的领导谈，也找了陈章太他们几个。我在这儿才干了一年的活，但几个领导挺开明的，说那就算科学院借用我吧。当时我们研究所在行政上属于国家语委管，在业务上我们是社科院的语言文字应用研究所，我这次就算是科学院与社科院联合派出的，派到德国去执行科学院的任务，科学院跟德国的夫琅禾费研究院(Fraunhofer Gesellschaft, 简称 FhG)有合作，我去做一个术语数据库，做一个叫做数据处理的术语数据库。我原来对术语也不懂。这个工作首先要求懂得数据处理；第二，这个数据库是多语种的，德文、法文、英文，所以我要懂这三门语言，我要做充分，我又要临时学习术语学，学习计算机术语学。我的兴趣是对语言进行分析，所以我的库做的跟别人不一样，别人做就是把意思写出来，德文怎么说，法文怎么说，英文怎么说，中文怎么说就完了，而我认为中文术语里词和词组是常用的东西。我要说它的结构是什么，强调文法而且说明它有没有歧义，在什么情况下会存在歧义，我要从同样的结构中看出它的不同，而且说明它为什么不同，所以我的研究方法跟他们的不太一样，我的方法会有更多的语言上的解释。所以我的库里面有 parser 这部分，要自动来分析结构，跟别人做得不太一样。在德国我学习了术语学，我主要研究歧义问题，研究术语歧义问题。当时我没有想到研究歧义对后来我的研究还是很有用的。我提出了 PA 理论，就是 Potential Ambiguity，叫做“潜在歧义”，我的意思是说我们很多语言学家研究歧义，找出一个歧义格式来他的任务就完成了，比如说，在中文中，一个 VP+NP₁ (的)+NP₂ 结构，NP₁ 加上一个“的”以后可能修饰 NP₂，NP₁ (的)+NP₂ 实际上等于一个大 NP。VP+NP 实际上就是一个句子 sentence，因为 VP+NP₁ 可能变成一个述宾结构，这个结构加上“的”可能修饰 NP₂，可能变成一个偏正结构，所以这种格式叫做歧义格式。比较具体的例子：消灭了敌人的主力部队。第一个意思：消灭了 / 敌人的主力部队（述宾结构），第二个意思：消灭了敌人的 / 主力部队（偏正结构），那么就出现了两种结构。朱

德熙说这就是歧义格式。我们国内的歧义研究到此为止，没有进一步深入研究，研究者提出好多歧义格式就完了，发现了歧义后任务就结束了。我做术语的歧义时，做 PA 论时，我发现了歧义问题，但我认为仅发现是不够的，因为并不是所有的具有这样格式的短语都是歧义的。比如说“消灭了敌人的主力部队”有歧义；但是比如说“摧毁了敌人的碉堡”这个短语只有一个意思，不可能是“摧毁了敌人的 / 碉堡”，只能是“摧毁了 / 敌人的碉堡”，所以不能说是凡有 VP+ NP₁ (的)+NP₂ 的结构都是歧义的，只有在某种情况下才有歧义。这个歧义是由于种种的复杂词汇的语义因素造成的。所以我说这个所谓的歧义格式是潜在的，不是直接表现出来的，真正的歧义我叫做 Instantiation，“实例化”。要实例化以后，变成实例以后才能判断歧义，而辨别实例受词汇的各种句法词法的影响。这是我的观点。我用这个观点来研究整个的术语库的歧义，研究得很好，然后变成“PA 论”的一个完整的系统。回国以后我也写了一篇文章在《中文信息》上发表⁸，我们在做自然语言的时候不是到处都能碰到歧义问题嘛，很严重啊，那么大家回头来看，需要一种理论来支持它。朱德熙先生的歧义格式不能说明所有的句子，只有少数是歧义的。后来我进一步讨论朱德熙的格式 VP+ NP₁ (的)+NP₂，把它例化以后真正有歧义只有 5%~10% 的句子，90% 的句子是没有歧义的。这是一种有很小歧义可能性的格式，而不能说他就是一种歧义格式，所以潜在歧义理论一旦搬到自然语言里，搬到普通日常生活语言中，它更明显，因为术语本身在命名时就要讲究没有歧义，术语要具有单一性，我研究术语歧义本身就发现了这么多歧义。我把处理术语歧义的本领拿来作自然语言，作普通语言，用我的理论，解释力就更强了，后来就逐步形成了我的 PA 理论。

当然现在国内有很多人作 PA 理论，我的博士生也在做。他们的工作等于是进一步用大量的例证来印证我的 PA 理论。我们在真正使用语言时，歧义本身的存在和发生，比如说为什么“消灭敌人的主力部队”会出现歧义呢？这是因为我们在日常生活中使用语言时像这种有歧义格式的量只有 5%~10% 左右，所以语用的歧义决定我们人的思想对歧义的认识。所以潜在歧义是现实世界在语用当中的反映，这种反映根深蒂固，我们人的语言知识根本不是人脑中天生具有的，也不是人聪明而想出来的，是人在社会中的语言实践决定了的，这是语用实践决定的。

张宜：你现在跟自然科学界还有联系吧！

冯志伟教授：因为我研究的是计算语言学，所以我不放弃跟理科的接触，所以我现在虽在文科单位，跟自然科学界还是有接触的。我是中国科学院自动化所模式识别国家重点实验室（NLPR）的学术委员会委员，同时也是全国自然科学基金会信息科学部的评审委员。我也参加自然科学界的活动。我在韩国科学技术院（KAIST）计算机系任兼职教授。我 11 月份还要去韩国科学技术院计算机系讲学。我是两个领域都在做。如果问我到底是个怎样的人，我说，我还是个语言学家，因为我是从语言学起家的。

张宜：也就是说您希望评价您的时候说您是位语言学家吗？

冯志伟教授：我希望是这样的。即使我做数学，我也是为了语言的目的。包括我在理科做 machine translation（机器翻译）全是语言工作，所以别人叫我填专业时，我在理科填应用数学，因为我学习过应用数学，我不能说我参加理科的评审会却说我是文科出身，但是绝大部分的情况下，我填的是语言学，因为我是个语言学家。

张宜：我看您的简历上写的是专攻语言学和信息科学。

冯志伟教授：当然了，我两方面都做嘛，但主要还是在语言学上。可以说语言学是我的第一个专业，也做一些信息科学，基本上是这种情况。

张宜：我还看见您在 1990—1993 年期间在德国特里尔大学作为客座教授，这是给他们讲学吗？

⁸ 见《潜在歧义理论用于自然语言处理》，《中文信息》1996 年第 1 期，总第 50 期。

冯志伟教授：是的。给他们讲计算语言学，另外也讲语言学。计算语言学嘛，主要讲计算语言学的基本原理了，语言学主要讲跟中文有关的。比如说我在那儿开过词汇学的课：汉语词汇学，汉语拼音正词法研究，中国古代散文选，中国当代散文选。我用德文授课，把中文 translate into German（译成德文）。我还用德文写了一本关于汉字结构的书在德国特里尔科学出版社出版。

张宜：我想您古汉语的底子也应该不错吧？

冯志伟教授：我的古汉语的底子是不错。因为小时候我进过私塾。我背过《古文观止》，《幼学琼林》，三、百、千：《三字经》，《百家姓》，《千字文》等我都背过。很小的时候背的，但是我不知道是什么意思，因为不背会，老师要打手心，为了不疼就要背，怕老师打手心。我那时也不知道将来对我有什么用，但是后来我从理科转到北大中文系，（古文底子）对我就特别有用。因为在中学时代我的兴趣还在理科上，我数学比较好，所以我看文科的东西不多，这些书虽然我会背，但还没有认真看到。《古文观止》我会背，但我没认真看过。转到北大中文系后，我要读这些东西了，一读才发觉我早都会背了，连老师都奇怪。考试的时候，我可以大段引证。我古文还不错，读古书是没有问题的，不过就是一直没有用到这个能力，所以现在基本上是文理两边跨着做了。现在年纪大了，也不太可能做大课题了。另外我的眼睛不太好，眼底黄斑有些问题，长期在计算机前工作，可能有些麻烦。本来我的专业是要用电脑的，现在因为有些毛病，所以不能长期用电脑，每天上机尽量不超过两个小时，所以到60岁以后，我就不再 programming（编程）了，我现在已不编程了。

张宜：可是我看您有好多著作啊、文章啊，到2000年的时候还很多嘛。

冯志伟教授：工作还在做。

张宜：那不都需要用电脑吗？

冯志伟教授：机器翻译需要我用电脑，我得写程序、调程序，这些工作我60岁以后就不再做了，但是论文还在写，我大部分是在研究一些算法。最近我在《语言文字应用》、《当代语言学》上发表的文章⁹都与算法有关系，就是一个事情怎么做，用什么算法效率才会高，基本上研究这些问题。《当代语言学》上我已连续发表了两篇文章，第三篇也要发表了。最后还有一篇，基本上是用自然科学方法来做文科，最近几年在做分析技术吧，因为年纪大了，这些问题也只能从理论上研究了，现在我在广播学院应用语言学系还有五个博士生，他们是计算语言学专业方向的。学生来源有搞自动化的，有搞数学的，还有外语的，两个中文的。

张宜：他们在各自的学习上是不是表现得很不一样？

冯志伟教授：大家互相学习。文科学生进来后要到计算机系补课，要学习数据结构跟离散数学，这两门课是必修的，是本科课，叫做博士生的本科补课，理科的学生进来要上中文系补课，语言学基础等，要考试的，这样慢慢地融汇起来。

张宜：冯老师在您的学术经历中对您影响比较大的人是谁？

冯志伟教授：岑麒祥、朱德熙先生对我的影响比较大。王力先生对我的鼓励比较大。我从法国回来后，首先去看王力先生，我跟他的关系比较好。文革时期见到他我向他鞠躬，他对我的印象也不错。他认为我是有良心的学生。另外，中国社会科学院语言研究所的刘涌泉先生对我的鼓励和帮助也很大，刘涌泉是我国机器翻译事业的创始人，每当我遇到困难、一筹莫展的时候，他总是坚定不移地支持我。他们都是我的恩师。

张宜：从学者的角度，他们能包容您，特别是在当时不被正常主流接受的时候能肯定您的想法。

⁹ 见《自然语言处理中歧义消解的方法》，《语言文字应用》1996-1、《机器翻译——从实验室走向市场》，《语言文字应用》1997-3、《中文信息处理专题研究（3）机器翻译，主持人的话》，《语言文字应用》1997-3；《标准通用置标语言 SGML 及其在自然语言处理中的应用》，《当代语言学》1998-4；《基于短语结构语法的自动句法分析方法》，《当代语言学》2000-2；《线图分析法》，《当代语言学》2002-4；《花园幽径句的自动分析》，《当代语言学》2003-4；《一种无回溯的自然语言分析算法》，《语言文字应用》2003-1。

冯志伟教授：他们认为我没错，他们认为我也不古怪。如果他们反对，我就没法儿待了。原则上他们还支持我，虽然岑先生觉得我有点儿怪，但还是不反对我做。

张宜：一件能改变您的事是什么？

冯志伟教授：应该是 57 年我在北大看到的 Chomsky 的文章。当时使我产生了用数学的思想去研究语言学的想法，这个想法促使我转入语言学，不然我跟语言学就不会有关系，促使我下这么大的决心，为此耽误了两年的课。

张宜：中国语言学的后来人应该感谢 Chomsky。

冯志伟教授：不能完全这样说，但这事还是有些好处。Chomsky 最近几年已经发展到带有很浓的心理学色彩。他要研究人为什么能说话的问题，他本身理论的运用色彩越来越不清楚了。但我认为他年轻时做了一些工作，他数学很好，他是宾夕法尼亚大学毕业的，是 Harris 的学生，他曾经做过希伯莱语的结构分析，他的博士论文是研究转换分析，他还是个语言学家。他的数学为什么那么好，可能是他自学的。我最早看 Chomsky 的文章，看到他的《语言描写的三个模型》，我顺着他这篇文章的参考文献中的《语言理论的逻辑结构》(*The Logic Structure of Linguistic Theory*)，进一步又找到另外几篇文章，这些文章的思想其实他早已经考虑成熟了，可是在 1959 年、1963 年才正式发表，《语法的形式特性》，*Formal Properties of Grammars*¹⁰，还有一篇《语法的某些形式特性》，*On Certain Formal Properties of Grammars*¹¹，这两篇文章都是数学非常深的。他讲的是一些代数结构的理论，这种数学的理论奠定了他后来理论的基础。Chomsky 这种研究语言的方法在人类历史上是绝无仅有的，所以他的思想在当时是不被接受的。Chomsky 的思想在 1956 年时就已经很成熟了。我看他那时的文章，他已经搞到了语法的形式分类，提出了 4 种文法：0 型、1 型、2 型、3 型。而且把文法和自动机对应起来，整个系统已经有了。但是他的东西没有人接受，他后来有名以后才登他早就考虑成熟的上面的那两篇文章。他的文章投稿就被退稿，认为是胡说八道，其实这里面语言学里非常重要的思想已经出现了。但是我们这些人还看不到，所以他是一位被退稿的投稿者，就在 Chomsky 非常困惑，没有办法的时候，有的人就劝告他，说你能不能把你的文章写的简单点，通俗点，因为大部分人很少懂数学，他就写了 *Syntactic Structure*（《句法结构》）¹² 这本书。其实他这本书是刚才提到的他这两篇文章的通俗读物。中国学者一般不读这两篇文章，中国的语言学者一般不会去读，要抽象代数学很好的人才能读懂。所以 Chomsky 的主要贡献是这两篇文章。这两篇文章太深奥了，所以是被退稿的两篇文章。直到 60 年代 Chomsky 有名以后，这两篇文章得以发表在《数理心理学手册》(*Handbook of Mathematic Psychology*) 上。

Chomsky 的 *Syntactic Structure* 《句法结构》写得很通俗，一般人都能读懂，但他的书里有很深奥的话都没有展开，展开以后怕人家看不懂，所以他写得非常简单，例子也是非常大众化的例子。他把书稿投到荷兰的 Mouton 出版社，这个出版社认为这个小册子写得挺有意思的，就给他出版了。出版以后，影响也不大。Chomsky 还是希望他的思想能被别人重视，他就请了他的朋友 Robert Lees 写一篇文章，非常强烈的攻击 *Syntactic Structure*。我认为他们是故意这么做的，但是不是商量好的我不知道。Robert Lees 跟 Chomsky 的关系很好，他为什么要凭空写这么一篇文章去骂 Chomsky 呢？我说不好，我猜他们是有准备的。Robert Lees 的文章发表在国际最高的语言学刊物 *Language* 上。于是 Chomsky 和 Robert Lees 就在 *Language* 上展开了争论，争论的结果 Chomsky 就成名了。可以说 Robert Lees 在某种程度上做了 Chomsky 的“托儿”，这是我的猜测，因为他们俩那么好的关系，怎么打得那么厉害，

¹⁰ Formal properties of grammars, In R. D. Luce, R. Bush, & E. Galanter, (Eds.), *Handbook of Mathematical Psychology*, Vol. 2, 323-418, Wiley, New York, 1963.

¹¹ On Certain formal properties of grammars, *Information and Control*, 2, 137-167, 1959.

¹² *Syntactic Structure*, Mouton & Co., S-Gravenhage, 1957. 中译本，《句法结构》，邢公畹、庞秉钧、黄长著、林书武译，中国社会科学出版社 1979 年版。

实际上他们演了一场戏，文字之争，然后引起学术界对他的重视，可能是这样的。

我做的计算语言学本来是个实用性学科，可以说我本人没有很高的能力，我也不准备赚很多钱。现在我也做一些应用型的研究，做一些机器翻译系统，有好多，我已记不清了。有汉译外的，FAJRA 系统，我做了法—汉、德—汉、日—汉、英—汉，还做了一个英—日，是帮助日本的 NEC 公司做的，这些系统中，只有帮助日本人做的这个系统实用化了，NEC 公司买了去装在一个 word processor（文字处理机）里面，其他所有系统我都是研究它的理论，我要通过这个系统来说明一种理论怎么做，研究算法和技术，我做机器翻译系统有一个特点，所有的系统都是我一个人来做，我没有别人帮忙。除了给日本公司做，因为是个工业系统，我一个人做不了那么大，所以请人帮忙做，其他系统都我一个人做。从编词典到设计语法，最后把语法变成 programming，工作量很大，这些系统也没法儿实用化，因为你要做一个实用化系统，词典起码要有几十万字，要处理一些非常特殊的规则，我的系统做得很有理性，规则都有我的语法观点，词汇量不能很大但要处理很多的多义词，所以我会在我实验的范围内把系统做得很好，但是这些系统显然不能拿到商业市场上去，他们可以随便拿一些东西就把我考倒，因为有一些东西我可能没有想到。所以有人管我的系统叫 toy's system（玩具系统），这是学术界对我的一些批评，说我做的东西是不少，但没有做出能卖很多钱的东西。

我是我国第一个做汉译外机器翻译系统的人。对此我有我的看法。第一，机器翻译需要做一些理论探讨，特别是有一些基础需要研究，如果所有人都去搞实用系统，都变成商人，都跟公司结合起来，你要做实用系统，你必须处理各种各样特殊的问题，甚至你为了要翻译得好，你要不惜违反语言学规则，做一些假动作。比如说，你在深层加一些特殊规则可以把机器翻译改得很好，甚至连诗歌都可以翻译，但这并不能说你的整个系统有翻译诗歌的能力，比如说 Shakespeare 的十四行诗或剧本《仲夏夜之梦》，如果我做很多手脚，每个句子的翻译都事先在词典中存起来，编上号码，我都修饰一番，然后我把整句输进去对号翻译，就可以翻译得很好，但实际上机器没这么大的本事。所以你要做商业炒作的话，肯定就没法在理论上研究机器翻译本身，因为商业利益驱动，老板要你很快拿下好卖钱，你没法做，你只能去做那些应付市场需要的假动作，当然用户不知道真相。所以我不是做生意的人，我也不需要那么做，除非公司来请我帮一点儿忙，但是我不愿意涉足商界，我做系统之后，我要写文章，这是我做学问的方法。我也碰到一些批评，有些人说做计算语言学的人，他们就为了写文章，写完了文章就万事大吉了。我认为写文章没有什么原则性的错误，需要这样的人，如果人人都去做商业性的东西，那么理论谁来做呢？我认为理论研究还是必须的。对于别人的批评，我从来不反驳。我甚至在课堂上说，我是做 toy's system 的 expert（专家）。我就是一个人这样的人，因为我的雄心壮志不是做了系统去卖钱，而是为了研究机器翻译的理论跟方法本身，这样我可以集中精力研究我的理论。所以我要说有的批评是对的，但是我也对有些批评不以为然，我认为作为一个研究者来说，自己的头脑要清楚。另外，我对学生的要求跟其他老师不一样，我的学生要读我们这个专业的经典著作。比如说我现在的博士生，进来以后要读国际主要的计算语言学的著作，不然怎么做博士？最近美国新出的 2000 年以后的，2001、2002 年出版的新著作要读完，我们要开研讨班要讨论，逐字逐句研究人家的内容，要下基本功，所以学生当然要苦一点儿，日子不好过。他们要读完几本英文的厚书，读了以后叫他讲出来。我不太主张学生一进来马上就跟着导师上课题，做完课题之后马上就毕业，这样的话，学生不可能读很多书，也可能有些本领，因为他做过课题，但他书读得少，当我的学生的好处是能多读些书。

张宜：您现在有几个博士生？

冯志伟教授：我有五个。

张宜：他们毕业了吗？

冯志伟教授：还没有毕业。

张宜：您是从 00 年开始招的吗？

冯志伟教授：我是从 01 年开始招的，但招得不好。考生计算语言学的最高成绩 13 分。其他科的分可能高，但我专业课的分太低啦。因为我在招生简章里说要读哪几本书，如果他们基本上不读，他们就凭他们自己原来的语言学知识来考，那肯定他们乱答一气，所以我没有给他们高分，这样我到 02 年才正式招到学生。

张宜：冯老师，您对您哪一本著作比较满意？

冯志伟教授：从社会反映比较好的就是《现代语言学流派》，好多学校用它作为教材，因为这本书比较通俗。我有一个原则，我不随便去批判人家，所以你看书里基本上没有评论，我就是说清楚来龙去脉。这本书本来不是给文科人写的，我只是在情报所搞机器翻译时，碰到各种各样的概念，Saussure, Chomsky 等，有些人不知道，他们是理科的。所以当时为了让他们知道什么是语言学，我有两个办法，一个是讲讲理论语言学，把大学中文系的语言学基础讲一讲。我讲过以后，他们可能还是不懂，因为他们并不知道理论；一个是把理论系统地讲一讲，所以你可以看出来这本书有些自然科学的手法，有一些论证，我基本上不评价，但是我要把一派说清楚。另外，我用到例句或引文的话，我要看原文，尽管我注上中译本，但我肯定查过原文，所以错误可能会少一点。当然这本书不是我的目的，因为不是我的专业，我曾经搞过语言学理论，但是后来不搞了。这本书可能影响好一些。

另外，在计算语言学方面我觉得我写的比较好的一本书是《自然语言的计算机处理》¹³。这本书是原来北京外国语大学许国璋先生在世时约我写的，他出了《现代语言学丛书》，他把这本书作为《现代语言学丛书》中的一本，这本书在写的时候，广东中山大学的王宗炎教授给我提了一些建议，他希望我在方法方面能够注重介绍，不要一概而论，所以这本书着重方法方面，加了一些基本的东西，如形态分析方法，句法分析方法等，技术性比较强。这本书我本人觉得不错，但是这本书的发行量很小，可能是印了 2 000 本以后就再也没有印，可能是因为社会需求量不大。只有懂这个的人才看它，一般搞语言学的人看不进去，很复杂，我写这本书的时候花了一些功夫。

再有，我还写过一本《数理语言学》。这本书是 1985 年出版的，83 年回国后写的，发行量还很大，大概印了 13 000 册，是上海知识出版社出版的。这本书主要讲数理语言学的基本原理，也是中国第一本数理语言学专著。这本书可能现在也买不到了，我在北京广播学院开了一门数理语言学用这本书，学生们大都 copy，陈章太先生有一本，因为出版后我送过他一本，他不用，后来人们知道他有一本就把它拿去 copy。早期的书就这些。

《现代术语学引论》¹⁴是另外一个领域，它已经不属于信息处理，因为术语学工作现在国家也在抓，国家有“全国科学技术名词审定委员会”，但是国家只注意个别术语的探讨，比如说“Severe Acute Respiratory Syndrome”到底是翻译成“非典”，还是“SARS”，有好多争议，理论上不研究，他们不注意单个术语的翻译在整个术语的概念体系中是否合理，是否和谐，他们也不注意术语知识本体（ontology）的研究。

张宜：您研究术语学跟您第二次去德国有关系吧？

冯志伟教授：对，以前我没有研究过术语学，因为到德国后的任务就是研究术语学。

张宜：那您觉得您现在从事的工作与术语学有关系吗？

冯志伟教授：有关系，因为机器翻译需要。

张宜：目前在国内除了您以外，还有研究术语学的人吗？我可不可以说您是第一人呢？

冯志伟教授：系统地研究术语学的第一人可能是我，但是国内研究术语学的人还有不少。

张宜：您觉得术语学将来会怎样发展？

¹³ 《自然语言的计算机处理》，上海教育出版社 1996 年版。

¹⁴ 《现代术语学引论》，语文出版社 1996 年版。

冯志伟教授：术语学是一门社会需要的学问，因为各个学科都牵涉到术语，特别是计算机学科。计算机词典是必不可少的，其他学科的术语辞典也是不可缺少的，新术语在不断的产生。

张宜：在现代社会在信息社会术语学发展得更快。

冯志伟教授：是啊。我曾经说过一句话——“术语是人类科学知识在语言中的结晶”¹⁵。人类每一次进步都要反映在词汇里，要用词汇来描述出来，必定要表现为术语，所以从术语上也可以看出科学的发展。

张宜：可不可以这样说，在现代意义上的术语学是应运而生的？可能古代已经有了术语，像《尔雅》，但像维斯特把术语学的理念建立起来，系统建立起来主要是基于上个世纪 20 年代科学发展到一定程度呢？

冯志伟教授：我认为术语学对我们国家来说是非常重要的。从行动上，国家现在也很重视。比如说全国科学技术名词审定委员会，我以术语学家的身份是它当中的一员，那么其他委员都是各个专业当中的领头人物吧，这个委员会是直属于科学院的，所以审定委员会的主任是科学院的院长路甬祥。在科学院的各个研究所里都需要术语，都需要协调术语，所以说科学院很重视术语，专门有这么一个委员会隶属于科学院。另外在标准化部门，在国家技术监督局下面还有专门的全国术语标准化委员会，专门搞工业方面的术语。工业术语跟自然科学的术语不一样。比如说机械上的零件、部件、飞机上某个部件的叫法，这些都是很有实用价值的。最简单的例子比如说做买卖，你要买双稳态触发器，这是计算机硬件上的一个零件，英文叫做 flip-flop，你到美国叫做 flip-flop，如果你到其他国家（比如，英国）去呢，flip-flop 可能就是单稳态触发器了。在这种情况下，如果你弄错了，你不知道在不同的英语国家有“单稳”和“双稳”的说法，那么你就会买错的。如果让你买 10 万个 flip-flop，结果你买错了，你好意思退货嘛？你买了之后退货也不好退呀，你写的就是这样，退货也不好退，但买了之后对你有毫无用处。所以说，术语的用处显然是很大的，语言学对社会生活的重大影响这可以说是个例子，弄不好会损失很多钱。对于工业部门来说术语很重要，它必须要把每个东西在各国语言中的叫法弄清楚，起码在做买卖方面，在科技交流方面是必不可少的，而且还要规范化，不能仅仅是工业部门和自然科学部门非常重视术语。可是这个问题语言学家不太关心，语言学家认为这是理工科的事，我们要研究人们日常生活中常用的语言，我们不研究术语。

张宜：那么可不可以从这个角度上来说，您作为语言学家很希望在语言学领域里有人重视并且研究术语学？

冯志伟教授：我认为是这样的，而且我认为术语学是应用语言学的一个分支，它应属于语言规划这个部分，而且这是真正可以规划的语言。日常生活语言很难规划。现在用字母词，比如说，我要去照 CT，有些语言纯洁主义者很反对，说我们中国话里怎么能有 C 跟 T，中国话里没有，你必须说成“计算机断层扫描法”，但是谁说啊。只有那些保守主义者才这样说，这些问题属于社会上的语言规范问题，那么在术语学里，在真正的科学技术领域里，术语是可以规范的，但在社会生活中是很难规范的。MTV 也是这样，你若让年轻人说“音乐电视”他们可能不听，他们偏说 MTV。社会上的语言规范要靠约定俗成来做。政府提倡什么，老百姓若不讲，你也没有办法。但术语学在自然科学和工程技术领域里可以通过政府法令规定实现，特别是新生术语刚刚产生，可以规范使用。所以我想术语是语言规范化里面其实可以真正做到的一个部门，就是可以通过政府部门或者是权威的力量可以在其中发挥作用的。什么人决定的，某个专家说的，我们就听他的。可惜我们搞语言规范时恰恰不考虑这些，我们现在没有重视术语学。

张宜：冯老师，我还想问您，您认为发挥汉语拼音在信息时代的作用这个具体表现是不是指您说的一句话叫“进一步扩大汉语拼音方案在语言信息处理当中特别是在网络上的使用范

¹⁵ 见《现代术语学引论》（冯志伟著）前言，语文出版社 1997 年版。

围”¹⁶。“扩大”我能理解，您认为怎样推行才比较合适呢？现在有什么实验吗？

冯志伟教授：原来中国文字改革时，毛泽东主席主张搞拼音化，走世界文字拼音化的方向。但拼音的实行后来在中国有些难度，因为从毛泽东说的话而做不到就可以看出难度。毛泽东说的话一般都能做得到。实际上就是因为我们有较强的文化背景，有几千年的文化背景，汉字文化的背景。汉字在日常生活流通当中使用量那么大，我们不能视而不见。我们只能做些小的改革。搞拼音文字，中国人的心理状态也接受不了大的改革。从国家来看，全国语言文字工作会议 1985 年以后拼音化方向就不作为国家语言文字工作的主要内容，就不再提了。现在的提法是“扩大并推广汉语拼音的使用范围，完善汉语拼音方案”。不再提拼音化了。这个问题我想也是考虑到我们做不到的事情我们不要强迫去做，而且我们老百姓也不太喜欢去做。

张宜：那也就是说在很长的一段时间里面拼音是不能取代汉字的吧？

冯志伟教授：是这样的。最多，最理想的是双文制的局面，但是这也要经过很长时间。

张宜：这与您说的计算机上实行“two-script system”（双文制）有关系吗？

冯志伟教授：有关系。这是个现实存在的问题。有些人对拼音有厌恶的情绪，再加上汉字优越论，实际上“汉字优越论”是错误的，文字没有优劣之分。

张宜：那说文字的表音表意表形，从表音到表形表意，这能否决定一种文字孰好孰坏呢？

冯志伟教授：我觉得很难说文字优劣。只要一种文字能被群众接受、存在以后，存在的就是合理的。如果它不存在就被淘汰，这就表明大家不喜欢用它，所以说语言的优劣跟文字的优劣的说法我都不太赞成。比如说我也能说好几种语言，但我很难说哪种是最好的。（我用它们）都可以表达我的意思，所以也不能说文字的优劣。我希望政府能采取比较稳妥的态度对待汉语拼音，不再提拼音化，这是符合现实的态度。但是现在要不要用拼音呢？我觉得还是要用。为什么呢？就是因为拼音终究是一种全国人民代表大会通过的具有法定性质的，而且国际标准化组织也推行的。比如说人名、地名必须用汉语拼音，所以它是被认可的一种汉字辅助工具，所以要发挥它的作用。而且拼音正词法，即分词连写的问题，必须要强调分词连写的问题，其实在拼音方案里已经有分词连写的因素。比如说我们要用隔音符号标明 pi' ao（皮袄），或者 xi' an（西安），我们用它来拼词而不是拼字，如果拼字就不存在隔音的问题，就说明汉语拼音方案在制定时已经考虑拼词了。既然要拼词，那么词是我们概念的基础，概念不是以音节或以字为基础的，所以这样一来智能信息处理是以词为基础的。在这种情况下，汉语拼音与词连写是时代的需要。所谓进一步的完善汉语拼音方案，我的意思是我们要把汉语拼音正词法完善一下，汉语拼音正词法是语委做的，后来成了国家标准了，技术监督局正式采用它，用 GB 号表示，即国标号。汉语拼音正词法是个软的正词方法，因为它规定一些 principles（原则），具体实施时有些东西就弄不清楚。比如说形容词和名词合在一起时，一般说要分写，如蓝一天，白一云。有人说不能分写。“青天”不能分，因为“青天”指的是包公，所以说这些原则在具体实施时就出现各种争论。这时非常急需的是一部具有权威性质的汉语拼音正词法词典，应该包括十几万或二十几万词，每一个词分不分写都写清楚，然后再根据它来做。其实这词典也好做，比如说通过教育部公布要各地中小学在教学中要使用这部词典。有一部这样的词典就好了。我所谓的“进一步完善”就是指这些。除了汉语拼音正词法之外，我们起码还要有一部词典，词典加正词法，解决老百姓在分词连写时的困难，那么逐渐全国按照这个来做。现在好多人反对连写，说分词正词法连写没法操作，于是干脆就不做，就一个一个字地拼，或者全部连在一块儿写，就像商店的名称，写在一块儿很难读，中国工商银行（zhongguogongshangyinhang），一大串儿，有些又一个一个字地写，这些都不符合拼音正词法。如果政府坚持，除了汉字以外，还有完善的拼音系统就好了。

张宜：如果电脑的拼音软件也能按照正词法来做就好。

¹⁶ 冯志伟《语言文字规范化对于语言信息处理的作用》（修订稿），《中国语文》1997年第5期，总250期。

冯志伟教授：是啊。要强调正词法，要有这方面的词典。我知道微软他们在做正词法拼音软件，哈工大有的人也在做。“微软拼音第二版”就是2002年出的这一版，跟过去的不一样，原来的微软拼音是比较差的，现在好多了，挑选的余地已很小了。它的智能性很好，它能自动判断，你打拼音它就能出来比较正确的汉字。

张宜：冯老师您现在用什么方法在电脑上打字？

冯志伟教授：早期我用过五笔字型。这几年我发现对我写文章有妨碍，我一边写东西一边还要想用什么码来输出汉字，影响我的思维，后来我就用“智能ABC”了。“智能ABC”是我们研究室和北大合作的，（这个系统）也有好多问题，我一直在用这个。

张宜：您如果向我推荐的话，您会推荐什么呢？

冯志伟教授：我自己虽然用“智能ABC”，但如果你要用好的拼音输入法的话，我推荐你用“微软拼音第二版”，“智能ABC”比较老了。

张宜：“狂拼”怎么样？

冯志伟教授：噢，是“智能狂拼”，是中文之星做的，也可以。但是就是说真正的“狂”不到什么程度，也有很多错；我还是认为“微软第二版”比较好，它可以根据上下文来推断，所以它有比较强的智能。我们的“智能ABC”有些智能，但还不行。如果政府下决心做一个可以超过微软的输入法，一打进拼音就能出来非常准确的汉字，做这么个软件这是最理想的，这一点可以做到的。日本就是个例子。日本在80年代就做到了。80年代以后日本全部是から汉字转换，から就是假名，假名是罗马假名，就是用普通键盘打拼音，电脑自动把它转成正确的汉字。罗马字母跟假名有一一对应的关系，假名跟汉字不是一一对应的。比如说

きしゃ，有“记者”、“火车”、“归社（回公司）”这三个意思，可以根据上下文辨别用哪个

意思。比如说“きしゃは きしゃで きしゃした”这个句子。第一个きしゃ是“记者”，因为

は前是主语，で前应是工具，所以第二个きしゃ是“火车”，した前是动词，所以，第三个

きしゃ是“回公司”。句子的意思是：“记者乘火车回公司”。日本就这样子解决了汉字的输入问题，日本人花了功夫，我们也可以研究利用中文的上下文进行识别，做出这样的软件。

据周有光教授说，日本80年代做出这样的软件后，就不搞编码了，到现在日本人已经用了10多年了。我国情况就很乱，因为始终没有一个定的东西，政府也不敢表态说这样的东西好不好，都说大家竞争，于是就出现了“万码奔腾”，这就比较麻烦。如果政府下决心集中力量做一个拼音输入法就好了。今年是汉语拼音方案公布45周年，国家准备在人民大会堂开会庆祝一下，政府的意思是不要忘记拼音的好处。

张宜：那么您是否想借这个机会把您刚才的想法再陈述一下呢？

冯志伟教授：我开过一次座谈会，我说过在讲汉语拼音如何好的同时，不要忘了汉语拼音正词法是汉语拼音的不可分割的组成部分，因为它也是必要的，所以在推行汉语拼音的同时也不要忘了推行汉语拼音正词法。另外我也提出过要编一部《汉语拼音正词法词典》。因为拼音正词法太抽象了，不好操作。如果找一些专家认真地编这部词典，把它说清楚就好了，这是我的一个想法。另外一个想法是我主张在信息传输当中可以采用汉语拼音。

张宜：由此我想问您另外一个问题，我记得您在一篇文章里写过：“中文信息处理中如果采用汉语拼音，其效率将可以同英文的信息处理系统相媲美”¹⁷，这是否可以说在信息处理中

¹⁷ 见冯志伟《语言文字规范化对于语言信息处理的作用》（修订稿），《中国语文》1997年第5期，总250期。

汉语拼音优于汉字？

冯志伟教授：我认为是。因为这是显然的。比如说尽管现在有很好的汉字输入法，但在收 e-mail 的时候，有时候就收到乱码，是不可避免的。因为只要错一个位，就可能是乱码，而输错汉字又是不可避免的。拼音是 ASCII 码，American Standard Code for Information Interchange（美国标准交换码），所以就不会出乱码。计算机不是根据汉字做的，而是根据 ASCII 码做的，所以原码是 ASCII 码就绝对不会出问题，就不会出错，提高效率。另外拼音的熵值比较低，只有四点几个比特，和英文的比特数比较接近。把汉字转成拼音，让拼音在机器上跑，在网上跑，最后再变成汉字。如果用现在的这个拼音，这不能完全用机器来做。比如说我有一个汉字文本，把它转成拼音，但拼音出了问题，拼音有同音字，这就不好办了。但如果用定型化的拼音码，作为一种码就不需要人的记忆，因为人记不住，用机器来记，靠一部词典就可实现用机器转换。机器上的码转换成一一对应的拼音，一一对应的拼音再转成汉字，也是 automatically（自动地）。我现在所谓的在机器上跑，在网上跑的拼音不是现在一般的拼音，而是拼音定型化以后的拼音，是区分了拼音同音字以后的拼音，是和汉字一一对应的拼音。如果有这么一套码，我把它叫做“单码”，单一的码。如果有单码的拼音，汉字转换了它，然后机器上全部跑它，到了跟人接触的时候，就全部转换成汉字。我们中国人喜欢汉字，也习惯汉字，中国人不愿跟拼音打交道，在计算机网络上才用拼音。我把这个叫做在信息网络时代在信息处理领域里的“双文制”，即 digraphia。因为双文制的思想很多人都提过，主张“双文制”的人早期有钱玄同，后期有茅盾，郭沫若。他们都有很激烈的言论，认为中国人应该是汉语拼音跟汉字共用。茅盾曾写过：“如果有一天我们小孩子既能用拼音写字，也能用汉字写字。外国人来了以后，我们让他们先学拼音文字，再学汉字该多么好”。他们是文学家，他们从感性出发，觉得这个很好，但怎么去实现，他们并没有说。这是他们的一种理想，我自己作为旁观者，我认为实现不了。他们完全是一种理想主义者，我觉得中国老百姓是不会用“双文制”的，因为中国人毕竟使用汉字。汉字虽然难学，但我们还是学会了，要改成“双文制”是不可能的。但是为了赶上信息时代的要求，提高工作效率，避免出错，我们在网上跑一套定型化的拼音文字，这套拼音跟人没有关系，完全由机器来做，这是可以做到的，这是我的思想。当然这些事情都要有人来干，而且政府也认可。你干了，它说你对了。假如干了一气，政府不认可，不实行那就白干了。因为工作量很大，必须要有政府的支持。有政府支持，别人才敢做。目前这个想法没有人响应，我是孤掌难鸣。在教育部召开的座谈会上我也说过这个问题，但没有人有响应，估计暂时不可能，这需要一个很长的过程才能得到人们的认识。汉字管理效率低。这个是一般人想象不到的。如果你在银行部门工作，你会深刻感到汉字存在的问题。汉字造表，各种统计速度非常慢，容易出错，所以很多外国人认为在信息时代汉字存在的问题很大，但我们中国人有爱国主义情绪，认为汉字好，你说有问题人们就不高兴，所以不敢公开说汉字不好。这事如果处理不好对国家安全会有影响。比如说在打仗的时候，如果信息传递出现错误，出现偏差，就要出问题。但现在这个问题还没有提到日程上来，人们没有认识到这个问题的重要性，他们没有做实际工作的经验，他们不知道在汉字使用当中出现的各种困难和错误，他们看到能从电脑里打出汉字就觉得已经很不错了。我所说的只能是建议而已，人们不妨去做一做。我不主张像郭沫若、茅盾那样在全民当中，包括小孩中推行“双文制”，没有这个必要，拼音作为一种辅助工具使用就可以了，但是在网络上跑我认为是要跑拼音。在人机界面上跟人有关的用汉字，这是我的一种主张，我把它叫做信息网络时代的“双文制”，这件事很难做到，它牵涉到语言文字改革的问题，所以只是一种建议而已，要做到恐怕还要相当长的过程。

张宜：冯老师，您说的机器翻译处理的三个过程：词汇阶段，句法转换阶段，还有一个是语义阶段，您说语义不便于形式表示，所以您说您只是做了一些小规模实验，这个大概是在 80 年代吧？那么现在怎么样了？

冯志伟教授：现在是这样的。到语义分析阶段是比较难，因为我们计算机的语义表示从我们现在人类的知识水平来看，主要是把句子表示成一阶谓词逻辑的表达式，即 **first order predicative logic**，简称 **FOPC**。一般是我们通过计算机分析以后加上语义信息，就可以做出任何一个句子的一阶谓词逻辑的 **FOPC** 表达式，知道这个句子的大意是什么。现在能做到的是对一个事件的表示，事情是什么事情，参与者是什么等，能做到一定程度。另外对于时间的表示，比如 **tense**（时态），过去、现在、将来等，用一阶谓词逻辑的表达式的一些研究成果可以表示出来，如果一个行为发生以后，那么这个行为所处的时间就很重要，这个在数学上是可以描述的。第三方面的问题就比较大，就是人的 **belief**，信念，比如说“我喜欢”，“我认为这件事是要失败的”，这件事不一定要失败是“我”认为的，这不是现实的东西，带有人的主观的东西，是一种 **belief**，这方面的数学研究做得比较差。在这方面的逻辑叫模态逻辑，它的研究成果不足以表达语言当中跟例句有关的语言现象，比较麻烦，所以现在逻辑学方面的研究要来真正的表达人类的自然语言还不到家。美国人从另外一个角度，不用一阶谓词逻辑，而用内涵逻辑，就是把一个句子分析完以后，比如 **Chomsky** 的短语结构分析是用树来表示的，把树通过内涵逻辑的方法把它变成一个内涵逻辑分析的表达式。这个方法美国有个叫蒙太格（**Montague**）的人做过，他只能做一个 **fragment**（很少的部分），大概有十几种语言现象。比如说一个长的名词词组能处理，量词，说明量的，**the, a, all, only** 这类表示覆盖范围的东西大概能处理，但只是一个 **fragment**。所以我认为目前在语义方面对语义形式化研究做得比较差，因为理论基础就那么高，实际上不可能做得很好。我估计这二十年来在语义研究方面会有些进展，就是说本身的理论基础就能看到这么多，一阶谓词逻辑，蒙太格语法。在数学上大概是这么多东西，这些东西目前的水平就是这样的，那么要有重大突破除非有新的理论产生，从新的角度来描述，所以从目前来说，机器翻译做到现在这个样子，即完全从规则出发，已经做到头了，我们只能做到 **60%到 70%** 的准确率，再往前几乎是步履维艰，非常困难。

张宜：那您认为要在哪个难点上有所突破呢？

冯志伟教授：现在美国人的一些成果很引人瞩目。他们几乎完全摆脱了人工编写语言规则的传统方法，他们通过统计方法来做，他们把语料库做成“树库”，对语料进行标注，每个句子都是一个树，然后从树里面获取各种各样的知识，叫 **tree bank**，树库。因为树库很大，有数十万词或一百万词。所以这些统计知识基本上能覆盖自然语言 **90%** 以上的内容，但这是一种数学表示方法，而不是语言学家能够懂得的方法，它们是一些数学的表和数学的统计数字。数学家用这些统计数字再去分析新的语言数据，这样效率会很高。**IBM** 公司做过五个模型，它们一、二、三个模型处理了单词的对应、单词位置的变换以及多义词等问题，其效率已经超过基于规则的 **SYSTRAN** 系统。

张宜：那要是把这部分和语言学的研究法合起来行不行？

冯志伟教授：可能会行。这种统计规则一般只能做相邻的，邻近的关系，比如说 **this book, a pretty girl** 等简单的可以，但是太长的句子，如 **the man who would like to go to Shanghai said...** 就不好分析 **man** 和 **said** 之间的关系了。怎么统计，怎么反映 **man** 和 **said** 之间有关系就很难。规则就是要解决这种问题，远距离的关系问题。这样可能会造成统计和规则相结合的分析系统，可能会有比较好的前景。去年美国人有一个机器翻译的比赛，看机器翻译的效率，他们有严格的评分标准，不用专家来评，而用数学系统来评，用传统的数学方法来做系统。德国 **Och** 是 **Aachen** 大学的一个年轻博士。他用统计方法在很短的时间内做了阿拉伯语和汉语到英语的若干个机器翻译系统。在比赛上他的了第一名，这是在 **2002** 年。他用的是统计的方法，叫做“最大熵模型”，这给做统计的人以很大的鼓舞，而且他做得很快。他说他可以在几小时之内做一个机器翻译系统，所以这种方法是很吸引人的，也许会有前途，但是那时候参与机器翻译的可能就不是语言学家，而是数学家了，因为很少用到语言知识了。

张宜: 您觉得这是将来的一个发展趋势吗? 就是数学家取代语言学家的的工作去做机器翻译?

冯志伟教授: 嗯, 有可能。但是有一个最大的问题就是对数学公式怎样解释。当然可能会出现一些懂得数学原理的语言学家, 这些语言学家对这些数学的原理进行解释, 他们能解释这个现象是语言当中的某种现象, 体现在数学公式中。然后让他们的统计方法具有人文性, 现在是完全没有人文性, 完全是算出来的, 不懂数学原理的语言学家会感到莫名其妙。我们设计者对里面的过程根本不清楚, 我们只是等待结果, 拿着结果去做机器翻译, 中间过程根本不参与。我们研究者所要做的只是把树库做好, 确保没有差错。另外这个方法与过去的方法完全不一样, 是新的方法, 这也是计算语言学中的一些比较鼓舞人心的事情。当然, 在基于统计的机器翻译系统的研制中, 也不能完全排斥语言学的知识, 目前已经出现了一些使用句法知识的基于统计的机器翻译系统了。只要机器翻译系统涉及到语言的句法分析, 语言学知识就是有用的了, 语言学家也就仍然可以发挥作用了。统计方法科学性很强, 但是语言是人讲的话, 具有人文性, 我一直主张把科学性与人文性结合起来, 探索有效的途径, 把统计方法和语言知识结合起来。最近美国成立了 Language Weaver 公司, 专门从事统计机器翻译系统的商品化, 他们就 very 注意统计方法和语言知识的结合。这个公司请 Och 做顾问, 开发了一个“学习软件”(Learner), 对双语并行语料库进行处理, 从语料库中抽取概率翻译词典、概率翻译模板以及概率翻译规则等语言信息, 这些抽取出来的语言信息, 统称为“翻译参数”, 这样的翻译参数实际上就是概率化的语言知识, 经过上述的处理, 语言数据就变成了概率化的语言知识。为了处理这些翻译参数, 这个公司还开发了一个统计翻译器, 叫做解码器(Decoder)。翻译参数和解码器成为了 Language Weaver 公司翻译软件的核心。解码器使用通过统计学习获得的翻译参数对新的文本进行机器翻译, 把新的源语言文本自动地翻译成新的目标语言译文, 提供给用户使用。他们非常明确地致力于把统计方法和语言知识结合起来, 当然, 他们获取的语言知识, 与基于规则方法的语言知识不同, 是概率化的语言知识。我们知道, Weaver 这个词在英文中的涵义就是“织布工人”, 从这个名称可以看出, Language Weaver 公司试图像织布工人那样一针一线地对机器翻译进行精雕细刻的研究, 这种精雕细刻的研究态度, 是值得称道的。

张宜: 冯老师, 如果将来机器翻译发展到如此高的阶段, 它能完全代替人工翻译吗?

冯志伟教授: 如果发展得很好的话, 有这种可能。但是你知道, 虽然准确率可能达到人工翻译的水平, 但主要是科技文章。一般文章, 特别是文学作品还不行, 因为这里面还有比喻, 容易出错。因为英文的比喻和中文比喻不一样, 这是一个问题。另外还有感情色彩, 机器再算也算不出爱恨、高兴、愤怒……。所以估计人的翻译永远不可缺少。翻译是一种高智能的活动, 培养一个外语系的本科生我们需要 4—5 年的时间, 而且只能是初步掌握翻译技能。所以要做到感情的计算很困难, 现在, 在计算机里已经开始考虑计算机情感问题。最近就要开一个计算机情感的国际会议, “情感计算”(affective computing) 就是要看一篇文章的情感是什么。这个对于中国来说很有用, 如果情感研究得好就可以用计算来过滤有关法轮功的错误信息。法轮功写文章攻击我们, 那么我们就可以过滤“李洪志”等词语。但我国批判法轮功的文章里也可能有“李洪志”的字样, 也可能滤掉, 所以计算机要总体上去权衡这篇文章的爱和恨, 是爱中国的还是恨中国的, 那么计算机是算不出这个的, 这时就需要情感计算了。现在我已经看到一些材料, 这方面已经有一些初步的理论。在美国有人在研究情感计算, 初步的研究主要在语音的合成方面, 一个 text (文本) 出现, 然后机器读出 text, 读出音来, 但是没有情感, 很呆板, 即使有抑扬顿挫, 也是机械式的抑扬顿挫, 体现不出感情, 要分析一段话的高兴程度, 英美已经在做这方面的研究, 叫做语音分析中的情感计算。

我们扩大范围的话。可以看到词汇里也有情感问题, 句法里也有, 篇章里也有, 比如篇章的总的立场问题。作为情感判断, 研究情感计算是很有用的, 它也有基本的计算方法。但是机器翻译要达到人工翻译的水平是永远不可能的, 计算技术的专家已经考虑过这些问题。

大约在三年以前，自然科学基金里有一个重点课题是自动化所申请的。他们申请情感计算，当时我也在座。我们都否定，因为我对情感研究方面没有任何的知识，我说太荒谬了，情感还能计算，这个题目本身就是错的，所以一分钱也没有给。今年他们又申请了，并派了一个口才比较好的同志去说服评审专家，这回同意了，并给了这个课题较高的资助。我们认为这个课题确实值得研究，为什么一开始被否定，因为我们对这一块不了解，今年还要开国际会议，请一些外国专家来。

张宜：冯老师，您对国内的一些翻译软件作何评价？

冯志伟教授：我们国内在机器翻译软件的译文质量上在国际上并不差，比较好的是“华建”。华建翻译质量还是比较高的，双向翻译：汉译英，英译汉，这个系统做得比较好，商业经营也做得比较好，所以整个公司处于比较景气的状态，既赚钱，又增值资本，是国内比较大的集团，资本也相当可观，在语言工程里，他们算是很好的了。一般语言工程大概有几亿资本的都很少，华建在中国也算是撑得起来的一个集团吧。我认为比较好的就这一家。它有个特点，基本上是搞句法分析和语义分析，每天都在翻译大量的资料。他们的研究所像个工厂似的，每天都要上班，年轻人都在那儿调试句子，翻译完了还要改。长年累月的这么做，这样就可以改善质量，所以华建的水平，我认为目前是比较高的。其他公司做了一些，但是没有它这么好的开发环境。华建他们等于有一个机器翻译的“工厂”了，有“车间”，每天有几十个“工人”上班。这些“工人”实际上都是语言工程师。而其他公司今天搞搞，明天可能就不搞了，这样子就是再聪明也搞不到什么程度。要有连续性，要持之以恒地搞开发，每天都要调试，随时发现错误，及时改正错误，这样才会好。所以我说像华建这样精良的机器翻译公司是目前世界仅此一家的公司。美国最好的机器翻译公司 SYSTRAN 可能也比不上它。哪有每天都来上班调试系统的机器翻译公司？很多公司都是些专家有空儿了就一来。所以我说华建会很有前途。华建也算是中科院的一个集团吧，它是独立的资本，但行政上跟中科院有关系，中科院给他们投资。其他的机器翻译软件我觉得一般，水平也差别不大。说穿了，我觉得只有 50%~60% 的准确率，不过他们商业上的炒作说自己可达到 80% 以上的准确率。他不这样说卖不掉。他说 50%，谁买他的。我认为机器翻译是整个语言产业水平的表现，因为机器翻译什么都要搞：词汇分析，句法分析，语义分析，外语生成，什么都要搞。机器翻译系统做得好，就代表着国家在这方面的水平。目前中国的翻译软件和计算语言学在实用方面在国际上不算差，但我们理论上比较差，因为中国人没有成套的理论被外国人引用，得到公认。比如 Chomsky 的理论，我们就在用。我们还没有这样的人出现，这是我们差的地方。但从实用角度看，我们的语言信息处理还是不错的，国家做过中国科技水平评测，大概说中国科技水平在实验上基于中等水平，但中文信息处理处于领先水平。这个估计基本上符合实际情况，这个水平外国也未必赶得上，好在计算语言学现在已被公认是语言学学科的一个分支了，这点很好，也招收博士生了。我年轻的时候不敢想，那时只希望自己的研究被认可。目前计算语言学在语言学界和计算机界两方面都得到承认。如在计算机界，计算语言学被认为是人工智能的一个分支；在语言学界则被认为是应用语言学的一个分支。所以计算语言学已经算是一个正式的、有模有样的 discipline（学科）了，变化很大。从 57 年到现在，经过 40 多年的变化，到现在观念已经完全不一样了，还有些年轻人立志要从事计算语言学研究，计算机专业和语言学专业都有这样的学生，已经不错，这比我年轻时所处的时代强多了。

张宜：冯老师，您对我做口述史怎么看？

冯志伟教授：过去我只听说过在别的学科里有这种做法，在语言学里还从未有过。比如我写《西方语言学流派》也是一种写人的语言学史的研究方法吧，因为学派集中在人的身上，然后通过语言学家的介绍，展示语言学的发展。像岑麒祥是作书面语言学史的，他要详细地占有材料，然后逐一加以分析，这样可能会有些第二手，第三手的转写性的材料，就会出现

一些差异，有些事情很难核实。

张宜：有些人说我这样做口述史，只听一人之说会带有很强的主观性。

冯志伟教授：问题在于你本身的分析能力，一个是你选择的专家要广一些，你要做哪几个学科，一定要有代表性，也就是你的抽样要正确，不要抽得太偏僻，那样就会有歪曲和遗漏，专家在口述当中谈的是个人观点，所以不免会有些个人看法，这在于你怎么去看待，你是认同他还是不认同他，那么你经过很多调查，你就可以比较出来，到底谁的比较符合事实，谁的有些问题。当然我想历史最重要还是客观，就是要真实反映情况，但专家在叙述时都带有主观色彩。比如我经历过的事情，我都会对它们有一些看法的，都会有一些感情色彩，不一定是历史本身，所以关键在于治史者本身如何去分析去把握。我觉得从收集资料来看，还是要收集第一手资料，比如索绪尔（Saussure）的《普通语言学教程》现在有好几个版本，包括当时听他课的学生的笔记本，他当时说的什么话等等。这些史料性的工作，也基本上带有口述的性质。具体什么课，谁在讲，谁在场，他的话是什么。有不同的笔记本，每个之间会有差别。主要是根据薛施蔼（A. Secheyne，索绪尔的学生）的笔记本整理出来的，后来有人又找了一些新的材料说明当时索绪尔没有这样说过，由此引发一些辩论。我们对索绪尔的理论怎么看与这些不无关系，所以要是第一手的材料当然是好的。我想这个做法不妨采用一下，而且现在条件比较好，我们有比较好的录音设备，我想我原则上还是同意这种做法，其实口述的东西也可以变成书面形式。

张宜：您的录音我要整理成文稿，整理好了之后，要拿给您过目，确认一下。今天您讲的哪些是可以公开的，哪些是要修改的，我们再商量，将来的东西应该是我们两个人的共同成就。谢谢冯老师接受我的访谈。