目　录

第二届全国语言文字应用学术研讨会开幕词…… 许嘉璐（1）

关于大规模真实文本语料库的几点理论思考…… 张　普（10）
汉语语素库的构造及其同语法信息词典的集成
…… 朱学锋　俞士汶　李　峰（31）
浅层句法分析方法概述 …… 孙宏林　俞士汶（41）
中文信息处理与现代汉语词类研究 …… 李　竹（56）
汉语词性自动标注软件兼类词鉴别规则库的
设计 …… 温锁林（73）
判断从属树合格性的五个条件 …… 冯志伟（82）
汉英机译系统译文质量的评测 …… 侯　敏　侯　敔（94）
汉语词汇语义系统结构
——语义图方法的初步探索…… 刘殿义（107）

中国语言文字使用情况调查准备工作中的
若干问题…… 苏金智（117）
语言调查中的语言态度问题…… 王远新（126）
生存还是消亡：汉语方言面临的抉择 …… 曹志耘（139）

新加坡华人的语言态度及其对语言能力和
语言使用的影响…… 陈松岑(149)
把广播电视主持人语言放到关系、系统和
社会政治变迁过程中考察…… 林兴仁(162)
简论社会用语的特征…… 王建华 袁国霏(172)
时尚词语探索…… 何伟渔(181)
语文词典收词中的观念更新问题…… 周洪波(191)
释"姓"与"氏"之文字构造与文化蕴涵…… 向光忠(195)
山东方言与儒教…… 董绍克(206)
中国满语文研究现状及发展…… 赵阿平(212)
贵州民族语底层地名命名依据和变易成因的
考察…… 金 美(222)

语言功能和可能规范…… 戴昭铭(235)
现代汉语规范化的规则本位和语用本位…… 施春宏(250)
语体与语言规范化…… 李熙宗 霍四通(265)
语感与语言规范…… 王培光(275)
由社区词谈现代汉语词汇的规范…… 田小琳(284)
港澳新词语的思考…… 邓景滨(294)
对于某些字音与数字用法的思考…… 李先耕(301)
对新外来词语的使用、规范等问题的观察与思考 … 杨 华(313)
动宾式类推及其规范…… 吴锡根(325)
"做"与"作"的使用与规范…… 马 彪(335)
语文应用的基础是规范与整理…… 安红岩 胡新化(343)
多重复句的分析和分号的使用…… 王大新(350)

对外汉语教学语法问题研究的基本态势 …… 张旺熹　崔永华(357)
对外汉语教学高级阶段精读课教学的一种尝试…… 蒋可心(367)
王力《古代汉语》亟须重新修订…… 富金壁(376)
“注音识字，提前读写”教改实验的特点与优势
——兼谈语文教学的两种思想…… 詹恒乙(388)
评《现代汉语词典》(修订本)的几项不足…… 曾子凡(397)

“实证性假设”与“启发性假说”
——语言相对论的双重读解…… 高一虹(413)
把语言放到信息范畴中来认识
——关于语言的哲学思考要点…… 奚博先(428)
现代汉语标题语言句法研究的价值与方法…… 尹世超(437)
方言学必须加强应用研究…… 李如龙(447)
20 世纪的中国应用语言学研究 …… 于根元(455)

第二届全国语言文字应用学术研讨会闭幕词…… 陈章太(463)
后　　记…… 佟乐泉(468)

附录：首届全国语言文字应用学术研讨会纪要
…… 《语言文字应用》特约记者(471)

第二届全国语言文字应用学术研讨会开幕词

许嘉璐

1995 年，我们在北京香山举行过第一届应用语言学研讨会。现在过去了将近三年，再次相会在北国美丽的城市哈尔滨。

回顾三年来，我国的应用语言学研究，正在不断取得新的进展。在语言教学方面，无论第二语言教学还是第一语言教学，无论是对外汉语教学还是民族地区的双语教学，都有新的收获。除发表了许多学术论文、出版了一批专著外，学术活动也很活跃。1997 年，语言文字应用研究所和中央教科所在佳木斯举行了“‘注音识字，提前读写’全国学术研讨会”，对外汉语教学界也举行了研讨会。在中文信息处理领域，国家“九五”重大科研项目“面向计算机的现代汉语词汇研究”已经启动，并且进展顺利；国家语委的科研项目“概念层次网络”理论的实用化工作已接近尾声；北京大学计算语言研究所的“现代汉语语法电子词典”已经完成；教育部全国高校古籍整理工作委员会的“古籍整理计算机辅助通用系统”工程已经启动。此外，在广播电视语言、法律语言研究等方面也都有可喜的成绩。特别应该看到的是，全国各行各业语言文字规范意识已经大大加强，语言文字学界进行应用研究的欲望也比过去强烈了。这次的研讨会，应该说是近三年语言文字应用研究成绩的一次具体反映。

下面我想就我国应用语言学的进一步发展谈三点想法，供同志们参考。

一　关于理论建设

我们人人都知道理论建设对于具体研究工作的重要性。近年来，语言学界也在应用语言学理论探讨方面有所建树，但是，至今我们还缺乏系统的、基于我国语言文字实际的理论研究成果。如果说，在几年前提出这个问题条件并不成熟的话，那么，现在如果再不把这个问题提到比较紧迫的日程上，就要影响到应用语言学今后的发展了。

为什么这样说呢？一方面，语言文字学在各个领域的应用，或者叫应用语言学的具体研究，近年来已经有了明显的进步，理论建设既有了一个比较扎实的研究实践的基础，也有了对理论指导的更迫切的要求；另一方面，我国应用语言学的范围也已基本确定，也可以说，具有中国特色的应用语言学已经出现，应该对它进行理论的总结。

大家都知道，应用语言学最初在西方出现的时候，仅指语言教学，而且指第二语言教学。经过了十几年，应用语言学在中国生根了，同时，内容也起了变化。现在语言文字学界已经基本取得共识，应用语言学，是以语言教学和中文信息处理为主干，包括了语言计划和语言文字在各个社会生活领域的应用研究的一门交叉性、综合性学科。这样一种理解是符合我国的语言文字实际和社会发展实际的。例如，中文信息处理，似乎应该归在计算语言学里，但是，汉语汉字的特点，就决定了它不像屈折语及其拼音文字那样，可以直接地或比较容易地形式化，然后变为算法；要实现用计算机对汉语汉字进行处理，就必须解决一系列汉语和汉字的一些特殊规律问题。因此，把中文信息处理归到语言文字应用的领域中是合适的。在应用语言学的范围等基本问题上大致取

得共识，就为其理论建设提供了必要的条件。

应用语言学的理论建设具有很重要的意义。首先，应用语言学已有的实践需要总结，今后的实践需要理论的指导。例如语言计划在当前以经济建设为中心、从计划经济向市场经济转化过程中，应该怎样制定和实施，就需要理论的支撑；又如，中文信息处理所需要的现代汉语研究成果和传统研究之间是怎样的关系，也需要理论的探讨。其次，理论建设涉及到给应用语言学定位的问题。也就是说，在还有不少人轻视语言文字应用研究的今天，要使应用语言学确立自己的学术地位，既要靠为现实及时提供有价值的研究成果，也需要从理论上加以阐述，以自己的理论体系赢得全社会的承认。

现在已经有一些同道对应用语言学理论进行了探讨，但是，过于分散，偏重微观。语言文字应用研究所在这方面做得多一些，于根元同志从史的角度有过论著，在他的即将出版的《语言哲学对话》中也有很多篇幅论及应用语言学。但是，系统地全面地进行理论研究，至今仍很缺乏。我建议，应用语言学会应该成为应用语言学理论建设的倡导者和组织者。请考虑是否可以在适当的时候举行关于理论研究的专题会议。

总之，从现在起，我们应该对理论建设问题重视起来，并且切实地做起来。

二　应用语言学进入大学课堂问题

大家都知道，我们的大学教学内容，至今还没有比较彻底地摆脱计划经济时期的影响，突出表现之一就是学科之间壁垒森严，重基础轻应用，基础教学多年一贯制。这不仅影响到从学校出来的人的知识结构，也深深地影响了他们的学术取向。就语言文字学这个狭小的领域说，要使应用性研究得到普遍的关注，使语言文字研究工作者的队伍壮大，重要的是“应用语言学”作为课程

进入大学课堂，对未来的语言文字工作者（其中大部分是教师）进行普及教育。但是，现在在近千所设有中国语言文学系的高等学校里，开设应用语言学的寥若晨星，“语言学概论”虽然普遍开设，但是其中很少有关于应用研究的论述。

要使一门新课程被普遍接受，有两个最基本的条件：1. 高质量的教材，而教材的基础当然还是高质量的理论研究；2. 教育主管部门的认可，而教育部门的态度首先取决于社会对该学科的需求和该学科自身建设的情况，这同样离不开理论建设。因此，为了使应用语言学成为我国正规大学教育的一门学科，也应该立即开展全面系统的理论研究。

应用语言学课程进入全国高校有关专业的教学计划，恐怕还需要一段比较长的时间。在此之前，一些条件比较成熟的学校，应该争取尽早地开设这一课程以及相关的或延伸的其他课程，为理论建设探路，为其他院校探路。某一个学校这样做，于其自身发展也是很有利的。最初开设这一课程的学校，很可能将来会成长为应用语言学理论研究的基地或龙头。

我建议，日后应用语言学会如果就理论建设问题进行研讨的话，可以把“应用语言学”课的大纲列入研讨的范围，集思广益，推进这一课程的开设和规范化；当条件比较成熟时，可以举办培训班或讲习班，为高等院校培养人才，为应用语言学理论建设准备队伍。

三　知识更新问题

应用语言学是一门综合性学科、交叉性学科。涉足这一领域，除了需要语言文字本体研究的基础外，还需要与之相关的其他学科的基本修养。由于过去大学没有有意地培养复合型、通用型人才，专业设置过细，培养出的人知识面过窄。所以像我们这一代人，对于语言学之外的其他领域的知识十分缺乏。王力先生在其

晚年一再遗憾地说，他的数理化知识少，影响了他在语言学领域的成就。近二十年来，科学技术又有了更大的发展，语言文字学和科学技术的关系比以前更加密切了，语言文字学界知识匮乏的矛盾也就显现得比王力先生在世时更为突出了。我希望语言文字学界的朋友们不要到80岁才产生王力先生晚年的遗憾。但是，人过中年，想再学其他学科，特别是学习理科，十分困难。因此我们寄希望于较年轻的同志，特别是四十岁上下的同志，因为他们是承上启下的一代。虽然过去有“人过三十不学艺”的祖训，但是时代不同了，我们都不同程度地受到过现代科技的熏陶，下决心学点新的东西还是可以的。

在众多有关学科当中，我认为以下一些是急需学习的：

1. 语言学内部自己不熟悉的分支。例如研究现代汉语的向古代汉语努力，研究古代汉语的向着现代汉语努力。研究古代音韵的研究研究文字和训诂，反之亦然。而进行具体语言研究的同志都要关心、学习和研究一些语言学理论问题，是自不待说的。

2. 计算机科学。要紧的是从仅仅把计算机当作写作工具、检索工具的阶段向前跨出一步，进到掌握计算机原理和软硬件基本知识，具备编写简单程序的能力的阶段。不达到这个阶段，就难以清楚地了解计算机对语言文字研究的需求，也就难以摆脱传统语言研究的模式、形成对语言文字观察思考的新思维习惯，也就不能进行面向计算机的研究。

3. 心理学。无论是语言文字本体，还是语言教学、法律语言、广告语言、广播电视语言等等语言文字应用领域，乃至计算语言学，无不和人的心理密切相关。现在关于第二语言教学、第一语言教学研究得还很不够，已有的成果，涉及心理分析的也并不多。而一些心理学家，对学习和运用语言时的心理过程比较重视，但是可惜，他们大多对语言文字学缺乏基本训练，因此研究结果常有功亏一篑或隔靴搔痒之憾。在语言文字应用的其他领域，情况

也大体如此。这说明，在心理学专家努力向语言文字学靠拢的同时，也需要语言文字学家掌握心理学的基本内容、方法和国内外研究动态。心理学，只读一些概论式的著作是不够的，对于实验心理学、儿童和少年心理学、认知心理学等等也应该学习。

4. 哲学。我在这里提到哲学，有些同志可能觉得过于迂阔。但是我认为，这是切切实实的建议。语言文字学，按传统学科分类，属于人文学科；但是它既是人类思维的工具，又是思维的外部表现，因而包含着丰富的哲学问题。在西方，哲学家们对语言一直有着特殊的敏感，给予特殊的关心。我国古代一些在语言文字学领域贡献颇多的学者，同时也是哲学家，至少在哲学上颇有见地。例如孔子、荀子、墨子、扬雄、郑玄，直至朱熹、戴震，难以一一列数。用哲学的头脑观察思索语言文字问题，语言文字问题又反过来丰富哲学的思辨，相得益彰。一百多年来，特别是近几十年来，我们在记录、描写语言方面取得了很大成绩，但是同时也几乎完全失去了对语言文字的哲学思考，这可能是我们对语言动态状况关心不够、在对语言事实进行解释方面进展不够理想、对语言文字缺乏宏观研究、难以走出传统圈子的重要原因之一。

对语言学界知识结构的缺陷，可能有见仁见智的不同。我借今天这个机会谈点意见，意在引起大家的议论，是货真价实的抛砖引玉。

把我上面所说的三个意思合起来，可以概括为这样一个思想：应用语言学，现在到了一个关键时刻。时势既需要我们抓紧对语言文字应用的各个领域、种种问题进行研究，尽快拿出社会所需要的成果，又需要在理论建设、进入教育体系和知识更新方面作出切实的努力，以便语言文字的应用研究能在不久的将来进入一个新的境界。说现在是个关键时刻，还有另外一层意思，这就是当前是弘扬应用语言学的最佳时期，如果我们不能及时抓住这个机遇，就可能延误语言文字学、特别是应用语言学的发展。

最后，我想向大家通报一下国家语委机构改革的情况。大家对这个问题很关心，因为国家语委未来的情况直接关系到国家语言文字工作，关系到语言文字学，特别是语言文字应用研究工作的发展。

国务院决定，国家语言文字工作委员会并入教育部，保留国家语委的牌子。具体的改革情况是：国务院规定教育部的职责一是全国的教育工作，二是全国的语言文字工作；在教育部内设语言文字应用管理司和中文信息管理司；国家语委为副部级行政单位；国家语委所属的事业单位语言文字应用研究所、语文报刊社、语文出版社、普通话培训测试中心依然保留。这一改革所产生的变化是：结束原国家语委党组的工作，撤消国家语委的机关党委，国家语委办公室的工作交教育部办公厅，按国务院给教育部下达的精简比例，精简国家语委的公务员。大家可以看出，作为管理全国语言文字工作的行政机构，国家语委的职能保留，权力保留，没有任何变化。同时，由于进一步转变政府职能，改革机关工作，今后语言文字工作会比过去更为加强。这样说是因为：1. 全国各地的语言文字工作机构会进一步健全完善，因为各地语委办不再是单列的机构，有教育厅就有语委办；2. 语言文字工作所需的经费将更有保证；3. 学校是语言文字工作的基础阵地，今后有关教育系统的工作将更顺，更方便；4. 语言文字应用研究所培养研究生的条件可能会有较大改善。李岚清副总理在去年语言文字工作会议上的书面发言里强调，语言文字工作必须加强。此后不久，在他和我谈话时再次强调，语言文字工作只能加强，不能削弱，但是机构要改革，人员要精简。我认为，现在国家语委的机构改革的情况、正是体现了他的讲话精神。教育部的领导和国家语委的领导，已经达成共识：今后的语言文字工作要更加依靠语言文字学及其应用研究，依靠语言文字学家，依靠语言文字学界。为此，就必须做到：1. 加强语言文字应用研究所的建设；2. 设计一个和

学术界保持密切联系的机制，采取一些切实的措施；3. 研究如何推进语言文字的应用研究；4. 研究如何发动全国语言文字工作者关注或投身于语言文字应用研究。

在我担任国家语委主任的四年多时间里，得到了全国语言文字学家、中文信息专家，包括香港、澳门、台湾学术界朋友们的大力支持，我从大家那里学到了很多知识，汲取了很多智慧；我更得到了国家语委全体工作人员，包括语言文字应用研究所各位年长的和年轻的学者的理解和支持。几年来，我一直是怀着难以言喻的幸福和感激之情进行工作的。现在我将怀着同样的心情离开这一令人怀念的岗位。请允许我借这个机会对在座的各位朋友表示由衷的感谢。

我自认为几年来是努力工作的，但是工作的结果和委内外朋友们的期望之间仍有很大距离，我自己也在幸福和感激感之外留下了不少遗憾和歉疚。例如，语言文字应用研究所的条件没有得到根本性的改善，该所的改革也没有到位，普通话培训测试中心的工作一直没有质的提高，全国的语言文字研究还没有发动和组织得差强人意，国家语委职工，特别是语言文字应用研究所人员的生活待遇还不理想，等等。另外，还有一些刚刚开了头的工作，我不能亲自主持进行了。例如全国语言文字使用情况普查、面向计算机的现代汉语研究基础工程、大型语料库的建设等等。这些都只好由教育部的领导同志和语委的新领导班子去完成了。我相信语委新的领导集体在教育部和国务院的领导下，一定会做得比我更好。我希望全国的语言文字学家，特别是从事应用研究的专家学者，继续并且更加加强和国家语委的联系，更加支持国家语委的工作。

我对我国的语言文字学的前景，对语言文字应用研究的前景充满信心。道理很简单，因为我们所从事的工作是国家与民族的需要，不但是今天的需要，更是明天的需要，也可以说是全人类

的需要；同时，我们有大批特别能吃苦，特别能耐得住寂寞的学人。我想，在座各位都怀有和我一样的信心。让我们在这里共同为语言文字学的未来祝福。

在这次会议上，我没有能提交学术论文。一是比较忙，二是语言文字应用研究非我所长。我只是语言文字应用研究的鼓吹者。吹鼓手说的关于剧情的话，总不如在台前演出的人说得准确详尽。因此，如果我说得不对，特别是关于应用语言学的话如果说得不对，尚望以吹鼓手的话待之，既批评，又宽恕。

衷心祝愿这次会议圆满成功，祝愿东道主黑龙江大学和吕冀平、戴昭铭等先生诸事顺遂，祝各位在黑龙江期间过得愉快。

谢谢大家。

（国家语言文字工作委员会　100010）

关于大规模真实文本语料库的几点理论思考*

张　普

一　关于语料库建设

我国的语料库建设始于20世纪80年代初期。那时的语料库叫语言资料库，建设的主要目的是为了给字词典的编纂提供例句或者给语言学家研究语言提供第一手资料。而信息处理领域的专家由于信息处理的需要，也差不多同时开始在计算机中建立语言资料库，用以自动获取语言统计知识，对语言进行计量研究。80年代中期，陆续有电子版的语言资料库及其统计成果投入使用，这种建立在计算机中的语言资料库简称语料库（corpus），它是大规模真实文本的有序集合，是利用计算机对语言进行各种分类、统计、检索、综合、比较等研究的基础，而“文本”（text）则是语言的符号串，文字信息的处理对象，是依据语言学的原则和数理统计的方法从自然语言中抽取出来的。（张普等1991）根据研究的需要，所抽取的文本的长度有时是其自然长度，有时是定长的。在从相对而言是无限的自然语言材料中抽取有限的文本时，有时是等密度的，有时是不等密度的。

从90年代开始，国际自然语言处理领域发生了一些重大变

* 本文承国家自然科学基金重点项目（项目号：69433010）资助。

化，其特征之一就是转向对大规模真实文本的研究和处理，以大规模真实文本为基础的语料库及其语言研究和知识自动获取受到高度重视，并且越来越走向深入和实用。1993 年清华大学黄昌宁教授在《语言文字应用》第 2 期发表《关于处理大规模真实文本的谈话》，指出国际计算语言学界已经把大规模真实文本的处理确定为未来一个时期的战略目标，这将会给语言文字的研究带来巨大的影响。他还认为这种变化和发展反映了现代语言学研究中经验主义思潮的复苏，在语法研究方面促动从宏观到微观的回归，给语言文字研究带来的巨大影响之一就是语料库语言学的崛起，该文引起语言学界的注意。1995 年清华大学出版社和广西科学技术出版社联合出版东北大学姚天顺教授主编的《自然语言理解》一书，其中有专门一章讲述“语料库语言学”。1997 年复旦大学出版社出版该校计算机系教授吴立德主编的专著《大规模中文文本处理》，该书在借鉴国外研究成果的基础上，以大规模中文文本为处理对象，系统地介绍了大规模真实中文文本信息计算机处理的理论和方法。

90 年代，汉语语料库（首先而且主要是现代汉语语料库）的建设和研究得到了蓬勃的发展。语料库的规模从百万级发展到千万级和上亿级，语料的加工深度从字一级发展到词法级、句法级、语义级和篇章级，不同级别的加工技术的成熟程度各不相同。据了解到目前为止，国内已经开发的不同加工深度的现代汉语熟语料库有二十余个。仅就北京语言文化大学而言，近十余年开发的各种语料库就有“现代汉语词频统计语料库”(1985 年)，“当代北京口语语料库”(1992 年)、“现代汉语语法研究语料库”(1995 年)，“汉语中介语语料库”(1995 年)，“现代汉语句型语料库”(1995 年)，与香港理工大学中文及双语学系联合建设的“现代汉语语料库”(1998 年)，与清华大学联合承担国家自然科学基金重点项目“语料库语言学研究的理论、方法和工具”也建设了“现

代汉语语料库”(1998 年)。由于计算机硬软件环境的发展和中文文本的电子版(包括光盘版和网络版)越来越普及,语料库的建设和开发相对而言越来越容易,而语料迅速扩充和膨胀也带来了另外一些问题,例如:语料中的明显错误和不规范用法应否修正问题;统计中的数据稀疏问题;垃圾语料带来的统计垃圾问题;汉语语料统计中的随语料增长的垃圾泛滥问题等等。(邱超捷、宋柔、欧阳龙根 1997)

本文对于语料库的建设和建设中的相关问题进行了一些反思,从普通语言学、社会语言学的角度,零星思考了一些与句法、语义、语用相关的理论问题,提出来与同行进行讨论,希望对今后的语料库建设能有所裨益。

二 关于交际

经典认为:语言是人类最重要的交际工具。现在应该再加上:也是人机之间最重要的交互工具。如果“对话”就是最重要的交互,那么,交互也就是人机之间的“交际”。

但是,什么是交际?交际具有什么性质?

交际总是双方的行为,交际首先分为语言交际和非语言交际。语言交际依靠语言作为载体来传递信息。我们仅探讨语言交际。

语言交际本是一种人类传递信息的行为和过程。信息的发出者(即说话人)的目的是传讯,行为是“编码”,信息的接受者(即听话人)的目的是受讯,行为是“解码”。通俗一点说就是“一个人”要把他所知道的消息告诉“别人”,“别人”要懂得“这个人”所说的消息。所以,交际就是一方表达,另一方理解。以电脑为“一方”或“另一方”,研究电脑如何表达人的语言是“自然语言生成”,研究电脑如何理解人的语言就是“自然语言理解”。因此,研究“自然语言处理”(包括生成与理解),不可以不研究语言交际,不可以不研究人脑的语言机制和模拟人脑的语言机制。

从表达方和理解方来看，现在交际行为至少有以下四种类型：

人表达←——→人理解

机器表达←——→人理解

人表达←——→机器理解

机器表达←——→机器理解

“交际语言学”认为交际是个极其复杂的问题，同样的交际主题，交际主体之一换个角色，由于其知识、教养、性格、心理素质、临时心绪等的不同，都会给交际带来截然不同的结果。（李岗1998）徐通锵先生（1997）认为：“所谓‘交际’，其实质就是交流对现实的认知。”他又说：“交际的过程既是相互交流认知活动的成果，也是人们自发地相互协调语言的结构，使之成为人们必须遵守的严密系统，以便把个人的认知活动的成果纳入社会共同创造的洪流。”

我们认为，交际活动或者说交际行为具有两重性，它既是一种社会行为，也是一种个人行为。交际活动是两重性的统一体，社会行为要通过个人行为来体现，个人行为要融入社会行为之中。据此，我们又有如下认识：既然是社会行为，就要遵从社会的习惯约定和为管理社会行为制定的规范，表达者和理解者都要遵从这些约定和规范才能达到交际的目的；既然是个人行为，并且是要让别人了解自己的认知成果，表达者又是自由的和自主的，因此，既会出错，也会创新，理解者既要容错，也要学习。交际过程中通过“问答”和“讨论”，作出“纠错”和“解释”是不可避免的。

这些认识是本文进行理论思考的最基本的也是最重要的出发点。

表达－理解，容错－纠错，解释－学习，对话－讨论，这些就是自然语言处理中计算机的最基本的也是最重要的智能活动或智能行为。

三　关于文本

语言交际又可以按照信息载体的形式分为口头交际和书面交际。

信息的第一载体是语言，第二载体是文字，第三载体是电磁波。现在一切载体都可以用数字化方式表示，数字是第四载体，是载体的载体，信息最终转化为数字。

对于电脑而言，现在有广义的“文本”，比如声音文本、图像文本、文字文本等，我们所说的大规模真实文本中的“文本”，是狭义的文本。我们遵从GB12000·1—90对“文本”的定义：“语言的符号串，文字信息的处理对象。”这个定义说明这里的“文本”指的是以文字形式记录的语言的文本，即书面语言。语料库通常就是指这种文本的有序集合。因此，口头交际是指利用有声语言的交际，书面交际就是利用文本进行交际。把口头和书面的方式带进来，前面说的交际行为的类型就从四种变成了八种：

（口头）		（文本）	
A. 人表达	←——→ 人理解	a. 人表达	←——→ 人理解
（口头）		（文本）	
B. 机器表达	←——→ 人理解	b. 机器表达	←——→ 人理解
（口头）		（文本）	
C. 人表达	←——→机器理解	c. 人表达	←——→机器理解
（口头）		（文本）	
D. 机器表达	←——→机器理解	d. 机器表达	←——→机器理解

文－语转换实际上是实现c+B，语音打字实际上是实现C+b，文本型的（书面）人机对话就是c+b或者b+c，口头型的人机对话就是C+B或者B+C，等等。

目前，自然语言处理的重点是放在文本方面，知识的获取、分析、表达、理解都是基于文本的，基于口头的处理也在进行，并