汉语言文学研究文库

计算语用学引论

刘根辉 著

Introduction to
Computational Pragmatics

華中科技大學出版社
http://www.hustp.com

汉语言文学研究文库

“华中科技大学文科学术著作出版基金”资助项目

计算语用学引论

刘根辉 著

Introduction to
Computational Pragmatics

华中科技大学出版社
http://www.hustp.com
中国 · 武汉

内 容 简 介

计算语用学是一门新兴的计算语言学分支学科，本书是有关计算语用学研究的入门级读本。作者在书中系统阐述了计算语言学、语用学、形式语用学、计算语用学等基本概念，侧重探讨了语用学与计算语用学关注的重点——言语行为与语境的形式化方法，以及基于语境的自然语言理解模型问题。全书结构清晰，逻辑严密，循序渐进，有助于读者比较全面地了解计算语用学这一方兴未艾的学科的发展态势，适合对语言学及计算语言学感兴趣的读者阅读。

图书在版编目(CIP)数据

计算语用学引论/刘根辉著. —武汉：华中科技大学出版社，2017.1
(汉语言文学研究文库)
ISBN 978-7-5680-2060-2

Ⅰ.①计… Ⅱ.①刘… Ⅲ.①计算语言学—研究 Ⅳ.①H087

中国版本图书馆 CIP 数据核字(2016)第 167911 号

计算语用学引论

Jisuan Yuyongxue Yinlun

刘根辉 著

策划编辑：周小方 杨 玲
责任编辑：李 祥
封面设计：原色设计
责任校对：马燕红
责任监印：周治超
出版发行：华中科技大学出版社(中国·武汉) 电话：(027)81321913
武汉市东湖新技术开发区华工科技园 邮编：430223
录 排：武汉正风天下文化发展有限公司
印 刷：武汉鑫昶文化有限公司
开 本：710mm×1000mm 1/16
印 张：10.5 插页：2
字 数：206 千字
版 次：2017 年 1 月第 1 版第 1 次印刷
定 价：38.00 元

汉语言文学研究文库

Introduction to Computational Pragmatics

序一

人类在数百万年的进化过程中逐步从原始猿类中进化分离出来，人类同其他动物最根本分界的特征有以下几个：

第一是直立行走。直立行走使人类可将手和脚进行分工，将双手解放出来，从事更丰富多彩的活动。直立行走使人大大提高了自己的视角，使眼界更远、更开阔，认识世界的能力更强。手脚的分工反过来刺激了脑的发育。

第二是能动手制作劳动工具，延伸自己的器官。动物只能被动地适应这个世界，而人类由于能制造工具，就摆脱了被动适应世界的处境，可以主动地改造世界，创造出自然世界原来没有的东西，并使人脑更加聪明和发达。

第三是在长期的实践中，人类产生了越来越丰富的语言。其他动物也有自己的传递消息的声音（语言）、动作（动作语言），如蜜蜂用舞蹈来告诉同伴花丛的位置与方向，鸟类的鸣叫、猿猴的不同叫声向同类发出“危险将至”、“此处有食”的信息。它们在欢愉、惊恐、痛苦、愤怒等不同情况下发出的声音是不同的，而同类是理解的，这种信号已具有了语言最初级的功能与特征。但这种“语言”十分简单、原始，语义含量极低。

人类是社会化的动物，在数百万年的进化过程中，在越来越复杂的实践活动中，由于许多活动需要多人配合才能实施，必须交流思想，就逐步形成了越来越丰富的口头语言。有了语言就突破了动物界代际重复的局限，上一代人积累的知识通过语言可以迅速扩散传播，可以顺利地传递给下一代，使下一代的认识能力、实践能力进一步提高，超过上一代。这是十分伟大的进步，大大加快了人类的进化速度。

口头语言逐步丰富，延绵使用了百万年之后，在大约一万年至数千年前形成了文字。文字可以精确地记录人类的实践，更大范围地传播人类的思想、理论和知识，产生了巨大的作用，使人类社会全面快速进步。因此，文字的发明被历史学家看作人类社会脱离野蛮时代、进入文明时代的标志之一。

正是以上三个基本特征的延续和发展，使人类最终成为地球的主宰者。语言的产生，其意义无论怎样强调都不过分。

斯大林曾指出，语言是思想的“物质的语言的外壳”。他的这一论断有积极意义，但又不全面。因为人类的思维形式有抽象（逻辑）思维、形象（直感）思维和创造性思维，而创造性思维中既有抽象（逻辑）思维，又有形象（直感）思维成分。人的抽象（逻辑）思维是建立在概念的基础上的，语言可以说是抽象（逻辑）思维的外壳。人脑的前额叶（其位置在人头部的前额内）很突出，而智力水平最高的类人猿——黑猩猩的前额低平。近三十几年来的研究表明，人的前额叶中至少有数千万个神经元，它们的激发只同抽象概念直接相关，而不管刺激来源于什么信息通道。换句话说，它们只同抽象概念相联系。从人类进化过程中不同时期的头盖骨形态变化中也可以发现，进化程度越高，脑容量越大，前额叶也不断增大。显然人脑语言中枢模块与前额叶的发展是同人类语言的发展相对应的。形象思维的发展比抽象思维早得多，在动物界早已有之。可以说有感觉和知觉的动物界主要是靠形象思维生存的。

我们应该看到，人用语言描述一个场景、一幅图像往往不准确，效率也不高。而用眼睛和视觉皮层配合一瞬间就可以准确完成。可见，形象的感知与思维极为重要，语言还是有很大局限性的。

人脑、人的思维是综合集成式的。人脑集中了动物进化过程中相当多的最优秀的神经模块。如人的脑干中甚至有爬行动物的神经模块，而丰富的脑皮层特别是前额叶的神经模块的许多部分却是人特有的。人的思维方式和工具多种多样，也具有综合集成的特点，为了解决社会实践中的问题，它们被综合应用、灵活应用。而语言是人与人之间思想交流的工具，是社会化的桥梁，地位特别重要。

科学技术是推动人类社会前进的最革命的因素之一，也是最根本的动力之一。正如钱学森院士所指出的，是科学革命推动技术革命，技术革命推动产业革命，产业革命推动社会革命。科学、技术、产业的发展使人类社会的生产力不断提升，当生产关系已经严重阻碍生产力发展的时候，就不可避免地发生社会革命，推翻阻碍生产力发展的生产关系和上层建筑，建立适应生产力发展的崭新的生产关系和上层建筑，使人类社会由较低级的社会向更高级的社会演进。

二十世纪是人类社会变化最剧烈，灾难最惨烈，进步也极为巨大的世纪。二十世纪人类创造的四项科技成果意义重大，它们是爱因斯坦相对论的提出和核能释放、航空航天科技、电子计算机及计算机网络的发明、DNA（脱氧核糖核酸）双螺旋结构的发现。

其中，电子计算机及计算机网络的发明意义更为深远。因为在计算机发明之前，科学技术的总的作用都是在延伸人类的体力、肢体和感知能力。如基于物理学和化学等学科的化石能（煤、石油、天然气）、热能、电能、核能的利用，蒸汽机、内燃机、电动机、火药、炸药、原子弹、氢弹等的发明，将人类的肌肉动力、人利

用的畜力(牛、马等)放大了十倍、百倍、千万倍。各种机器的研发成功地将人类从繁重的体力劳动中解放了出来。而望远镜、显微镜、电子显微镜、雷达、声呐、射电望远镜等的发明使人类看到更广阔、更遥远、更细微的世界。而电子计算机及其网络的发明,则开始延伸人类的大脑和思维,开始将人类从繁重的脑力劳动中解放出来,因而具有划时代的伟大意义。

第一台电子计算机于1946年在美国宾夕法尼亚大学莫尔电子学院诞生,主要用于科学计算(计算火炮的弹道)。此后,计算机的性能不断提高,应用领域不断扩大,经过70年的发展,每秒钟运行10亿亿次的我国高性能超级计算机——神威·太湖之光计算机已经面世,其运算速度居世界第一位。计算机的应用类型从科学计算到数据处理、过程控制、数字化通信、人工智能等,几乎渗透到人类社会生活的每一个角落,深刻地改变着人类社会的形态和面貌,改变着人类的思维方式。计算机网络将整个世界联系起来,将偌大一个地球变成了地球村,使信息传递的空间距离几乎消失。

电子计算机发明十年之后,一门旨在用人造系统模拟人类思维和智能行为的,横跨计算机科学、数学、物理学、心理学、生理学、脑科学、逻辑学、哲学、社会学、行为科学、语言学等多学科的崭新的学科——人工智能应运而生。1956年暑假,在美国达特茅斯学院召开了世界上第一次人工智能学术会议,宣告这个学科的诞生。参加会议的有卡内基·梅隆大学的认知心理学家、诺贝尔经济学奖得主赫伯特·西蒙、心理学家艾伦·纽维尔、麻省理工学院数学家明斯基等16人。这以后人工智能的研究领域被确定为自然语言理解、模式识别、知识表达、问题求解、机器学习、机器定理证明、机器翻译、专家系统、机器人、计算机视觉、语音与声音合成、数据库智能查询、自动程序设计、人工智能程序设计语言等。近十几年来,数据挖掘、大数据的智能化处理、无人化智能平台与机器人成为研究热点。

自然语言理解是人工智能领域最"古老"的也是最困难的研究方向之一。早在20世纪50年代初,就有人开展同语言计算机处理相关的研究,如美国乔治城大学与IBM公司研发"俄-英自动翻译"试验系统等。随着计算机工作者同语言学工作者的精诚合作日益加深,自然语言处理、计算语言学这种提法应运而生,它的范畴超出了人工智能工作者最先提出的"自然语言理解"范畴,并且具有了广泛应用的意味。

华中科技大学文学院以尉迟治平教授为首的中文系研究团队高度重视计算语言学方面的研究工作,20世纪八九十年代,在中文系建立中文信息处理实验室。1998年底学校领导决定要我参加计算语言学研究工作,支持文科发展。2000年华中科技大学文学院申请"语言及应用语言学"博士点,获得国家学位委员会批准。2001年,文学院成立计算语言学研究所,委任我担任第一任所长。

刘根辉是毕业于中文系的青年学者，他以惊人的毅力勤奋学习数学、计算机和人工智能等领域的知识，掌握了计算机及网络的应用技术，并以优良成绩考入华中科技大学人工智能研究所控制科学与工程学科模式识别与人工智能专业，成为我的第一位计算语言学方向的博士研究生。经过六年多的艰苦努力，他跨过文科、理工科两大领域的界线，获得工学博士学位，成为跨领域的两栖型学者，并被提升为副教授。当前，国内这样的人才十分稀缺。我们深深感到，只有培养更多既有深厚文科根基，又有扎实的数理与计算机科学知识和技能的两栖型学者，计算语言学或自然语言理解领域的堡垒才能最终被攻克。

在为刘根辉选择研究方向的时候，我们有多种选择，为什么最终选定计算语用学为突破口呢？我在计算机、人工智能领域工作了四十三年，发现我国学者中相当多的人总在做跟踪性研究，而且培养出了一种“奴性”：外国人没有碰过（或碰得少）的问题，他们不敢碰；外国人没有说过的话，他们不敢说；外国人的缺陷或错误，他们不敢反驳、纠正。这使得国内许多领域的学术研究“殖民化”，民族自尊心、自信心和创造性被严重压抑。我们必须突破这种氛围！怎么突破呢？我的父亲李国平院士认为：“多学科的边缘领域是原始创新的富集区，应该到那里去干。”他还认为：“一般的老师，总是将学生带到自己最熟悉的领域之中，在别人论文的缝隙中做工作、讨生活。好的老师是将学生带上一条前途远大的康庄大道，尽管自己并不一定很熟悉，但他有决心引导年轻人去探索。”他还告诉我们：“对洋人的好东西，我们要认真学习，但不能有任何奴性。那些外国人我们都较量过，没有什么了不起的！”四十年前，就是在父亲这种学术思想的引导下，在当时国内还不敢提人工智能，很少有人从事相关研究的困难情况下，我起草了国内第一份人工智能发展规划报告，并完成了第一项将人工智能技术应用于重大武器装备的科研工作。在刘根辉的博士论文选题上，我们让这种传统再现了。

爱因斯坦曾经说过，一个蹩脚的科学家是在木板上寻找一个最薄的地方，然后密密麻麻地打上许多个洞。而一位优秀的科学家是在木板上选择一个最厚的地方，扎扎实实地打上一个深洞。

我们认真分析了自然语言理解和计算语言学的发展过程和趋势。经过全球许多科学工作者的艰辛努力，计算语言学在语音学、词汇学、语法学、语义学、语料库语言学等方面都已经取得了重大进展，形成了不少理论和方法，甚至研发出了一些应用系统，而当时国际上计算语用学的论文很少，国内尚无人问津，而且汉语中又有许多特殊规律需要探索。计算语用学需要大量前期工作的积累，是综合性极强、难度极大的研究工作。因此，选择它作为主攻方向符合我们的上述理念。

经过多年的积累和艰苦探索，刘根辉博士不负众望，出版《计算语用学引论》这本专著，我们都感到无比欣慰和愉悦。这本著作是我国第一部计算语用学专

著，由于作者独具匠心，鼓励后继，这本专著又是一本大学本科生、研究生和学者探索计算语用学的入门引导书，其作用十分重要。

本人认为该书具有以下特点：

第一，作者系统深入地综述了计算语言学、自然语言理解领域较为完整的发展过程、当前状况和未来趋势，读者可以获得十分清晰、系统的概念和信息。

第二，作者介绍了计算语用学的起源、发展和主要的研究方向，论述了汉语语用学的研究发展趋势。充分考虑了我国学者研究汉语语用学的需要。

第三，在所有的计算机应用问题中，形式化都是关键的一步，舍此无法使用计算机解决任何问题。作者在第三章研究了语用的形式化问题，结合形式语用学发展概况，阐述了国内形式语用学研究的发展思路。

第四，作者深入探讨了计算语用学的研究途径和方法，就语用推理、溯因推理、信任推理、动态环境以及话语意义的计算模型等五个热点问题进行了深入探讨。

第五，作者在 Austin、Searle 的语言行为理论的基础上，从认知角度出发提出了言语行为的形式化模型系统。

第六，作者在探讨语境的形式化与自然语言理解模型的基础上，给出了有关语境的一系列形式化定义，提出了基于语境的自然语言理解实现框架，构造了一个基于语境的自然语言理解模型系统。

全书结构清晰，逻辑严密，循序渐进，有助于读者比较全面地了解计算语用学这一方兴未艾的学科的发展态势，适合对语言学及计算语用学感兴趣的读者阅读。尤其让我感到高兴的是作者具有强烈的创新精神，提出了一些国内外学者没有提过的见解、理论、方法、方案。并且知行统一，构造出模型系统，做出了开创性的贡献！这种既敢想敢干又严谨求实的科学态度和钻研精神是应该大力提倡的！

李德华

（华中科技大学人工智能研究所、计算语言学研究所所长，二级教授，博士生导师）

2016 年 6 月 21 日于华中科技大学人工智能研究所

序二

本书是国内第一部系统探讨计算语用学研究的学术专著。

在语言学的诸多分支学科中，语用学是一个十分晚起的新兴学科。语用学的两大支柱，一是语义，二是语境。长期以来，对这种特定语境中的特殊语义，即“语境义”，或称“言语义”，西方现代语言学并不关心，甚至是排斥的。相反，倒是中国古代的训诂学家十分重视语境义，他们撰写了大量的经史子集的注疏，这种往往被讥为“寻章摘句”的工作，实则正是随文释义，探求必须能贯通上下文的语义。尽管有时不免陷于烦琐，或耽于玄想，或失于误读，但注疏家说解的主旨是试图寻求在母本中的语境义是没有疑问的。中国古代注疏绵延累积两千年，载籍汗牛充栋，可见这种重视语境义分析的传统深厚悠远，在世界文化之林中罕有其匹。

中国古代学者将传统汉语言文字学称为“小学”，下分“训诂”“文字”“声韵”三个分支学科，大致分别相当于现代语言学的语义学、文字学、音系学。语法学在《马氏文通》之前并没有成为独立的学科，有关虚词和句法的讨论归属于训诂学。

清嘉庆年间，谢启昆撰《小学考》，在“训诂”“文字”“声韵”之外另立“音义”一门。谢启昆在《〈小学考〉序》中对“音义”的性质进行了阐述，说：“卷首恭录‘敕撰’。次‘训诂’，则续《经义考》《尔雅》类而推广于《方言》《通俗文》之属也。次‘文字’，则《史篇》《说文》之属也。次‘声韵’，则《声类》《韵集》之属也。次‘音义’，则训读经史百氏之书。‘训诂’‘文字’‘声韵’者，体也，‘音义’者，用也，体用具而小学全焉。”这段话的意思是说，“训诂”“文字”“声韵”是通释一般的语义、文字、语音的门类，“音义”是解读经史百家文本中具体语义、文字、语音的门类，“训诂”“文字”“声韵”是本体，“音义”是其应用，体用俱全，“小学”才能构成一个完整的学科。“敕撰”单设，出于意识形态，不必厚责古人，增立“音义”则不能不说确实是慧眼独具。《小学考》中著录的“音义”书，之前都依附于母本文献，并不被认为是“小学”著作。例如唐陆德明《经典释文》在《四库全书总目》列于经部“五经总义类”，何超《晋书音义》列于史部“正史类”，殷敬顺《列子释文》列于子部“道家

类”，谢启昆将这些著作分别剥离出来，归入小学，应该说已经具有了某种程度上的“语用”意识。

实际上，音义与注疏两类都是释读特定文本中的语境义的，性质相同。谢启昆之所以没有把注疏合于音义归入小学，是因为注疏是夹于母本随文释义，不像音义是单本别行自成一书。但是，汉唐注疏原本也就是单本别行的，后人为了阅读方便，才将注疏散入母本，夹于相应词句之下。因此，可以说音义是注疏的原始形态，注疏是音义的转换变形，二者都是寻章摘句，随母本按字、词、句析释语境义，性质是完全相同的，应该合为一类。从魏晋古注到清儒新疏，经过两千年的漫长发展，训诂学重视语境义的实践提炼出了以语境义为准的学科门类，但是由于西方现代语言学的引入，中国传统语言文字学十分迅速地实现了向现代语言学的转型，“小学”自体的发展基本上已经中断，音义学最终没能从传统学脉中自主萌生蜕化成具有中国语言文字学特色的“语用学”。

西方语言学如果从古希腊语言学算起，已经有两千年的历史，而现代语言学发端于索绪尔及结构主义语言学，至今不过一百多年。在索绪尔的语言观的影响下，语言学家区分语言和言语，认为只有抽象的语言系统才具有研究价值，语法学、语义学、音系学成为语言学研究的主流。但是，后来学者发现即使操着标准的语音，遵循规范的语法规则，正确地使用词语，听者还是可能无法正确理解言者想要表达的意思；反之，虽然语音蹩脚、语法有误、词义扭曲，但也可能并不妨碍听者正确理解言者说的话。可见，将言语排斥在语言学研究领域之外，会造成许多困扰，只有兼顾语言的抽象的形式化理论和具象的实用性运用，才能构成完整的语言学学科体系，正如谢启昆所云：“体用具而小学全焉”。

“语用学”作为科学术语，最早是在二十世纪三十年代由哲学家提出的，但并没有引起语言学家足够的重视，直到二十多年后，相关研究才逐渐取得了长足的进展。1977 年，《语用学学刊》(*Journal of Pragmatics*)在荷兰阿姆斯特丹出版发行，以此作为标志，语用学才成为语言学的一门独立的分支学科。

计算语言学使用计算机技术模拟人的语言能力，从而理解、分析、处理人类实际运用的自然语言。对应于人类的语言学的音系学、语法学、语义学，计算语言学也有语音识别、语法分析、语义理解等研究领域。在汉字文化圈，汉字信息处理也是计算语言学的一个重要研究领域。从二十世纪五十年代至今，计算语言学在以上各个方面都取得了许多重要成果。如果参照我们上面所回顾的，我国传统语言文字学从训诂学、文字学、声韵学到另设音义学门类，现代语言学从语法学、音系学、语义学到新立语用学的学科发展历程，计算语用学正应该是计算语言学进一步发展的必然方向。

计算语言学期待计算语用学的问世，计算语用学的建立为计算语言学发展之势所必趋，刘根辉博士的《计算语用学引论》的出版，可以说正是应运而生，顺

适学科发展之需求，符合学术演进的规律。

语用学的研究取向不再只停留在静态的语音、词汇、语法等语言知识获取的层面上，而是更关注动态的交际过程，研究会话双方如何实施有效的言语策略，成功达到交际的目的。言语交际过程由一连串会话组成。每一段话都以上一段话为前提，又对下一段话有所期待，回话对这个期待做出回应，以对方的期待为前提完成会话，同时又设置自己的期待，开始己方的一轮会话。这些会话如此彼此勾连，又服从双方各自的交际目的，从而构成一个互相关联的整体。会话必须瞻前顾后，这个言语交际必须关照的上下文，就是语用学的核心概念——“语境”。另外，所谓“语境”还包括非语言的元素，包括交际的时间、地点、环境、氛围等客观元素，会话人的情绪、态度、会话经验、生活经历、文化教养、心理素质、思想观念等主观元素，交际双方的身份、地位、亲疏、好恶、敌友等关系元素，可以统称为“背景语境”。在某种背景语境下，交际双方应该采取怎样的会话策略，存在着约定俗成的社会规约，违反这些规约，就是语用失误。

音义学的创设的学术背景，是清乾嘉时“因声求义”的训诂方法的成熟与盛行。两汉人注音方式是“读若”“读如”“读为”“读同”等，段玉裁《说文解字注》在示部“禱”篆下说：“‘读若’者，皆拟其音也。凡传注言‘读为’者，皆易其字也。注经必兼兹二者，故有‘读为’，有‘读若’。‘读为’亦言‘读曰’，‘读若’亦言‘读如’。字书但言其本字本音，故有读若，无读为也。读为、读若之分，唐人作正义已不能知。”这段话已经将《小学考》“训诂”和“音义”两类的区别说得很清楚了：训诂书虽然也注音，但仅仅是“拟其音”，音义书注音也有“但言其本字本音”的“拟其音”的，但更重要的是“易其字”，即破假借，通过注音指明本字，揭示语义。不单是“读为”“读若”，其他注音方式，如反切、直音等也都如此，分为“拟其音”和“易其字”两类。乾嘉时以戴震和段玉裁、王念孙师生三人为杰出代表的皖派学者不仅揭破了这个“唐人作正义已不能知”的千年隐秘，并上升到理论的高度，提出探求语境义的方法——“因声求义”，最著名最精练的表述就是王念孙在《〈广雅疏证〉叙》所说的“就古音以求古义”。这种治学方法形成了段王之学的鲜明特色，盛行于世。谢启昆作为乾嘉时代的著名学者，自应受到时代学风的影响，充分认识到音义书的重要学术价值，从而创设为小学的一个新的分支学科——音义学。

上面所说，一个是言语会话，一个是文本阅读，核心都是“语境义”，可以说二者构成了语用学研究的两个层面。戴震《〈转语二十章〉序》说：“人之语言万变，而声气之微有自然之节限，是故六书依声托事，假借相禅，其用至博，操之至约也。学士茫然莫究。今别为二十章，各从乎声，以原其义。夫声自微而之显，言者未终，闻者已解，辨于口不繁，则耳治不惑。……俾疑于义者以声求之，疑于声者以义正之。”将言语会话和文本阅读的不同性质辨析得十分清楚。

言语会话时，交际双方共处于同一语境中，对话题心照不宣，根据上下文自

然地更相授受，对此时此地的背景语境所要求的会话规约默从而不逾矩，“言者未终，闻者已解”，这是言语会话的常态。只有在不同的阶层、不同的族群、不同的语言之间，即所谓“跨文化交际”的情况下，由于对会话的社会规约的隔膜，才可能出现语用失误。

文本阅读与言语会话不同，读者处在作者的语境之外，读者要想正确而准确地解读文本，首先必须了解并进入作者彼时彼地的语境中去，在读者与作者所处时代相距悠远，语言发生了巨大变化的情况下，这是非常困难的，这就是戴震所说的“学士茫然莫究”。段玉裁《〈广雅疏证〉序》说：“圣人之制字，有义而后有音，有音而后有形。学者之考字，因形以得其音，因音以得其义。治经莫重于得义，得义莫切于得音。”汉字形音义三位一体，字形标示字义，常规的用字，字音、字义与词音、词义相同，读者据字形知字义明字音，也就会知词义明词音，文本上下文贯通无碍，语义畅晓明晰。如果用假借字，字音与词音相同，但字形标示的字义与词义完全无关，借字替代的是本字，本字的字形标示的字义才与词义相通，链接借字、本字、词语的媒介是“音”。作者创制文本时假借字只是记音符号，笔下所录的虽是借字，口中暗诵的是本字音，心里悬设的是本字字形标示的词义，读者如果置身于作者相同的语境中，也可以通过“音链”作为媒介辨识本字直达词义，从而读通文本。但是，如果读者与作者的原始语境隔膜，由于本字是隐性的，阅读所见只是显性的借字，就会将借字字形标示的字义当作词义，从而为上下文语义窒碍感到困惑，或者强解迂释，曲圆一说，扭曲上下文语义，这就是所谓“望文生训”，是文本阅读之大忌。如果语言发生了变化，借字音、本字音、词音有了差异，导致借字、本字、词语的音链断裂，读者即使明知此处可能是假借字，但在不同的语音系统中丢失了音链，段玉裁所说的“因形以得其音，因音以得其义”将会因为古今音异而失效。汉魏经师训释先秦文本就遇到了这样的问题，但是先秦人阅读这些文本的经验和知识，通过师长辈辈教授而为他们所掌握，知道何字是借字，其本字是何字，于是采用“读为”的注音方式指明本字音，帮助读者跳过“因形以得其音”，直接“因音以得其义”，辨识这个“语境义”，这就是“音义”之学。必须指出的是，音义家所注的并非先秦古音，而是自己口中的汉魏今音。“读为”是在汉魏语境中重建本字和词语的音链，借字和本字的音链仍然是断绝的，二者今音并不一样，注借字音是“拟其音”，注本字音是“易其字”，作用南辕北辙，绝不能将“音义”书的注音误为古音，或者将“易字”音误为借字音。音义学是在后世不同的语境中凭借先世传下的阅读特定文本的知识来解读文献的，如果遇见超出这些具体知识的情况仍会“茫然莫究”。唐宋以降，去古日远，由家学师说传承的阅读先秦具体文本的知识会逐渐失真或残缺，晚出的音义著作不免良莠不齐，得失参半。只有读者自觉地有效地进入作者当时的语境中，才能科学地进行文本阅读，获取准确无误的语境义。以戴震、段玉裁、王念孙为代表的乾嘉学者，总

结了历代声训和音义学的经验，提炼出“因声求义”的科学理论，提出了一整套可操作的研究方法。文本阅读时如果从上下文语境看语义扞格不通时，就应该考虑因声求义，这就是戴震所说的“疑于义者以声求之”。读者首先应该考察作者创制文本时的背景语境，研究作者当时的语音系统，确定每个汉字的声韵调，即所谓“音韵地位”，作为因声求义的先决条件，这就是王念孙所说的“就古音以求古义”。然后就可以进入作者创制文本时的语境，恢复借字、本字、词语原本的音链，据借字审定其古音，循音链推得本字音，再穷尽比对符合此音韵地位的所有汉字，考察每个字的全部义项，筛选出音义匹配与上下文语境密合的本字，这就是戴震所说的“疑于声者以义正之”。整个过程破借字，循音链，求本字，明词义，就是段玉裁所说的“因形以得其音，因音以得其义”，富有学理，操作性强，实在是文本阅读的不二法门，但其运用并不容易，要“求古义”须先知“古音”，所以戴震、段玉裁、王念孙都精研声韵，苦心孤诣，是著名的古音学家，这是一般学者无法企及的。

上面讨论了语用学的两个层面：对于言语会话，语用学研究的是如何根据语境采取恰当的会话策略，避免语用失误；对于文本阅读，语用学研究的是如何从上下文语境准确理解语义，获取正确的信息。言语一发即逝，文本可永久保存。会话虽然是两人以上多人参与，但常态为隐私活动，外人无从知晓其内容，文本多是作者自说自话，但其目的却是供人共享。正因为文本具有这种知识性、传媒性、共享性，所以它成为文献的主要形态，存储着人类古往今来全部的知识，是人类获取知识的主要信息源，学术研究主要就是利用以往的文本，挖掘新的知识，创造新文本。因此，言语会话的研究具有很强的应用价值，文本阅读的研究则还具有很强的学术价值，应该也是语用学研究的重要内容。

言语和文本是人类语言活动的两种形态，两种形态并非彼此孤立，毫无关涉，互相可以转换，言语笔录即转为文本，文本诵读即转为言语。在互联网高速发展的今天，人类以往常态以言语形态出现的语言行为以空前规模以文本形态出现。文本撰写原本是作家和学者的行当，劳神耗时，撰写困难，出版周期长，专业性强，受众也就有限，书价日贵，读者日少，文本阅读日趋小众化。但随着自媒体的蓬勃兴起，文本写作成了寻常百姓也可以做的简便易行的事。作者随手撰写，随时发表，读者随意浏览，也可以随心回应，读者与作者即时交流，使文本阅读具有了会话的性质。因之，文本写作和文本阅读日益平民化、普遍化，所以自媒体又被称为“公民媒体”、“个人媒体”。如果说自媒体的发布和回复还只是具有会话的性质，那么，社交媒体(social media)则使言语会话的形式真正实现了文本化，会话参与者借助互联网交际，交际过程与言语会话完全相同，各自根据上下文交互发言，只是口耳相授的有声语词变成了闪烁于荧屏的可视化的有形文字。社交圈内每个人都既是文本写作者又是文本阅读者，而圈内其他人也都

可以围观共阅，或者插入会话发表意见。会话的隐私性被消解，会话文本往往通过社交媒体公开发布，目的就是展示自我言语，供公众阅读。另一方面，由于社交场合网络化，造成了会话的隐秘化（不是上面说的隐私性），言者可以隐身、化名，那些当面不好说、不愿说的话，在平常不能说、不敢说的话，都可以在社交媒体上倾诉、发泄，有时甚至会话人近在咫尺也要通过文本来传递信息。信息时代会话和文本无缝对接，互相融合，极大提升了文本阅读的价值，给语用学带来新任务和新挑战。

传统媒体文本的载体是纸张等固态物质，文本必须首先数字化才能运用计算机进行阅读和处理，自媒体和社交媒体所产生的文本基于互联网，天然就是数字化文本，适于机读。同时，这些文本呈爆炸式增长，体量庞大，增长率高，传播迅捷，更新急速，具有传统媒体文本前所未见的“日日新，又日新”的特征，是典型的“大数据”（big data），阅读这种时刻不停刷新的海量文本是人工绝对无法承受的，也只有计算机才能胜任文本阅读和处理的繁重工作。

计算语言学利用计算机技术使电脑模仿人脑，模拟人类的语言能力，进行言语会话和文本阅读，取得了丰硕的成果，但始终为一个老大难的问题——排歧——所困扰。自然语言中大量词语有多重意义，语言越发展，多义词就越多，而且一个多义词的义项也就越多，另外还可能有由于各种修辞而产生的临时语义。人类在言语会话或文本阅读时会运用他全部的人生智慧和全部的言语经验排除歧义，筛选出正确的语境义，一般不会产生误解。但对于计算机来说，要想获得人类千百万年累积的语言能力和几十年磨砺出的语言直觉，实非易事。计算语言学家一般会建立一个机读词典，尽其所能收录词语及其所有义项，以仿构人类的语言知识，然后根据语法结构、词语搭配规则等设置条件从词库中筛选词条词义，以模拟人类的语言能力，这是基于规则的方法；还有基于统计的方法，根据概率来择词定义，这是模拟人类的言语经验。但无论是基于规则，还是基于统计，都不能完全彻底地排除歧义，说明应该还有一种人类的语言能力没有进入计算语言学家的视野，这就是从上下文判断辨识语境义的能力。乾嘉学者由于秉承历代训诂学家重视语境义的传统，又具有丰富的实践经验，对语境义的求取方法有很深入的讨论，例如王念孙《〈广雅疏证〉叙》在论述“因声求义”时说：“有字别为音，音别为义，或望文虚造而违古义，或墨守成训而鲜会通。易简之理既失，而大道多岐矣。今则就古音以求古义，引申触类，不限形体，苟可以发明前训，斯凌杂之讥亦所不辞。其或张君误采，博考以证其失；先儒误说，参酌而寤其非。”这段话的意思是说，综合考察字形和字音可以确定词条，但“音别为义”，必须从词的众多义项中筛选语境义，不能只据字形“望文虚造”，也不能仅凭词典“墨守成训”。正确的方法是“会通”文意寻求“古义”，也就是说要贯通上下文确定语境义，否则将“大道多岐”，迷失在众多义项中。求取语境义的先决条件是“就古音

以求古义”，考察并进入文本作者当时的语境中去。在这段话中王念孙特别强调“引申触类”，“博考”“参酌”，即应该尽量搜集其他同类性质的文本，同一个词，处在同样的上下文框架中，互相参见比较，用来作为推定语境义的证据。这是典型的归纳逻辑。王念孙强调所说不辞“凌杂之讥”，显然有枚举归纳证成假说的方法论的意识。从乾嘉学者“因声求义”的训诂实践看，他们列举的例证一般不是白文文本，都会有古训、谐声、重文、声训、注音、通假、异文、连文、对文等实证，这种以同类语境文本的已知语境义推定阅读文本的待考语境义的方法，就是现代归纳逻辑的类比推理。乾嘉学者的这些论述和方法是两千年中国传统语言文字学的经验总结，是专门之学，需要经过学习和训练才能掌握。推及一般，可见人类从上下文判识语境义的能力是后天养成的经验，与人类通过遗传天生就有的语言能力性质不同，必须建立专门的学科，模拟人类这种语言经验能力，以济计算语言学之穷，这就是计算语用学。以往计算语言学基于统计的方法也是模拟人类后天的语言经验，但其取向不同，优选的是概率高的义项，而计算语用学是排除冗余的歧义，筛选适合于上下文的义项，这个义项多半概率较低，至于临时义，在计算机词典之外，其概率为零。但是，计算语用学并不是与计算语言学相对相反，而是相辅相成的计算语言学的一个不可或缺的分支学科。

综上所述，计算语用学是计算语言学发展的必然需求，也是计算语言学不可缺少的有机组成部分。刘根辉博士的《计算语用学引论》作为国内第一部研究计算语用学的专著，系统地探讨了计算语用学的研究对象、研究目标、研究方法等诸多方面，为计算语用学建立了一个比较系统的理论框架，同时对计算语用学的主要研究论题——语用推理、溯因推理、信任推理、动态语境以及话语意义的计算模型——做了详细的介绍和深入的讨论分析，并且构建了一个基于语境的自然语言理解模型系统，验证了理论框架的可实现性，有力地推进了中国计算语用学的建设和发展。

计算语言学与普通语言学不同，计算语言学面向计算机，研究如何使用计算机理解、分析和处理人类的自然语言，包括言语和文本两种形式。因此，计算语言学不仅要探讨理论问题，更要提出解决问题的技术方案，并在计算机上实现。要做到这一点，首先必须将要处理的对象形式化，建立数学模型，才能使用计算机语言编制程序对形式化的语料进行处理。不言而喻，比起语法、词汇、语音，语境的形式化显然更困难，更富有挑战性。刘根辉博士的这本书对语境形式化作了较为深入的探讨，给出了面向自然语言理解的语境及相关概念的形式化描述，构建了一个基于语境的自然语言理解模型系统，为语境形式化研究提供了思路。

以上所述，是我们觉得在阅读本书时值得注意的几点。

互联网改变了人类社会的方方面面，也改变了人类的言语行为。传统媒体的文本由作家或学者精心制作，结构完整，逻辑严密，自成一个完整文档，上下文

前后呼应，脉络明晰，便于提取具有标记性的语境信息。但自媒体和社交媒体文本，由于由言语转化而来，带有会话的鲜明烙印，话题游移不居，每人发布的各篇文字，文本分离呈碎片化、离散化，一篇文字可以和上下篇没有逻辑关系，而与并不相邻的另一篇遥相呼应，甚至与本网站、本社交圈、本会话无关，而是从互联网上的某一篇文本引用导入话题，互为上下文关系，语境由语篇扩散到整个网络。同时，由于自媒体和社交媒体的平民化、即时性，其文本不可避免具有不规范、猎奇性的特点，常有语法偏误、字形错讹、语序颠倒、词义扭曲的现象，会造成文本阅读的问题。这些文本中还大量使用各色各类的所谓“网络语言”，诸如缩略语、双关语、同音别字、方言词、外语、字母词、表情符号、字母画画等等，都不见于正常的言语和文本，即使是人，如果不是经常上网也会阅读困难。这些词语最大的特性是时尚性，流行时如潮涌席卷网络，流行过如雪化销声匿迹。时过境迁，孤立地阅读这些文本可能如观天书，必须将往时网络作为语境才能解读。显然，计算语用学已经应该引入云计算(cloud computing)的观念和技术。这是时代向计算语言学提出的新课题，我们期待刘根辉博士下一部计算语用学著作能对诸如此类的问题做出更精彩的研究。

2016 年 4 月 21 日于华中科技大学