

语言信息学

Linguistic Informatics

吕公礼◎著

中国社会科学出版社

语言信息学

Linguistic Informatics

中国社会科学出版社

图书在版编目(CIP)数据

语言信息学/吕公礼著. —北京: 中国社会科学出版社, 2016. 10
ISBN 978-7-5161-9390-7

I. ①语… II. ①吕… III. ①语言学—研究 IV. ①H0

中国版本图书馆 CIP 数据核字(2016)第 280447 号

出 版 人 赵剑英
责任编辑 熊 瑞
责任校对 周 昊
责任印制 戴 宽

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010-84083685
门 市 部 010-84029450
经 销 新华书店及其他书店

印刷装订 北京君升印刷有限公司
版 次 2016 年 10 月第 1 版
印 次 2016 年 10 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 21.25
插 页 2
字 数 303 千字
定 价 82.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社营销中心联系调换
电话: 010-84083683
版权所有 侵权必究

导 论

人类已经进入信息时代，这是一个无须特别强调的事实。而信息时代究竟具有什么特征、其本质是什么、带来什么重大问题和挑战、其演化发展趋势应如何认识，却是需要深入思考和探索的课题，其重大理论和现实意义怎样强调都不为过。的确，这个时代的任何事物都容易被贴上信息的标签，信息在当代学术文化中也极易成为崇拜、泛化，甚至炒作的概念，从而冲淡其应有的价值。显然，这种趋势发展到极致就是所谓的信息主义，这也不时引发有识之士的反思。（肖峰，2008）然而，环顾我们生活的方方面面，信息的确无时没有、无处不在。正如杨（Young，1987：1）所言，“无论喜欢与否，我们依赖信息而生存，信息如同呼吸的空气一样对我们生死攸关，像任何生成代谢所需的能量一样不可或缺”。环顾我们的世界，计算机、手机、网络、电邮、电商、短信、微博、微信等早已成为人们生活的一部分。而更需我们关注的是，现代信息和通信技术正在深刻影响着我们的认知世界的方式、时空意识、人际交往乃至思维方式，进而塑造出一种全新的社会文化生态。所谓信息化、数字化、虚拟化、快餐化、移动化、掌上化、微小化、碎片化，便是这种塑造的具体通俗表现。西方未来学家托夫勒（Toffler，1980）曾提出“浪潮理论”（the wave theory）来概括人类社会的演化发展。按照“浪潮理论”，第一次浪潮（the first wave）以农业革命为标志，第二次浪潮（the second wave）以工业革命为标志，而第三次浪潮（the third wave）的标志就是我们身处其中的信

息革命。无论从深度、广度、速度或规模来看，第三次浪潮带来的变化都是史无前例的，而这种变化的独特之处就在于，信息社会在很大程度上是对工业社会的逆转乃至颠覆。布朗和杜吉德（Brown & Duguid, 2000）曾用“endism”和“6—D”来概括上述变化，前者的基本意义可概括为“终结论”，具体包括媒体的终结、大学的终结、民族国家的终结等。“6—D”中的D代表英语中的一类语词前缀，如“de-”和“dis-”，意为“去……化”，构成的具体语词如“demassification”（去规模化）“decentralization”（去中心化）等。信息革命代表的第三次浪潮主要发生在人类社会的非物质层面，但这次浪潮同样对社会组织机构影响深远，“endism”和“6—D”可谓信息革命带来的具体变化。这种深刻巨大的变化并非没有原因。如果公路、铁路、航空构成了现代社会的物质流动通道，那么网络和移动通信则已成为现代生活、社会、政治、经济、文化、教育、科技、军事等活动的信息中枢和流动通道。如此无所不在，如此普遍深刻的影响，我们似乎找不到比信息更为恰当的词汇来描述和概括了。毫无疑问，现代社会的种种形态和问题大都可以归结为信息问题。信息绝非昙花一现的时髦词汇，它凝结和概括了现代社会的本质特征和基本生态，是具有普遍意义的核心范畴。

信息范畴凝结了深刻的理论内涵，包含着普遍的实践价值，承载着深远的时代意义，这是确信无疑的。信息作为重要研究课题，其重要价值和意义也无须特别强调。人们在这方面经常面临的首要问题在于，信息的思考和探究究竟始于何时，在学术文化中确定一个起点并非易事。2011年，科学历史名家詹姆斯·格雷克的*The Information: A History, a Theory, a Flood*（《信息简史》，人民邮电出版社2013年版）问世。翻阅这部畅销书我们会发现，信息观念其实同人类文明一样久远。如果以通信信息论的产生为起点算起，那么信息的科学探索已有半个多世纪的历史了。随着计算机科学、互联网和现代通信技术的发展，通信信息论代表的技术主义信息范式也得以确立，其影响随之扩展到众多人文社会科学领域。进入21世纪以来，信息和移动通信技术加速发展。如上所述，这些发展产生了诸多新的问题和挑战，技术主

义的信息范式日益显现出极大的局限性。在现有的信息理论中,大量新生信息文化现象越来越难以得到充分的科学解释,信息研究面临新的重大选择和深刻转变。在我们看来,就方法论而言,我们至少面临四种选择和转变。首先是描述与解释的选择。我们把信息社会的种种现象细细观察,然后加以归纳和概括,这是描述性研究,可达到知其然之目的。在描述的基础上,我们致力于探究现象背后的机制和原理,对其做出科学解释,可达到知其所以然之目的。显而易见,信息社会更需解释性研究,从描述性向解释性研究的转变,是21世纪信息研究更为迫切的方法论选择。实际上,描述和解释是科学研究的普遍方法论环节。毋庸置疑,科学研究需要现象的观察和描述,但不能停留在观察和描述,而要揭示现象背后的机制和原理,才能对现象做出真正的科学解释。苹果从树上掉下是现象,牛顿据说是从这一现象的观察开始,进而发现万有引力定律的。然而,牛顿仅靠继续观察更多苹果从树上掉下,要发现万有引力定律是不可能的。牛顿的独特之处在于,他善于超越观察到的现象,并对其发生的原因进行解释,进而发现解释任何微粒间引力的普遍定律。回到信息社会,信息爆炸、信息超载、虚拟空间、注意力经济、文化快餐化、生活碎片化、活动去中心化、机构终结说可谓信息时代产生的基本现象和问题。显然,这些现象和问题早已为人们反复观察和熟知。在方法论的选择上,继续观察已经反复观察和熟知的现象不再是主要任务。而这些现象为何发生,其发展的趋向如何,应如何解释和应对,才是更具科学价值和实践意义的课题,而这需要超越经验观察和简单描述,揭示现象背后的深层机制和普遍原理。这里我们思考最多的问题就是,信息时代对人类究竟意味着什么?对此,人们听到的欢呼声肯定更多。而在我们看来,人类进入信息社会,这固然是巨大进步,但也陷入了两难的困境。西方文化中有两个寓意相反的典故最能说明这种困境:一个是皮格马利翁(Pygmalion),另一个是弗兰肯斯坦(Frankenstein)。两者都是创造之物,但最终结果与创造者的期待截然不同,前者成就了创造者的期待,而后者却走向创造者期待的反面。在大自然面前,人类为了生存和发

展，曾经不断追求攫取和拥有，但过度攫取引发了大量生态灾难，其结果不是皮格马利翁式的，而更像是弗兰肯斯坦式的。在信息社会，人类以追求信息的无限拥有为目标，显然面临同样的挑战和困境。在描述互联网知识的未来时，西方信息哲学家弗洛里迪（Floridi, 1996）曾撰写了一篇文章，冠以这样的问句标题：“互联网：组织知识的未来是弗兰肯斯坦还是皮格马利翁？”（Internet: Which Future for Organized Knowledge, Frankenstein or Pygmalion?）这样的追问显然不是简单观察和描述，而是对上述困境的深刻反思。信息社会究竟怎样认识，民间和学界都有大量的讨论和说法。其中，“世界触手可及”、“人人都是通讯社，个个都有麦克风”等可谓当代中国最为经典的流行语，只是这些大都为欢呼赞美之词。在欢呼和赞美之余，稍作理性思考，我们也许会更多感受到其中的隐忧。人们自然要问，在网民数以亿计的时代，在人人沉溺于手机的社会，信息接收者岂不是要面对数以亿计的“通讯社”和“麦克风”？在信息时代，互联网和个人终端（个人计算机和手机）高度普及和发展，加之无数的网站和信息平台，形成了纵横交错相互链接的复杂网络信息世界。在这种信息世界中，数以亿计网民参与其中，每位网民个体都构成潜在的网络节点。这种节点显然不再是终端，而是中继和生成起点。节点上的主体不仅被动下载接收信息，而且通过评论、分享、上传、转发等形式，每时每地不断主动生成和传递新的信息。用通信信息论的术语来表达，这种节点既是信宿，也是信源。在信宿与信源的不断变换中，信息不断传递、生成、增长和膨胀。显而易见，官方媒体传播的信息量已然达到海量，加上来自数以亿计的“通讯社”和“麦克风”信息，任何信息受体要面对的信息量之大，显然已经不是超载和爆炸之类的词语所能描述。前面提到托夫勒（Toffler, 1980）的“浪潮说”，而关于信息革命掀起的浪潮，布朗和杜吉德（Brown & Duguid, 2000: 12）也有生动阐释：“如果说之前我们游泳时水太少，而现在游起来连保持漂浮（stay afloat）都很难做到。第三次浪潮正快速演化为海啸（tsunami）。”

毫无疑问，我们渴望更多的信息，信息也为人类所不懈追求和拥

有。亚里士多德说，人生来就有求知的愿望。在我们看来，求知就是获取信息，而人类认知世界获取信息是与生俱来的秉性。而我们要说的是，不光人类如此，一切有机物都以信息获取为本质特征。控制论的创始人维纳（Wiener, 1954: 17）说：“有效地生存意味着拥有足够的信息来生存。”实际上，为了表达这样的认识，乔治·米勒（George A. Miller, 1983）还发明一个英语词语“informavore”。这个单词中的-vore 源自拉丁语，意为“以……为食的动物”。例如，“carnivore”意为“食肉动物”，而“informavore”就可解释为“以信息为食的动物”。乔治·米勒（George A. Miller, 1983）对此提供了如下背景说明：“正如身体靠摄取负熵生存一样，心智靠摄取信息而存在。在最为普遍意义上，所有高等有机物是摄取信息的动物（informavore）。”人类是信息动物，追求和拥有更多的信息也是人类的本质特征，这是毫无疑问的。而问题的关键在于，当信息浪潮海啸般涌来时，人类是否有足够认知能力来接收和处理，以便获取真正需要的信息，进而迅速做出有效决策？对于这类问题，有识之士常常给出更为负面的回答。所谓信息趋向超载甚至爆炸，生活日益碎片化，文化不断快餐化，社会机构去中心化之类的说法，大多带有负面色彩。我们在此要做的自然不是继续列举这些现象，而是要科学地解释：信息为何超载和爆炸？文化为何快餐化？生活为何碎片化？诸如此类的变化对人类带来何种挑战？显然，这些都涉及深层动因和机制，需要确立普遍原理，才能做出科学解释。我们还要进一步强调的是，求知就是获取信息，而人类获取信息，归根结底是追求真实信息。然而，当信息浪潮海啸般涌来时，我们是否有足够能力迅速加以甄别？我们究竟有多少自信，认为我们能够在海量信息中辨别真伪？真实离我们更近还是更远了？我们渴望更多的信息，因为我们渴望真理。真理的确需要更多信息支持，但越来越多的事实也表明，当我们拥有过多信息时，真理却反而可能被淹没。我们更不应该忘记的是，真实原本并不一定裸露在外。在更多情况下，真实甚至被有意遮蔽和掩盖，谎言和虚假信息也常会大行其道。在植物和动物世界，伪装是重要的生存之道，捕食者常常要面对大量

虚假信息。高级生物由低级生物演化而来，这一规律自然会延续下来。最后还要强调的是，我们总是渴望和追求更多的信息，但我们不能不认识到，信息并非只有唯一的呈现方式。相同的信息可能以不同形态呈现出来，就是同一真实信息也会有多种版本，经由语言表达和传递的信息更是如此。显而易见，我们的确比以往拥有更为充裕的信息，我们也常常引以为豪。但当我们认识到，这些信息不少是同一信息的多样呈现，就很难说那是我们所需要的信息量。信息超载和爆炸都有信息充裕之意，但如果充裕的信息并非真实信息，而是浮泛虚假信息，就很难说是我们所需要的信息量。我们的确正陷入各种信息悖论。凡此种种现象，人们早有感受，也早有思考和探索。但这些现象背后有无某种深层机制，如果有又是什么，我们已经探索很多，但仍有大量的疑问和谜团等待解开。要消除疑问、解开谜团，我们需要对现象加以观察描述，但我们更需要透过现象，发现现象背后的机制和原理，从而对观察到的现象做出科学解释。机制和原理虽然超越现象，却有烛幽发隐、以简驭繁之力，其理论价值和实践意义也许更为深远。我们前面提到万有引力的发现，而就复杂性而言，信息问题无疑远超引力问题，更需解释性研究。

人类生存和发展是物质和能量活动，但人类活动不限于物质能量活动，而且上升为信息活动。信息活动是人类生存和发展的高级活动。在此意义上，信息活动植根于人类活动的深厚广阔背景之中。这种活动固然是现时的存在，呈现为共时状态。所谓共时就是信息社会的现状，而现状是忽略其随时间动态演变的一种抽象规定。然而，信息社会的共时状态要得到充分解释，最终要回归人类、动物乃至有机物漫长的历时演化，从中探寻合理的解释。信息活动是人类活动的高级形态，信息社会出现的种种现象和问题，其机制和原理无疑需要在人类活动的历时演化中来探索 and 发现。由此我们进入了信息研究的第二个方法论转变，那就是共时到历时的转变。从历时演化来探索信息，可以说是信息研究的一种普遍选择。柯伦（Coren, 1998）出了一本著作，题为 *The Evolutionary Trajectory: the Growth of Information in the History and Future of the Earth*

(《地球信息增长：历史与未来》)。正如这本书标题所示，作者探索信息增长，而增长时段覆盖了地球、地球生命及人类文明产生和演化的全部过程。在本书研究中，我们对演化自然有自己独特的理解，具体将在第一章详细展开探索。

概括起来看，回归人类活动，并在其历时演化的链条中探寻信息的本质，进而探究信息社会种种现象的发生机制和原理，实际上是回归信息活动的具体完整状态。这种回归从根本上讲就是回归人类信息活动的自然本体状态，而这种回归意味着超越通信信息论代表的技术主义范式的局限，超越技术主义范式塑造的非自然信息社会文化形态。显然，这种回归包含着第三个重要方法论选择，那就是从信息的技术形态向人文形态的转变。说起技术与人文的对举，我们不能不提到信息论的另一位先驱维纳。在其控制论的创建过程中，维纳始终把机器和包括人在内的动物等有机体一并讨论，试图对其控制和通信进行统一概括。其中，人的控制和通信一直被视为控制论的主要研究对象。维纳（1961）出版了 *Cybernetics, or Control and Communication in Animal and the Machine*（《控制论或关于在动物和机器中的控制和通信》）一书，可谓这种统一性思想的集大成。其实，维纳在1954年出版的 *The Human Use of Human Beings*（《人有人的用处》）一书中，关于人类的控制和通信早有大量通俗论述。继上面的相关论述（Wiener, 1954: 17）外，维纳还说：“有效地生存意味着拥有足够的信息来生存。因此，通信和控制属于人类内在生活的核心部分，甚至如同属于人类在社会中的生活一样。”在维纳的心目中，人也是的确是在具体完整意义上理解的，如人的行为、人的目的、人的生物构成、人的社会属性、人的心智和思维、人的大脑神经系统和人类自然语言等，都在通信和控制的研究范围之内。不过，维纳毕竟是数学家，关注的主要是工程控制问题。他关于控制和通信的思想无疑首先源于机器，人的控制和通信在很大程度上是机器控制的类比和延伸，这样来类比显然有把人类还原为机器之嫌。维纳的根本局限在于，他与香农一样，同样在技术意义上界定和刻画信息。在科学形态意义上，香农和维纳是通信信息论的

奠基者，可概括为“Shannon-Wiener”信息范式。为了讨论方便，我们在之后的论述中统称为经典信息论。这里我们要指出的是，这种范式本质上是技术主义的，因而是技术主义信息范式。我们在本书研究中强调转向人文，是要超越技术主义信息范式，真正从人的自然本体形态出发，在人类信息活动的具体完整性中认识信息，从中建构新的信息范式。在我们看来，人类信息活动的具体完整形态在自然语言信息交流中得到最为完整的体现。在此意义上，转向人文实际上就是转向自然语言。本书冠以“语言信息学”的标题，其根本动因和意义就在于此。为此，我们确立了一个更为概括的命题：信息活动是人类活动的高级形态，而语言信息交流活动是人类迄今最高级的信息活动形态，语言与信息的内在联系在此形成。不过，这样的命题和概括自然会引来不少疑问。观察和探究信息的视角很多，如物理、数学、化学、计算、工程、生物、社会、心理、认知、逻辑、传播及哲学等（闫学衫，2007）。信息研究转向语言，理论上究竟有何必然性？有何可靠的思想依据？有何独特优势和价值？面对这些问题，我们首先想到的答案是，任何讨论、研究、思想和理论都不能离开语言来进行，这是一个再也浅显不过的道理。信息和哲学活动凌驾于多数科学活动之上，可谓真正的元思想活动，但离开自然语言，任何完整、深刻和系统的信息和哲学思考、研究和阐释都无法进行。在很大程度上，哲学是在语言中展开、建构、演化和呈现的，哲学转向语言探寻发展之道，就是对自身本质的最好诠释。可以说，哲学是最不应忽视语言的一门学问了。哲学如此，信息更无例外可言。当然，本书并不打算以浅显的道理来立论。我们极力主张信息的语言取向，其根本依据在于，语言和信息原本就内在地联系在一起。纵观语言和信息研究思想传统，语言和信息历来相生相伴。这一点可从詹姆斯·格雷克的 *The Information: A History, a Theory, a Flood* 《信息简史》中找到大量证据。这是一部信息历史专论，但书中充满了关于语言的论述、语言的故事甚至语言的起源。在科学信息意义上看，自经典信息论诞生起，语言与信息就联系在一起了。在香农（1948）发表的那篇经典论文中，语言信息量是

探讨的重要问题。关于通信和语言的关系，控制论的创始人维纳则有明确的论述。维纳（1954：74）在论及通信与语言的关系时指出：“自然而然，没有一种通信理论能够回避语言的讨论。事实上，从某种意义上来说，语言就是通信自身的别称，也是用以描述通信借以发生的编码的名称。”按英文本义解释，通信（communication）就是信息传递。上面提到柯伦（1998）的那本书，其中的“后生物进化”和“信息”两个章节都对语言有专门论述。在信息论的先驱和理论大家的思想中，语言都被视为核心问题，而在信息科学、认知科学和语言学高度发展和融合的时代，语言和语言学对信息研究日趋重要，其理论和实践价值更不应该被轻视甚至被完全忽视。

不难理解，由于时代的局限，语言与信息的关联在早期信息论中大多较为肤浅。随着现代认知科学的发展，语言不再被视为思想的外壳，而是大脑信息变换的有机组成，语言的语音形式只是其外化形态而已，语言与信息完全有机融合，语言是人类认知和信息过程不可分离的部分。在现代脑神经科学中，众多脑成像技术（如 PET, MRI, fMRI, MEG, NIRS, EEG 等）被用于大脑探测，极大地推动了人类心智研究。然而，心智并非大脑神经元和系统的简单加和，而是其涌现特征（吕公礼，2011），无法简单还原为神经的物质能量状态。脑科学技术探测到的只是大脑神经系统的物质能量状态，而不是心智本身的机制和结构。实际上，解剖大脑是看不到心智的。心智无法离开大脑神经系统而存在，但心智是非物质形态，而大脑神经系统是物质能量形态，两者之间存在着巨大的鸿沟。怎样跨越这种鸿沟，无疑是认知科学面临的巨大挑战。如果说心智和认知就是信息过程，那么信息就是非物质的存在，因而不是具体可见的研究对象。在这个意义上，如何“窥视”和把握信息过程，也是信息研究面临的根本问题。在我们看来，回归自然语言是解决这一问题的必然选择。语言是心灵的窗口，可谓这一方面最为朴素的认识。在认知科学高度发展的 21 世纪，我们比以往任何时候都更为深刻地感受到，这一朴素认识其实包含着深刻的科学道理。我们可以毫不夸张地说，探索并解开心智之谜，语言也许是

最真实、最完备、最重要的天然之窗，也是认识人类信息活动具体完整性的最为天然的通道。在人类科学发展的长久历史中，从大脑认识语言，一直是传统的思路和方法。随着认知语言学和心理学的发展，这种传统思路正在改变，从语言来洞察大脑和心智成为新的重要选择。我们的相关研究（吕公礼，2012）表明，认知心理学家早期从认知动词探索心理和思维。近期，认知心理学家则从跨语言比较来发现认知动词的差异，从中探寻不同文化主体的心理差异。在认知语言学传统中，语义学家威尔兹彼卡（Wierzbicka, 1972, 1992, 1996）从普遍语义基元（semantic primitives）来刻画人类思维，戈达德（Goddard, 2003）进一步探索了思维跨语言表现的共性和变异，福特苏（Fortesue, 2001）则通过心智活动言语表达的跨语言比较，探索心智的民俗模式（folk models of mind）。语言类型学家更是试图构建所谓概念空间（conceptual space）和语义图（semantic maps）。克罗夫特（Croft, 2001: 364）甚至认为，语义图表现的是人类的心智地图，它从世界语言的事实中解读心智，其方法提供的信息为迄今最先进的大脑扫描技术所无法做到。本书从语言探索信息，建构语言信息论，其根本理论依据和独特科学意义就在于此。

自然，同样由于时代的局限，也由于范式的惯性效应，经典信息论诞生之初，多被直接用于语言，特别是语音、字母或字形信息的概率统计刻画。因此，经典信息论所谓语言信息，实际上是语言形式层面的数学刻画。这样的界定和研究其实并不奇怪。我们只需回视香农（1948）发表的那篇经典论文的标题 *A Mathematical Theory of Communication*（《通信的数字理论》），就已知其大略了。这是一篇关于通信的数学理论，研究的是纯粹的通信工程问题，而香农在文中也明确指出，消息的语义方面与通信工程问题没有关联（These semantic aspects of communication are irrelevant to the engineering problem）。更为关键的问题是，经典信息论以热力学的熵（entropy）来界定信息，定义的显然不是信息本体，而是对不确定性的数量类比。正如麦凯（Mackay, 1969: 56）所言，这种界定“把信息与信息内容相混淆——即把事物与事物的度量相混淆”（“a con-

fusion of the concept of information with that of information-content—the confusion of a thing with the measure of a thing”)。这种信息界定用于语言信息研究,只能限于语音和文字概率特征的刻画,对语言的语义内容信息难有作为。经典信息论的局限性表明,建构信息新范式,从信息形式转向内容是又一重要方法论选择,也是本书研究要做出的第四个选择。关于经典信息论的局限性,当时圈内清醒者不乏其人。信息论的另一位先驱韦弗(Weaver, 1948: 27)就曾明确提醒道:“在该理论(通信信息论)中,信息一词是在特殊意义上使用的,不应与日常用法混为一谈,尤其不应与意义混为一谈。”这表明,通信信息论的信息与意义并没有关联,自然也与语言的意义内容无关。实际上,经典信息论之后也的确被扩展到人文科学领域,但后来大多证明难以走通。我们知道,20世纪前期,信息社会的步伐迅速加快,而经典信息论的诞生恰逢其时,为信息社会的发展其提供了理论范式。之后,经典信息论日益成为显学,加之计算机所展现的强大技术力量,无论是香农本人的说明,还是韦弗的提醒,显然都淹没在之后出现的信息革命浪潮之中。科学新范式和范畴诞生之初,跟风甚至狂热是难免的,结果往往会掀起巨大和持久的思想学术浪潮。但风平浪静之后总会出现反思,经典信息论也不例外。20世纪60年代末,心理学中曾发生了一场“信息革命”,后来林赛还致力于“用信息概念系统地改造传统心理学”(闫学衫, 2007)。然而,多年以后,心理学界开始出现反思。米勒(2003)回顾了认知革命的发展历史,其中提到自己早年在信息论应用扩展中遭遇的困境。他曾试图把信息论运用于心理学,却无法超越香农对书面文本字母序列的分析。米勒认识到,信息度量以概率为基础,而无论是概率还是相应的对数,都无法解释语言的心理过程。

现代语言学和符号学早已形成了一个基本认识:语言是形式与内容的统一,是语形与语义的统一,语言靠两者的统一表征和传递信息。对于语言形式背后的语义内容,任何信息研究都不可能永远视而不见。经典信息论之后,逻辑学家和语言哲学家卡尔纳普(Carnap, 1942)和巴希勒和卡尔纳普(Bar-Hillel & Carnap, 1953)提出了语义信息论,

显然是信息研究反思之后的发展。关于语义信息论产生的背景，荷兰逻辑学家欣蒂卡（Hintikka, 1970）的相关论述颇能说明问题。逻辑学家和语言哲学家为何尝试超越统计信息论并另辟蹊径呢？欣蒂卡（Hintikka, 1970: 3）作出了这样的解释：“语义信息论产生的一个根本动因在于，通信信息论所依托的统计信息对‘最重要意义上的信息（information in the most important sense of the word）’研究没有多大作为，甚至可以说没有多少联系。”按照欣蒂卡的进一步解释，所谓“最重要意义上的信息”就是“有意义的句子和其他类似符号组合对能够理解这些句子和组合的人所传递的任何（内容）”。显然，欣蒂卡的论述包含了语言信息交流的基本因素，他所说的“最重要意义上的信息”就是自然语言传递的信息内容。语义信息论试图超越语言形式层面，切入形式背后的语义信息内容，其积极意义值得肯定。但仔细审视我们会发现，语义信息论其实并未彻底摆脱技术主义的羁绊。从产生背景来看，语义信息论可以追溯到波普尔（Popper, 1934/1959）科学哲学论。我们知道，科学哲学关注的是科学理论中证据对假设的确证度问题，而确证度用概率来加以刻画。就概率逻辑而言，语义信息论与经典信息论可谓一脉相承。这其实也并不奇怪，巴希勒和卡尔纳普都出身哲学和逻辑学，数理逻辑和概率逻辑为当时普遍推崇的思维工具，他们的思想和研究不可能不受影响。在哲学层面，当时分析哲学和逻辑实证主义兴盛，数理逻辑和概率逻辑在此过程中形成。这类逻辑形态大都以忽略符号意义内容来建立，属于纯粹形式化的人工符号系统。对于自然语言语义内容，形式化系统的先天不足是显而易见的。现代哲学的“语言转向”催生了形式逻辑系统，并试图用它改造甚至取代自然语言，但之后日常语言哲学兴起，表明改造和取代难以成功，忽略意义内容的纯形式化之路并不可取。哲学的曲折发展表明，语义信息论对于“最重要意义上的信息”同样难有作为。

实际上，语义信息论的探索并不限于科学哲学和逻辑学。英国物理学家和信息学家麦凯（Mackay, 1969）曾出版 *Information, Mechanism and Meaning*（《信息的机制和意义》）一书，对香农和韦弗为代表的经

典信息论进行了全面评述,并试图建立一种包含意义的信息理论,而作者的主要动因是“关注人类大脑中表征并相互交流的信息,而不是基础命题逻辑模式中界定的信息”(preoccupation with information as represented and utilized in the brain and exchanged between human beings, rather than as formalized in logical pattern of elementary propositions)。在经典信息论之外,信息研究究竟何去何从,麦凯的想法显然与欣蒂卡的理解颇有相通之处,只是因其学术思想背景所限,他们都缺少真正的语言学思维,其信息界定和理论并未走出多远。20世纪80年代产生的情景语义学是语义信息研究的另一理论。情景语义学的产生背景是真值条件语义学模型论的局限性,却有明确的信息研究动因,被称为“体现‘信息’精神的意义理论”(邹崇理,2002:190)。按照德夫林(Devlin,1991:14)的陈述,情景语义学以认知主体间(cognitive agents)的信息流(information flow)为研究对象。这表明,关于通信信息论的技术倾向的局限,情景语义学者不乏清醒认识。不过,德夫林所说的信息流实际上只限于主体对环境信息的感知和表征,其局限性也显而易见。例如,主体看到树桩,根据年轮与树木生长的关联,获得树木生长年龄信息。从认知科学和语言学目前的认识水平来看,情景语义学所说的人类认知和信息流动显然过于简单。主体观察树桩,这作为理论起点倒也方便,讨论和刻画也较为容易。但这样的认识显然大大低估了人类信息活动的复杂性。更重要的局限在于,德夫林(Devlin,1991:2)声称要建立“信息的数学理论或模式”(a mathematical theory/model of information),这与香农(1949)发表的经典论文“通信的数学理论”(A Mathematical Theory of Communication)比较,显然只有一字之差。在英文中,“communication”一词的本义就是“the transmission or exchange of information”(信息传递或交换)。这样看来,两者的意义几乎相同。虽然致力于建立新的信息理论,比之经典信息论代表的数学和技术范式,情景语义学并未实现本质性超越。

上述几种语义信息研究只是学术探索和思想模式,要真正全面了解相关研究态势,学术机制(如研究机构和刊物等)也是一个重要参

照。2001年，作者赴美学习，也曾对国外的语言信息研究抱有很高期待。当时看到，国际著名学府斯坦福大学设有“语言与信息研究中心”（Center for the Study of Language and Information，简称 CSLI），心中曾经涌现不小的兴奋。语言与信息作为专业领域放在一起，设立专门的研究机构，并定期出版系列研究成果，至少说明这个课题和领域在学科建制上是成立的。另外，一些重要国际专业刊物如《语言、逻辑及信息学刊》（Journal of Language, Logic and Information）也说明该领域的认可程度。作者于是不顾囊中羞涩，用手里有限的美元买了几本 CSLI 出版的系列书籍。然而，粗略翻阅就发现，其中充斥着大量的形式符号和数学公式。显然，CSLI 的研究技术主义的色彩依然十分浓厚，而作者当时的失望之情是不难想象的。作者从事语言学教学和研究多年，即使凭直觉来判断，这些研究理解的语言信息也与作者心目中沉淀下来的认识相去甚远。之后作者越来越强烈地意识到，语言信息研究现有的路子难有作为，另辟蹊径是不得不做出的选择。

毫无疑问，回归人类信息活动自然本体，转向自然语言是必然选择，而要建构科学的语言信息理论，显然不能停留在语言的初步了解，而是要真正回归语言学理论，从中寻找合理的思路 and 模式。这当然不是说，语言学与信息研究从来没有关联。语言从人类信息活动中演化而来，语言是人类信息活动的高级形态，语言与信息天然地联系在一起，语言活动包含着人类信息活动的物质能量、生理、心理、社会、文化经验的全部，因而代表人类信息活动的具体完整形态。语言学是语言的科学研究，因而在很大程度上也是信息的科学研究。不过，这里我们仍然要认识到，现代语言学同样产生于现代西方思想文化传统，其主流研究也深受形式主义技术传统的影响。因此，回归语言学首先需要对语言学自身发展进行一番梳理，找到合理的模式并加以必要重构。实际上，语言学与信息科学早期结缘，可以追溯到发生于 20 世纪 50 年代的认知革命。这场革命与语言学结缘，首先因为它由语言学家乔姆斯基发起，而进一步的发展则与计算机科学和人工智能相连，包括自然语言生成和理解、机器翻译、人机互动及计算机为中介的交流