

关于控制论与动态语言知识更新的思考¹

张 普

北京语言文化大学

E-mail:zhangpu@blcu.edu.cn

摘要: 本文从控制论的角度重新审视动态语言知识更新问题。主要的视点有:

- 1、关于“信息和通讯作为组织化机制”;
- 2、关于“通讯和控制的年代”;
- 3、关于“牛顿时间和柏格森时间”;
- 4、关于“反馈”和“稳态”;
- 5、关于“种族信息量的测定”;
- 6、关于“本书的教训之一”和“反内稳定的因素”;
- 7、关于“学习”和“自生殖机”;
- 8、关于“白箱”和“黑箱”;
- 9、关于“可信的程度只能到达头几位数字”

关键词: 语言信息处理; 语言知识动态更新; 控制论

On Cybernetics and Dynamic Updating of Language Knowledge

Zhang Pu

Beijing Language and Culture University

Abstract: This paper reviews the problem of dynamic updating of language knowledge from the point of view of Cybernetics. The main points are as follows:

1. Information and communication as organizational mechanisms;
2. The era of communication and control;
3. Newtonian Time and Bergsonian Time;
4. Feedback and Homeostasis;
5. Measurement of amount of information of a race;

¹ 本文承北京语言文化大学对外汉语研究中心重点项目资助。

6. One of lessons learned from this book and anti-homeostatic factors;
7. Learning and self-reproducing machine;
8. “White box” and “black box”;
9. Creditability can only be reached at the first few digits.

Keywords: language information processing; dynamic updating of language knowledge; Cybernetics

引 言

我们将“控制论”与动态语言知识更新联系在一起，并非要哗众取宠。

我们较为仔细地阅读了 N. 维纳的《控制论（或关于在动物和机器中控制和通讯的科学）》一书，认真地将这两个概念放在一起，并进行思考。

事情的起因是陈原先生的《语言与控制论札记》²，该文使我们注意到语言学与控制论的紧密联系。他说：“维纳作为语言学家的儿子——他的父亲是语言学家——论述作为信息系统的语言问题，值得注意。”而维纳本人在出版《控制论》和《人有人的用途》时都写明：“献给我的父亲莱奥·维纳——曾任哈佛大学斯拉夫语言学教授，我最亲密的诤友和挚爱的论敌。”

到 1982 年更有弗朗克在《控制论语言学》中明确地提出“语际语言学可以成为控制论的一个分支。”他说：“语际语言学的主题因此属于控制论的主题——更准确地说，属于语言控制论的主题。”³

我们并不一定是“语言控制论”或“控制论语言学”的拥护者或者反对者。我们只是希望实事求是地探求控制论对动态语言知识更新的理论支持或方法论支持，我们相信控制论的科学性，相信控制论对于“动物和机器”在宏观上的控制和把握。而我们的动态语言知识更新的重要机制之一，正是要用机器来模拟人的语感，为机器建立语感模型，关涉“动物和机器”二者。

1. 关于“信息和通讯作为组织化机制”

维纳说：“至于说到社会学和人类学，十分明显，信息和通讯作为组织化机制不但对于个体是重要的而且对于集体也是重要的。”⁴

我们认为：语言交际是人类社会的重要的“组织化机制”，语言正是人类社会的“信息和通讯”的工具。语言或者说言语既是个体的行为，也是社会的行为。语言交际既受社会的约定俗成的行为规范的制约，也受个人表达的创造意愿的激励。语言对于人类的个体和集体都是十分重要的

语言交际既受社会的约定俗成的行为规范的制约，也受个人表达的创造意愿的激励。

² 参见陈原《语言学论著》卷三 729 页，辽宁教育出版社。

³ 参见 Frank h.g. <Sprachkybernetik>(1982)，转引自陈原《语言学论著》。

⁴ 参见维纳《控制论》（第二版）19 页，科学技术出版社，1963 年

语感也有公众语感和个人语感之分，语感就集体而言是一个民族有始以来的语言经验的积累，就个体而言是一个人一生的语言经验的积累。就时间序而言，语感不是一个定数，而是一个变数。社会在发展，语言在变化，语感也在变化。所以，第三代语料库向动态语料库发展、收集一个民族积累的语言经验是理所当然的，只是受计算机软硬件发展的限制，前人不具备制作的条件。

所谓“语言经验积累”的结果，就是每一个体都在“信息和通讯”的组织化行为机制中，感受到了作为信息和通讯载体的语言符号的流通度，知道怎样表达是大家接受或认可的表达，也知道怎样在大家都认可的表达中夹进个人的语言创造。组织化机制对于集体的语言交际而言是形成公众语感，对于个体而言是影响个人语感中的共同语感。所以维纳说“信息和通讯”作为人类社会的一种“组织化机制”，是十分重要的，“不但对于个体是重要的而且对于集体也是重要的。”⁵

我们所谓的“公众语感”，就是一种“集体语感”，是大家或者绝大多数人的语感。我们把绝大多数人的语感假定（目前，这种“假定”仅仅是“猜测”，将来也许还是“猜测”。但是，我们的出发点是：说客观流通度是绝大多数人的语感，总比说客观流通度是一个人或几个人的语感，风险要小一点）为最接近语言符号的客观流通度。这种语感是约定俗成的。

我们以前表述过：“我们是否可以认为：说到底理性主义也是经验主义，是语言学家的个人语感经验，这种个人语感由于其经历、素养等的影响，有时候对于语言的描写，与客观事实很接近，或者说是与公众语感很接近。而经验主义是通过文本的抽样来尽量公平地表述公众语感，或者说经验主义描述的是公众的经验，理性主义描述的是权威性个人（或者一些权威性个人）的经验。”⁶

我们的整个研究工作主要就是寻找客观流通度的最大近似值，然后寻找流通度与公众语感的最大近似值。所谓语感模型或者说模拟语感就是建立在这种最大近似的基础上的。

2. 关于“通讯和控制的年代”

维纳说：“如果十七世纪和十八世纪初叶是钟表的年代，十八世纪末叶和十九世纪是蒸汽机的年代，那么现在就是通讯和控制的年代。”⁷

“我们研究着这样一种自动机，它不仅通过能量流动和新陈代谢，而且通过印象和传入消息的流动以及由传出消息引起的动作的流动和外界有效地联系起来。”⁸

“在自动机运转的时候，它的操作规则本身会按照过去通过接受器的数据的情况而多少发生变化，这就象是学习的过程。”⁹

“我们决定把这个关于既是机器中又是动物中的控制和通讯理论的整个领域叫作

⁵ 参见维纳《控制论》（第二版）19页，科学技术出版社，1963年

⁶ 参见张普《关于第三代大规模真实文本语料库的几点理论思考》，载《自然科学基金重点项目结题报告》（内部）清华大学1998年。

⁷ 参见维纳《控制论》（第二版）39页，科学技术出版社，1963年

⁸ 参见维纳《控制论》（第二版）42页，科学技术出版社，1963年

⁹ 参见维纳《控制论》（第二版）43页，科学技术出版社，1963年

Cybernetics(控制论), 这个字我们是从希腊字 κυβερνητης 或‘掌舵人’变来的。……我们也想提到这个事实: 船舶的操舵机的确是反馈机构的一种最早而且最发达的形式。”¹⁰

我们认为: 虽然今天已经有了信息时代和网络时代的说法, 但是原则上并没有超出维纳的“通讯和控制的时代”的命题, 控制论在今天仍具有指导意义, 只是通讯的技术更加先进, 信息传播的速度、领域、方式前所未有。

网络帮助了信息传播, 电脑帮助了人脑做信息处理。网络和电脑也是一种“具有为善和作恶的巨大可能性”的“技术发展”, 如果说个人电脑是增强了人类个体的“信息和通讯”的“组织化机制”的话, 网络就是增强了人类集体的“信息和通讯”的“组织化机制”。因此, 电脑时代、信息时代、网络时代更需要控制论。

信息处理, 就是因为信息需要处理。不处理就将出现信息爆炸、信息泛滥、信息失控。今天信息已经到了“多多不益善”的时代, 出现了“过犹不及”。所以要研究信息加工、信息提取、信息筛选等等。过去可以认为“信息的积累就等于知识”, 今天信息的超量积累, 对于个体来说却不等于知识, 而是信息垃圾或垃圾信息, 只有经过加工、提取、筛选后的信息才是知识或者说是有用的知识。

对大规模真实文本的信息加工、筛选和提取需要动态语言知识更新的技术支持, 我们曾经提出过一个“基于社会传媒的网络语言规划模型”, 模型的上下两大部分是: “语言自动控制体系”和“语言自动学习体系”。在这两大体系之间的是“语言知识自动反馈体系”, 另外还有在社会传媒之中的“主页和文本的自动检测体系”。这四大体系构成了一个学习—反馈—控制—检测模型, 可称为L F C C模型, 简称语言的学习—控制模型, 即L C模型。¹¹

显然, 我们追求的目标是一个网络时代的或者说是信息时代的语言信息处理“自动机”, 它具有学习、反馈、控制语言的能力, 不断的“新陈代谢”, 把我们同社会交际这个语言“外界有效地联系起来”。我们不认为这是一件轻而易举的事情, 维纳计算过不等于我们计算过, 更不等于语言的学习——控制模型即L C模型可以计算成功。但是我们认为这件事符合控制论的基本思想, 我们需要并研究的是一个网络时代的“反馈机构”, 是网络语言的“操舵机”。

3. 关于“牛顿时间和柏格森时间”

维纳说: “由牛顿时间可逆到吉布斯的时间不可逆这个转变是有哲学方面的反响的。柏格森曾经强调指出物理学的时间和进化论与生物学的时间的不同: 前者是可逆的, 其中没有什么新事物出现; 后者是不可逆的, 其中总是发生着新奇的事物。”¹²

“因此, 近代自动机跟生命体一样, 都存在于柏格森的时间中。按照柏格森的观点, 我们没有什么理由认为生命活动的基本方式一定和模拟生命体的人造自动机有所不同。”¹³

我们恰恰把语言看作类似是有生命的东西。语言有活语言和死语言, 语言有自己的新

¹⁰ 参见维纳《控制论》(第二版) 11—12页, 科学技术出版社, 1963年

¹¹ 参见张普《关于网络时代语言规划的思考》, 载《语言研究》1999年3期。

¹² 参见维纳《控制论》(第二版) 38页, 科学技术出版社, 1963年

¹³ 参见维纳《控制论》(第二版) 44页, 科学技术出版社, 1963年

陈代谢，语言的各个成分都会经过诞生、成长、衰老、死亡的历程，语言本身也是不断进化的，没有一成不变的语言。语言的活动方式符合生命活动的基本方式，语言的时间应该是柏格森的进化论与生物学的的时间。

我们恰恰也把第三代语料库视为类似是有生命的东西。我们认为第三代语料库应该是动态语料库，是历时语料库，是活语料库。我们非常赞同并曾经引用过徐通锵先生的观点，他说：“时间观是语言研究方法论的一个重要基础，要改进语言理论的研究，如仍旧保持索绪尔的时间观，那是不会有什么成效的。”“索绪尔的语言理论就是建立在他的共时时间观的基础上的。”¹⁴我们还认为：就语言的发展而言，历时中包含有共时，共时中包含有历时。

15

我们恰恰也把基于第三代动态语料库获取的语言知识作为类似有生命的东西。我们曾形象地把第三代动态语料库称为“水”，基于这种语料库获取的语言知识称为“鱼”，我们说“活水养活鱼”。因此，我们也十分赞同台湾简立峰先生提出的“活的字典”的概念，这不是我们一般意义上所说的形容某人知识渊博的“活字典”，而是进化论与生物学时间概念上的字典。

我们在1999年提出的“LC语言学习—控制模型”正是一种符合维纳所说的现代自动机理论的模型，是一种把人和语言交际这个外界有效地联系起来的模型。信息的传递与控制是这一代自动机的理论基础。

这种模型正是符合柏格森时间的生物进化论的模型。进化的过程是从简单走向复杂，主要研究语言的基因、遗传、变异、偶然变异和突变等等。统计理论、概率、测度等等是重要的研究手段和研究方法。维纳说：“不够清楚的地方也许就在于灵敏自动机的理论是一个统计的理论，通讯工程的机器，根据单独一次输入而产生的动作是不会使人感到兴趣的。这种机器如果要能充分发挥作用，它就必须对全部输入都作出令人满意的动作。这也就是说，对一类从统计上预期要收到的输入作出统计上令人满意的动作。”这也正是LC语言学习—控制模型追求的目标，该模型就是不希望根据“单独一次输入”的语料进行统计分析，而是要不断动态输入语料，对语言的遗传和变异进行历时的观察。

4. 关于“反馈”和“稳态”

维纳：“我们不要忘记反馈的原理在生理学上还有一个重要的应用。在很多场合，一定形式的反馈不仅是生理现象中常见的例子，而且它对生命的延续也是绝对必要的，我们在所谓稳态（homeostasis）情形中可以看到这点。”¹⁶

“高级动物的生命，特别是健康的生命，能够延续下去的条件是很严格的。体温只要有摄氏半度的变化，一般就是疾病的征候；如果有长时间的五度变化，就不能保持生命。血液的渗透压和它的氢离子浓度必须保持在严格限度内。体内的废物在浓度达到有毒以前必须排泄出去。此外，白血球和抵抗感染的化学防疫作用必须保持适当的水平；心搏率和

¹⁴ 参见徐通锵《语言论》P 64、P 73，东北师范大学出版社1997年。

¹⁵ 这一观点我们还会另文论述。

¹⁶ 参见维纳《控制论》（第二版）115页，科学技术出版社，1963年。

血压必须既不太高也不太低；性生殖周期必须符合种族的生殖需要；钙代谢必须既不使我们的骨质松化，也不使我们的骨质钙化，等等。一句话，我们内部组织中必须是一个由恒温器、氢离子浓度自动控制器、调速器等等构成的系统，它相当于一个巨大的化学工厂。我们把这些总起来叫作稳态机构。”¹⁷

“任何一本关于控制学的教程，都应当透彻详尽地讨论稳态的过程。”

我们很赞同并引用过陈原先生的论述：“约定俗成是语言文字最惯用的‘规律’，语言文字在使用过程中发生变异，自动调节和人工调节，达到一种有序的稳态。”¹⁸“非规范化同规范化是矛盾的统一体。矛盾斗争的结果，达到一个‘内稳态’（homeostasis），就是自我平衡。‘内稳态’最初是从生理学上提出来的，有人译作‘稳态’……‘内稳态’的学说后来在控制论、信息论上得到了广泛的应用。维纳在他的控制论里提出了两个重要概念，一个是‘反馈’，一个就是‘内稳态’。”¹⁹

我们也曾经有过关于“内稳态”的提法，在《关于网络时代的语言规范化》一文中，我们曾说：“在控制和检测这两个体系方面，实际上要伴随一定的政府行为，这是与以往的语言规范化相同的做法，但是政府行为还要依赖于一些时常更新的应用软件（比如“语言巡警”或“语言巡逻兵”、“语言清道夫”、“语言教师”等），以维护网络语言的“内稳态”和健康发展，这是与以往的语言规范化不同的做法。”²⁰

我在同一篇文章中还说过：“网络语言规划的研究将涉及理论语言学、社会语言学、心理语言学、认知语言学等，当然还要涉及计算机技术、网络技术、通讯技术、多媒体技术等，甚至还应该涉及系统论、控制论。限于笔者功力，本文着重从人文科学的角度而且主要是从理论语言学和社会语言学的角度进行思考，提出拙见，以供技术专家和商家扬弃。”

现在我们正在从控制论的角度重新审视动态语言知识更新。动态语言知识更新的关键也是反馈和稳态。机器自动学习新的知识已经很难，如何将那些个体产生的语言变异学习进来并归结为语言的突变，是自动学习系统的任务。但是更不容易的是如何反馈这些学习来的变异。

在学习和反馈的过程中，我们绝对不认为可以做到百分之百的正确或准确，我们认为不可能也没有必要这样追求。但是为了维持动态语言知识更新体系的“稳态”，就必须还要研究这个系统的“体内的废物在浓度达到有毒以前必须排泄出去”的功能，要研究语言自动更新的“恒温器”、“氢离子浓度自动控制器”、“调速器”等等，这些构成语言自动更新体系的“稳态机构”，这是更难和更重要的研究，只有突破了这一点，才能实现启动知识自动地更新为滚动知识，滚动知识又自动地成为新的启动知识，语言知识的“遗传”和“变异”才得以在进化中实现。

流通度和语感模拟，也许还有别的，就是在进行这方面的探讨。

¹⁷ 参见维纳《控制论》（第二版）115—116页，科学技术出版社，1963年。

¹⁸ 参见陈原《语言学论著·卷三》P344。

¹⁹ 参见陈原《语言学论著·卷一》P536。

²⁰ 参见张普《关于网络时代语言规划的思考》，载《语言研究》1999年3期。

5. 关于“种族信息量的测定”

维纳说：“人的相互通讯包括全部复杂的语言和文献以及其他许多东西。”²¹

“不管一个种族用什么通讯方式，这个种族所使用的信息量总是可以测定的，并且可以把对种族有用的信息量同对个人有用的信息量区别开来。当然，对个人有用的信息并不就是对种族有用的信息。除非它能改变个人对于其他人的行为，而且这种行为还可能没有种族的意义，除非别人能把这个行为同其他的行为方式加以区别。因此，要决定某一种信息是属于全族的还是纯粹对个人有用的，就要看个人所采取的行动方式是否被种族中其他成员看作特定的行动方式，以及是否能逐一影响这些成员的行动等等。”²²

我们曾提出“流通度”的概念，其中许多说法是与维纳的说法相一致的，或者说维纳的论述差不多就是对语言的变异和语感的论述。重要的是他说“这个种族所使用的信息量总是可以测定的”，我们希望通过测量流通度来对语感加以数学界定、加以量化，使得“能不能说”、是否已经“被理解”、“被认可”、“被传播”变得可以通过流通度的计算进行判定。进一步还想把“流通度”的知识或者说“流通度”的获取方式教给电脑，使电脑通过获得“流通度”来（相似地）获取“语感”，或者来模拟语感，从而获得自学能力。所以流通度理论不仅是在语言学方面使人的“语感”得以量化，更重要的是在信息处理方面有可能使计算机真正获得语言的自学能力，使智能化进入一个新的发展阶段。²³

我们认为“流通度”是一种语言事实在社会交际中的流行通用的程度。人们对一种语言现象的流行通用程度的感觉，也就是所谓“语感”。流行通用程度高，听得多，习惯成自然了，就感觉能说，否则，就觉得不能说。

我们还指出过：“共同语感G总是稳态的语感，是操同一语言的全体人的语感。差别语感总是动态语感，是操同一语言的部分人的语感，这‘部分人’有三种情况：少部分人、大部分人、一半人。刚好一半人说的的时候是偶然的，少部分人、大部分人说的的时候是必然的。只有当一种语言现象为大部分人或者绝大部分人所接受时，这种语言现象才可以说是比较成熟的，已经流通了。”

我们还说过：“我们实际上是主张建立一个动态的大规模真实文本的书面语的语料库。把语料库的建设和使用从静态推向动态，把文本的选择和抽样原则从分布原则推向流通原则，把对语言成分的一般性的统计分析推向对语感的推测性统计分析和验证，从而探索使电脑可以逐步获得语感并随时增强和调整语感的路径。”

我们实际上是通过“对‘种族信息量’流通度的动态测定，来相似于捉摸不定的语感。我们相信控制论中“不管一个种族用什么通讯方式，这个种族所使用的信息量总是可以测定的”。维纳所说的“对种族有用的信息量”和“对个人有用的信息量”大致对应于我们的“共同语感”和“个人语感”。

²¹ 参见维纳《控制论》（第二版）156页，科学技术出版社，1963年。

²² 参见维纳《控制论》（第二版）157页，科学技术出版社，1963年。

²³ 张普《关于语感与流通度的思考》，载《语言教学与研究》，1999年第2期。

6. 关于“本书的教训之一”和“反内稳定的因素”

维纳说：“对于社会所有这些反内稳定的因素来说，通讯工具的控制是最有效也是最重要的。”²⁴

“本书的教训之一就是，任何组织所以能够保持自身的内稳定性，是由于它具有取得、使用、保持和传递消息的方法。在一个过于大的社会里，社会成员无法直接相互接触。因此，出版物（包括书籍和报纸）、无线电、电话网、电报、邮递、剧院、电影院、学校、教堂就都成了取得、使用、保持和传递信息的工具。它们除了具有作为通讯方法这个内在重要性以外，还有其他的次要作用。”²⁵

“什么通讯系统比所有别的系统应该对社会内稳定性更有贡献，它就直接被掌握在那些最醉心于争权夺利的人的手中，而我们已经知道，这种争夺是社会中的主要的反内稳定性的因素。”²⁶

我们在研究“流通度”时，和维纳同样重视“出版物（包括书籍和报纸）、无线电、电话网、电报、邮递、剧院、电影院、学校、教堂”这些“取得、使用、保持和传递信息的工具”，我们从今天的角度称为“公众传媒”，我们略去了学校、教堂等等。我们曾经说过：“流通度是一个普遍的概念，是一种普遍存在的现象，它存在于两大类媒体之中：书面文本媒体（报纸、杂志、图书等）和口头文本媒体（广播、电影、电视等），也存在于因特网上。首先是媒体有流通度属性和属性值，其次是刊载于这种媒体的文本有流通度属性和属性值，再次是依据这种文本所获取的语言知识也具有流通度的属性和属性值。”

我们还说过：“我们发现所谓个人语感和公众语感都是对于语言成分的流通度的一种感觉。因此我们提出用流通度相似于公众语感的量化方案，我们说‘相似’，是因为流通度本身并不是精确的，它已经略去了一些影响语言流通度的因素，例如：个人之间的语言交际也是一种发表和流通，但是受众和影响有限，与大众传媒的复制和流通相比较，对于语言流通度的影响就忽略不计了。此外，还有学校教育、父母教育等。所以，依据大众传媒的流通度计算出来的语言的流通度，客观上只是流通度的一种最大接近值，这一点我们还要另文详述。而公众语感就是绝大多数人对这种流通度的感觉。所以，我们有： $I_p \approx C$ 。这样我们就可以从流通度的量化进入语感的量化，进而计算语感，模拟语感。这是极其关键的一步，对于计算机的智能化、人性化具有相当深刻的意义。”²⁷

维纳已经看到了“传递信息的工具”对于“任何组织所以能够保持自身的内稳定性”的重要，他已经历了各种传递信息的工具，除了电视和因特网。我们为了获得研究语言的稳态的数据，确定了当代的六大常规媒体——报纸、杂志、图书、广播、电影、电视和网络这种新的媒体作为测量流通度的主要依据，而略去其他一些次要因素，应该也是符合控制论对通讯工具与内稳态或反内稳态的关系基本估计的。

²⁴ 参见维纳《控制论》（第二版）160页，科学技术出版社，1963年。

²⁵ 参见维纳《控制论》（第二版）160页，科学技术出版社，1963年。

²⁶ 参见维纳《控制论》（第二版）161页，科学技术出版社，1963年。

²⁷ 参见张普《信息处理用动态语言知识更新的总体思考》，载《语言文字应用》2000年第2期。

7. 关于学习和自生殖机

维纳说：“学习的能力和生殖自己的能力是我们公认的作为生命系统的特征两种现象。”²⁸

“如果行为方式有遗传不变性，那么在各种有了变异的行为类型中，那些被传播开的方式，总可以发现它们对于种族的继续生存有某些好处，因而能使自己稳定下来，另外一些对种族继续生存有害的行为方式也就会被消灭。与个体的个体发育的学习比较，上述结果就是某种种族的或系统发育的学习。个体发育和系统发育的学习都是动物根据周围环境来调节自己的方式。”²⁹

“个体发育和系统发育的学习，特别是后者，不只是适用于所有的动物，而且适用于植物，适用于所有从任何意义上看来是有生命的有机物。”³⁰

我们过去曾说过：“新的语言成分总是有人最先在局部范围内开始使用，传播开并稳定后，就成了被大家接受的新词、新义、新用法，如不然，就逐渐被淘汰，个人或极少数人仍使用，就被视为生造。语感的量化和可计算，应该能够提供一个阈值，从而较为科学地确定——一个新的语言成分何时算是“约定俗成”，被公众（或者说是被绝大多数人）接受了。”³¹

我们还引用过徐通锵先生的话：“变异在其产生之初，在社会人群中呈无序的、随机的分布，如果变异成分的某一变异形式在言语社团中被某一社会人群接受并开始传播，那么无序的变异就进入有序的行列，意味着演变的开始。如果使用这种变异形式的社会人群在言语社团中具有某种特殊的地位，那么这种变异形式就可能成为其他社会人群的仿效对象。”³²

我们认为语言的生命力就在于这种稳定中的变化。这些变化的端倪就隐藏在大规模的真实文本（无论是经典的还是非经典的文本）之中，甚至就隐藏在那些非规范现象里。一切新词、新义、新用法一开始总是不在约定和规范之中的，通过“对话”和“讨论”，利用“已知”对“新知”作出“解释”或“纠错”，新知一旦被大家接受并广为传播，最终将进入约定或规范，这就是语言发展的辩证法和规律。³³

我们相信维纳论证的“系统发育的学习”“适用于所有从任何意义上看来是有生命的有机物”的这一论点，他强调了“所有”和“任何”，而我们恰恰强调了语言的有生命和生命力。我们论证的“动态语言知识更新”正是将语言看成是不断产生变异的“活”的东西，将动态流通语料库看成是“活”的东西，将启动语言知识库和滚动语言知识库都看成是“活”的东西。我们论证的“个人语感”，就是语言的“个体发育学习”的关键，“公众语感”就是语言的“系统发育学习”的关键。维纳的下述论断只要稍稍做些替换，几乎就是为语言的更新而描述的：“那些被传播开的方式（替换为“语言变异”），总可以发现它们对于

²⁸ 参见维纳《控制论》（第二版）167页，科学技术出版社，1963年。

²⁹ 参见维纳《控制论》（第二版）167页，科学技术出版社，1963年。

³⁰ 参见维纳《控制论》（第二版）167页，科学技术出版社，1963年。

³¹ 参见张普《关于大规模真实文本语料库的几点理论思考》，载《语言文字应用》1999年第1期。

³² 参见徐通锵《语言论》P69、P71，东北师范大学出版社1997年。

³³ 参见张普《关于大规模真实文本语料库的几点理论思考》，载《语言文字应用》1999年第1期。

种族的继续生存（替换为“语言交际”）有某些好处，因而能使自己稳定下来，另外一些对种族继续生存（语言交际）有害的行为方式（语言变异）也就会被消灭。

8. 关于“白箱”和“黑箱”

维纳说：“我们已经成功地做到的是：制造一个能模拟任何非线性变换器的特性的白箱，然后把它与一给定的黑箱变换器进行类比，方法是给两者加上同一随机输入并把它们的输出以适当方式联接起来，使它们不需要人的干预就能达到一种适当的结合。”

他还说：“请问这个过程，与另一些过程——基因作为一个样板，从氨基酸和核酸的一种比例不定的混合物中，形成与它相同的另一些基因分子，或者，病毒把从它的寄主的组织和体液中形成的其它同种病毒分子引变成自己那种类型——从哲学观点看来是否有很大不同。我并不完全指望这些过程在细节上也是相同的，但是我却相信从哲学的观点看来，它们是非常类似的现象。”³⁴

我们的看法是：所谓语感——无论是个人语感还是公众语感，类似于是一个“黑箱”，社会组织和其每一个个体是如何依据话语“通讯”经验的积累来获得语感能力，如何约定俗成地实行语言“控制”，我们是看不到的。但是“出版物（包括书籍和报纸）、无线电、电话网、电报、邮递、剧院、电影院、学校、教堂”等这些“取得、使用、保持和传递信息的工具”——我们今天称之为媒体（还有维纳没有或不可能提到的电视、因特网等）的东西却是看得见、摸得着的，我们可以有多种途径测知其流通的情况，这类似于是一个“白箱”。然后将“白箱”和“黑箱”进行相似性类比，将白箱的量化数据传递到黑箱。

我们的类比方法当然不同于维纳的类比方法，我们同样也“并不完全指望这些过程在细节上也是相同的，”但是我们一样也“相信从哲学的观点看来，它们是非常类似的现象。”

9. 关于“可信的程度只能到达头几位数字”

维纳说：“在社会科学中，极难使被观察的现象和观察者之间的耦合减到最低限度。相反，观察者能够对他所关心的现象施展巨大影响。虽然我十分尊敬我的那些人类学家朋友的智慧、本领和诚实目的，但是我并不认为他们所考察的任何社会以后将永远不变。”

“另一方面社会学家没有从永恒的、与时间地点无关的角度来冷静观察他的科目的那种便利。”

“总之，不管我们在社会科学中的研究是统计性的或是动力学性质的——这种研究一定具有两可的性质，它们可信的程度只能够到达头几位数字，一句话，它们不能给我们提供大量的可以验证的、有意义的信息，如同我们在自然科学中可以希望得到的那种信息一样。我们不能忽视这些信息，但我们对这些信息的可靠性不要抱着太大的希望。”³⁵

我想这两段话可以作为我们对于研究结果的预期。动态语言知识更新的研究结果，其

³⁴ 参见维纳《控制论》（第二版）178页，科学技术出版社，1963年。

³⁵ 参见维纳《控制论》（第二版）163—164页，科学技术出版社，1963年。

可信程度如果可以到达头几位数字，就已经是我们非常满意的了。

我认为：如果人脑不能保证学进来的东西百分之百的正确，如果人脑的知识每天都有改变，如果每个人的语感都有不同之处，如果同一个人的语感也会有变化，我们如何能够要求电脑的处理语言时的“准确率”“召回率”“正确率”达到百分之百呢？动态语言知识更新的验证标准如果有的话，也会是动态的、相对的。

我还是希望用控制论中维纳的话来做结尾：

“我们不必提出这样的问题：能不能制造一部能下冯·诺意曼所说的最优棋局的机器。就是最好的人脑也无法做到这一点。但是另一方面，无疑地可以制造那样一种机器，不管下得好下得坏，它总是按照规则来下棋。

我认为可以制造一部比较粗糙但决不平凡的机器来实现这一目的。”³⁶

参考文献

- [1]维纳《控制论》(第二版),科学技术出版社,1963年。
- [2]赵元任 YUAN REN. CHAO, Meaning in language and how it is acquired, Transactions of the Tenth Conference, April 22, 23 and 24, 1953, Princeton, n. j.
- [3]陈原《语言与控制论札记》,载陈原《语言学论著》卷三729页,辽宁教育出版社,1998年。
- [4]徐通锵《语言论》,东北师范大学出版社,1997年。
- [5]张普《关于大规模真实文本语料库的几点理论思考》,载《语言文字应用》1999年第1期。
- [6]张普《关于语感与流通度的思考》,载《语言教学与研究》,1999年第2期。
- [7]张普《关于网络时代语言规划的思考》,载《语言研究》,1999年3期。
- [8]张普《关于第三代大规模真实文本语料库的几点理论思考》,载《自然科学基金重点项目结题报告》(内部),清华大学,1998年。
- [9]张普《信息处理用动态语言知识更新的总体思考》,载《语言文字应用》2000年第2期。
- [10]隋岩、张普《1997中文报纸媒体流通度分析》,载《计算语言学文集》清华大学出版社,1999年。
- [11]邢红兵《信息领域汉英术语的特征及其在语料中的分布规律》,载《术语标准化与信息技术》,2000年第3期。

³⁶ 参见维纳《控制论》(第二版)164页,科学技术出版社,1963年。