

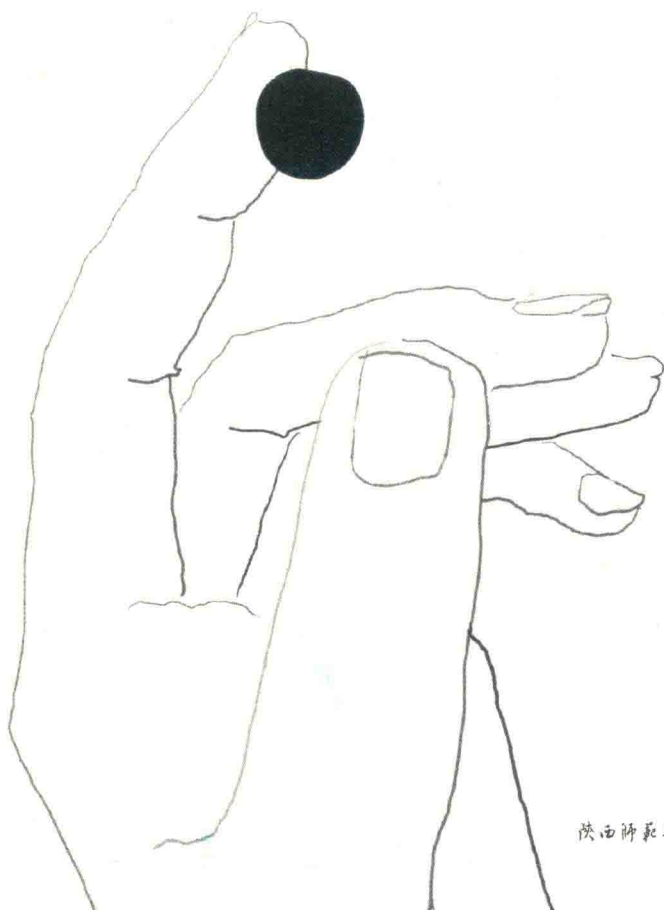
World & Words

超信息论

与人工智能的语言学原理

物与词

史雷鸣 著

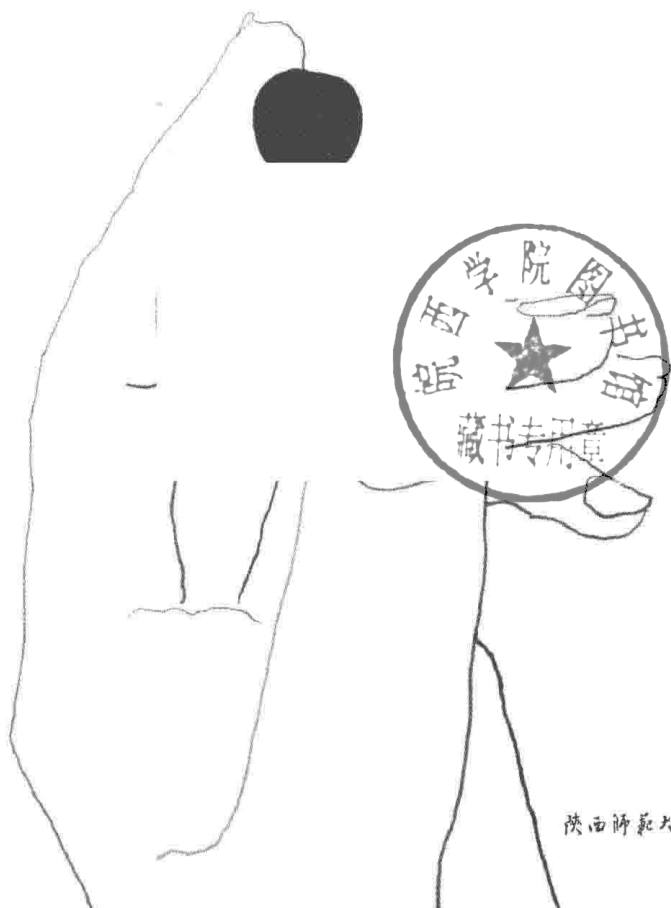


陕西师范大学出版社

超信息论
与人工智能的语言学原理

物与词

史雷鸣 著



陕西师范大学出版总社

图书代号: ZH19N0025

图书在版编目 (CIP) 数据

物与词: 超信息论与人工智能的语言学原理 / 史雷鸣
著. — 西安: 陕西师范大学出版总社有限公司, 2019.1
ISBN 978-7-5695-0545-0

I . ①物… II . ①史… III . ①人工智能语言—研究
IV . ① TP312.8

中国版本图书馆 CIP 数据核字 (2018) 第 300677 号

物与词: 超信息论与人工智能的语言学原理

WU YU CI CHAO XINXILUN YU RENGONGZHINENG DE YUYANXUE YUANLI

史雷鸣 著

责任编辑 / 张建明 刘 荣

责任校对 / 李 曼

装帧设计 / 贾 嫣

出版发行 / 陕西师范大学出版总社有限公司

(西安市长安南路 199 号 邮编 710062)

网 址 / <http://www.snupg.com>

经 销 / 新华书店

刷 / 西安市建明工贸有限责任公司

开 本 / 787mm × 1092mm 1/32

印 张 / 3.5

字 数 / 50 千

版 次 / 2019 年 1 月第 1 版

印 次 / 2019 年 1 月第 1 次印刷

书 号 / ISBN 978-7-5695-0545-0

定 价 / 20.00 元

读者购书、书店添货或发现印装质量问题, 请与本社营销部联系调换。

电话: (029) 85307864 85303622 (传真)

未来正来，一张进入人工智能时代的入场券

突破信息论

理解智能本质

序 言

世界复杂庞大，横亘于时光中，一切的原因造成今日的结果，而此刻无数的事件正影响着未来。

我们得以认识这个世界，至少是认识这个世界的一部分，是因为认识了其中的因果。

科学就是对这种因果的解读与理解。

人工智能历经了几十年的发展，我们都能感到一个人工智能的时代更在逐渐揭开序幕。从阿尔法狗战胜人类的围棋冠军，到自动驾驶和 AI 辅助的医疗诊断。

但人工智能到底是什么？我认为能够说清楚智能，才是人工智能发展的一个里程碑。

这显然是一个复杂的系统的问题。计算机专业和

数学专业，都很难阐述清楚这个问题。尽管人工智能的实现依赖计算机系统和数学。

在很多年前，我开始为我的博士研究阶段选择方向和课题的时候，我选择了从语言学角度理解建筑、艺术和世界的方向。从那个时候开始，我努力用语言学和符号学建立对建筑和艺术的重新认识。那个努力的过程持续了四年。而研究的最后阶段，我的结果就是一个系统的完整的语言哲学，和语言科学理论。

在那个阶段，我建立了物质本身具有语言属性，人类的语言更多的是对物质本身的语言的模型的建立这样的一个基本理论基础。于是，一个拉通了科学、艺术，自然语言的线索出现了。语言是世界的全局性存在，甚至物理化学和生物学不过是对所指征的物质世界的语言建模。

在这个广义的语言学的理论建立过程中，一个新的认识论也伴随着浮现。现象、语言、本质三个要素，构成了广义语言论的新的认识论的三个核心要素。并且这三个要素彼此依赖。

在这个理论基础上，对信息论的理解，就有了哲学和科学层面的新的支撑和认识。用广义的语言论重新梳理并且解释信息论，我们可以发现信息论建立之初的哲学基础的不完备。也因此有了用语言论全面的重新理解和构建信息论的新版本的基础和可能。因此，在随后的研究中，我转向了语言与信息论的关系、语言与计算的关系、语言与人工智能的关系的研究和拓展。

而与此同时，语言 and 智能本身的关系，成了一个重要的突破方向。在冷静地从整体以及细节上的分析之后，我认为智能就是语言对现象和变量的命名与分析计算过程和结果。

智能不是一个实名词。智能就像波浪对于水一样，是另一些事物的过程和状态以及结果。就这一点而言，对语言本身的理解，的确有助于我们跳出之前的认识误区的错误概念的误导，智能不是一种物化的存在，是一些具有语言能力的叙述结构所产生的语言学的计算能力和对其他语言的解读能力。

这种语言现象和结构，在我所完成的广义语言论中已经进行过分析和论述。而这本书的重点是在这个

基础上，用语言哲学和语言科学重新建构对人工智能的认识，并且直接通过语言学的数学化，建立用语言学混和数学形成的人工智能的科学原理和技术原理以及路线。

达尔文的进化论，使得我们理解了生物的整体性与进化演化，构建了生物学的统一的最基础原理。而广义语言，对于人工智能甚至于知识整体都有类似的基本原理和定律的意义。

本书，尝试了将语言放置在因果律和演化的基础上，从而将语言组织成为因果关系的演化进化网络。这个网络的拓扑，是有一定的确定性的。而这个拓扑关系中，隐含着语法和算法以及思维的路径。并且，这个语言网络因此成了一个拓扑化的数学模型，并且因此而彻底数学化、几何化。语言整体的数学化，形成的数学和拓扑结构，成为一种可计算的科学化的超级函数。并且可以用数据库得以高效实现。

这种理论，将人类的语言作为世界的地图和模型，

用人类语言构建人工智能的语言库和语言算法，继承人类的知识体，并可自我进化。这样的软件和以后的硬件化，意味着存储是计算的重要组成部分，计算是计算，存储也是计算，检索也是计算。因此，在软硬件两个层面上，都存在对冯诺依曼体系突破的可能。

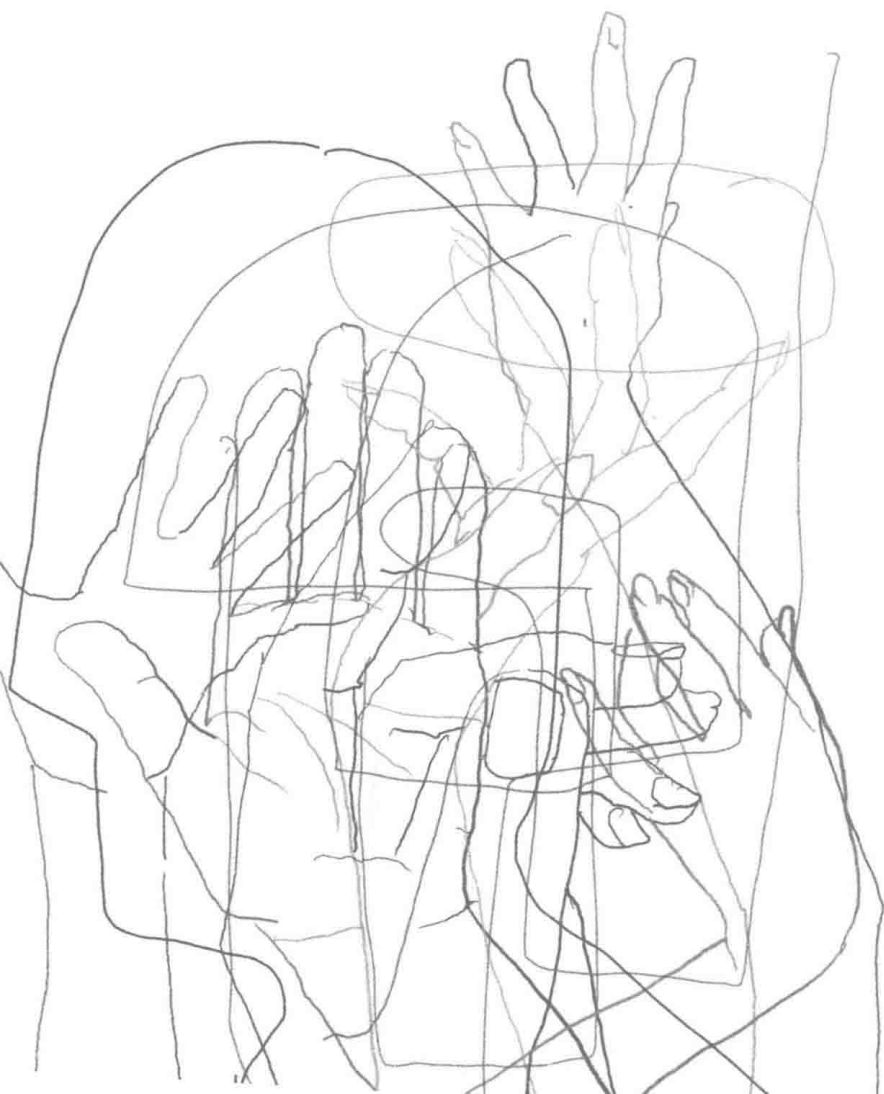
法国著名哲学家福柯，有一本著作《词与物》。而我们将在这本书中更加深入的理解词与物的关系。这本书叫作《物与词》，也是对福柯等哲学家，以及一百年的语言哲学发展的致敬。

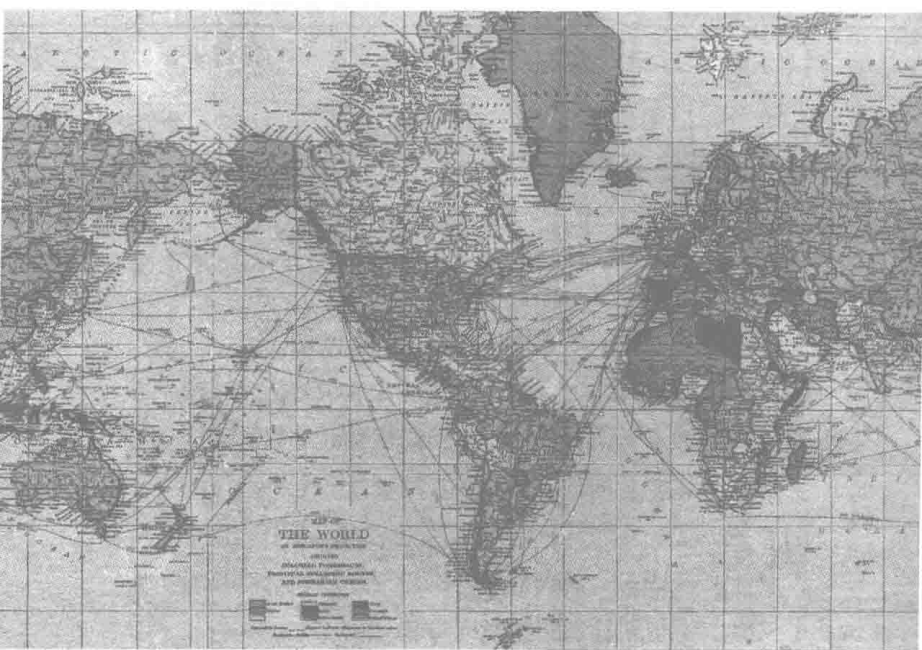
写下这样一段算是导读的文章作为序言，主要是为了解释，为什么这本书叫作“超信息论与人工智能的语言学原理”，以及介绍这个基于语言学的人工智能原理研究的前因和前期的理论基础。

把这本书，献给我的父母、爱人，献给世界上对万物充满好奇的孩子们。献给即将到来的人工智能时代，或者说，机器智能的时代。那是一个新的纪元。

史雷鸣于西安

2018年12月18日





目 录

- 一 智能是一个语言学问题 / 1
- 二 智能是一个词语构成的知识体网络 / 9
- 三 词语库是一个基于树状混合网状的拓扑结构 / 19
- 四 数学是语言的分支以及语言的离散数学内核 / 45
- 五 人工智能的哲学和语言问题 / 59
- 六 生命本质与机器智能时代 / 75
- 七 开放的树网词语库的构建与开放的机器智能 / 81
- 八 词语拓扑网络逻辑硬件和 CPU / 89
- 九 结语 / 93



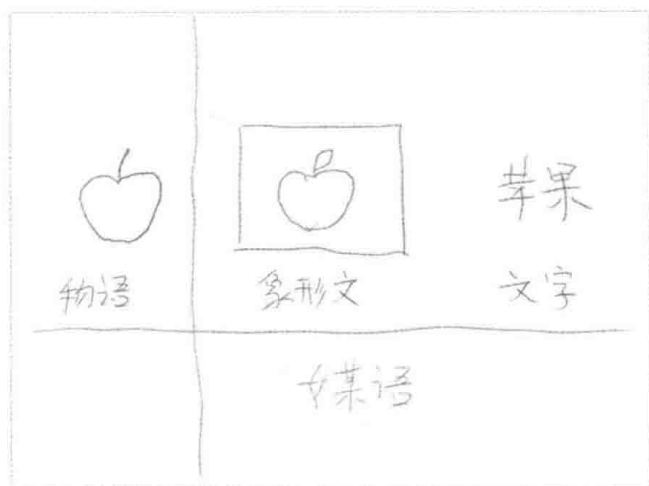
智能是一个语言学问题

1. 世界是语言的

世界之所以能被描述，是因为其具有可被测量观察的属性和变量，形成了现象，当人类观察和定义到这种现象及其变量，创建了媒体符号予以指征，形成了媒体语言。对应的，被指征的存在的事物，其就是自身的语言，这种存在以能量和物质的方式构成了世界，它们是实在的物的语言，我们可以称之为物语。人类构建的文字，音乐，绘画，雕塑等语言，依赖于媒体表达。我们可以称之为媒语。计算机代码，也是一种媒语。其基础构成的0和1，可以用电，也可以用磁，也可以用光作为载体和媒体。而其约定的指征的意义，又以0和1两种符号作为表达。

所有的语言都是符号系统。语言学的基础是符

号学。作为符号学的开端，索绪尔创建了符号的模型和基础理论。他将符号分为了符形和符意两个部分。符形是它的形式或者载体，符意是其携带或者指征的意义。中文“树”和英文“tree”，它们所差别的就是符形，而其所指代的意义都是木本植物。人类的媒语是一种约定，尤其是文字语言。媒语也是一种符号系统。而自然存在的事物也是如此，其存在的物质能量本身就是其符形，其属性或者所具备的物理化学特性



就是其符意。人类对自然的研究，正是不断地对自然存在的拆分和破译，不断地创造新的媒语指征揭示所对应的物语。例如原子、电子、夸克。物理和化学等自然科学，是对世界的解读和使用媒语为之构建的语言学模型。整个自然科学就是解读自然的物语并用媒语为之建立模型的语言学成就和结果以及过程。当然，图表、公式、数字，以及动画，都是一种媒语。

2. 知识是语言的

显而易见，人类的知识体系主要是以文字语言、图表、公式、数字构建的。图表公式和数字，是一些特殊的符号语言。还有一些知识，存在于文化和习惯中，以及口语与记忆中，或者存在于经验中。它们也是一种符号体系。只不过是未被提取或者整理成严密和公共化的语言体系。即使是模糊的经验、工艺流程、行为，甚至我们的思考都是一种程序。程序，本身是一个语言序列。动物的行为植物的生长代谢，都是一种或者多种程序的混合。程序，是语言的动态序列。因此，人类所有的知识，都是语言或者类语言。

正是因为人类创造了媒语，可以将很多经验、认识，用媒体语言记录和传播，因此成为知识汇集记录转载传播的知识池，这种可以高效记录和传播的外部媒体载体语言，成为人类大脑和意识的外部存储器、网络、输入与输出设备，甚至外部的计算器，大大促进了人类的文化和文明的发展。或者说，构成了人类文明的基础。

人类构建的高等级知识以学科的方式被创建，而事实上每一个学科的构建和成熟，都以形成文字语言的理论和学说呈现。每一个学科都以文字和图表以及公式的方式，完备地叙述和呈现自身。当然，本论文也是如此。

3. 智能为什么是一个语言学问题

现代人和原始人存在着巨大的智力差别。原始人和现代人的孩子在出生时其实差别甚小。或者说现代人出生的时候其实就是一个原始人甚至是一个小野兽。甚至一万年前的人类的大脑容量比现代人更大。他们后天的差异是造成这种差别的核心原因。他的经历和受教育的差异，构成了智力的差异。而现代人所受的

教育，核心是体验和词语化。受教育的过程就是一个构建词库、扩充词库和不断强化优化词语之间的关联性和解释关系的过程和体验。朝鲜和韩国的国民的基因体质差别微乎其微，而他们的受教育与经验之后的词语库和词语解释关系的差别，构成了他们作为社会的整体以及个人的知识能力的较大差异。类似的这样的很多证据，可以帮助我们理解到，智能，取决于你认识多少事物，掌握多少词语，以及词语之间的关系构建，以及它们之间的解释关系的差异。

人类的思考，是以文字和图表为主要的方式进行的。我们一再解释图表是另外一种特殊的语言和符号系统。但其依然是一种语言系统。而语言是一个演进累积逐步构建的系统。人类的语言的源头，都有其最初的简单的普遍的常规的源自世界的自然现象及其变化和变量的原型。最初的语言不过是对最基础的自然现象的特征和变量和变化的指征。在此基础上，人类不断的细分现象及其变量，发现最初的朴素的逻辑，并以此为基础不断扩展发展而来。人类的整体的智力也伴随着这个语言和词语库的发展而发展。物理学、化学，即使今日