

# 语言认知与语言计算 – 人与机器的语言理解

王少楠<sup>1,2\*</sup>, 丁鼎<sup>3,4\*</sup>, 林楠<sup>5,6</sup>, 张家俊<sup>1,2</sup>, 宗成庆<sup>1,2,7</sup>

1. 中国科学院自动化研究所模式识别国家重点实验室, 北京 100190
2. 中国科学院大学人工智能学院, 北京 100190
3. 浙江大学生物医学工程教育部重点实验室, 生物医学工程与仪器科学学院, 杭州 310027
4. 之江实验室, 杭州 310000
5. 中国科学院行为科学重点实验室(中国科学院心理研究所), 北京 100101
6. 中国科学院大学心理学系, 北京 100049
7. 中国科学院卓越创新研究中心, 北京 100190

\* 通信作者. E-mail: shaonan.wang@nlpr.ia.ac.cn, ding\_nai@zju.edu.cn

国家自然科学基金(批准号: 61906189, 62036001, 31771248), 之江实验室重大科研项目: 2019KB0AC02

**摘要** 语言理解是认知科学和计算机科学交叉领域共同关心的问题, 但两个学科在选择具体研究问题时却十分不同。认知科学领域的研究侧重解析大脑的工作机制, 更多地关注于描述大脑对语言的响应, 缺乏对大脑语言功能整体化、系统化的研究, 而计算机科学家在选择研究问题时重点关注实际应用效能, 往往忽略了对语言最本质规律的研究。那么, 如何实现两种思路的交叉融合, 为智能语言计算模型的构建和语言认知机理的研究带来新的机遇和启发呢? 本文首先简要回顾了认知科学和计算机科学在语言理解方向上的研究问题、发展历程和研究方法, 重点阐述研究现状和面临的挑战, 之后对比认知科学和计算机科学领域对于语言理解问题的主要观点, 分析两者之间的异同。最后对现有语言认知和语言计算两个领域的交叉融合方法进行归纳和总结, 并对未来发展趋势予以展望。

**关键词** 语言认知, 语言计算, 人的语言理解, 机器的语言理解, 交叉研究

## 1 引言

语言是一个多层次的符号系统, 包括了语音、词法、句法、语义、语用等不同层面, 它可以将最基本的语言符号组合起来构成更复杂的、无穷尽的符号序列, 从而灵活地表达意义。因此, 语言也被认为是人类思维的载体, 是人类交流思想和表达情感最自然、最直接、最方便的工具。

正是由于语言的多样性、灵活性和复杂性, 研究人脑的语言理解机制, 以及构建能够理解语言的机器系统是十分困难的。在计算机科学诞生初期, 许多语言理解研究的先驱者尝试进行跨越计算机科学、语言学、认知科学(包括心理学、语言学、神经科学等)的交叉学科研究, 并且在人脑语言理解机理和语言计算模型之间建立联系<sup>[1~6]</sup>。然而由于问题的复杂性, 几十年来学科间的研究逐渐分离, 形成了计算机领域的自然语言理解, 认知科学领域的心理语言学、神经语言学等细分领域。本文中, 认知科学主要指认知心理学、认知神经科学这两个领域, 尤其是心理语言学、语言的认知神经科学这两个分支<sup>[7,8]</sup>。

图 1 给出了认知科学和计算机科学在语言理解方向上研究问题和研究方法的关联。可以看出，两者重点关注的研究问题和采取的研究方法存在很大差异。计算机科学领域主要采取理性主义方法（以基于规则的方法为代表）和经验主义方法（数据驱动的建模方法，以统计机器学习方法和神经网络方法为代表），从技术的角度更多地关注应用研究，即如何构建智能系统使其理解自然语言，从而完成各项实际应用（如机器翻译、人机对话、自动文摘等）。这里所谓的“理解”其实更多地是以应用目标为导向的“处理”，所以在很多场合下人们通常用“自然语言处理”统称这一学科方向。认知科学领域采取神经影像和行为分析方法，更多地关注人类语言理解的心理和神经基础，比如不同脑区在语言理解中的功能以及神经活动如何编码不同层次、不同类型的语言信息。相同的是，在语言理解问题上，两者都以语言学作为学科基础，利用计算建模的手段分析问题。总体而言，语言认知和语言计算在各自的方向上都取得了丰硕的成果，新的理论和方法不断被提出，并得到了成功的应用。

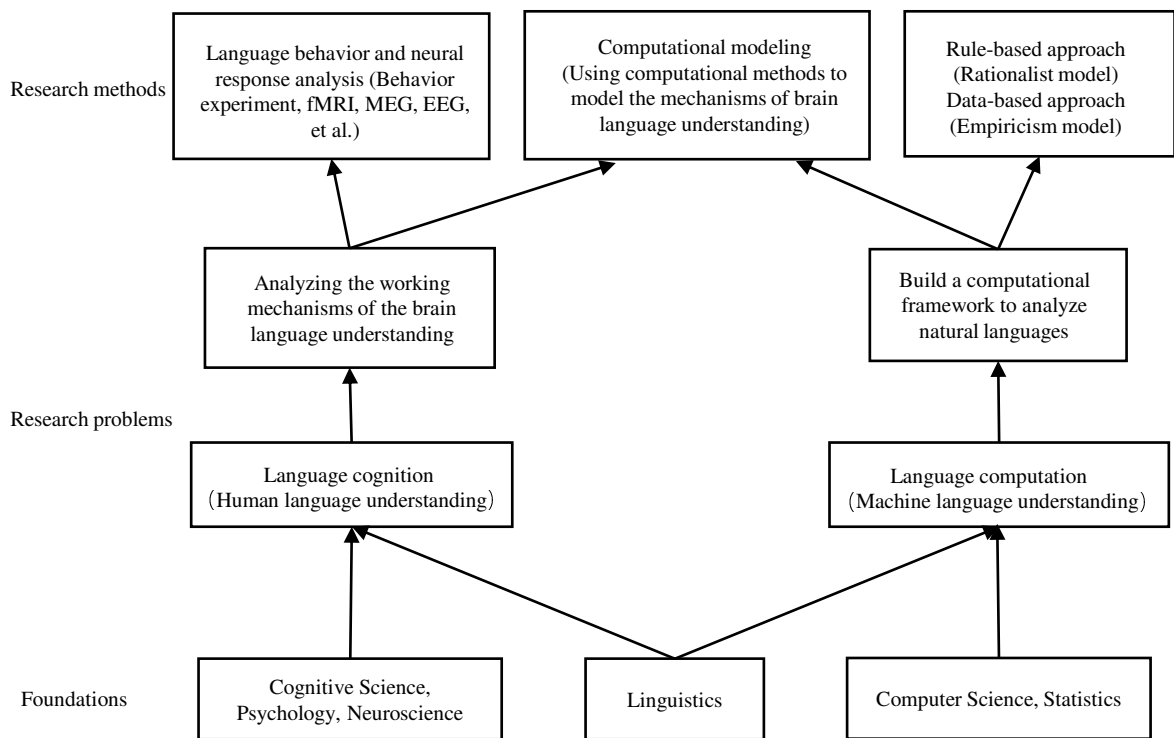


图 1 认知科学和计算机科学在语言理解问题上的关联

Figure 1 Connections between cognitive sciences and computer sciences on the language understanding problem

近年来，计算资源的提升和深度学习算法的进步带动了人工智能和计算机科学领域的快速发展，计算机在国际象棋、知识竞答、围棋、视频游戏等任务中击败人类职业选手。在自然语言理解领域也出现了 Siri、Waston 等自动人机对话与服务系统，以及较为实用的机器翻译系统<sup>[9~13]</sup>。但是，目前的人工智能系统依赖大规模训练数据和计算资源，缺乏基本的常识知识，在学习能力和泛化能力上与人类智能还存在很大差距<sup>[14~18]</sup>。因此，人脑作为智能实现的唯一范例，已经再次引起计算机科学家的重视，借鉴人脑神经机制和认知行为机制发展类脑智能已成为国际上前沿的研究热点。

另一方面，包括认知神经科学、认知心理学在内的认知科学迅速发展。对大脑语言加工过程的无

损、实时监测正在成为可能。通过各种精巧的实验设计,认知科学领域的研究人员对语言加工的脑基础进行了多方面的考察,并取得了很多有价值的发现<sup>[19~23]</sup>。传统的认知科学实验由于较多依赖严格的实验控制,使得这类研究在生态性、全局性等方面有较大不足。然而,近年来,越来越多的研究开始采用高生态效度的实验范式,并采用先进的数据分析方法分析人脑在高生态效度任务下的信息加工机制。在语言研究方面,研究人员也逐渐开始借助计算模型方法在自然文本的实验刺激条件下来研究人脑的语言理解过程<sup>[24~31]</sup>。

随着最近几年认知科学和计算机科学等领域交叉研究工作的快速积累,研究人员开始总结计算机科学的方法,尤其是深度学习模型,可以给认知科学领域的研究带来哪些变化<sup>[32~37]</sup>,以及认知科学发现的结论如何帮助构建人工智能模型<sup>[38~42]</sup>。上述工作探讨了宏观上两个领域结合的可能性,但是并没有讨论在细分问题上如何结合两者开展工作。为了给认知科学家和计算机科学家在语言理解方向上开展跨学科研究提供借鉴,本文在语言理解问题上对已有交叉研究工作进行了总结和展望。

综上所述,在语言理解问题上,计算模型可以帮助认知科学家更加定量地研究、建模人脑的语言认知机理;同时,如果可以深入了解人脑语言理解的工作机理,将有助于计算机科学家构建更加智能的语言计算模型。因此,为了推动人和机器语言理解研究的新一轮发展,结合认知科学和计算机科学开展跨学科交叉研究势在必行。下文首先围绕人脑语言认知(第2节)和机器语言计算(第3节)的定义、主要研究问题、研究现状和研究方法、以及已有研究的局限性这几个方面展开介绍;接着本文将语言认知与语言计算的主要观点和概念进行对比(第4节),在不同层面上分析了两者的异同;然后本文总结归纳了已有语言认知与计算融合的工作(第5节),在此基础上,分析了已有融合方法的局限性,并且重点对未来的可行研究方向进行展望(第6节);最后对全文进行总结(第7节)。

## 2 关于语言认知机理的研究

### 2.1 语言认知的定义

本文中提的语言认知指人脑对语言的理解,泛指个体在接受语音、文本等信息时,从听觉、视觉等感知觉信息中提取抽象的符号信息的过程。语言认知是一个复杂的过程,不同层次和不同单位的语言结构,以及不同类型的语言信息的加工机制各有不同,所依赖的脑网络也非常复杂。比如语音理解过程中需要听觉系统对语音的基本声学特征进行编码,随后涉及词汇识别、语法语义分析等多个步骤,最终才可以实现语言理解。

### 2.2 主要研究问题

语言是复杂序列,人脑是一个复杂系统,因此对人脑语言加工机制的研究是非常具有挑战性的。一方面,语言中有不同大小的单元,如果研究人脑对语言的加工,应当以对哪种语言单元的加工作为切入点呢?大脑在加工语言的过程中,是否也是以某种单元作为最核心的加工单元的呢?这是下面介绍的第一个研究问题。

另一方面,大脑的信息加工过程非常复杂。不同类型的信息往往是由不同的脑区加工的,而且加工次序有先有后。对语言加工脑区和语言加工时间进程的研究分别概括为下面的第二个和第三个研究问题。

最后,无论是对脑区还是加工时序的观测都还只是对大脑语言加工进行现象上的描述,这些现象背后蕴含着哪些认知、计算机制呢?这是下面介绍的第四个研究问题。

### 1. 语言认知的单元和维度

语言信息非常复杂, 语言学中定义了大量不同大小、不同类型的语言单元, 比如音位、音节、语素、词、词组、语句等。语言认知研究中关心的一个关键问题是, 这些语言单元只是语言学家为了便于研究所提出的一些概念, 还是确实是进行语言理解所依赖的加工单元。在正常的语言理解过程中, 大脑是以怎样的语言单元进行分析? 大脑是否会为不同类型的语言信息(比如语音、语法、语义)构建不同的神经表征? 这些都是语言认知研究所关注的问题。比如, Liberman 等人<sup>[43]</sup>认为音位是语音加工的基本单元, 而且认为音位识别是大脑运动系统的功能, 而 Greenberg<sup>[44]</sup>和 Hickok 等人<sup>[45]</sup>认为音节是更核心的加工单位, 而音位加工可能只在特定任务下进行。再比如, Townsend 等人<sup>[46]</sup>认为大的短语甚至语句是语义理解的基本单元, 但是众多联结主义者认为词是大脑的基本加工单元, 短语和语句结构几乎不影响大脑的语言加工<sup>[47]</sup>。

### 2. 定位不同类型语言信息的脑网络

语言是大脑的功能, 但是这项功能到底依赖于大脑的哪些部分? 神经科学研究发现大脑可以从结构和功能上进行分区, 而最早的关于大脑功能分区的证据来自于失语症研究, 这些研究会在下一节进行介绍。失语症研究以及现代神经影像学研究发现语言不是单一功能, 而是包含了许多功能模块<sup>[48~50]</sup>, 因此当前语言认知研究更加关注于定位具体功能模块所涉及的脑网络。

### 3. 语言信息加工的时间进程和控制

在大脑进行语言加工的过程中, 不同模块的加工次序是怎样的? 比如大脑是不是会先解析语法结构, 然后再理解语义<sup>[51]</sup>? 每个步骤的加工大约需要多少时间? 比如大脑经过多长时间可以识别一个词汇中的不同特征<sup>[52]</sup>? 大脑加工语言的步骤和次序是自动化发生的、一成不变的过程, 还是需要注意力、工作记忆等认知功能的影响和调节<sup>[53]</sup>? 这些也是语言认知研究所关注的内容。

### 4. 语言信息的神经编码形式和计算机制

对脑区和加工进程的研究主要是在现象上描述语言加工。这些现象到底是如何产生的? 从计算的角度来说, 大脑中进行计算的“数据结构”是什么, 又是通过什么算法来操作这些数据结构的? 对加工机制的研究势必涉及数学模型, 研究难度也很大。目前一类模型是试图直接解释大脑的神经响应<sup>[54~56]</sup>, 另一类模型是模拟语言行为, 比如模拟语言习得过程<sup>[57]</sup>。

## 2.3 发展过程和现状

早期的语言认知研究主要来自对失语症的研究, 这些研究大约开始于十九世纪中叶, 主要分析病人脑损伤与语言行为之间的关系。近五十年来, 脑电图 (Electroencephalogram, EEG)、脑磁图 (Magnetoencephalography, MEG)、正电子发射断层成像 (Positron Emission Computed Tomography, PET), 功能磁共振 (functional magnetic resonance imaging, fMRI) 等技术的成熟为研究正常大脑的语言功能提供了有力手段。这些基于行为或者神经科学实验的研究获得了大量成果, 结合上一小节中的研究问题, 这里仅以四方面的研究作为例子进行介绍。

第一方面的研究是关于语言理解的单元。一种极端的观点认为语句(或者大的短语)是加工的基本单元, 听或者读语句的过程只是信息读取的过程, 到达语句边界时才会对信息进行加工和整合; 另一种极端的观点认为语言加工是即时进行的——每一时刻大脑都会对当前时刻获取的信息进行充分加工, 没有必要在一些重要的语言边界进行集中处理。第一种观点的证据是语言理解严重依赖于上下

文,比如 George Miller 发现,在噪音环境下,同一个词汇如果放在语句中就可以被更好地识别<sup>[58]</sup>。再比如,听文章的过程中,如果突然中断文章,让听者回忆之前听到的内容,听者往往只能准确回忆当前语句之内的词汇<sup>[59]</sup>。第二种观点也有很多证据支持。比如,如果播放语音并且让人跟读,有些人可以以 300 毫秒左右的速度跟读,也就是说跟读者自己发出的语音只比听到的语音慢一个词左右。这种情况下,如果听到“tomorrane”这个词,但是根据上文信息这个词应当是明天的意思,则跟读者有较高概率说出“tomorrow”,而不是“tomorrane”;在缺乏上文信息的情况下,跟读者则不会做这种校正处理<sup>[60]</sup>。这一现象说明,跟读者在即时整合上文信息,而不是等到语句边界才进行加工。另一项研究发现,如果一个人在屏幕上看到四个物体(比如马、苹果、桌子、报纸),当听到“这个小孩在骑”的时候,这个人的目光经常会落到“马”这个物体上。这也说明人的语言加工是预测性的,人会根据上文即时生成预期。如果综合上述两种观点,则可以认为大脑既会即时进行预测性加工,并且也会在语句或者短语边界处进行额外整合。

第二方面的研究是关于语言加工所包含的模块,以及这些模块对应的脑区或者脑网络。早期的失语症研究发现当某些脑区由于外伤或者病变原因受损时,语言功能会受到损伤。尤为重要的是,失语症研究发现语言并不是单一功能,而是由不同脑区负责的复杂功能的复合体。比如,布罗卡失语症患者不能产出语言,但是可以听懂语言;威尼克失语症患者可以说出语言,但是听不懂语言。这两种失语症说明语言的产出和理解这两种功能分别由不同的脑区负责,因此可以选择性受到损伤。更为细化的研究发现,有些病人损伤了对名词的识别能力,但是保留了对动词的识别能力;另一些病人则恰恰相反——这说明动词和名词在大脑中的加工方式不同<sup>[61,62]</sup>。甚至有些病人损伤了辨别音素的能力(比如无法辨别/ba/和/da/),但是听觉词汇理解能力基本正常;另一些病人则相反——这说明对音素的辨别与听觉词汇识别涉及不同的脑区<sup>[45]</sup>。这些现象都说明语言理解涉及很多个模块,以至于对特定脑区的损伤只影响部分的语言功能。基于功能磁共振等方法的研究也同样揭示了这种功能分区,这些方法可以在正常人处理不同信息或者执行不同任务的过程中观测不同脑区的激活程度。磁共振研究发现语法与语义加工所激活的脑区有所区别<sup>[63~65]</sup>,即使在语义系统之内处理不同类别的概念时激活的脑区也有所差异,比如看到各种工具时激活的脑区与看到各种动物时激活的脑区不同<sup>[66,67]</sup>。一般而言,失语症研究发现一些关键脑区的损伤可以影响某种功能,但是磁共振研究一般发现该种功能其实涉及分布更为广泛的脑网络。比如传统的失语症研究一般认为颞叶损伤更可能导致名词理解问题,额叶损伤更可能导致动词理解问题,但是近期磁共振研究表明其实动词和名词的加工都涉及非常复杂的脑网络,而且网络内部整体的连接性质与两类词汇的加工相关<sup>[68,69]</sup>。

第三方面研究是关于词汇的识别与加工。大量研究发现,词汇识别是一个并行的过程。比如听到英语音节/kæp/的时候,一般认为大脑会并行激活这个音节开头的词汇,比如 cap, captain, caption 等。这些词中词频较高或者与上文更为一致的词汇的激活更强。心理学家是如何得到大量词汇并行激活这一结论的呢?一类经典的实验采用了跨模态启动效应进行研究,这些实验发现听到/kæp/之后,人对视觉呈现的 captain, caption 等词汇的识别都会比听到其它音节(比如/da/)之后更快。而且,听到一个词之后,与这个词语义相关的词汇也会得到激活,比如听到 captain 之后,人对 ship 等词汇的识别也会更快<sup>[70]</sup>。从神经科学的层面上,词汇会诱发一个潜伏期约为 400 毫秒的脑电响应 N400, N400 的幅度与词频及上文信息密切相关。比如,ship 诱发的 N400 的幅度依赖于前面一个词,如果前面一个词是 captain, ship 诱发的 N400 幅度相对较小;如果前面一个词是不相关的词汇(比如 apple), ship 诱发的 N400 幅度相对较大<sup>[71]</sup>。一般认为,词汇识别是一个并行的过程,在听到词汇的部分信息之后,大量词汇都被激活,但是随着听到信息的增加,与新的信息不符的词汇又被抑制,直至大脑最终确定一个可能的词汇<sup>[70]</sup>。大脑之所以并行加工可能的词汇是因为语言中充满歧义而且信息呈现速度

又非常快，如果等到所有歧义都消除之后再开始加工，则不但可能导致反应速度过慢，而且可能导致信息已经超出了大脑的工作记忆容量而被遗忘。类似的问题在语句加工中更加常见，比如“the horse raced past the fence fell”这句话，看到最后一个词之前我们会认为 raced 是句子的谓语动词，但是看到最后一个词之后我们会发现之前的理解不对，fell 才是句子的谓语动词。这类包含歧义，以至于理解过程中要变更对语句结构的分析的语句称为“花园幽径”语句。目前，有的理论认为人脑也会同时构建多种可能的句子结构，然后不断进行筛选；但是另一些理论认为大脑会先构建一种最可能的句子结构，如果发现这个结构不对再对句子重新分析。

第四方面的研究是关于复杂环境下的语音理解。上世纪五十年代，英国科学家 Colin Cherry 发现注意力对复杂环境下的语音理解起到至关重要的作用<sup>[72]</sup>。Cherry 的研究以及后续研究发现，如果在实验中同时播放两段不同的语音（两段语音来自不同的说话人，或者不同的空间方位）并且让听者把注意力集中于其中一段语音，那么听者可以很好地理解他们所注意的语音，但是事后无法回忆他们没有注意的语音的内容。心理学家往往还会定量分析各种因素对语音识别的影响，比如在不同噪声强度下测量语音的识别率，然后绘制语音识别率随噪声强度变化的心理曲线。美国科学家 George Miller 发现语音识别率不但与噪声强度、听者注意力有关，也与听者的先验语言知识有关<sup>[58]</sup>。在噪音环境下，相对于随机词串，人可以更好地识别合乎语法的句子。认知神经科学实验进一步表明，注意力和先验知识都可以直接调节听觉皮层对语音声学特征的加工，在这种调解之下，听觉皮层的神经活动主要编码了被注意的语音。上述研究表明了注意力和先验知识在语言理解中的重要作用。

上面四方面的研究表明预测性加工、语言结构加工、并行加工、注意力和先验知识等都是人类语言认知的重要特点。

## 2.4 研究方法

宏观来说，认知科学、生命科学的研究可以分为假设驱动的研究和数据驱动的研究。语言认知的研究也可以粗略划分为这两类研究。

### 1. 假设驱动的研究

大多数语言认知研究是假设驱动的研究 — 研究人员在实验前会明确实验所要验证的假设（一般简称 H1）以及与其对立的假设（H0），并且明确怎样的实验结果与 H1 或者 H0 的预期一致。如果最终实验结果与 H1 的预期一致，与 H0 的预期不一致，则 H0 被证伪（即支持 H1）；相反的，如果最终实验结果与 H0 的预期一致，与 H1 的预期不一致，则 H1 被证伪。比如，研究人员想要验证的假设是听觉皮层不仅编码了语音的声学特征，还编码了音位信息。根据这个假设，研究人员要设计实验来区分声学特征与音位特征，然后分析后者是不是影响了听觉皮层的响应。下面用两个具体例子来说明假设驱动的研究。

上文提到，大脑加工语言的单元是一个有争议的问题。一项研究提出的假设是（H1）：大脑可以并行编码多个层级的语言单元，而且编码某个语言单元的神经活动应当在时间上与这个语言单元同步 — 语言单元出现时，编码这个单元的神经活动也出现；语言单元结束时，对应的神经活动也结束。与这个假设相对立的假设（H0）是大脑只根据单一层级（比如词）进行加工或者不同层级的语言单元并不表现为与语言单元同步的神经响应。如果这个假设成立，那么编码音节、词、词组、语句等不同大小的语言单元的神经活动将具有不同的更新速率。比如，如果语音中每秒钟有 4 个音节，那么音节的响应也会每秒钟变化 4 次。如果每 2 个音节组合成为一个词组，每 2 个词组构成一个语句，那么词、词组的神经响应应当每秒钟更新 2 次，语句的神经响应应

当每秒钟更新 1 次。于是,研究人员按照上述思路设计了实验,实验发现听语音的过程中人脑的 MEG/EEG 响应确实包含 4Hz、2Hz 和 1Hz 的成分,对应于假设的音节、词组、语句的神经响应 [73]。这一实验结果支持了 H1 假设。

另一个例子是关于词义的研究。一个词汇的语义可以通过两种方式习得。一种方式是直接建立词汇与其所指代的客观实体之间的关系,比如看到番石榴这种水果,然后被告知这是“番石榴”。这种学习方式直接建立了“番石榴”这个语言符号与它所指代的物体的感觉特征(视觉特征、味觉特征等)之间的关系。另一种方式是通过文字描述习得词汇的意义,比如看到字典上说:番石榴为桃金娘科植物,果实可吃等。该研究提出的假设(H1)是:通过上述两种方式习得的语义分别在大脑不同区域进行编码。与之相对的假设(H0)是一个词汇在大脑中的表示与习得途径无关。为了区分这两种习得方式,该研究比较了正常视力者与先天盲人对颜色的加工——这两组人群都可以通过语言来获得颜色的概念,但是只有正常视力者可以直接建立颜色词与视觉颜色信息之间的对应关系。这项研究发现一些脑区对颜色的编码方式在两组人中一致,但是另一些脑区对颜色的编码方式只在正常视力者中存在。这一结果也支持 H1 假设 [74]。

## 2. 数据驱动研究

假设驱动的研究往往是针对性很强的研究——针对特定假设专门设计实验来加以验证。与假设驱动的研究相对的是数据驱动的研究。数据驱动的研究是探索性的,并不是先提出假设,而是通过收集实验数据来探索有怎样可能的结果。假设驱动的研究实验目的明确,所以较为容易获得稳定的结果,比如上述第一项实验为了验证神经活动与层级语言结构同步这一假设,采用恒定速率来播放语音,从而使得数据的分析可以得到简化(只需要观察频谱中特定频率的响应);第二项实验为了区分词义的两种获取方法,选取了两组人群进行比较。但是,假设驱动的研究和数据驱动的研究往往很难严格区分,没有探索性的研究则很难提出假设;没有假设则很难确定分析数据的哪个方面。比如,上述两项假设驱动的研究也都含有数据驱动的成分,第一项研究并没有确定在哪些 MEG/EEG 通道可以获得响应,第二项研究也没有事先假设将在哪个脑区发现实验现象。

### 2.5 已有研究的局限性

语言认知研究初步揭示了人脑语言理解的部分规律,但是距离真正解析人脑语言理解机理还有很大差距。当前面临的主要问题是,在理论上侧重局部问题上的定性解释,在方法上依赖小样本和严格的实验控制,导致了研究结论的生态性和全局性不足,概况为如下三点:

#### 1. 缺乏对定量机制的探讨

大多数认知科学研究是在现象层面进行描述,即使对于机理的讨论往往也是定性、主观的。比如上文提到大脑处理一个词汇时会产生 N400 这个脑电响应。大量研究针对这一响应进行了研究,澄清了各种性质的上文信息如何影响 N400 的幅度,但是这些研究只是说明上文信息可以影响 N400 响应,没有回答这种影响反映了怎样的计算机制。认知科学文献会讨论 N400 的多种产生机制,比如一种假说认为如果 N400 代表了当前词汇是否可以被大脑预测——可以预测的词产生较小的 N400;另一种假说认为 N400 代表了当前词汇与上文整合的难易程度——如果容易整合则 N400 幅度较少。但是这些假说是一些定性的语言描述,并没有明确大脑的预测和整合反映了怎样的计算机制。

## 2. 往往针对特定的语言现象

认知科学实验往往采用严格控制的实验设计, 针对特定的, 甚至是非常细节的语言现象展开研究。由于要严格控制实验变量, 所以实验中的语料往往趋于一致, 因此实验结论很可能只适用于实验中所涉及的高度一致的语料。Poeppel 与 Embick<sup>[75]</sup> 曾经指出语言学和神经科学研究之间存在研究尺度的不匹配, 语言学往往关心非常精细的问题 (比如某个词的用法、某个语句的句法结构到底应当如何划分), 而神经科学关心相对宏观的问题 (比如大脑的哪部分加工语法)。但是, 即使神经语言学的研究往往也只采用相对一致的、典型的语料进行研究, 因此结论的普适性不强。

## 3. 研究结论难以整合

与上一问题密切相关的另一个问题是严格控制的实验导致研究趋于碎片化, 一项研究只关心一种特定的语言现象。如果每项研究关注一种语言现象, 而语言中只包含有限的几种语言现象, 那么就可以通过整合不同的局部研究得到一个总体结论。但是, 由于语言过于复杂, 无法统一地划分为若干基本现象, 而且实验手段过于多样, 导致各种研究结论之间整合非常困难。比如, 认知实验研究发现不同类型的语言材料可以激活不同的脑区、诱发不同类型的脑电响应, 但是如果把这些研究汇总起来, 能够告诉我们大脑是如何一步一步地理解一个哪怕非常简单的语句吗? 其实并不能。

# 3 关于语言理解的计算方法

## 3.1 语言计算的定义

本文中提的语言计算是指机器对语言的理解过程, 强调研制具有表示语言能力的模型, 建立计算框架来实现这样的语言模型。以中文为例, 语言计算的过程包括对文字的识别和表示, 字、词、短语、句子和篇章的结构及语义解析, 文本符号与外部世界的关联分析等, 最终达到使机器理解语言的目的。以下我们所说的语言计算可与自然语言处理中的基础问题研究 (也称作自然语言分析<sup>[76]</sup>) 类比, 如词法、句法、语义和篇章分析以及知识表示与计算等, 并不涉及应用技术研究。另外, 目前常用的相关术语还有“计算语言学”、“自然语言理解”、“自然语言处理”, 相关定义的解释可以参考《计算机科学技术百科全书》<sup>[76]</sup> 和《统计自然语言处理》<sup>[77]</sup> 一书。

## 3.2 主要研究问题

要想让机器理解自然语言, 首先要把语言文字本身的信息编码为计算机可以处理的形式, 这一过程被称为文本表示任务。为了进一步解析文本内容所表达的信息, 需要分析文本的结构和语义信息, 即进行结构分析和语义分析任务。至此, 机器知道了不同语言符号间的关系, 还需要将语言符号与外部世界和知识相关联, 才能像人类一样理解自然语言。下面分别介绍这几个主要研究问题。

### 1. 文本表示方法

语言是由小的元素分层递归地组合成较大的单元, 依次组成词汇、短语、句子和篇章。在语言表达中, 词汇是最基本的语义单位, 而词汇的组合需要依据特定的规则, 有限的规则可以将不同概念进行组合从而构造出无穷尽的文本单元。如何设计词汇语义的编码形式, 如符号<sup>[78]</sup>、函



数<sup>[79]</sup>、向量<sup>[80]</sup>、张量<sup>[81]</sup>等?如何构建高效的词汇语义学习方法<sup>[82~84]</sup>?如何将词汇编码进行整合从而构成更大粒度文本单元的含义<sup>[85~88]</sup>?这些都是语言计算的关键问题。

## 2. 结构分析方法

结构分析一般分为句法结构分析和篇章结构分析任务。其中,句法结构分析研究的是句子中词语之间的组合和依赖关系,篇章结构分析研究的是段落或篇章中句子之间的组合和依赖关系。这两类分析可以消解输入文本中结构的歧义,分析输入文本的内部结构,也为文本的语义分析提供了结构化信息,被认为是实现语言理解的重要环节<sup>[89~91]</sup>。这个方向的主要研究问题包括如何设计或选择语法的形式化规则以及如何设计自动分析算法。这类工作中代表性的有 Klein 等<sup>[92]</sup>提出的基于规则的短语结构分析方法和 Chen 等<sup>[93]</sup>提出的基于神经网络的依存结构分析方法。

## 3. 语义分析方法

对于不同的语言单位,语义分析的任务不同。对于词汇来说,语义分析关注如何进行词义消歧和如何识别词汇间的语义关系(包括反义、同义、部分整体和事件关系等);对于句子来说,语义分析包括语义角色标注、语义解析、文本间语义相似性计算和蕴含关系的识别;对于篇章来说,语义分析包括如何进行指代消解和识别篇章句间关系。上述语义信息和语义关系的识别和计算是理解文本内容的基础,如何构建高效的语义分析方法<sup>[94~96]</sup>是目前语言计算的难题。

## 4. 知识表示与符号关联方法

本文的知识泛指世界知识、历史知识、常识性知识、各学科门类的专业知识等。知识的表示就是对知识的一种描述。目前的模型将知识以符号形式<sup>[97]</sup>或分布式向量<sup>[98]</sup>进行表示,通过检索或映射到统一的表示空间中实现语言符号和知识的关联。其中,如何设计知识的编码形式并自动学习知识的表示是这类研究的关键<sup>[99~101]</sup>。另外,无论是人还是机器,要想理解语言符号中编码的含义,都需要将其和世界知识进行关联。否则就如同“中文房间”所描述的一样,房间中的人不会中文,无法真正的理解接收到的中文信息,但是他可以使中文母语者认为他能流利地说中文,营造出一个智能的印象<sup>[102]</sup>。因此,如何将语言符号和知识进行关联也是语言计算研究的核心问题。

### 3.3 发展过程和现状

文本是序列化和结构化的语言表达。如何表示文本含义、对文本语义和结构进行自动化分析是语言计算研究的关键环节,也一直是机器语言理解面临的一个主要挑战。几乎所有的自然语言处理任务,例如机器翻译、自动问答和人机对话等,都依赖于对输入语言序列的语义表示和计算。

自然语言处理几十年的发展历程中,文本表示方法经历了从离散符号表示到连续向量表示的体系转变。离散符号表示将词语视为离散的符号,每个词语可以表示为维度等于词表规模的 one-hot 向量,其中某一维为 1 其余维都是 0。这种表示体系下,句子和篇章通常采用词袋模型来表示。1954 年 Harris 在《Distributional Structure》的文章中提出词袋的概念,在随后的几十年中词袋模型一直是文本表示的主流模型<sup>[103]</sup>。这种基于离散符号的文本表示方法只能采用字符串匹配的方式提取特征并计算语言单元之间的相似性,从而容易导致数据稀疏问题,并且无法捕捉词语之间的语义相似性。分布式连续向量表示便于语义计算和度量,在理论上可以解决词语、句子和篇章之间的语义鸿沟问题。Harris 和 Firth 分别于 1954 和 1957 年提出并明确了词语的分布式假说:一个词语的语义由其上下文决定,即上下文相似的词语其语义也相似<sup>[103,104]</sup>。矩阵分解和神经网络是学习词语分布式向量

表示的两大主要模型,其中,神经网络是近年来学习分布式向量表示最主流模型。2003年,Yoshua Bengio 等人<sup>[105]</sup>提出神经网络语言模型,采用低维连续的实数向量表示每个词语,并以此为基础学习  $n$  元语法模型,标志了分布式文本表示的开端。Tomas Mikolov 等人<sup>[80]</sup>在2013年提出包括 CBOW 和 Skip-gram 两种模型的 Word2Vec 方法,极大简化了词语的分布式向量学习方法,从而可以充分利用海量无标注文本数据高效学习词语的低维连续向量表示。2017年,谷歌团队提出的 Transformer 模型<sup>[106]</sup>,通过词汇间两两计算的方式更加高效地组合词汇的语义从而获得文本的语义表示。此后,基于 Transformer 架构的大规模预训练模型如 BERT<sup>[107]</sup>、TransformerXL<sup>[108]</sup>、GPT3<sup>[109]</sup>等在各项语言处理应用中大放异彩,进一步奠定了分布式向量表示的统治地位。

分布式文本表示模型极大地便利了自然语言的表示和计算,成为深度学习应用于自然语言处理任务的基石,推动了文本理解和机器翻译等应用的突破性发展。但是,现有方法缺乏对细粒度文本语义和结构化信息的建模,无法有效处理词汇的歧义、反义、引申义以及句子和篇章的结构歧义等语言现象<sup>[110~112]</sup>。而语义分析、句法分析和篇章分析方法研究如何表示语言单元的含义、组合和依赖关系,为文本表示提供了结构化信息。

语义解析是语义分析中最具代表性的任务,它研究如何将自然语言解析成完整的,机器可以识别或计算的语义表示(如逻辑表达式)。具体包括如何获取句子中词汇的语义,词汇间的语义关系,以及将词汇表示组合为句子的语义表示<sup>[113]</sup>。同多数自然语言处理任务一样,语义分析也经历了基于规则的、统计的和神经网络的方法的转变。尽管目前基于神经网络的语义分析模型的性能有了很大的进步,但是该方法仍然强烈依赖大量人工标注的数据,通过学习统计规律进行预测,无法真正实现语义理解。除了上述语义分析外,理解文本的含义离不开对其结构的分析。在句子级别,结构分析指将输入的单词序列分析出合乎语法的句子的句法结构,主要分为短语结构分析和依存结构分析。短语结构分析将句子分析为层次化的短语结构树,理论基础是短语结构文法,这是1957年 Chomsky<sup>[114]</sup>创立的语言转换生成理论的一部分。依存结构文法主要用于刻画词语之间的语义依赖关系,最早于1959年由 Tesnière 提出<sup>[115]</sup>。在篇章级别,结构分析研究的是段落或篇章中句子之间的组合和依赖关系,目标是从整体上分析篇章结构而从理解篇章含义。计算语言学家一般认为句法和篇章分析是实现语言理解的必要环节,只有正确地分析了文本的结构,才能理解文本的含义。但是,近几年在各项语言处理任务上的结果表明,显式的建模文本结构并不能显著提升模型的性能<sup>[116]</sup>。这使得研究人员重新思考文本结构分析的作用,以及现有语言模型是否可以隐式地学习到某种句法规则,人为定义的句法结构对于机器语言理解是否必要等问题。

语言理解除了需要对词、短语、句子等浅层含义进行表示和分析之外,还需要更高层次的表达能力,如与“知识”的联系,感知上下文和推理的能力。传统的知识表示方法采用基于逻辑符号的计算框架,如以 Prolog 为代表的基于一阶谓词知识表示的逻辑推理语言和以马尔可夫逻辑网为代表的概率图模型等<sup>[117]</sup>。这些方法擅长进行精确推理和计算,但是在知识不完备情况下缺乏泛化和近似语义计算能力。近年来,随着深度学习的发展,知识表示逐渐采用基于神经网络的方法来学习实体和关系的表示。这种基于表示学习的方法泛化性能更好,更有利于信息处理中广泛存在的不确定性计算问题,但是无法进行精确的语义计算和推理<sup>[118~121]</sup>。因此将基于符号逻辑的精确语义计算与基于表示学习的近似语义计算进行结合是未来知识表示方法的研究趋势。

### 3.4 研究方法

同计算机科学领域多数研究方向一致,自然语言方法主要分两大阵营,一种是理性主义方法,另一种是经验主义的方法,或称数据驱动的方法。基于规则的方法是理性主义方法的代表,这种方法认

为人的很大一部分语言知识是与生俱来的,由遗传决定的;数据驱动的方法包括统计机器学习方法和神经网络方法(或称深度学习方法),这种方法假定人类复杂的语言结构是可以通过后天训练学习到的。以下分别对这两种方法进行简要介绍。

### 1. 基于规则的方法

由于自然语言在本质上属于人类社会因交流需要而产生的符号系统,其规则和推理特征鲜明,因此早期自然语言处理的研究首要采用规则方法。该方法通过对一些代表性语句或语言现象的研究得到对人的语言能力的认识,归纳语言使用的规律,以此分析、推断测试样本的预期结果<sup>[77]</sup>。此处以上下文无关语法在句法分析问题上的应用为例介绍基于规则方法的基本原理<sup>[122]</sup>。例如,我们有如下上下文无关语法:

$$G = (V_n, V_t, S, P)$$

$$V_n = \{S, NP, VP, N, V, Det\}$$

$$V_t = \{\text{一个, 小孩, 追赶, 一只, 猫}\}$$

$$S = S$$

$$P : \{S \rightarrow NP VP, NP \rightarrow Det N, VP \rightarrow V NP, Det \rightarrow [\text{一个}], Det \rightarrow [\text{一只}], N \rightarrow [\text{小孩}], N \rightarrow [\text{猫}], V \rightarrow [\text{追赶}]\}$$

基于上述规则,通过自顶向下或自底向上分析法可以构造句法树,此处仅以自顶向下分析法为例进行说明。首先,从初始符号 S 开始,自顶向下地进行搜索,选择语法 P 中适用的规则来替换搜索目标,并用语法规则的右边部分同句子中的单词相匹配,如匹配成功,则抹去这个单词。然后,继续对输入句子的遗留部分进行搜索,一直分析到句子结尾为止。具体来说,在上述实例中,搜索步骤为:

- (a) 选择  $S \rightarrow NP VP$  规则,无匹配单词,遗留句子“一个小孩追赶一只猫”;
- (b) 选择  $NP \rightarrow Det N$  规则,无匹配单词,遗留句子“一个小孩追赶一只猫”;
- (c) 选择  $Det \rightarrow \text{一个}$ ,匹配“一个”,遗留句子“小孩追赶一只猫”;
- (d) 选择  $N \rightarrow \text{小孩}$ ,匹配“小孩”,遗留句子“追赶一只猫”;
- (e) 选择  $VP \rightarrow V NP$ ,无匹配单词,遗留句子“追赶一只猫”;
- (f) 选择  $V \rightarrow \text{追赶}$ ,匹配“追赶”,遗留句子“一只猫”;
- (g) 选择  $NP \rightarrow Det N$ ,无匹配单词,遗留句子“一只猫”;
- (h) 选择  $Det \rightarrow \text{一只}$ ,匹配“一只”,遗留句子“猫”;
- (i) 选择  $N \rightarrow \text{猫}$ ,匹配“猫”,无遗留句子。

根据这样的搜索过程,我们可以把这个句子“一个小孩追赶一只猫”的句法树表示为图 2 所示的结构。

### 2. 数据驱动的方法

人类语言毕竟不是形式语言,规则模式往往隐式存在语言中,规则的制定并不容易;另外,自然语言的复杂性使得规则很难既无冲突又能涵盖全部的语言现象。数据驱动方法省去了人工

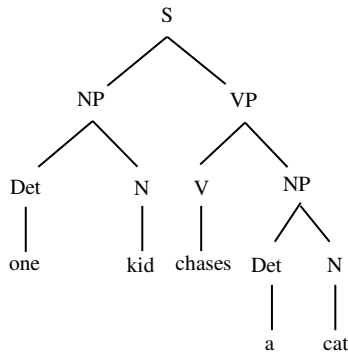


图 2 句法树示例

Figure 2 Example of a syntax tree

编制规则的负担，在模型生成方面自动生成特征或者评估特征的权重，具有较好的鲁棒性。一般来说，数据驱动的方法包括统计方法和神经网络方法，下面分别对两种方法进行介绍。

基于统计的语言计算方法使用大规模真实语言数据，往往需要借助人工的帮助（标注数据和筛选特征等），利用统计学方法发现语言使用的规律及其可能性（概率）大小，进而以此为依据来计算测试样本的可能结果<sup>[77]</sup>。此处以语言模型为例介绍统计方法的基本原理，代表性的有  $n$  元语法模型。

语言模型的目标是计算出字符串作为一个句子出现的概率，也就是计算句子  $s$  中每个词在不同历史情况下出现的概率的乘积： $P(s) = p(w_1) \times p(w_2|w_1) \times p(w_3|w_1w_2) \cdots p(w_l|w_1 \cdots w_{l-1})$ 。其中，产生第  $i$  个词  $w_i$  的概率是由已经产生的  $i - 1$  个词  $w_1w_2 \cdots w_{i-1}$  决定的。当简化上述方法，仅利用前面  $n - 1$  个单词预测下一个词汇时，就是  $n$  元语法模型。例如，二元语法模型计算概率  $p(\text{一个小孩追赶一只猫}) = p(\text{一个} | \langle \text{bos} \rangle) \times p(\text{小孩} | \text{一个}) \times p(\text{追赶} | \text{小孩}) \times p(\text{一只} | \text{追赶}) \times p(\text{猫} | \text{一只})$ 。其中， $\langle \text{bos} \rangle$  为开始符。一般采取最大似然估计来计算  $p(w_i|w_{i-1})$  的条件概率： $P(w_i|w_{i-1}) = \frac{c(w_i, w_{i-1})}{\sum_{w_i} c(w_i, w_{i-1})}$ 。

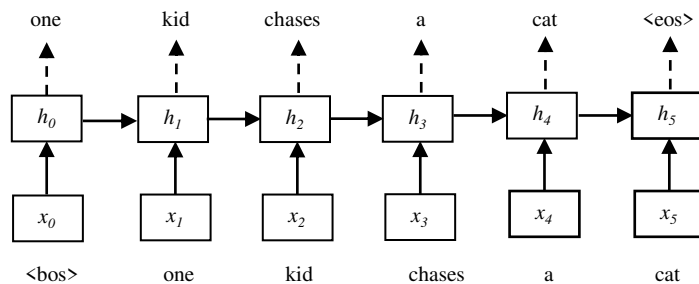


图 3 循环神经网络示例

Figure 3 Example of a recurrent neural network

基于神经网络的方法主要研究如何设计任务相关的神经网络结构和优化神经网络参数。此处，我们以文本表示问题为例，通过循环神经网络模型为例来介绍基于神经网络方法的基本原理。循环神经网络是神经网络的一种，擅长处理序列信息和长距离依赖关系。例如，语料中有以

下内容“一个小孩追赶一只猫,……”。为了对上述文本进行表示,需将文本中语句依次输入到循环神经网络中。如图3所示,循环神经网络模型依次处理句子中的每个词汇,在时刻 $t$ 将对应词汇 $w_t$ 编码为一个实值向量 $x_t$ ,和上一个时刻产生的隐层向量 $h_{t-1}$ 进行组合,得到这一时刻的隐层向量:

$$h_t := \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b),$$

其中,符号“:=”表示“定义为”; $\sigma(z) = 1/(1 + \exp(-z))$ ;  $W_{xh}$ ,  $W_{hh}$  和  $b$  是模型参数。模型的目标函数是最大化上文预测下文词汇的概率,常用的目标函数是最小化模型的交叉熵损失:  $\text{Loss} = -\sum w_t \log y_t$ 。其中,  $w_t$  是真实的下文词汇,  $y_t$  是利用隐层向量  $h_t$  预测的下文词汇:  $y_t = \text{Softmax}(V_{hy}h_t + c)$ , 其中  $V_{hy}$  和  $c$  是模型参数。

利用上述目标函数在大规模无标注文本上学习到模型的参数,使得模型可以通过上文词汇预测下文词汇。在此过程中,语言模型得到了词汇的表示向量,同时,句子的表示向量可以取模型最后一个时刻得到的隐层向量或者是所有隐层向量的最大或平均池化结果。

### 3.5 已有研究的局限性

现有语言计算模型远没有达到像人一样理解语言的程度,当前面临的主要问题是,模型结构缺乏理论基础且参数训练依赖大规模计算资源,具体可以概括为如下四点:

#### 1. 文本表示形式单一

现有文本表示模型不区分不同类型信息(如不同词性的词汇、不同粒度的文本单元),统一将其编码为相同维度大小的稠密向量。这种编码方式在构建神经网络模型时十分高效,但是由于忽略了不同类型文本自身信息量的大小,要想编码所有信息,就必须对所有类型文本统一使用最大信息量的编码方式。这种采取“参数数量换模型结构”的方式,使得模型变得十分庞大,存在大量参数冗余。

#### 2. 模型缺乏可解释性

尽管深度学习方法大幅度提升了自然语言处理中各种任务的性能,但是由于模型中向量维度的含义不可解释,无法解析网络中的每一次运算对语言单元进行了何种操作,也无法给出模型得到错误结果的原因。这使得模型得到的结果并不可靠,同时也阻碍了模型结构的设计和性能的进一步提升。相反,如果模型具备可解释性,可以像人一样给出得到某个结论的原因,那么就可以针对性的改进模型结构,进而提升模型性能。比如,在判断两个句子是否表达了相同的含义时,模型可以给出是或否的结论,同时给出两个句子中哪些文本片段显著影响了这个结论。这些都可以作为判断模型合理性的依据。

#### 3. 缺乏自主进化学习的能力

现有方法针对每一种任务构造训练数据集,采用“训练—测试”的开发方式,无法直接利用新的训练样本修正已训练的模型,在某一任务中训练的模型也难以应用于其他任务中。不同的语言任务都需要模型具备语言表征和理解的能力,因此新任务可以利用已经训练好的模型,在此基础上学习新任务特异化的信息。类比人的自主进化学习能力,即从简单的任务开始学习,在此基础上不断学习新的任务和修正已有的知识,一个智能系统也应该具备持续不断学习的能力,以达到模型自主进化的效果。

#### 4. 依赖大规模单一模态训练数据

现有语言计算模型严重依赖于训练样本的质量和规模，对于训练语料中没有出现过的词汇、语言结构和语言表达束手无策。反观人类在学习语言概念时，往往是从多种模态样本中获取信息，并且仅通过几个样例就可以学会一个新的词汇或语言表达。例如，学习“长颈鹿”这个概念时，我们从文本中得知“它是世界上最高的哺乳动物，显著特征有长脖子等”；当看到一张画有“长脖子、长腿、身体上有斑点图案的动物”时，即使之前从未见过，我们也可以快速识别出这是“长颈鹿”。这说明多种模态信息是相互补充和验证的，因此综合利用多种模态信息开发小样本学习算法势在必行。

## 4 语言认知与语言计算对比

由上述章节可以看出，语言认知和语言计算领域定义的人和机器语言理解的过程基本一致，都是从基本语言单元的编码开始，继而将其组合为更大粒度的文本单元，最终与外部世界相关联。不同的是，两者的实验方法和研究问题的关注点有很大区别。语言认知研究多采取“提出假设—设计实验—数据验证假设”的实验手段，而语言计算研究大多采取“采集数据—构建模型—任务验证模型性能”的方法。在研究问题上，语言认知研究偏向于揭示问题背后的原因，而语言计算研究则重点关注解决问题的方法。表 1 总结对比了两个领域的主要观点和概念。可以发现，语言认知研究大脑语言理解的基本单元和维度，其在大脑中的空间表征形式和时间进程，以及世界知识和认知功能对于语言理解的调控作用，着重强调“是什么”和“为什么”的问题。语言计算研究如何让计算机高效地切分和表示词汇，将词汇表示组合为句子和篇章表示，以及分析文本中各个元素的功能和关系，主要关心的是“怎么做”的问题。

上述差异为语言认知和语言计算的结合带来了巨大的挑战和机遇。挑战在于语言认知和语言计算的研究相互分离。语言认知的研究主要关注人类语言背后的认知规律及其神经基础，由于技术条件的限制，目前仍然偏重于定性的分析和宏观规律的探索，缺乏可应用于实践的、全面且定量的规律或模型。比如，对于词汇语义的表征，语言认知领域虽然已经发现了诸多语义表征维度，揭示了其大体上的神经关联，并初步建立了一些理论模型，但由于技术上的限制，仍无法精确记录单次出现的词汇所诱发的神经活动，因而对于语义信息的底层神经表征规律和计算方式也很难被直接揭示。同样的，计算语言学家选择研究问题时重点关注实际应用效能，往往忽略了对语言最本质规律的研究。构建语言计算模型时仅考虑其在下游任务中的性能，而忽视了模型的可解释性和类人相关的特性。因此，作为一个不具备人类特性的“黑盒子”，语言计算模型也很难用于人脑语言理解的建模中。

下面以词汇表示问题为例，进一步阐释上述挑战。在人脑词汇（概念）表征问题上，Wang 等人<sup>[123]</sup>研究了抽象词汇在大脑中的表征。他们通过构造两种不同的词汇表示类型（基于语料库统计共现的词汇表示和基于语义特征的词汇表示），测试了两种大脑抽象概念表征的经典认知理论——两个理论分别认为抽象词汇的表征是以语言形式通过上下文关联来表示的，或者由语义特征来表示的。实验结果表明，两种类型的词汇表示在大脑中都发现了显著性效应，不同的是基于语料库的抽象词汇表示与负责高级语言加工的脑区相关，而基于语义特征的抽象词汇表示呈现分布式表征的特点，与大脑多个脑区相关。因此，他们总结抽象词汇在大脑中的表征分为两种模式，并且由不同的脑网络负责。这个研究揭示了大脑编码抽象词汇的语义维度可能同时包括语言学 and 语义特征。这对于进一步解析大脑理解语言非常重要，但是这种分析过于宏观，因而无法直接指导计算模型的构建。计算机科学家希望了解大脑编码词汇概念时需要的具体语义维度，以及编码形式。比如，“知识”这个概念，由大脑中

表 1 语言认知和语言计算领域的主要观点、概念  
**Table 1** Main concepts in language cognition and language computation areas

	Language cognition	Language computation
Language processing unit	Language is organized according to different levels of structure (such as morphemes, words, phrases, sentences, discourses, etc.), with different levels of information (phonetic, grammatical, semantic, etc.)	Word is generally used as the basic unit, and they are integrated through language models such as neural networks
Word representation	The neurons or neural networks related to the word are activated. It is generally believed that the neural representations of multiple words can be activated at the same time	Encoding words into a form that can be processed by computers (such as symbols, vectors, matrices, graphs, etc.)
Word form analysis	Morphology information encoding	Remove the affix to get the root (plays → play) or transform the complex form of the word into the most basic form (are → be)
Information integration	The process of assigning syntactic structure and semantics to the corresponding input words	Analyze the semantic and syntactic information and their relationships, then combine word representations to form the representation of sentences and discourses through a composition function
Multimodal information	World knowledge (common sense) and language understanding scenarios can affect language processing, which may involve the interaction of brain language networks with sensory and motor networks	Associate language symbols with other modality information
Task effect	Language comprehension is regulated by cognitive functions — if cognitive functions such as attention and working memory are adjusted through experimental tasks, cognitive neural processing of language will be affected	The parameters of the computational language model are determined by the objective function — once the objective function of the task changes, the parameters of the computational language model will change accordingly

哪些神经元进行编码? 这些神经元分别代表了什么信息? 这些神经元之间存在什么关系? 它们的连接

模式是什么? 上述问题对应的答案可以直接启发构建新的词汇表示模型架构。

相反的, 语言计算模型在编码词汇含义时, 是将词汇表示为维度不可知的实值向量, 通过向量间的关系编码词汇的信息。这样做的目的是让计算机可以高效地处理语言符号, 从而完成各项语言任务。虽然这种维度不可解释的向量化表示方法对于计算模型完成语言任务十分重要, 但是并没有揭示语言本身的规律, 对于大脑如何编码词汇概念这一问题也没有直接的启示。认知科学家更希望了解的是, 利用哪些特征构造出的语义向量可以解释人的行为学数据? 计算模型中编码了什么信息、进行了何种操作导致其在下游任务中的出色表现, 这种计算过程是否也可以解释人的语言处理机理? 计算模型中哪个计算模块是对于语言建模必不可少的, 人脑中是否也有对应的处理模块? 等一系列有助于解释大脑运行机制的问题。

综上, 语言认知和语言计算领域重点关注的研究内容和思考方式不同。正因如此, 笔者认为两者的差异可以为双方带来新的启示和见解。例如, 人脑在语言理解过程中, 不仅自下而上将词汇编码组合为句子和篇章的编码, 而且利用注意力、工作记忆等认知功能自上而下对语言编码过程进行调控。相反, 语言计算模型是一个静态过程 — 不受外部环境影响, 且对一段文本计算一个固定的编码结果。而如果可以借鉴人脑动态的编码机理, 融入人脑记忆、注意力等模块构造新的语言计算模型, 这样模型也许会学习到更多通用知识, 更容易迁移到另一个场景中。又如, 已有实验表明, 在机器语言理解过程中, 加法是将词汇表示整合为短语表示时非常有效的组合方式。这种从数据中自动学习到的语言组合规则也许可以从底层计算方式上为人脑词汇组合过程的研究提供一些新的思路。更多两者结合的思路会在第 5 节和第 6 节进行介绍。

## 5 认知与计算的融合

近年来, 越来越多的研究人员开始关注跨越认知科学和计算机科学领域进行交叉学科研究。下面介绍语言认知和语言计算领域相互启发和融合的相关工作。

### 5.1 利用语言计算方法的语言认知实验

近年来, 越来越多的研究人员开始借助语言计算方法, 在自然文本的实验刺激条件下来研究人脑语言理解过程。这类方法为建立精确到词的神经表征模型提供了极大的便利, 进而, 不仅可以用于分析传统实验数据, 更在分析自然语言加工数据时展现了巨大的潜力。具体来说, 这类方法采集词汇、句子或者篇章的神经活动数据, 然后使用语言计算模型编码实验刺激, 最后利用编码后的刺激研究大脑语言理解的问题。下面介绍此类方法的典型工作。

在大脑如何表征概念语义的问题上, 2008 年 Mitchell 等人<sup>[54]</sup>在《科学》上发表文章, 他们发现人在阅读名词时的 fMRI 数据可以由某些动作词汇的统计规律建模得到。具体来说, 他们采集了被试阅读 60 个名词刺激 (图片 + 词汇文本) 时的 fMRI 数据, 在包含万亿字符的语料库中分别计算这 60 个名词和 25 个感觉运动相关的代表性动词 (如, 看、听、说、吃等) 的共现频率, 并当作名词的 25 维表示向量。然后利用留二交叉验证 (leave-two-out cross validation) 的方法训练这些表示向量预测 fMRI 数据的回归模型 — 每次交叉验证都会预测 2 个测试词汇的 fMRI 数据, 并和真实的 fMRI 数据进行对比作为测试准确率。结果表明, 回归模型对 fMRI 数据中名词所诱发的脑激活模式具有显著大于随机值的分类准确率, 提示了词表示模型和 fMRI 数据之间存在直接的可预测的关系, 以及大脑表征名词语义时显著依赖感觉运动属性。这个工作改变了以往只能通过语义类别间的对比来考察概念表征的实验范式, 开辟了可以研究单个词汇概念的数据驱动新方法, 启发了后续很多的研究工作。



另一个代表性工作是 2016 年 Huth 等人<sup>[55]</sup>发表在《自然》上的文章,他们利用语言计算模型更加全面的研究不同的语义信息是如何在大脑中进行编码的。具体来说,他们采集了被试收听 2 个多小时叙事性故事(共包括 10,470 个不同的单词)时的 fMRI 数据,在语料库中选择了 985 个描述不同主题的基本词汇作为不同的语义属性,接着通过在大规模文本中统计故事中词汇和 985 个基本词汇的共现频率,对每个故事词汇构造了一个 985 维向量。然后他们训练岭回归模型使得 985 维词汇向量预测 fMRI 数据中的每个体素。其中,模型中得到的“985 × 体素个数”大小的参数矩阵即为 985 个语义属性的大脑表征。结果表明,不同被试间的大脑语义表征是十分类似的,不同语义特征在特定的脑区中进行编码。不同于上述 Mitchell 等人对 25 个感觉运动属性在大脑中表征的研究,Huth 等人的工作首次全面研究了 985 个语义特征在全脑不同体素中的表征模式,再次证明了借助语言计算模型在研究大脑语言理解问题上的作用。

在大脑对语言信息的计算机制这一问题上,Brennan 等人<sup>[124]</sup>研究了大脑是采用线性还是层级的结构来加工句子的。他们首先采集被试在收听《爱丽丝梦游仙境》第一章材料时的 fMRI 数据,然后利用一系列语法模型(包括了线性语法和层级语法模型)计算同样材料里句子中每个词汇的句法复杂度,最后将句法复杂度指标与 fMRI 数据做回归分析。其中,句法复杂度可以由词汇在某种语法结构中出现的概率计算得到。上述研究的假设是,如果人在收听并理解故事时采用层级加工的计算方式,那么相比利用线性语法模型计算的复杂度指标,利用层级语法模型计算的复杂度指标与 fMRI 数据的相关性应该更强。结果表明线性效应广泛分布在大脑语言网络中,而只有大脑左侧颞叶的特定区域负责加工层级结构信息,提示了大脑颞叶在语言理解过程中是采用层级结构的加工方式处理信息。

语言认知的研究也常借鉴语言计算模型的运行机制提出假设。这类工作中一个代表性的例子是 Li 等人<sup>[125]</sup>利用多种语言认知模型和神经网络模型来研究大脑理解代词指代关系时的神经机制。由于不同语言间代词表达形式不同,如中文代词在发音时不区分性别(“ta”)而英文则会区分(“she”, “he”, “it”),所以为了探究人脑在理解指代关系是否采取了不受语言影响的通用解析策略,他们分别采集了中文母语被试和英文母语被试在收听中文或英文版《小王子》全文时的 fMRI 数据。然后采用认知科学常用的通用线性模型(generalized linear model, GLM)计算与代词解析相关的脑区。他们发现中英文听力材料都会显著激活左前颞中回、左后颞中回前和角回脑区。为了进一步探究大脑解析指代关系时的计算机制,他们首先构造了 5 种针对代词指代消解问题的计算模型,分别是基于句法理论的 Hobbs 模型、基于篇章理论的 Centering 模型、基于记忆理论的 ACT-R 模型、基于神经网络 ELMo 结构和 BERT 结构的指代消解模型。随后检测了上述模型在解析《小王子》中指代关系的有效性,并将上述模型中计算的代词指代概率与 fMRI 数据做相关。结果表明,只有基于记忆理论的 ACT-R 模型可以显著预测中英文实验材料对应的神经激活数据,表明了大脑在解析代词指代关系时采取了与语言无关的通用记忆检索策略。Wehbe 等人<sup>[126]</sup>提出将循环神经网络语言模型(Recurrent neural network language model, RNNLM)和大脑在阅读时的工作机制进行类比。他们发现人脑在阅读故事时和 RNNLM 在处理句子时的工作机制在一定程度上是类似的。另外,Schrimpf 等人<sup>[127]</sup>对比了 43 种最先进神经网络模型和多种神经活动数据的关联。他们发现以语言模型作为目标且基于 Transformer 网络结构的模型可以显著预测神经响应、行为数据和下一个词汇的神经响应,说明了大脑的语言系统针对预测处理进行了优化,且 Transformer 网络结构与神经活动预测能力相关。

## 5.2 语言认知启发的语言计算方法

基于神经网络的深度学习方法在近年来备受推崇,它在某种意义上确实模拟了生物大脑的认知功能。但是这种方法并不是基于大脑的工作机理建立起来的数学模型,它同样难以摆脱对大规模训练

样本的依赖性,在泛化性和学习能力上与人类智能存在很大差距。本文中提的认知机理启发的语言计算模型旨在通过研究大脑的语言认知机理,分析认知机理与机器语言计算之间的关联,最终设计更加智能的语言计算模型,完成各项语言处理任务。

由于目前对大脑语言理解机理的研究远不如其他认知功能深入,因此,认知启发的计算方法大多集中在视觉认知和机器学习领域,在语言领域的工作较少。本文将已有认知启发的语言计算方法归纳为下述四类:

1. 借鉴大脑表征、学习、注意力、记忆等认知机制,设计或改进计算模型,使得(部分)模型具有与大脑类似的结构,从而提升模型处理下游任务时的性能。

例如,受到人在阅读句子时选择性地注视或跳读某些词汇的启发,Wang 等人<sup>[128]</sup>提出一种受人类注意力机制启发的句子表示模型。该方法利用眼动信号的预测因子(词汇惊异度 *surprisal* 和词汇类比标签)构建注意力模块,并将其结果作为权重引入到句子表示学习模型中。结果表明,注意力模块会给重要的词汇赋予更高的注意力权重,且权重结果与人类阅读时间是显著相关的。另外,注意力模块的加入可以显著提升句子表示的性能。

此外,Liu 等人<sup>[129]</sup>利用人的注意力机制改进图片描述生成模型的性能;Sun 等人<sup>[130]</sup>提出了基于记忆增强的小数据词表示学习方法;Han 等人<sup>[131]</sup>提出一种基于情景记忆激活与记忆巩固的连续学习方法。

2. 将大脑神经活动、神经影像或行为数据看作一种额外模态的数据,可以提供不同于已有数据的信息,因此在模型训练过程中将其与计算任务的数据相融合,以此提升模型的性能。

例如,Klerke 等人<sup>[132]</sup>提出一种多任务学习的方法将眼动数据引入句子压缩任务中。他们利用三层双向循环网络模型,底层网络用于预测眼动时间,最上层网络用于预测句子压缩。结果表明,这种多任务学习的方法可以有效地将眼动数据引入句子压缩任务中,提升模型的性能。Tiwalayo 等人<sup>[133]</sup>将人类预测下一个词汇的概率分布与语言模型预测下一个词汇的概率分布进行融合,有效地提升了语言模型的性能。

此外,Malmaud 等人<sup>[134]</sup>将预测眼动时间引入到阅读理解任务中;Barrett 等人<sup>[135]</sup>将眼动数据作为一种特征加入到词性标注和命名实体识别模型中;Mishra 等人<sup>[136]</sup>将眼动数据应用于提升情感分析模型的质量。另外,Fereidooni 等人<sup>[137]</sup>将 fMRI 数据引入到词汇表示学习上;Roller 等人<sup>[138]</sup>和 Wang 等人<sup>[139]</sup>将人的行为数据(词汇关联评分)引入到多模态词汇表示学习中。

3. 从模型底层架构上模拟生物神经元或神经环路的结构和工作机理,以此构建更加智能的计算模型。

例如,果蝇大脑利用 Kenyon 细胞从多种感觉模态接收信息,特定神经元会控制这些细胞的激活和抑制状态,因此果蝇大脑对输入的信息是稀疏的高维表示。Liang 等人<sup>[140]</sup>将上述信息编码过程形式化,并将其应用于词汇表示学习任务中。实验结果证明了该网络可以学习到单词的静态和上下文相关的语义表示,且性能与其他表示学习方法相当。另外,该方法将词汇表示为稀疏二进制哈希码的形式,相比其他方法需要更少的计算资源。

另外,这种认知启发的方法常见于通用计算方法的研究<sup>[141]</sup>,如记忆网络<sup>[142]</sup>、神经图灵机<sup>[143]</sup>、胶囊网络<sup>[144]</sup>和可塑权重巩固(Elastic Weight Consolidation, EWC)算法<sup>[145]</sup>等。

#### 4. 借鉴或使用认知科学的研究方法来解析神经网络模型编码的信息。

例如, Chien 等人<sup>[146]</sup>利用神经科学领域常用的时间尺度映射(timescale-mapping)方法来研究长短时记忆模型(long short term memory, LSTM)模型中每个神经元编码的信息。他们通过观测正常句子和随机替换上文片段时模型中每个神经元对下文句子的激活值来推断神经元的功能。背后的逻辑是,如果神经元的功能是编码短时间尺度的语言信息,那么当替换上文一小段文本时,其在下文句子的激活值变化应大于编码长时间尺度语言信息的神经元。研究发现,约15%的神经元用于编码长时间尺度信息,且这类神经元可以分为两类,分别是负责的连接各个神经元的控制器(controller),以及负责整合长距离信息的整合器(integrator)。

此外,受到神经科学研究神经元编码机制方法的启发,Lakretz 等人<sup>[147]</sup>研究了 LSTM 模型中每个神经元在完成的任务时的工作机制。Ivanva 等人<sup>[148]</sup>借鉴神经科学探针任务的设计方式,提出在机器学习方法中设计探针任务时应遵循的准则。

## 6 分析与讨论

现阶段,研究人员在利用语言计算模型预测大脑活动数据,以及通过人脑语言理解机制启发语言计算模型两个方向上,已取得了初步成果,但是仍然缺乏语言认知和语言计算相结合的细粒度和系统性的研究。例如,在人脑语言理解方面,人是如何从最基本的语言单元出发,逐渐构建更大的单元,直至最终理解语言,这方面的机理并不清楚,也缺乏系统有效的建模方法。在机器语言理解方面,计算模型在多项语言处理任务上取得了“拟人”的准确率,但是在常识推理能力、自主学习能力、泛化能力、学习效率、可解释性和可靠性等方面与人类智能还有很大的差距。如何借鉴更深层次的大脑语言理解机理构建更加智能的语言计算系统尚没有明确的解决方案。

笔者认为,未来在研究人的语言理解问题上,如图4所示,计算理论驱动的语言理解认知实验大有可为,即根据计算模型的结构或结果提出研究假设,然后在行为学或脑活动数据上进行验证。具体有以下五个重要研究方向。

### 1. 收集多语言多模态神经活动数据

已有的语言认知研究大多局限于使用单一数据采集方式(如仅使用 fMRI 或 MEG)来研究某一种语言的特定语言现象,往往导致得出的结论鲁棒性低和可重复性差的问题。因此,未来的语言认知研究一定是要在多种语言和多种类型的数据上进行多重验证。尤其对于结合计算模型的研究来说,数据的规模和质量直接决定了结果的可靠性,因此利用侵入式和非侵入式的多种工具针对不同语种采集大规模高质量的神经活动数据十分重要。同时,数据的开放和共享也在逐渐成为语言认知研究的趋势,这将会极大促进语言认知的研究。

### 2. 启发新的认知机制假设

语言计算模型的运行过程在一定程度上是透明的、全局的,其计算过程也是可见的,即模型学习到的词汇表示、将词汇表示组合成短语和句子表示的计算方式、预测推理出某一结果的计算步骤都是可观测的。从计算机制层面上解释大脑的工作原理是认知科学的重要任务,笔者认为未来可以深入探索计算模型中的表征和计算模块是否确实可以解释一些脑区在语言加工过程中的神经活动。如果一个脑区的神经活动可以被某种计算模型所解释,那么可以认为这个脑区进行了

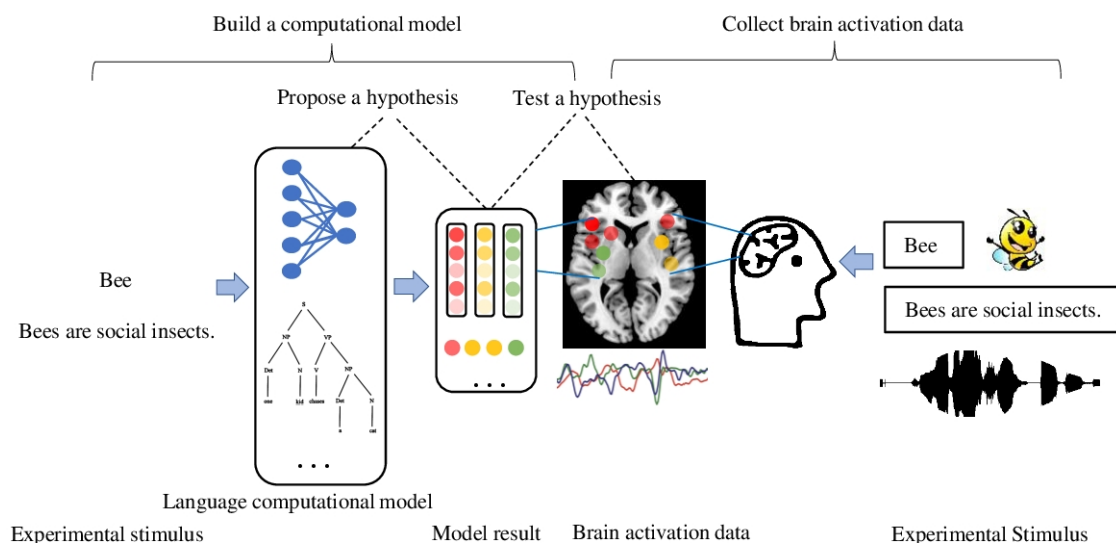


图 4 计算理论驱动的自然语言理解认知实验示意图

Figure 4 Schematic diagram of the cognitive experiment for natural language understanding driven by computational theory

模型中所清晰可见的计算功能。换言之，我们可以把不同语言计算模型中的各个模块当做大脑计算机制的假设，用认知科学实验进行验证。

### 3. 关联多种语言变量和认知功能

语言理解的过程十分复杂，不仅涉及多种语言变量的加工，如词法、句法、语义等，同时也与多个认知功能密切相关，如执行控制、注意力、记忆等。以往的研究往往通过严格控制实验变量来消除其他语言变量和认知功能的影响，在一个实验中只研究某一个语言变量或某种认知功能的作用。笔者认为，结合计算模型进行语言认知实验可以消除上述研究限制。例如，利用计算模型的方法可以分离不同实验变量，在自然文本采集的神经活动数据基础上，研究不同语言变量和认知功能的作用。随着基于神经网络方法的语言计算方法性能不断提升，使得用模型分离不同语言特征越来越准确，从而使得在同一批数据上计算出大脑不同区域在视听觉感知、多模态信息融合和语言理解中的其他功能机制成为可能。

### 4. 解析大脑语言理解的底层计算机制

已有的语言认知研究大多基于语言学理论，而语言学与神经科学研究的问题之间存在很大差距，如对于大脑如何操纵最基础的语言单元，语言学主要研究短语结构和语义组合，而神经科学主要关注神经震荡和同步。这就导致了目前语言认知的研究缺乏神经基础，无法匹配神经科学发现的结论。随着脉冲神经网络（Spike neural network）和震荡神经网络（Oscillatory neural network）的不断发展，未来的计算模型一定可以融合神经科学的结论来模拟底层神经元的工作模式，同时也可以根据语言学理论对语言单元进行操作从而完成语言理解的任务，这样就为关联语言学和神经科学的研究提供一种新的解决方案。

### 5. 探索语言学习、演化的机制

早在 80 年代, 认知科学家就采用联结主义模型来探讨怎样的模型, 怎样的数据可以模拟人的语言习得过程。但是当时的联结主义模型的计算能力相当有限, 只能解决一些案例性的, 简单的语言任务。当今, 深度神经网络的语言处理能力已经比 80 年代有了质的飞跃, 而且也有了更多婴儿语言习得过程中的语料记录, 因此可以重新尝试利用计算模型探索语言学习、演化的机制。甚至还可以深入探讨语言计算模型中的循环连接、卷积操作、点积注意力机制、反向传播算法等是否也是人脑语言理解计算过程中的必备环节。

在机器的语言理解问题上, 如图 5 所示, 语言认知研究提示人脑可能具备许多高效的加工方式, 在表征、学习记忆的认知机理, 神经元的工作机制, 行为学和影像数据等方面启发构建新一代语言计算模型具有巨大潜力。笔者认为以下五个方面将成为未来重要发展方向。

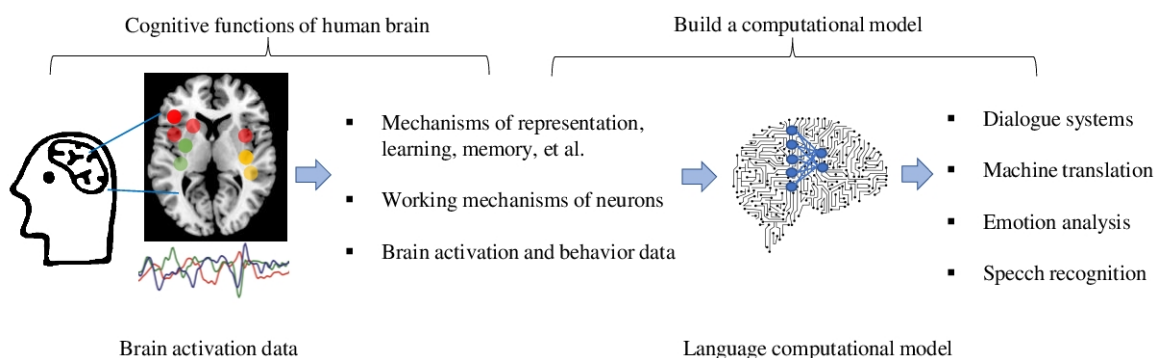


图 5 人脑认知功能启发的语言计算模型示意图

Figure 5 Schematic diagram of a language computational model inspired by the cognitive function of the human brain

### 1. 文本的表征和组合方式

在编码概念含义时, 大脑会针对不同类型的概念使用不同的表征方式。例如, 当阅读名词和动词时, 大脑会激活不同的脑网络, 呈现分布式编码的特点; 当观察到特定或熟悉的人时, 大脑中某一特定的神经元会被激活。在编码句法结构时, 大脑会对不同类型短语使用不同的组合方式, 采用层级编码方式如树型结构来指导词汇的组合顺序, 并且采用并行加工的方式来编码多个层次语言单元 (词汇、短语、句子等) 的信息。上述编码方式使得人脑存储和计算文本含义十分高效, 也与人类具有“举一反三”和快速学习的能力息息相关。未来的语言计算模型可以借鉴这种机制, 结合符号化和分布式的表征方式, 采取多样性、层级性和并行性的组合方式来学习文本表示和组合模型。

### 2. 语言的连续学习

人类在孩提时期就具备连续学习和小样本学习的能力。例如, 给 2 岁儿童看一张长颈鹿的图片, 他就可以识别其他的长颈鹿图片, 且这种能力可以迁移到识别其他动物图片, 以及寻找图片对应的文本等不同场景的任务中。这种学习能力与人类的记忆系统密切相关, 学习新信息的结果便是记忆。不同类型的信息由不同的记忆系统负责加工和存储, 如声音和图像等信息由感觉记忆负责存储, 并在较短的时间内进行保持。人脑复杂结构化的记忆系统保障了海量数据的高效组织和有需要时快速的提取。这些都是计算模型可以借鉴的。

### 3. 语言的交互学习

现有最好的通用语言计算模型使用预测下一个词汇作为目标函数,在海量文本中进行训练,在多项任务中取得卓越的性能。不同的是,相比分析和记忆语言结构,人类往往通过与他人进行交互来学习语言,这是一种更有效的学习和提升语言能力的方式。借鉴这种交互学习方式,未来的语言计算模型除了利用文本信息作为监督信号,也可以从其他模型的结构或者输出结果中得到反馈,在彼此不断的交互中进行学习。

### 4. 多模态信息融合

与交互学习密切相关的是多种模态信息的综合处理。人类的语言学习环境是一个多模态的系统,相比单模态的信息,人类更加擅长处理多模态的信息,并且对于多模态信息的处理速度要快于单一模态的信息<sup>[149]</sup>。但是如何让计算模型高效融合来自不同模态的信息是一个难题。因此,笔者认为在多模态信息融合上,脑启发的计算模型是未来一个重要研究方向。例如,根据“Hub and Spoke”理论<sup>[150]</sup>,概念是由视觉、听觉、嗅觉、体感等多种模态信息共同表征的,且存在一个语义中枢编码模态无关的信息,来关联不同模态的信息。不同模态之间的信息是相互补充且可以相互验证的,组合在一起可以表征更加丰富的信息。多模态信息的融合机制亦可借鉴上述“中枢”机制,设计一个模态无关的独立模块用于整合和关联不同类型的信息。

### 5. 计算模型的可解释性

认知科学设计实验来解析人脑的工作机理,这种研究方法和认知实验数据也可用于解析或评估语言计算模型的工作机制,启发新的模型可解释性方法。例如,借鉴语言认知实验常采用的对比分析方法,设计两组实验材料使其仅在某一个语言属性上有差异。如一组高句法复杂度的句子和一组低句法复杂度的句子,且两组实验材料在句子长度、语句含义等方法基本一致,将其输入到计算模型中,观测模型的效果是否与句法复杂度相关。如果网络中的某些结点在编码高复杂度句子时的激活明显强于低复杂度的句子,则这些神经元负责编码句法信息,反之则不负责编码句法信息。

## 7 结束语

语言认知的研究是认知科学和脑科学的核心问题之一,不仅对揭示人类语言智能的基础、揭示大脑工作原理的奥秘有重要意义,也有助于推动脑启发的语言智能技术的发展。与此同时,语言计算研究的新思想和新技术亦可以为脑语言认知的研究提供重要的借鉴和支持。因此,未来语言认知和语言计算的研究必将产生更紧密的结合,这方面的交叉研究的前景非常值得期待。

## 参考文献

---

- 1 Hebb D O. The organization of behavior; a neuropsychological theory. A Wiley Book in Clinical Psychology, 1949, 62: 78
- 2 Hinton G.E., McClelland J.L., Rumelhart D.E. Distributed representations. In Explorations in the Microstructure of Cognition (MIT Press), 1986, 7710-9
- 3 Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. In: Proceedings of the National Academy of Sciences, 1982, 79(8): 2554-2558
- 4 McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133

- 5 Machinery C. Computing machinery and intelligence - AM Turing. *Mind*, 1950, 59(236): 433
- 6 Chomsky N. Three models for the description of language. *IRE Transactions on Information Theory*, 1956, 2(3): 113-124
- 7 Yang Y. *Psycholinguistics*. Science Press, 2015 [杨玉芳. 心理语言学. 科学出版社, 2015]
- 8 Kemmerer D. *Cognitive neuroscience of language*. Zhejiang Education Press, 2017 [戴维·凯默勒著; 王穗莘, 周晓林, 等译. 语言的认知神经科学. 浙江教育出版社, 2017]
- 9 Hoy M. B. Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical Reference Services Quarterly*, 2018, 37(1), 81-88
- 10 Strickland E. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 2019, 56(4), 24-31
- 11 Wang W, Zhang J, Li Q, et al. Incremental learning from scratch for task-oriented dialogue systems. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, 3710-3720
- 12 Tan Z, Wang S, Yang Z, et al. Neural machine translation: a review of methods, resources, and tools. 2020, ArXiv:2012.15515
- 13 Zhu J, Wang Q, Wang Y, et al. NCLS: neural cross-lingual summarization. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2019, 3045-3055
- 14 Sun J, Wang S, Zhang J, et al. Distill and replay for continual language learning. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020, 3569-3579
- 15 Otter D. W., Medina J. R., Kalita J. K. A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*. 2020
- 16 Geirhos R., Temme C. R. M., Rauber J., et al. Generalisation in humans and deep neural networks. 2018, ArXiv:1808.08750
- 17 Chen G, Sun M, Liu Y. Towards a universal continuous knowledge base. 2020, ArXiv:2012.13568
- 18 Qiu X., Sun T., Xu Y., et al. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 2020, 1-26
- 19 Jin P, Lu Y, Ding N. Low-frequency neural activity reflects rule-based chunking during speech listening. *Elife*, 2020, 9: e55613
- 20 Luo C, Ding N. Cortical encoding of acoustic and linguistic rhythms in spoken narratives. *Elife*, 2020, 9: e60433
- 21 Sheng, J., Zheng, L., Lyu, B., et al. The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, 2019, 29(8), 3232-3240
- 22 Lin N, Xu Y, Yang H, et al. Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. *Brain Structure and Function*, 2020, 225(3): 995-1008
- 23 Lin N., Yang X., Li J., et al. Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. *Cognitive, Affective, Behavioral Neuroscience*, 2018, 18(2), 273-283
- 24 Sun J, Wang S, Zhang J, et al. Towards sentence-level brain decoding with distributed representations. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019, 33(01): 7047-7054
- 25 Sun J, Wang S, Zhang J, et al. Neural encoding and decoding with distributed sentence representations. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, Vol. 33, 7047-7054
- 26 Shain C, Blank I A, van Schijndel M, et al. fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 2020, 138: 107307
- 27 Wang S, Zhang J, Wang H, et al. Fine-grained neural decoding with distributed word representations. *Information Sciences*, 2020, 507: 256-272
- 28 Toneva M, Stretcu O, Poczos B, et al. Modeling task effects on meaning representation in the brain via zero-shot MEG prediction. In: *Proceedings of the International Conference on Neural Information Processing Systems*. 2020, 33
- 29 Deniz F, Nunez-Elizalde A O, Huth A G, et al. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *Journal of Neuroscience*, 2019, 39(39): 7722-7736
- 30 Jain S, Huth A G. Incorporating context into language encoding models for fMRI. In: *Proceedings of the International Conference on Neural Information Processing Systems*. 2018: 6629-6638
- 31 Anderson A. J., Binder J. R., Fernandino L., et al. Predicting neural activity patterns associated with sentences using

- a neurobiologically motivated model of semantic representation. *Cerebral Cortex*, 2017, 27(9), 4379-4395
- 32 Richards B. A., Lillicrap T. P., Beaudoin, et al. A deep learning framework for neuroscience. *Nature Neuroscience*, 2019, 22(11), 1761-1770
- 33 Saxe A., Nelli S., Summerfield C. If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, 2021, 22(1):55-67
- 34 Cichy R. M., Kaiser D. Deep neural networks as scientific models. *Trends in Cognitive Sciences*, 2019, 23(4), 305-317
- 35 Kietzmann T. C., McClure P., Kriegeskorte N. Deep neural networks in computational neuroscience. In *Oxford Research Encyclopedia of Neuroscience*. 2019
- 36 Nikolaus Kriegeskorte. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science*. 2015, 417-446
- 37 Lindsay G W. Convolutional neural networks as a model of the visual system: past, present, and future. *Journal of Cognitive Neuroscience*, 2020: 1-15
- 38 Lake B. M., Ullman T. D., Tenenbaum J. B., et al. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 2017, 40
- 39 Poo M. Towards Brain-Inspired Artificial Intelligence. *National Science Review*. 2018:785-785
- 40 Kasabov N K. Time-space, spiking neural networks and brain-inspired artificial intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019
- 41 Wang S, Zhang J, Lin N, et al. Investigating inner properties of multimodal representation and semantic compositionality with brain-based componential semantics. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 5964-5972
- 42 Cope A. J., Vasilaki E., Minors D., et al. Abstract concept learning in a simple neural network inspired by the insect brain. *PLoS Computational Biology*, 2018, 14(9), e1006435
- 43 Liberman A M, Cooper F S, Shankweiler D P, et al. Perception of the speech code. *Psychological Review*, 1967, 74(6): 431
- 44 Greenberg S. Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 1999, 29(2-4): 159-176
- 45 Hickok G, Poeppel D. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 2007, 8(5): 393-402
- 46 Townsend D J, Bever T G, Bever T G. Sentence comprehension: The integration of habits and rules. MIT Press, 2001
- 47 Frank S L, Bod R, Christiansen M H. How hierarchical is language use? In: *Proceedings of the Royal Society B: Biological Sciences*, 2012, 279(1747): 4522-4531
- 48 Kay J., Lesser R., Coltheart M. Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, 1996, 10(2), 159-180
- 49 Blank I., Balewski Z., Mahowald K., et al. Syntactic processing is distributed across the language system. *Neuroimage*. 2016, 127, 307-323
- 50 Pylkkanen L. Neural basis of basic composition: what we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society B*. 2020, 375(1791), 20190299
- 51 Friederici A D. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 2002, 6(2): 78-84
- 52 Laszlo S, Federmeier K D. Never seem to find the time: evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience*, 2014, 29(5): 642-661
- 53 Fodor J A. The modularity of mind. MIT Press, 1983
- 54 Mitchell T M, Shinkareva S V, Carlson A, et al. Predicting human brain activity associated with the meanings of nouns. *Science*, 2008, 320(5880): 1191-1195
- 55 Huth A G, De Heer W A, Griffiths T L, et al. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 2016, 532(7600): 453-458
- 56 Ding N, Simon J Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 2012, 109(29): 11854-11859
- 57 Marslen-Wilson W, Tyler L K, Seidenberg M. Sentence processing and the clause boundary. 1978
- 58 Miller G A, Heise G A, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 1951, 41(5): 329



- 59 Jarvella R J. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10(4): 409-416
- 60 Marslen-Wilson W D. Sentence perception as an interactive parallel process. *Science*, 1975, 189(4198): 226-228
- 61 Caramazza A, Hillis A E. Lexical organization of nouns and verbs in the brain. *Nature*, 1991, 349(6312): 788-790
- 62 Silveri M. C., Di Betta A. M. Noun-verb dissociations in brain-damaged patients: Further evidence. *Neurocase*, 1997, 3(6), 477-488
- 63 Skeide M A, Brauer J, Friederici A D. Syntax gradually segregates from semantics in the developing brain. *Neuroimage*, 2014, 100: 106-111
- 64 Friederici A. D. The brain basis of language processing: from structure to function. *Physiological Reviews*, 2011, 91(4), 1357-1392
- 65 Pyllkanen L. The neural basis of combinatory syntax and semantics. *Science*, 2019, 366(6461), 62-66
- 66 Binder J. R., Desai R. H., Graves W. W., et al. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 2009, 19(12), 2767-2796
- 67 Chen Q., Garcea F. E., Mahon B. Z. The representation of object-directed action and function knowledge in the human brain. *Cerebral Cortex*, 2016, 26(4), 1609-1618
- 68 Yang H, Lin Q, Han Z, et al. Dissociable intrinsic functional networks support noun-object and verb-action processing. *Brain and Language*, 2017, 175: 29-41
- 69 Binder J. R., Conant L. L., Humphries C. J., et al. Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, 2016, 33(3-4), 130-174
- 70 Cutler A. Native listening: Language experience and the recognition of spoken words. *Mit Press*, 2012
- 71 Kutas M, Federmeier K D. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 2000, 4(12): 463-470
- 72 Cherry E C. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 1953, 25(5): 975-979
- 73 Ding N, Melloni L, Zhang H, et al. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 2016, 19(1): 158-164
- 74 Wang X, Men W, Gao J, et al. Two forms of knowledge representations in the human brain. *Neuron*, 2020, 107(2): 383-393
- 75 Poeppel D, Embick D. Defining the relation between linguistics and neuroscience. *Twenty-first Century Psycholinguistics: Four Cornerstones*, 2005, 1: 103-118
- 76 *Encyclopedia of computer science and technology*. Tsinghua University Press, 2005 [计算机科学技术百科全书. 清华大学出版社有限公司, 2005]
- 77 *Zong C. Statistical natural language processing*. Tsinghua University Press, 2008 [宗成庆. 统计自然语言处理. 清华大学出版社, 2008]
- 78 Chomsky, N. *Aspects of the Theory of Syntax*. MIT Press. 2014
- 79 Vilnis L, McCallum A. Word representations via gaussian embedding. 2014, ArXiv:1412.6623
- 80 Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. In *Proceedings of the International Conference on Learning Representations*. 2013
- 81 Smolensky P. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 1990, 46(1-2): 159-216
- 82 Palatucci M, Pomerleau D, Hinton G E, et al. Zero-shot Learning with Semantic Output Codes. In: *Proceedings of the International Conference on Neural Information Processing Systems*. 2009
- 83 Wang S, Zhang J, Zong C. Associative multichannel autoencoder for multimodal word representation. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018: 115-124
- 84 Peters M E, Neumann M, Logan R, et al. Knowledge Enhanced Contextual Word Representations. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2019: 43-54
- 85 Mitchell J, Lapata M. Composition in distributional models of semantics. *Cognitive Science*, 2010, 34(8): 1388-1429
- 86 Ling W., Luis T., Marujo L., et al. Finding function in form: Compositional character models for open vocabulary word representation. 2015, ArXiv:1508.02096
- 87 Wang S, Zong C. Comparison study on critical components in composition model for phrase representation. *ACM*

- Transactions on Asian and Low-Resource Language Information Processing, 2017, 16(3): 1-25
- 88 Wang S, Zhang J, Zong C. Empirical exploring word-character relationship for chinese sentence representation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2018, 17(3): 1-18
- 89 Shen Y., Lin Z., Jacob A. P., et al. Straight to the tree: Constituency parsing with neural syntactic distance. 2018, ArXiv:1806.04168
- 90 Li Z., Cai J., He S., et al. Seq2seq dependency parsing. In: *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, 3203-3214
- 91 Liu J., Cohen S. B., Lapata M. Discourse representation structure parsing. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 2018, 429-439
- 92 Dan Klein, Christopher D. Manning. Accurate Unlexicalized Parsing. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2003: 423-430
- 93 Chen D, Manning C D. A fast and accurate dependency parser using neural networks. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2014: 740-750
- 94 Navigli R. Word sense disambiguation: A survey. *ACM Computing Surveys*, 2009, 41(2): 1-69
- 95 Kamath A, Das R. A survey on semantic parsing. 2018, ArXiv:1812.00978
- 96 Van Dijk T A. Semantic discourse analysis. *Handbook of Discourse Analysis*, 1985, 2: 103-136
- 97 Miller G A. WordNet: a lexical database for English. *Communications of the ACM*, 1995, 38(11): 39-41
- 98 Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. In: *Proceedings of the 27nd International Conference on Neural Information Processing Systems*. 2013: 926-934
- 99 Chen X., Jia S., Xiang Y. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 2020, 141, 112948
- 100 Wang Q., Mao Z., Wang B., et al. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(12), 2724-2743
- 101 Guo S, Wang Q, Wang L, et al. Knowledge graph embedding with iterative guidance from soft rules. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018, 32(1)
- 102 Harnad S. Searle's Chinese Room Argument. *Encyclopedia of Philosophy*, Macmillan. 2005
- 103 Harris Z S. Distributional structure. *Word*, 1954, 10(2-3): 146-162
- 104 John R. Firth. A synopsis of linguistic theory. *Studies in Linguistic Analysis*. 1957
- 105 Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model. *The Journal of Machine Learning Research*, 2003, 3: 1137-1155
- 106 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017
- 107 Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2018, 4171-4186
- 108 Dai Z, Yang Z, Yang Y, et al. Transformer-xl: Attentive language models beyond a fixed-length context. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2019, 2978-2988
- 109 Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners. In: *Proceedings of the International Conference on Neural Information Processing Systems*. 2020
- 110 Shwartz V., Dagan I. Still a pain in the neck: evaluating text representations on lexical composition. *Transactions of the Association for Computational Linguistics*, 2019, 7, 403-419
- 111 Loureiro D., Rezaee K., Pilehvar M. T., et al. Analysis and evaluation of language models for word sense disambiguation. *Computational Linguistics*, 2021, 1-55
- 112 Faruqui M., Dyer C. Improving vector space word representations using multilingual correlation. In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, 462-471
- 113 Berant J, Chou A, Frostig R, et al. Semantic parsing on freebase from question-answer pairs. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2013, 1533-1544
- 114 Noam Chomsky. 2002. *Syntactic structures*. Walter de Gruyter, 2002
- 115 Lucien Tesniere. *Elements de syntaxe structurale*. Paris: C. Klincksieck, 1959
- 116 Bowman S. R., Manning C. D., Potts C. Tree-structured composition in neural networks without tree-structured architectures. In: *Proceedings of the 2015th International Conference on Cognitive Computation: Integrating Neural*

- and Symbolic Approaches-Volume. 1583. 37-42
- 117 Zhao J., Liu K., He S. Knowledge graph. Higher Education Press, 2018 [赵军, 刘康, 何世柱. 知识图谱. 高等教育出版社, 2018]
- 118 Ji S., Pan S., Cambria E., et al. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*. 2021
- 119 Trichelair P., Emami A., Cheung J. C. K., et al. On the evaluation of common-sense reasoning in natural language understanding. 2018, ArXiv:1811.01778
- 120 Mostafazadeh N, Chambers N, He X, et al. A corpus and cloze evaluation for deeper understanding of commonsense stories. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016, 839-849
- 121 Sap M, Shwartz V, Bosselut A, et al. Commonsense reasoning for natural language processing. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. 2020: 27-33
- 122 Feng Z. Formal model of natural language processing. University of Science and Technology of China Press, 2010 [冯志伟. 自然语言处理的形式模型. 中国科学技术大学出版社, 2010]
- 123 Wang X, Wu W, Ling Z, et al. Organizational principles of abstract words in the human brain. *Cerebral Cortex*, 2018, 28(12): 4305-4318
- 124 Brennan J R, Stabler E P, Van Wagenen S E, et al. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 2016, 157: 81-94
- 125 Li J., Wang S., Luh W. M., et al. Cortical processing of reference in language revealed by computational models. 2020, BioRxiv
- 126 Wehbe L, Vaswani A, Knight K, et al. Aligning context-based statistical models of language with brain activity during reading. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2014, 233-243
- 127 Schrimpf M, Blank I A, Tuckute G, et al. The neural architecture of language: Integrative reverse-engineering converges on a model for predictive processing. 2020, BioRxiv
- 128 Wang S, Zhang J, Zong C. Learning Sentence Representation with Guidance of Human Attention. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. 2018: 4137-4143
- 129 Liu C, Mao J, Sha F, et al. Attention correctness in neural image captioning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017. 31(1)
- 130 Sun J, Wang S, Zong C. Memory, show the way: Memory based few shot word representation learning. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2018, 1435-1444
- 131 Han X, Dai Y, Gao T, et al. Continual relation learning via episodic memory activation and reconsolidation. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2020, 6429-6440
- 132 Klerke S, Goldberg Y, Søgaard A. Improving sentence compression by learning to predict gaze. In: *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*. 2016, 1528-1533
- 133 Eisape T, Zaslavsky N, Levy R. Cloze Distillation: Improving Neural Language Models with Human Next-Word Prediction. In: *Proceedings of the Conference on Computational Natural Language Learning*. 2020, 609-619
- 134 Malmaud J, Levy R, Berzak Y. Bridging Information-Seeking Human Gaze and Machine Reading Comprehension. In: *Proceedings of the Conference on Computational Natural Language Learning*. 2020, 142-152
- 135 Barrett M, Bingel J, Keller F, et al. Weakly supervised part-of-speech tagging using eye-tracking data. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2016, 579-584
- 136 Mishra A, Dey K, Bhattacharyya P. Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2017, 377-387
- 137 Fereidooni S, Mocz V, Radev D, et al. Understanding and Improving Word Embeddings through a Neuroscientific Lens. 2020, BioRxiv
- 138 Roller S, Im Walde S S. A multimodal LDA model integrating textual, cognitive and visual modalities. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2013, 1146-1157
- 139 Wang S, Zhang J, Zong C. Learning multimodal word representation via dynamic fusion methods. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018, 32(1)
- 140 Liang Y, Ryali C K, Hoover B, et al. Can a Fruit Fly Learn Word Embeddings? In: *Proceedings of the International*

- Conference on Learning Representations. 2021
- 141 Hassabis D, Kumaran D, Summerfield C, et al. Neuroscience-inspired artificial intelligence. *Neuron*, 2017, 95(2): 245-258
- 142 Weston J, Chopra S, Bordes A. Memory networks. 2014, ArXiv:1410.3916
- 143 Graves A, Wayne G, Danihelka I. Neural Turing machines. 2014, ArXiv:1410.5401
- 144 Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. In: Proceedings of the International Conference on Neural Information Processing Systems. 2017, 3859-3869
- 145 Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks. In: Proceedings of the National Academy of Sciences, 2017, 114(13): 3521-3526
- 146 Chien H. Y. S., Zhang J., Honey C. Mapping the Timescale Organization of Neural Language Models. In: Proceedings of the International Conference on Learning Representations. 2021
- 147 Lakretz Y., Kruszewski G., Desbordes T., et al. The emergence of number and syntax units in LSTM language models. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics. 2019, 11-20
- 148 Ivanova A. A., Hewitt J., Zaslavsky N. Probing artificial neural networks: insights from neuroscience. In: Proceedings of the International Conference on Learning Representations Workshop. 2021
- 149 Holler J., Levinson S. C. Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 2019, 23(8),639-652
- 150 Patterson K, Ralph M A L. The hub-and-spoke hypothesis of semantic memory. *Neurobiology of Language*. Academic Press, 2016: 765-775

## Language cognition and language computation – human and machine language understanding

Shaonan WANG<sup>1,2\*</sup>, Nai DING<sup>3,4\*</sup>, Nan LIN<sup>5,6</sup>, Jiajun ZHANG<sup>4,1,2</sup> & Chengqing ZONG<sup>5,1,2,7</sup>

1. *National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China;*
2. *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China;*
3. *Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou 310027, China;*
4. *Zhejiang Lab, Zhejiang University, Hangzhou 310027, China;*
5. *CAS Key Laboratory of Behavioural Science, Institute of Psychology, Beijing 100101, China;*
6. *Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China;*
7. *CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China*

\* Corresponding author. E-mail: shaonan.wang@nlpr.ia.ac.cn, ding\_nai@zju.edu.cn

**Abstract** Language understanding is the key scientific issue in the fields of cognitive science and computer science, but the two disciplines are very different in choosing specific research questions. Research in cognitive science focuses on analyzing the working mechanism of the brain and describing the brain's response to language, while little research study the brain language system as a whole. In contrast, computer scientists focus on the efficiency of the practical applications when choosing research questions, and often ignore the most essential laws of language. Then can we combine these two disciplines to bring new enlightenment and insights for building intelligent language models and studying language cognitive mechanisms? The following first reviewed the research question, history, and methods of language understanding in cognitive science and computer science, focusing on the current research status and challenges. Then this paper compared the commonalities and differences in language understanding between cognitive science and computer science. Finally, this paper analyzed the existing work that combined insights from language cognition and language computation and prospected the future development trend.

**Keywords** Language cognition, language computation, human language understanding, machine language understanding, interdisciplinary research



models and learn and generalize like humans.

**Shaonan WANG** received her Ph.D. degree in pattern recognition and intelligent system from Institute of Automation, Chinese Academy of Sciences (CASIA) in 2018. She is currently an Associate Professor at CASIA. Her research interests include studying how humans represent and understand language meanings; how can human learn and generalize language from few examples; how to build intelligent language



titled "Text Data Mining" and published over 200 papers. He is currently a Fellow of CAAI, Fellow of CCF and senior member of IEEE. He serves on journals ACM TALLIP as Associate Editor and IEEE Intelligent Systems as a member of editorial board. He served many top-tier international conferences, such as ACL-IJCNLP'2021 as conference chair, ACL-IJCNLP'2015 and COLING'2020 as PC Co-Chair.

**Chengqing ZONG** received his Ph.D. degree from the Institute of Computing Technology of the Chinese Academy of Sciences (CAS), in March, 1998. He is currently professor in the Institute of Automation of CAS. His research interests include machine translation, dialogue systems, and linguistic cognitive computing as well. Dr. Zong authored a book titled "Statistical Natural Language Processing", co-authored a book



tal paradigms to characterize neural encoding of speech and language during natural comprehension tasks.

**Nai DING** received his Ph.D. degree in electrical engineering from University of Maryland in 2012, and was a post-doc researcher at Psychology Department of New York University from 2012 to 2015. He is currently a research professor at Zhejiang University. He studies the neural mechanisms underlying auditory perception, speech perception, and language comprehension. He has developed a number of new experimental