

基于记忆的汉语句子语义处理

崔 跃

提要 自然语言处理过程是以语言表面结构为线索,结合特定的语言使用环境和背景知识,确定各种语言单位的语义功能,最终建立起符合语言使用环境的、确定的意义结构的过程。因此,建立意义结构是自然语言处理的核心。句法知识是处理过程所依赖的线索。在特定语言使用环境和背景知识的参照下,确定语言单位的恰当的语义功能是自然语言处理的关键。语义功能的确定和意义结构的建立必须综合运用句法、语义、语用等多方面的知识。因此,心理语言学研究将语言处理作为一个基于知识的信息处理过程。

知识单元的存储、提取,特定知识对于语言处理过程的制约作用,以及语言处理过程中各种不同性质知识单元之间的相互作用及联系,就成为人们心理语言学语义问题研究的重点问题之一。在这个问题上包含两方面的研究内容。第一,句法、语义、语用等方面知识的存储和使用是以相应的记忆结构为基础的。同时,处理过程中产生的中间环节也需要不断地存储、提取和更新,这伴随着一个动态的记忆过程。因此,自然语言处理具有不同性质的记忆结构关系非常密切。从记忆结构和记忆过程入手,不仅可以发现知识组织和使用的规律和特点,而且能够为智能化自然语言处理系统基本结构的建立提供恰当的机制。第二,在自然语言处理过程中知识使用方式是复杂的,独立或相互作用、并行或串行。对知识使用方式的探讨可以揭示意义结构建立过程的各个环节间的关系、认知资源分配方式、推理过程等特殊加工机制的特性。

由于汉语不仅在文字外部特征上,而且在语言结构体系上都具有自己的特点,因而汉语语言文字的处理更多地依赖于语义和语境信息的约束,更有利于揭示意义结构建立过程的各种特性。汉语同西方语言的明显区别在于它没有变格,即不能通过词形变化来确定一个名词属于什么格,一个动词属于什么时态。汉语的理解过程是通过多途径进行的,句法、语义、语用知识并非是在对每一句话的处理中都同时起作用的。在不同情况下,它们的重要性并不相等(李家治、陈永明,1985a,1986,陈永明,1989)。而且,汉语的句法结构是由不同句法标志的单词按照一定顺序构成的。汉语的词序是表达语法意义的重要手段。正如汉语语素的次序对词义具有重要意义一样,汉语的词序对于句子意义也具有非常重要的意义。词序表面上反映一定的句法结构、逻辑事理和语言习惯,但在本质上也是由语义决定的。因此,语义信息在汉语句子处理的心理语言学研究中具有特殊的地位。

一 汉语句子处理与动态记忆结构

在有关汉语词汇处理的心理语言学研究中,一般将各种词汇信息的组织和存储用心理词典来表示。但是在句子处理过程中,仅仅有心理词典是不够的。句子中的每一个词都属于一个特定语义场,或服务于特定的语义表达的需要,它只是建立句子意义结构的入口或索引。然而,每个词的具体意义和作用必须放在特定的语言使用环境中才能确定。因此,心理词典只有和存储世界知识的知识库有机地结合在一起,才能实现对句法、语义、语用信息的综合运用。这种有机结合的心理学基础就是汉语句子处理所依赖的动态记忆。

认知心理学认为,人的记忆系统可大致分为感觉记忆、短时记忆和长时记忆。句子加工与短时记忆和长时记忆关系密切。动态记忆有两方面的含义。一方面,心理词典和世界知识在记忆中进行组织和存储,并且随着人的言语活动的进行,不断地发生结构调整,是知识自动获取和自适应过程的基础;另一方面,随着句子处理过程,需要不断激活一些信息,并且与输入信息相结合,生成句子意义结构的片断,并在记忆中进行短暂的、快速的、准确的存储和处理。

我们在动态记忆中心理词典和世界知识的组织和存储方面采用了两条途径进行。第一条途径是以框架结构为主体,参照语义记忆的层次网络模型和扩散激活模型、概念依赖理论和格语法进行的。在机器理解汉语的研究中首先对心理词典的组织进行了初步探索(李家治等,1982;李家治、陈永明,1985b,1986),同时还对语义记忆中的信息提取特点进行了研究(陈永明、彭瑞祥,1985;陈永明等,1992)。这些研究不仅建立了我国第一个实验性的计算机汉语处理系统,而且突出了句法、语义知识在汉语处理过程中的综合作用机制,强调推理能力在句子处理中具有重要作用。研究还发现,(1)记忆中语义信息的提取依赖于概念之间的语义联系程度,这种语义联系程度不单纯是由概念在语义研究中所描述的语义场中的层次位置所决定,而要依赖于人对概念范畴之间语义关系的认知度;(2)人对概念范畴之间语义关系的认知度,影响语义网络结构的细节,认知度的变化是记忆结构动态过程形成的关键。通过这些发现,我们认为,汉语的句子处理系统中仅仅包含句法和语义信息是不够的。汉语句子处理需要一些情节性知识的补充。而情节性知识的传递需要更多的上下文信息进行约束,而这种约束在单句中是不易实现的。因此,我们逐渐尝试将句子处理与语段处理结合在一起进行,并且尝试将词典、语义记忆和情节记忆融合为一体,形成一个比较完整的动态记忆结构。基于事实-事件网络的汉语篇章理解系统就是在这种背景下产生的另一个实验性的汉语计算机处理系统(崔耀、陈永明,1994)。该系统的知识库包含两个相对独立,但又具有多种联系的网络。一个是事实网络,其中存储了表示具体和抽象事物的词的概念意义和它的部分句法信息;另一个是事件网络,其中存储了表示活动及活动之间关系的情节性知识。事件网络中,各个节点中都包含一个相对固定的语义框架。语义框架的填充必须通过事实网络才能进行。将事实网络和事件网络紧密联系在一起的是各种语义格关系。在两个网络中节点之间根据语义联系强度进行搜索和信息提取。该系统通过人机交互方式可以在网络中添加新节点,可对网络结构进行动态调整。在这个实验性系统的建立过程出现的困难是事实-事件网络中,节点间语义联系的类型及其强度的确定、基于事实-事件网络的推理的性质和作用机制、网络中节点的激活和抑制过程的协调关系等。这些问题直接反映了汉语处理中动态记忆的结构和处理特点,并且涉及到动态记忆结构中信息的快速存储和处理问题。

从1995年起,我们将动态记忆的两个方面结合起来进行研究。以汉语段落处理中的预期

性推理为研究对象,重点考察了语义联系的强度和类型以及人的认知资源对于处理过程的影响(崔耀、陈永明,1996,1998)。认知心理学认为,人的在进行认知活动时所能使用的资源是有限的,但是人可以通过学习和不断的实践,使一些认知活动达到高度熟练的程度,形成特殊的认知技能,从而使有限的资源得到充分而有效的利用。人的言语活动就是这样一种高度熟练化的过程。言语活动中,人们对语义知识、情节知识进行快速、准确的存储和提取。最初,研究者们用短时记忆来描述信息的暂时存储和提取问题,但是随着研究的深入,在言语研究中越来越多地用工作记忆(Working Memory)的概念来讨论信息的存储、提取、处理与认知资源的关系问题。这样做的目的是为了突出在记忆中发生的动态过程,它同时涉及到短时记忆和长时记忆(崔耀、陈永明,1997,崔耀,1997)。研究发现,汉语信息处理过程中仅仅使用传统语义网络中的语义联系类型是不够的。同时,人对概念间语义联系的认知度是一个动态变化的维度。例如,人对于一个概念或活动的熟悉程度是随学习或接触的频度的增加而非线性地增加的,不同的生活经历决定了这是熟悉程度的变化存在很大的个体差异。这样根据一些常识性知识而形成的语义网络并不一定能够真实地反映每个人对同一对语义关系的熟悉程度,因此也就无法正确反映不同人对同一对语义关系的认知度,进而影响语义处理过程对认知资源的消耗。因此,我们用熟悉性、典型性、可区分性作为人对语义联系认知度的衡量指标。通过心理语言学实验发现,这三种语义联系类型的强度差异直接影响阅读过程中推理活动的认知负荷,影响推理的进行,影响语义知识的组织和存储,并且影响基于事实-事件网络的汉语处理系统的处理效率。根据心理语言学的研究结果,我们在上述汉语处理系统中增加了认知资源控制机制,其目的是为了克服原系统中只注重信息的激活过程,而缺乏对被激活信息的恰当、适时地抑制。

第二条途径是以联结式方法进行(雷晓军,1990)。知识表征和组织的联结式模式主要有分布式联结模式和局部联结式模式。我们在汉语句子语义处理中主要采用了局部联结式模式。这种模式将传统串行符号加工系统的许多内容与分布式并行加工方法结合在一起,能够处理非常复杂的问题。将这种模式用于语义网络之后,网络节点的加工是同时进行的,节点之间的联结不再表示概念之间的语义关系,而是一种数字关系。我们建立的并行加工的语义网络首先描述的是一种静态的知识,即概念的内在构成关系。然后通过动作关系、状态关系、空间关系和时间关系作用于概念而形成一种动态的知识结构。

综上所述,自然语言表达的意义大多是一种动态知识的集合体。因此,语义问题更应该进一步从知识在动态记忆结构中的表达及其与自然语言处理过程的关系来考察。

二 汉语句子处理中语义知识的使用过程

汉语句子处理必须综合地使用句法、语义、语用知识。对于各种知识的作用形式和机制在心理语言学研究中有很大的争论。其焦点是句法加工和语义加工是否是独立的,各种知识的使用是否同时进行。在这个问题上至今没有统一的结论和解释。我们的心理语言学实验也同时支持存在句法分析独立进行的结果(江新,1997)和句法、语义分析交互作用的结果(陈永明、崔耀,1994,1995;崔耀、陈永明,1996,1998)。尽管如此,我们在汉语句子处理系统的建立方面主要采用了句法、语义、语用信息交互作用、并行处理模式。在汉语句子处理中下面三个层次的描述是必不可少的:(1)成分层次:句子的表层句法结构,由词、短语等范畴及其间关系来表达;(2)功能层次:成分间的语法关系,由主语、宾语等范畴来表达;(3)格层次:深层语义关系,由格角色来表达。人在理解句子时,每读到一个词都会激活一些不同层次上的认知操作,这些操作

交互进行,这种多层次分析方法特别适用于汉语句子的处理(李栗,1990)。而传统的自然语言处理范式是:首先对输入的句子进行句法分析,得到一个句法树,或一个功能结构,然后在语义规则和推理的作用下,将这个句法树或功能结构转化成一个深层的意义结构。但是,在人们日常交际中发生的言语活动往往是不符合句法规定的。“先句法,后语义”的范式不能对这种情况进行有效分析,必须使用并行处理方法(雷晓军,1991)。上下文对句子歧义的解析过程,不仅说明句法、语义、语用知识是对句子理解具有交互作用的模式,而且说明人在句子语义处理过程中可以快速地、自动地使用预期,并推动对下文词义的确定和整合过程(陈永明、崔耀,1994)。这个整合过程需要一定的时间。后继实验表明,人看完一个句子后,整合过程并不立即结束,还会持续一段时间(陈永明、崔耀,1995)。

根据上述动态记忆和知识使用过程的研究,我们正在试图通过综合的方法,完善现有的汉语处理系统,已经利用联结主义方法实现了汉语句子的格角色分配(张东松等,1996),并正在尝试将它与基于事实-事件网络的汉语处理系统结合在一起。同时,正在进行汉语处理过程中各种推理机制的研究,强化系统中的抑制过程和自适应过程,并且扩大和充实网络的内容,使系统有更强的语言处理和概念学习能力。

参考文献

- [1] 崔耀《阅读理解中的长时工作记忆》,《心理学动态》1997年第1期
- [2] 崔耀、陈永明《一个实验性的汉语篇章理解系统》,《中文信息学报》1994年第3期
- [3] 崔耀、陈永明《阅读理解中的预期推理》,《心理学报》1996年第3期
- [4] 崔耀、陈永明《工作记忆与阅读理解》,《心理科学》1997年第1期
- [5] 崔耀、陈永明《语义联系类型和强度对预期推理生成的制约作用》,《心理学报》1998年第1期
- [6] 陈永明《自然语言理解的方法和途径》,《心理学报》1989年第3期
- [7] 陈永明、崔耀《句子先提述的参与者在可提取性上的优势现象》,《心理学报》1994年第2期
- [8] 陈永明、崔耀《先述参与者的优势及句子不同成分的可提取性》,《心理科学》1995年第1期
- [9] 陈永明、彭瑞祥《汉语语义记忆提取的初步研究》,《心理学报》1985年第2期
- [10] 陈永明、雷晓军、宋均《句子加工和语义记忆结构的实验研究》,《心理学报》1992年第2期
- [11] 江新、彭聃龄主编《汉语句子和课文的理解》,《汉语认知研究》1997年
- [12] 雷晓军《语义网络的并行处理-概念结构中的继承和识别处理》,《视觉与听觉信息处理实验室年报》,北京大学1990年第1期
- [13] 雷晓军《自然语言理解的并行处理》,《心理学报》1991年第2期
- [14] 李家治、陈永明《句法、语义和推理在理解汉语中的作用》,《思维科学》1985a第4期
- [15] 李家治、陈永明《机器理解古汉语的初步探索》,《心理学报》1985b第3期
- [16] 李家治、郭荣江、陈永明《机器理解汉语-实验 I》,《心理学报》1982年第1期
- [17] 李栗《汉语句子的多层次分析》,《心理学报》1990年第2期
- [18] 张东松、陈永明、喻柏林《汉语句子的格角色分配的一种神经网络方法》,《心理学报》1996年第1期

(崔 跃 中国科学院心理研究所,邮编:100012)