

汉语语句形式结构到语义结构的理解自明度^{*})

韦向峰¹ 张全¹ 熊亮^{1,2}

(中国科学院声学研究所 北京 100080)¹ (中国科学院研究生院 北京 100049)²

摘要 计算机理解语句的方法之一是从形式结构出发得到语义结构。本文从计算机的角度给出了语句形式结构的表示方法,基于概念层次网络理论说明了语句语义结构的描述单位,介绍了从语句形式结构得到语句语义结构的分析方法。用理解自明度来反映分析结果的可靠性,并给出了描述分析过程中各子过程的理解自明度的计算公式。理解自明度是衡量系统分析性能的一个重要指标,也有助于系统解模糊能力的改进和提高。

关键词 理解,自明度,形式结构,语义结构

The Degree of Self-knowledge to Understand the Chinese Sentences from the Formal Architecture to the Semantic Architecture

WEI Xiang-Feng¹ ZHANG Quan¹ XIONG Liang^{1,2}

(Institute of Acoustics of Chinese Academy of Science, Beijing 100080)¹ (Graduate School of Chinese Academy of Science, Beijing 100049)²

Abstract For the computer program, one of the methods to understand Chinese sentences is uncovering the semantic architecture through the formal symbolic architecture. This paper describes the formal architecture of Chinese sentences in the opinion of the computer program, introduces the units of semantic architecture of Chinese sentences based on the theory of HNC (Hierarchical Network of Concepts), and explains the process of analyzing Chinese sentences from the formal architecture to semantic architecture. In this process, the degree of self-knowledge to understand Chinese sentences indicates the reliability of the results of analysis. The degree of self-knowledge to understand Chinese sentences is not only one of the important criterions to evaluate the performance of the analysis system, but also lend itself to improve and enhance the ability to eliminate ambiguities of the system.

Keywords Understanding, Degree of self-knowledge, Formal architecture, Semantic architecture

1 引言

在汉语中,语法单位从小到大通常划分为语素、词、短语、句子、复句、句群、段落和篇章。从计算机处理的角度看,切入点可以是语素或词。因为它们是构成更大一级语法单位的基础,不管是分析短语还是分析句子都必须分析字和词。但从计算机理解的角度看,句子也是很好的切入点。因为句子是“前后都有停顿、带有一定的语调、表示相对完整意义的语言片段”^[1],理解语言的基本内涵就是能够揭示语言所表示的意义。

计算机要理解语言,仅仅进行形式分析是不够的,必须结合意义进行分析。汉语是一种意合型语言,其语言单位的划分主要依靠概念之间的相互关联和作用。计算机擅长的是形式计算和处理,计算机理解语言的过程应该从形式得到语义,从而获得语言单位在意义上的结构和表示。

理解有一个程度的问题。“图灵标准”的目的是检测计算机是否具有类似人类的智能,对人来说具有可操作性,但对计算机来说以假乱真地以人的方式给出问题的答案还是一个理想目标,目前还难以实现。现在急需解决的问题是,计算机需要制定一个自己评判自己的标准,即在计算机理解过程的各个步骤中,计算机可以给出哪些是它自己认为是确定性的结果,哪些是不能确定的结果,哪些是可能性的结果,可能性是

多大。

为了从汉语语句的形式结构得到语义结构,使计算机理解语句,并衡量理解过程的可靠性和准确度,本文给出了描述汉语语句的形式结构和语义结构的方法,并介绍了从汉语语句形式结构到语义结构的计算机分析过程,具体给出了描述此过程的理解自明度的计算公式。

2 语句的形式结构描述

从计算机处理的角度看,汉语语句是一个顺序排列的符号串,根据编码知识计算机可以区分出汉字字符和非汉字字符,根据符号定义计算机可以识别出英文字母、阿拉伯数字、标点符号等。这种以符号特征为基础来描述语句结构的方法称为语句的形式结构描述。

在汉语语句中,汉字是最常见的符号,其次是标点。标点是划分语句形式结构的主要依据。汉语语句一般是以“句号、问号、叹号”为结束标志。以“逗号、分号”为结束的语句称为小句,也有不加结束标记的语句(如新闻标题中的语句)。

给定一个符号串 T ,以“句号、问号、叹号”为分隔符可以将其切分为一个个语段^[5] D_i ,以“逗号、分号”为分隔符可以将语段切分为一个个语串^[5] S_j ,语串则由一个个字符 Z_m 构成。因此,可以给出如下式子:

$$T = \sum D_i, i = 1, 2, \dots, N_1 \quad (1-1)$$

^{*} 本文承国家 973 项目“自然语言理解的交互引擎研究”(2004CB318104)、中科院声学所知识创新工程项目“HNC 语言知识处理理论及技术”的资助。韦向峰 博士,助理研究员;张全 研究员,博士生导师。

$$D_i = \sum S_j, j=1, 2, \dots, N_2 \quad (1-2)$$

$$S_j = \sum Z_m, m=1, 2, \dots, N_3 \quad (1-3)$$

文本字符串 T 由若干个语段构成, 语句由若干个语串构成, 而语串则由汉字、英文字母、阿拉伯数字等元素符号组成。从符号上看, 一个语句的形式结构可以描述为符号序列 $Z_i (i=1 \sim N)$ 。在汉语的 Z_i 中, 大部分是汉字, 小部分是标点, 还有些是阿拉伯数字、英文字母或特殊符号。语句符号串中的符号在形式上是一维的, 没有关联性, 但在意义上却相互关联。在一个有意义的语句中, 每一个符号都有其特殊含义, 并且与其它符号之间存在或强或弱的联系, 即它们在语义上是相关的。

现代汉语中, 一个字常常和其它字组字成词, 组词成句。按词的字数分词语分为单字词、双字词和多字词。将常用的词装入计算机中形成字词库, 根据字词库可以把语句切分成词语。但即使完满地完成了语句分词的任务, 由于词与词之间的语义关系没有揭示, 计算机依然无法理解语句的含义。

为了揭示词语之间的联系, 必须建立词语之间的关联符号。从这些符号中推导出意义的关联, 从而得到语句的意义结构。因此, 要使计算机理解语句, 就需要从语句的形式结构找出语义结构, 揭示语句的意义结构以及它们之间的关联性。

3 语句的语义结构描述

揭示意义的关联性可以从概念入手。一个词语可能对应多个不同的意义(概念), 但在语句使用条件的约束下其意义是确定的。语言符号在形式上并不反应其意义上的关联性, 但却可以通过概念性的、关联性的符号系统来反映和揭示符号之间的意义关联。这个符号系统称为概念基元符号体系^[2], 例如“思考”这个词的概念符号为“vg80”, 而“想法”这个词的概念符号为“r80”, 两词语形式上不相关, 但意义上强相关, 都属于思维活动概念(80)。

从概念基元符号体系出发, 根据“作用-效应”链和判断概念得到 7 种语句类型基元, 根据 7 种语句类型基元和概念基元符号体系的延伸概念, 定义出 57 组基本句类。基本句类及其组合可以描述语句的语义类型和基本构成。语句的语义类型称为句类, 句类的具体构成称为句类表示式, 句类表示式中的主要元素是语义块。语义块是构成语句的基本语义单位, 在语句中可以是字、词、短语或句子, 甚至句群。例如“他给我一本书”, 这个语句的句类表示式为“T0J=TA+T0+TB+TC”, 有 4 个语义块, 分别是“他”、“给”、“我”和“一本书”, 符号分别为 TA、T0、TB、和 TC, 分别表示“转移”概念的发出者、转移概念本身、转移概念的接收者、转移的物或信息。

因此, 一个语句的语义结构可以用一个句类表示式来揭示, 一个句类表示式由若干个语义块构成, 语义块的个数和关系等约束知识是句类知识决定的。句类知识是关于各种句类的特征的描述, 包括语句格式知识、语义块构成知识、语义块之间的概念关联知识、句类转换和语义块变换知识^[6]。因此, 句类表示式和句类知识共同给出了语句的语义结构。设语句 S 对应的句类表示式为 SC , 则 SC 可以用式(1-4)表示如下^[3]:

$$SC = [fKn] + GBK1 + EK + GBKm (m=2-4) \quad (1-4)$$

其中, 辅块 fK 在语句中可能不出现; 广义对象语义块 $GBKm (m=2-4)$ 在无省略的语句中必然存在; EK 则是语句的特征语义块, 指出了语句所属的语义类别。

4 从形式结构到语义结构

把计算机理解语句定位于获取语句的语义结构, 理解过程就是如何把一个文本字符串从式(1-3)变换得到式(1-4), 即根据一定的变换法则将字符串进行切分或组合得到语义块, 并给出语义块之间的关系和语义块的内部构成。这个理解过程分为以下五步。

首先, 把可以组合的字符串先行组合得到分析处理的基本单元。经过这一阶段的预处理, 字符串被转换为以下处理单元: 字、双字词、多字词、字段、数字段、字母段和标号段。字段是存在切分歧义的汉字串, 其中的切分歧义留待后续处理阶段解决。数字段和字母段分别是连续出现的数字和英文字母。标号段是配对使用的标号中的汉字串。

第二, 为了从字符串形式的结构得到概念符号表达的语义结构, 需要把字词映射为概念基元符号体系表示的概念符号。以常用字词为基础, 从概念空间考察它们的用法和意义, 我们积累了 41049 条词语的义项和 2333 条汉字的义项, 存储在词库和字库中。词库和字库记录了字词的概念类别、概念符号、使用频度、句类代码、语法约束知识、语义块约束知识等。因此, 从字库和词库可以得到字词的概念符号。

第三, 从概念符号感知语义块。语义块的感知分为两个部分, 一是特征语义块(记作 EK)的感知, 二是广义对象语义块(记作 GBK)的感知。特征语义块的感知可从具有概念类别符号“ v ”的词语入手。这些词语大多数是动词, 有可能构成语句的核心, 因此形成了特征语义块的候选集。对候选词集进一步筛选, 排除不可能的词语, 按可能性大小排队。对于含有“把、被、向、对”等语义块指示符(属于 1 类概念)的语句, 通过字词的概念符号可以感知 1 类概念, 1 类概念和 v 概念形成对广义对象语义块的切分。例如“ $GBK1 + 102 + GBK2 + v$ ”, 其中 $GBK1$ 和 $GBK2$ 分别表示第一和第二个广义对象语义块。对于不含语义块指示符的语句, 需要根据 v 概念及其句类代码在句类检验阶段处理。

第四, 根据感知结果假设并检验句类代码和语义块。对特征语义块候选集中的词语, 按优先顺序逐个得到在字库中的句类代码, 从而得到句类代码候选集。从句类代码候选集中取出一个句类代码, 假设它是正确的, 根据句类代码相关的句类知识, 对各广义对象语义块及其之间的概念符号进行检验计算, 若符合则通过检验, 假设的句类代码可作为语句的句类代码; 若不符合则不能通过检验, 需回溯到第三步再处理。

第五, 分析语义块的内部构成。语义块的一般构成可以分为要素(或核心)和说明(或修饰)两个部分。特征语义块的构成分析在第三步中已经基本完成。广义对象语义块的构成可以分为: 简单构成、串并联组合构成、词语概念组合构成、包含语句或语句蜕化形式的复杂构成。简单构成指只由一个字或词构成。串并联组合构成指含有“的”、“和”、“、”等串并联标志的字词组合。词语概念组合指没有串并联标志的多个字或词的组合。对复杂构成的分析, 与第三第四步非常类似。

从形式结构到语义结构的分析过程被称为句类分析。在句类分析的过程中, 到处充满了“多义选一”的消解模糊工作, 如从多个 v 概念中选出正确的特征语义块、从多个句类代码中假设检验出正确的句类代码等等。因此, 需要对这个过程的解模糊性进行评价, 量化考查句类分析的处理能力。考查标准应从计算机掌握的知识 and 解决问题的程度来制定, 而不

应从人的角度和标准出发。程序能做到的就给出结果和依据,不能做到的就说明不能做到,同时给出详细的过程分值,使程序具有“自知之明”。

5 理解自明度及其计算

长期以来,人们把“图灵检验”^[7]作为检测计算机是否具有智能的标准。图灵检验设想有两个黑箱,一个黑箱中是人,另一个黑箱中是计算机,如果输入相同而两个黑箱的输出使人无法分辨哪一个是计算机,那么就认为计算机具有了智能。图灵检验具有可操作性,是一个长期的理想目标,但从各方面来看,目前计算机离这个目标还十分遥远。反过来想,如果一个毫不理解中文的人在一个黑箱中完全机械地执行各种指令和操作,将中文翻译成英文,其翻译结果和翻译专家的结果无法区分,那么我们是否能说这个毫不理解中文的人理解了中文这种语言呢?这个黑箱就是塞尔设想的“中文屋”。所以理解标准的制定应当考虑是从人的角度出发还是从计算机的角度出发、是只看输入输出还是要看全过程、是用单一指标考核计算机是否像人一样还是要结合计算机自身特点多方面考核。

从计算机自身来看,它应当对自己能做什么、不能做什么、做到什么程度、依据什么做到的有一个详细的清楚的认识,即具有“自知之明”。自明度就是指计算机自知之明的程度。自明度从程序的操作过程、各种数据的使用、中间结果以及最终结果等多方面描述输出结果,而不是单纯地以人的标准只看计算机的输出结果。自明度从两个层面描述了输出结果的可靠性:第一个层面是结果的正确率,第二个层面是得到此结果的过程和条件。自明度是动态的,随着被考核过程的层次和数据的逐级深入和细化,自明度的数据集也逐渐增多且更加全面。

就本文定义的计算机理解语句的过程而言,其理解过程是从语句的形式结构得到语句的语义结构的过程。这个理解过程的自明度是第4节介绍的5个子过程的自明度的加权平均。对每一个子过程而言,在计算机所知范围内考查影响句类分析结果的各种数据,考查可能会产生模糊和多个分支的数据项,并计算它们正确率。由于是在计算机所知范围内考虑,因此正确率的标准是计算机的标准而不是人的标准。例如按人的判断是错误的结果,如果是计算机按正确的规则和步骤得出的话,那么该结果也是正确的。句类分析的结果主要包括句类代码、语句格式、语义块及其构成。句类代码主要来自特征语义块,语句格式主要来自*l*概念的语义块指示符和特征语义块,语义块及其内部构成主要依靠句类检验和词语的概念符号。因此,语句理解自明度将主要考查字段的自明度、特征语义块的自明度、语义块指示符的自明度、句类代码的自明度、广义对象语义块及其构成的自明度。

5.1 字段分析的自明度

字段是指形如“ABC”的字串,按词库匹配时既可以切分为“A/BC”(BC是词),也可以切分为“AB/C”(AB是词),还可以切分为“ABC”(ABC是词)。当字段中的字数超过3个时,这种字词的切分组合数目会更多。那么如何衡量最终切分结果的正确性呢?如果按照人的标准,在通常情况下有固定切分,但在特殊应用场合中可能取另一种切分,切分标准难以统一。按照计算机的标准,如果切分结果通过句类检验,那么该切分就被认为是正确的;如果不能通过句类检验,那么该切分就被认为是错误的。因此,假设字段所有的切分组合数为

N ,通过句类检验的切分组合数为 n ,则字段分析的自明度可表示为:

$$M_1 = \begin{cases} 1 - \frac{n-1}{N} & (1 \leq n \leq N) \\ 0 & (n=0) \end{cases} \quad (1-5)$$

自明度 M_1 的取值范围从0到1。当没有切分组合能通过句类检验时,计算机不知道取哪一个切分组合,此时 n 取值为0,自明度 M_1 为0。当通过句类检验的切分组合只有一种时,计算机找到了一种确定的切分组合,此时 n 取值为1,自明度 M_1 为1。通过检验的切分组合数 n 越大,计算机找到正确的切分组合的可能性越小,字段分析的自明度 M_1 也越小。

5.2 特征语义块的自明度

特征语义块来自感知处理后的 v 概念词语(包括单字词)候选集。经过句类假设检验后, v 概念词语候选集中的词语可能是语句的特征语义块,也可能不是。由于候选集中的词语是经过优化排序的,因此词语在候选集中的次序对其是否作为特征语义块是有影响的。假设 v 概念词语候选集的词语总数为 N ,通过句类检验的词语总数为 n ,通过检验的词语在候选集中的序号为 i (从1到 N ,序号越小优先级越高),则特征语义块感知的自明度可表示为:

$$M_2 = \begin{cases} 1 - \frac{1}{nN} \sum_{p=1}^n (i_p - 1), & (1 \leq n \leq N) \\ 0, & (n=0) \end{cases} \quad (1-6)$$

其中, i_p 表示第 p 个通过检验的词语在候选集中的序号。对于特征语义块候选词语的 v 概念感知子过程的自明度,也可以深入考查和计算。假设词语的 v 概念出现在第 i 个义项、第 j 个概念类别符号中,词语的义项总数和概念类别总数分别为 N_1 和 N_2 ,那么特征语义块核心的 v 概念感知子过程的自明度可以表示为: $M_{21} = (i-1)/N_1 + (j-1)/N_2$ 。

5.3 句类代码的自明度

句类代码来自特征语义块的候选词语,但是一个词语可能对应多个句类代码,因此句类代码候选集要大于特征语义块核心候选集。假设句类代码候选集的句类代码总数为 N ,通过检验的句类代码在候选集中的序号为 i ($1 \sim N$),通过句类检验的句类代码总数为 n ,则句类代码自明度的表示式与式(1-6)相同。

$$M_3 = \begin{cases} 1 - \frac{1}{nN} \sum_{p=1}^n (i_p - 1), & (1 \leq n \leq N) \\ 0, & (n=0) \end{cases} \quad (1-7)$$

其中, i_p 表示第 p 个通过检验的句类代码在候选集中的序号。

5.4 广义对象语义块自明度

在感知处理阶段, l 类概念中广义对象语义块的指示符被感知。假设广义对象语义块的指示符出现在词语的第 i 个义项、第 j 个概念类别符号中,词语的义项总数和概念类别总数分别为 N_1 和 N_2 ,那么广义对象语义块指示符的自明度可以表示为:

$$M_{41} = (i-1)/N_1 + (j-1)/N_2 \quad (1-8)$$

在句类假设检验阶段,根据假设的句类代码可以确定广义对象语义块的总个数,并得到句类对广义对象语义块的约束知识,从而完成对广义对象语义块的检验。假设句类代码要求的广义对象语义块的总数为 N ,通过检验的广义对象语义块总数为 n ,则检验广义对象语义块的自明度的表示式可以表示为: $M_{42} = n/N$ 。对于广义对象语义块检验的两个子过

(下转第161页)

的泛化能力。

结论 本文将半监督学习技术运用于药物活性预报中,并且用特征选择的方法克服了结构特征参数中无关和冗余特征对预报性能的影响。实验证明,半监督学习比单单在标记样本上进行完全监督学习具有更好的性能,同时特征选择能够提高半监督学习在药物活性预报中的精度。进一步的工作包括如何提高半监督学习的效率、如何选择合适的机器学习方法以提高半监督学习的泛化能力,从而提高化合物活性预报的效率和精度。

参考文献

1 Xu L, Wu Y, Hu C, Li H. A QSAR of the toxicity of amino-

- benzenes and their structures. Science in China (Series B), 2000, 43(2):130~136
- 2 Blum A, Mitchell T. Combining labeled and unlabeled data with Co-Training. In: Proceedings of the 1998 COLT, Morgan Kaufmann Publishers, 1998. 92~100
- 3 Goldman S, Zhou Y. Enhancing supervised learning with unlabeled data. In: Proceedings of the 17th ICML, San Francisco, CA, Morgan Kaufmann, 2000. 327~334
- 4 Zhou Z H, Li M. Semi-supervised regression with Co-Training. In: Proceedings of the 19th IJCAI'05, Edinburgh, Scotland 2005. 908~913
- 5 Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research, 2003(3): 1157~1182
- 6 Moody J, Utans J. Principled architecture selection for neural networks; Application to corporate bond rating prediction. In: NIPS 4, Morgan Kaufmann Publishers, Inc., 1992. 638~690

(上接第 144 页)

程——要素检验和结构检验,也可以用类似的方法描述子过程的自明度。

在分析广义对象语义块的内部构成阶段,需要确定组成广义对象语义块的各个词语之间的概念关联。假设组成广义对象语义块的词语总数为 N ,判定出的词和词之间的关系总数为 n ,则广义对象语义块构成的自明度可以表示为 $M_{43} = n/N$ 。对于词与词之间的组合层次关系,以及出现语句嵌套的广义对象语义块等子过程的自明度,也可以根据子过程的分析方法和分析步骤计算。

5.5 小结

理解自明度的计算是以从形式结构到语义结构的句类分析过程为基础,其基本标准是句类检验能否通过。通过计算机程序的句类检验的,就作为正确的结果计入自明度,不能通过句类检验的只能作为候选集合总数的一部分来影响自明度的大小。自明度计算公式实际上反映了从候选集中找到通过句类检验的唯一元素的可靠性,即自明度越趋向于 1,找到的可靠性越大;当自明度为 0 时,表示无法找到。自明度描述了构成理解过程的子过程结果的可靠性,可以逐级深入。整个句类分析过程的理解自明度可以用下面的式子表示:

$$M = (M_1 + M_2 + M_3 + M_{41} + M_{42} + M_{43}) / 6 \quad (1-9)$$

结束语 在计算机理解汉语语句的过程中,既需要从外部制定符合要求和预期的指标,也需要从系统内部给出过程数据和结果的分析评价。语句理解自明度就是计算机内部的

一个自我评价指标,它不仅考核过程结果的正确性,而且考核过程结果的数据来源依据和可靠性。自明度从数据的使用和结果等方面说明了解过程的理解过程的解模糊程度,从语句形式结构到语义结构的过程中经常要从多个候选中选出正确的,在这个选择过程中自明度越高说明系统解模糊的能力程度越强。

本文把从语句形式结构到语义结构的理解过程划分为五个子过程,对每一个子过程中重要的可能出现模糊的数据给出了自明度的计算公式。这些公式只是粗略地反映了理解过程中计算机进行“多义选一”操作的效果,还没有反映出详细的支撑知识和规则。对于如何记录理解过程中使用的规则和知识,并评价它们对理解过程的影响,以及如何计算它们的自明度,将是本文进一步深入研究的内容和方向。

参考文献

- 1 房玉清. 实用汉语语法[M]. 北京:北京语言学院出版社,1992
- 2 黄曾阳. HNC(概念层次网络)理论——计算机理解自然语言的新思路[M]. 北京:清华大学出版社,1998
- 3 黄曾阳. 语言概念空间的基本定理和数学物理表达式[M]. 北京:海洋出版社,2004
- 4 晋耀红. HNC句类分析的“自知之明”. 见:第一届 HNC 与语言学学术研讨会,武汉,2001
- 5 池毓焕. 汉语动词形态困扰的分析与处理[D]:[博士学位论文]. 北京:中国科学院声学研究所,2005
- 6 苗传江. HNC(概念层次网络)理论导论[M]. 北京:清华大学出版社,2005
- 7 [英]玛格丽特·博登,编著. 人工智能哲学[M]. 刘西瑞,王汉琦,译. 上海:上海译文出版社,2001
- 8 徐波,孙茂松,靳光谨,主编. 人中文信息处理若干重要问题[M]. 北京:科学出版社,2003

(上接第 147 页)

成了统计数据存在一定不完美,对于这一类的准确划分仍需要深入的配合句群的研究工作展开。

其次,辅块省略标记符号还原为辅块时问题仍然存在,主辅两可块的问题有待深入解决,也造成了统计数据存在一定不完美。例如:这条新闻电视台播过。可还原为“电视台||播过||这条新闻”,也可还原为“这条新闻||~在电视台||播过”。国外纺织企业已普及无梭织机。可还原为“国外的纺织企业||已普及||无梭织机”,也可还原为“在国外~||纺织企业||已普及||无梭织机”。

再次,对于规则的判断,需要用到具体的语义块构成以及语义块之间的信息,其中涉及到的深层次问题仍然存在,规则的具体形式化工作仍需进一步结合计算机形式语言的特征深入展开。

最后,对于“主谓谓语句”的统计数据是建立在人为搜索语料后得出的,这其中必然含有人为的遗漏、误添,实际的统计数据在此基础上一定存在一定的偏差,更精确的数据还需要在更大规模的真实语料的基础上展开。

总结 本文在现有的 HNC 研究所取得的成果的基础上,用 HNC 理论中的句类理论和语句格式知识来分析“主谓谓语句”这一现代汉语语言现象,并归纳得出了一套解决现代汉语主谓谓语句的语句格式判断规则,该规则采用 HNC 理论句类知识和语句格式知识为核心,利用语言概念空间中的概念及其关联知识来解决主谓谓语句给计算机带来的模糊不清问题,并试图通过规则得出判断结果。初步实验表明,这些规则是高效可行的。今后的工作将集中在规则进一步验证和将规则的学习机制引入知识库的构造过程的研究上。能否很好地解决这些问题将是整个计算机能否解决主谓谓语句带来的语义模糊的关键所在。

参考文献

- 1 朱德熙. 语法答问. 北京:商务印书馆,1985
- 2 徐青. 现代汉语. 修订版. 上海:华东师范大学出版社,1977. 295~302
- 3 黄曾阳. HNC(概念层次网络)理论. 北京:清华大学出版社,1998. 6~9
- 4 张全, 萧国政. HNC 与语言学研究. 武汉:武汉理工大学出版社,2001
- 5 黄曾阳. 语言概念空间的基本定理和数学物理表达式. 北京:海洋出版社,2004