

0 绪 论

今天，没有人会怀疑计算机处理语言文字的迫切需要，但是，也很少有人能说清楚自然语言处理将会有什么样的乐观前景。当人们考虑计算机如何不断适应处理本国语言文字问题的时候，面对汉语时不禁会增添一分困扰和责任感。目前汉语研究侧重在结构形式，况且还缺少独立、完整、有共识的理论体系，计算语言学研究成果也限于句法分析和统计方法。就中文信息处理的发展看，迫切需要加强现代汉语和语义理论研究，尤其是汉语语义形式化的研究。

如果要清醒地总结国内语言学及语言信息处理研究领域的经验教训的话，与其只对所取得的种种成果的短效性和局限性感到不满，还不如对久久未能取得突破性进展的方法论上的缺憾作一点反思。

随着计算机科学的深入发展，越来越明显地表明在程序语言、自然语言、数理逻辑这三个不同领域之间存在着种种对应性和相似性。语言学的研究将语言现象总结为规律；数学的研究将语言规律和内在关系抽象为数学描述并使分析过程形式化和可计算化；计算机科学的研究将这些数学表达式及计算规则在计算机上具体实现。完成这样一个大流程需要三个领域知识的结合（汉语、数学、数理逻辑和抽象代数）计算机科学。概括地说就是两个步骤，第一，汉语表达式的量化与数学抽象；第二，抽象表达式在计算机上计算和实现。这两步根本不同于现在直接用数据结构表示汉语表达式的中文信息处理的方法。原有的研究方法不能确切地解释汉语语义，因为它没有涉及语句的内涵义，而把同一个表达式在不同环境中的多次出现都看作同一个外延义。

本书研究现代汉语的计算语义。计算语义研究的是一种可计算

的语义表示形式。要使汉语语义分析过程像函数程序一样成为一个计算过程，最让中外学者关心而又感到困惑的是：汉语语义的表示形式是什么？表达式的计算规则是什么？如果汉语语义的抽象表示形式能与国际上的现代语言学、计算机科学理论、数理逻辑这三个不同学科交融，那么演算规则完全可以从这些领域中的现有理论成果中得到移植和借鉴，因此汉语语义的抽象表达形式是首要的研究内容。本书旨在定义汉语句子的语义抽象表达式，它兼有对人的可读性、机器的自动生成和适用于中文信息处理的特点。

汉语语义分析的中心目标是分析句义。动词，是句子结构中组成成分的核心，也是句子意义中的重心。如果句子对应高阶谓词逻辑公式，那么动词就对应了逻辑公式中的原子成分——谓词。因此，研究汉语动词的数学抽象与语义计算，无论是从汉语语言学角度还是从数理逻辑角度来说，都是公道的，而且又是衔接这两个学科的最好结合点和切入点。

语句 S 的抽象本质表现在逻辑公式 α 中，公式 α 在模型 M 及可能世界 W （具体时间、空间等语言环境）下得到一个解释，这个解释与语句 S 的句义是完全同构对应的，因此关键点在于如何由 S 自动生成 α 。汉语语序、移位、缺省等特点，决定了不能凭借形态标记使 $S \rightarrow \alpha$ 一步到位，得分两步完成：由语句到函子的生成过程与函子典式到高阶谓词逻辑式的转换过程。

句法上，句子谓语动词及其配置的名词性成分在价（Valency）和格（Case）上有一致性。抽象形式上，动词的论元取的值都是词语项，可以是名词性成分，也可以是空语类变元。空语类变元在句法上有可能找回的，也有不可能找回的，但是在逻辑上，这些缺省成分都得逐个找出来补全，才能完整地体现名词性成分与动词的逻辑语义关系，这也是句中实词（notion words）所表示的意义。因此，需要寻找一种结构上的表示形式，使动词论元的值既能用词语项表示，又是绝对完备的。它形式上是词语，还不是逻辑谓词，但这两者完全是对应

的。

于是有必要定义一个中介形式：函子 (functor)。它的初始形式是由语句的谓词动词及其配置成分组成的，其中缺省成分将添补完备。初始函子归约为典范函子，后者将论元位置上的词语项转换为指称对象的集合形式。于是典范函子就具有逻辑语义的基本涵义。典范函子最终组合成高阶谓词公式。一阶谓词逻辑中，谓词的论元只能是个体变项 (常量或变量) 高阶谓词逻辑中允许函数、命题、谓词作为项出现在论元位置上。高阶谓词公式包括了原子谓词公式 (典范函子转换形式) 算子 (时 tense、体 aspect、模态 modality , 它们对应了汉语虚词 function word) 和联结词。因此，函子是汉语语句与其逻辑式之间的中介形式，它耦合了一个从汉语句子到函子的面向句法的生成过程与另一个从函子典式到高阶谓词式的面向逻辑的转换过程，从而使构造汉语语句到逻辑公式的自动转换成为可能。

函子兼有逻辑谓词和动词核心结构的特点，因此兼有分析语义和理解语义的作用。函子及其自动提取技术可以直接应用于动词词义自动分析、派生词典生成、同义词典生成、语言词典转换成数据库等方面。借助于扩展函子和词义模式，可将语句语义信息生成信息库。这些都预示着函子形式比逻辑公式本身具有更多的灵活性和更大的实用价值。

求解核心的逻辑因子——函子，应保证论元项的完备性。如遇到论元缺省的情况，需要自动添补。本书提出 5 条复合规则以及差分规则，其中包括论元继承优先原则、弱扩展函子规则、函数变目提升论元规则等。缺省问题实际上是从句法映射到语义的问题，一直引起中外学者的关心。大家越来越认识到内涵的重要性，认为缺省是汉语区别于印欧语的本质特征之一。我们关心的是有没有一种科学的、精密的、形式化的方法能把句子的表层结构与它的语义解释联系起来，给出汉语缺省的定位和自动添补方法，并基于句法研究成果建立汉语的计算分析模型，以便让计算机理解和处理汉语语句的缺

省。本书所提出的求解汉语语句缺省成分（逻辑成分和句法成分）的方法，不仅是构造论元项完备函子的关键步骤，也是计算机理解汉语时最为关键的高难度技术之一，同时也是目前关于汉语篇章情景理论方面形式化、计算机处理的一种新的尝试。本书构造了一个语言环境模型 $M = (\delta, \Sigma, \Delta, M0)$ ，它们分别描述了情景、上下文、知识情景、客观世界知识。在此环境模型下，利用求解规则和策略成功地求解出简单句、复合句中诸函子的缺省成分。

本书对汉语词语义给出内涵抽象性定义和外延指称实体义，有效地解释了吕叔湘等语言学家长期争论的两个例句：“他的老师教得好”和“他的老师当得好”中关于统领关系 *possession relation* 在语法语义上的一致性问题的解释。给出一个至今为止令人信服的形式化解释。使用内涵与外延这一组逻辑语义概念，可以解决汉语新构词法和句法研究中一些令人至今迷惑不解的问题，对于正确分析汉语语义以及构造新的语法理论有重要的意义。

用汉语计算语义理论的实现系统——汉语逻辑函子自动提取系统（HYLJ 系统）能输入汉语语句，生成函子和扩展函子，进一步完善系统，从而最终生成高阶谓词公式。系统能自动提取一系列汉语句法信息、词义信息，并生成相应的数据库；作为理解汉语句子的信息也可供实际应用系统查询。系统由动词形式规范管理模块、动词形式规范应用接口模块、函子提取模块、信息填充模块四个部分组成。其中，管理模块能生成、维护、查询有关动词信息的数据库，包括动词功能库、动词配价库、动词释义库、名词语义类型库等。

1 自然语言处理与语言理论的研究和发展

“我们日常工作的信息，绝大部分是以语言文字作为媒介来传播、交换和记载的。因此随着计算机的推广应用，由数据处理、信息处理发展到知识处理，对语言文字处理的要求无论从深度和广度上说，都越来越高。这个问题在西方国家虽然也存在，但并不突出，因为计算机从诞生之日开始，就是以处理西方语言为基础的。换言之，他们无须经过呼吁和宣传，随着计算机推广应用的发展，很自然地都会主动地研究和解决计算机如何不断地适应自己国家的语言文字的问题。可惜，我们的汉语与西方语言的差别很大。能够处理西方语言的计算机面对汉语却显得无能为力。”（陈力为，1997）

1.1 理性主义和经验主义

自然语言处理综合了语言学、计算机科学、数理逻辑、心理语言学、人工智能、数理统计等多种学科的研究成果，通过建立形式化的计算模型来分析、处理自然语言。自然语言处理的基本方法，或者说基本思路、基本策略，主要有两种：理性主义的和经验主义的。

理性主义的方法是基于乔姆斯基（N. Chomsky）的语言原则（principles）。理性主义认为，人的很大一部分的语言知识是与生俱来的，由遗传决定的。按理性主义的研究思路，自然语言处理主要研究人的语言知识结构（语言能力，language competence）。实际的语言数据（语言行为，language performance）只提供了这种内在知识的间接证据。

经验主义的方法是基于先农 (Shannon) 的信息论 将语言事件赋予概率。经验主义认为, 人的知识通过感官输入, 经过一些简单的联想(association) 与通用化(generalization) 的操作而得到的。人并不是生来就有一套有关语言的原则和处理方法。按经验主义的研究思路, 自然语言处理的研究对象是实际的语言数据。在计算语言学中, 许多研究试图从大量的语言数据中获取语言的结构知识。

1.2 概率统计和规则分析

概率统计和规则分析是语言处理的两种主要手段。传统语言学崇尚规则方法, 这是因为语言学家首先是从规则着手的, 而不是从统计角度来认识和处理语言的。计算机诞生之后又发现这些规则实际上是面对语言的使用者(人) 的, 面向机器的规则分析始终不尽如人意。机器翻译由于强调整理解, 单纯靠规则方法, 曾经一度陷于低谷。统计方法带来了可喜的积极成果, 使机器翻译的正确率可达 60%, 汉语切分的正确率达 70%, 汉语语音输入的正确率达 80% 这是对传统语言学的严重挑战。

语音输入系统 ViaVoice 的正确率据称达到 92%。我们用新闻体语料做了一个小实验 发现错误还是比较明显: “硕士”错成“设施、摄氏、特使”这类同音、近音词。用最新的 98 版本实验, 仍会出现错误 如“ / 能占 冷战 结束以来, / 有一次, 一个北大女生去清华的三会 (舞会 跳舞, / 并以 已 举行两次会议, / 这女生惊呆了 以致一为 以为 他有什么机芯 居心), / 表情更苏木(肃穆) / ”。所有括号内的词是正确内容 而且都在该软件的“候选集”中出现 也就是说 系统选取的是最优解, 但不一定是正确解。那么如何纠正这些错误呢? 当然首先是要检查出这些错误, 再从“候选集”中召回正确的词。这两个过程对于概率统计方法来说恐怕已经是勉为其难了。公司纵然有成熟的语音技术和巨大的语音语料库、新闻体的语料库, 也会感到统

计方法局限性较大，需要寻求用语言学知识 / 语言规则来发现和纠正错误。这样的分析结论在日语输入中得到同样有力的印证。日语输入正确率原先也不高，因此，语音技术加上语言知识处理合成一体之后才成为好用的产品。

其实，据有关专家介绍，语音技术的正确率目前至多达到 80%，语音语料库还有很大的潜力可挖，语料的语言模型也仅限于新闻体，日常口语的对话体还很缺乏，诗词更谈不上。作为统计方法的母体——语料库来说，究竟要有多大，是否应有一个切实可行的尺度，而不是假定无穷大的理想尺度。

局部性的有效的语言知识及其规则分析可能是优化技术的核心。这种优化仍然是有限的（例如：“机芯（居心）/ 捍卫我（俄）领土完整”中的纠错是很难通过部分语言知识而取得明显效果的，但是这样做至少可以提高百分之一到五的正确率。

1.3 自然语言处理与语法理论的研究和发展

自然语言处理是应用计算机技术处理自然语言的学科，是计算机科学与现代语言学相结合的产物。语言学是一门领先的学科，它的发展会影响其他学科。计算机科学与技术是高新技术，两者结合之后形成这一新学科，它不仅应用性强而且理论性也很强。自然语言处理不仅有重大的学术意义，而且对社会经济的发展也有现实的或潜在的经济价值。同时自然语言处理也向传统的理论语言学发出了挑战。各种立足于自然语言自动处理的句法分析理论和方法大批涌出。它对于传统的句法学冲击最大。

乔姆斯基的形式语言理论是影响最大的早期计算语言学的句法理论。乔氏定义了 0 型语法、上下文有关语法、上下文无关语法和正则语法四种类型的形式语法。他的转换生成语法引起了语言学界的

一场革命，使得语言学和计算机科学之间有可能互相渗透。近年来出现了不少新的语法理论，其中较多被大家接受的是乔氏的管辖与约束理论 the theory of government and binding 简称 GB 理论与盖兹达 (G. Gazdar) 等人提出的“广义短语结构语法 (Generalized Phrase Structure Grammar 简称 GPSG)”以及 J. Bresnan & R. Kaplan 提出的词汇功能语法理论 (Lexical Functional Grammar 简称 LFG)。它们可以代表当代语法理论发展的水平。

1987 年出版的塞尔斯 (Sells, 1987) 和霍洛克斯 (Horrocks, 1987) 的两本书中只讨论了三种语法：管辖与约束理论即 GB 理论、广义短语语法和词汇功能语法。这三者的共同基础是数理逻辑，共同特点是由此展开的形式化方法都达到相当的高度。这三种至今还很辉煌的语法理论所必须具备的背景，给我们在方法上以很大的启示，为什么国内的现代语法理论没有这样的辉煌呢？因为我们还没有奠定这样的基础。

1.3.1 GB 理论 the theory of government and binding)

GB 理论认为，世界上任何一种语言都有两大系统构成，一个是规则系统，一个是原则系统。

在乔姆斯基的‘标准理论’时期，有一大堆规则，如语类规则、语义规则、转换规则、音系规则等。到了 GB 理论，规则系统比较简单，只有“移位 α ”(move α) 变换规则 (transformational rule)。该规则的主要内容是允许对句子成分作‘移位’ (movement) 、“加接” (adjunction) 、“删除” (deletion) 、“滋生” (overgenerate) 等处理，但是任何句法结构的派生必须得到原则系统下各个原则的“认可 (licensing)” 否则变换无效。与规则系统相比较，原则系统内容相当丰富。原先在乔姆斯基‘标准理论’中的许多规则到 GB 理论则被更一般的原则系统所取代，以求对人类语言有更强的解释力。原则系统主要包括“X 标杠理论” (X-bar theory) 、“论旨理论 (θ -theory 也称“ θ -理论”)” 、“格位理论” (Case theory) 、“管辖

理论”(government theory)、“约束理论”(binding theory)、“控制理论”(control theory)、“界限理论”(bounding theory 等。

“X 标杠理论”是用来代替原先的短语结构规则的，是对语类结构或者说词组 后来也有人把句子也包括在内，下同 规则描写法的一种限制。按照“ \bar{X} 标杠理论”，一切语类结构都是向心结构 都有一个中心语 head)。整个语类结构的中心语是最高的节点，往下的一个个节点（如果有的话）均层层低一级。如果把最高层次标为 X^n ，那么往下的便是 $X^{n-1}, X^{n-2}, \dots, X^1, X^0$ 。一切语类结构都按这样的规则构造。最高层次也可看作是语类结构的最大投射。显然，从另一个角度看，“X 标杠理论”也可以看作检验一个语类结构是否合格的条件。

“论旨理论”是规定动词的论元(argument)与动词的论旨角色(θ -role)之间论旨关系的理论原则。说得形象一些，如果把动词所表达的语义比喻为一出戏的剧情的话，论元好比是演员，论旨角色就好比是剧中的角色。论旨理论就好比是有关演戏规则的理论。按“论旨理论”里的“论旨准则”(θ -criterion)规定：每个论元都可以作论旨角色，但是一个动词的某一个论元只能作该动词的一个论旨角色；反之，动词的每个论旨角色必须由动词的论元来担任，但是动词的每个论旨角色只能由该动词的一个论元来担任。

“格位理论”不是指的菲尔墨的“格理论”。这是与“论旨理论”有密切关系的一种理论。在英语里，将菲尔墨说的“格”写作小写 c 字母开头的 case，“格位”则写作大写 C 字母开头的 Case。“格位理论”包含这样一些内容：一是“格位过滤”(Case filter) 凡是句中具有实际的语音形式的名词组，必须指派它格位，也就是说这样的名词组必须被赋予格位，但只能赋予一种格位。二是“格位指派语参数”(Case-assigner parameter) 规定哪些句成分可以指派哪些格位。三是“格位被指派语参数”(Case-assignee parameter) 规定哪些句成分可以获得格位。

“管辖理论”是要说明语类结构中的各个成分是否在同一个管辖区之内，以及在管辖区内什么是主管成分（governor 亦称主管语），什么是受管成分（governee，亦称受管语）。在一个语类结构中，如果甲、乙两节互相统制（c-command 或 m-command），而甲节是中心语（head，亦称词首），那么甲管辖乙。管辖的概念很重要，因为管辖是赋格的必要条件，只有主管成分才能把格赋予受管成分；管辖概念还可用来规范论旨角色的传递，即如果甲管辖乙，则甲能把论旨角色传递给乙。

“约束理论”研究的是：名词组之间的指称关系。指称理论有 A、B、C 三个原则（principles A, B, & C）原则 A，照应词（anaphor，指反身代词，如英语的 himself 这样的词，汉语的“自己”）必须在约束范围内受约束；原则 B，代名词（如英语的 he, she, him, her，汉语的“他、她”）在约束范围内是自由的（free）即代名词不得在约束范围内受约束；原则 C，指称词（R-expression，如专有名词）总是自由的，即指称词不得受约束。

“控制理论”研究的对象是一种隐含性的空语类（用大写的 PRO 表示，有代词性，所以也称为“大代语 big-PRO”）具体研究 PRO 怎样受先行词（antecedent）约束。

“界限理论”则研究在一些移位或转换情况下，如何用界限节点（bounding nodes）来对移位或转换加以限制，使得语迹（trace）与移走的成分之间的间隔，不超过一定的界限，使语迹受到先行语的严格管辖。

1.3.2 广义短语结构语法

简称 GPSG 由 Gazdar et al(1985)提出的，其主要目的是废止生成语法中转换规则的应用，只用表层结构。这样做是为了将结构简单化，消除由转换规则而引入的一些生成能力过强的问题。它的语义解释是采用模型理论，并用将语义解释与每个结构直接相配的方式来实现结构与语义之间的相互制约，相互补充。广义短语结构语

法的一个显著特点就是把 Montague 语法对语义的描写引进句法描写中。GPSG 理论建立在模型论基础上。模型是世界的再现，它决定语言成分的所指。内涵逻辑 (intensional logic) 是对自然语言语义进行模型论描写的一个行之有效的工具。内涵逻辑表达式通过内涵逻辑中的语义规则可以把它所代表的语义映射到模型中的所指上去。GPSG 语义解释通常是通过把句法信息翻译成内涵逻辑表达式来完成的。句法信息包括成分结构、句法特征和作为终端节点的词项。这三种成分构成了 GPSG 的树形结构。GPSG 的内涵逻辑模型包括三个逻辑结构成分：个体集合 E 可能世界集合 W 和赋值函数 F 即 $\langle E, W, F \rangle$ 。内涵逻辑表达式的所指域取决于该表达式的义类。所谓所指域即可能的所指范围，它不但可以是一个集合，而且还是由一个集合向另一个集合映射的函数。GPSG 理论还包括一系列的特征。特征同样可以由内涵逻辑来表示，并在结构中进行运算。主特征翻译和句法范畴的翻译不同，它们不是由下而上通过函数实现来层层翻译的。只有在结构中处于最高层次的主特征才参与语义解释。GPSG 是现代国外语法学流派中很有影响的一个学派。当代科学技术的发展给语言学提出了新的要求。GPSG 的形式化特点正迎合了这一时代潮流。GPSG 对于机器翻译、人工智能等领域的研究具有重要参考价值。

1.3.3 词汇功能语法 lexical functional grammar

词汇功能语法，更确切的翻译应为“词汇函项语法”（徐烈炯，1990）。“function”一词语言学上常译为“功能”但在数学上译为函数、函项，“functional”又译为函数式、泛函。这里采用了数学上的函数、谓词这样的表达方式，其实后者译法更妥。不过，现在大家都熟用“功能”本书仍采用。

词汇功能语法是 70 年代末期发展起来的形式语法理论。它的创始人是 R. Kaplan 和 J. Bresnan。该理论由转换生成语法理论结合

计算理论和心理语言学而来，它在语法理论研究的客体对象、方法等问题上继承了生成语法学派的观点。词汇功能语法体系的构造方法始终遵循形式明确的原则，注重数学方法的应用。词汇功能语法的特点是把许多重要的信息放在词汇部分处理，充实对词项的描述，减轻句法的负担。另外，它把主、宾语等概念看作语法描述中的基本概念，用函数式来表示谓语与主语，以及谓语与宾语之间的关系，并提出语法函项（grammatical function）。词汇功能语法认为动词、一部分形容词和一部分名词在句子的语法结构中的作用相当于数理逻辑中的谓词，它们的词义可以决定它们各自需要哪些主目（论元）。以谓词为中心的句义就能知道句中该有哪类词，语法功能是什么。谓词的管辖范围以及谓词的主目等都是确定语法结构的重要因素。谓词管辖之外的语法成分才用语法规则来解释。函项信息在 LFG 中的重要程度已经超过了语法范畴，成为整个理论的主要基础。LFG 用到的功能主要是语法功能，为了从功能结构得到句子的语义解释，还用了谓词功能。LFG 力图按照公理学方法构造自己的理论体系。公理学方法就是从某些基本概念和基本命题出发，依据特定的演绎规则推导一系列的定理，从而构成一个演绎系统。它注重数学方法的应用。

LFG 继承了关系语法的观点，取消了转换，不需要 GB 理论中深层结构到表层结构的转换规则，从成分结构求功能结构。他们认为句子可由两个相互独立的层次来描述，成分结构层次和功能结构层次。成分结构用以描写句子成分之间的结构关系，而功能结构则用来描写句子主语、谓语等等的语法功能关系，并强调词汇在决定所从属词汇中的作用。同时，词汇功能语法把句子分析的可计算性^[2]作为它的一个基本目标。由于词汇功能语法认为这两个层次之间没有一一对应的关系，它采用归一算法来将这两个层次的信息综合。归一算法的中心思想是把来自不同部分的信息，相互补充、相互限制和积累而成为一个和谐的整体。它的基本表示结构是特征结构。

1.3.4 系统功能语法 (systemic functional grammar)

系统功能语法是 60 年代末兴起的重要语言学派之一，创始人是著名语言学家韩礼德 (M. A. K. Halliday)。系统功能语法包括“系统语法”和“功能语法”。系统语法着重说明语言作为系统的内部底层关系，它是与意义相关联的若干个子系统组成的系统网络 (system network)。功能语法则说明语言是社会交往的工具。语言系统的形成正是人们在长期交往中为了实现各种不同的语义功能所决定的。系统功能语法把语言的实际使用作为研究对象，它的理论已被广泛应用于各个领域，如：(1) 帮助各种年龄段的人们学习语言，并在一系列情景语境中有效地使用语言；(2) 了解语言与大脑的关系，帮助人们克服教育方面的和病理性的语言紊乱如“失语症患者”“弱智者”；(3) 了解聋哑人的语言和设计助听器；(4) 使用计算机分析和生成话语；发展一种能译码和编码的软件，以及指导和解释其语法的语义表达法；设计能生成和理解言语以及在书面语篇和口语语篇之间转换的系统；(5) 探索涉及语言的各种实际活动，如词典和语法书的编纂等。

1.3.5 链语法 (link grammar)

链语法是斯里特 (D. Sleator) 和汤佩雷 (D. Temperley) 于 1993 年在《用链语法剖析英语》(Parsing English with a Link Grammar, Technical Report of Carnegie) 一文中提出来的，它试图用链的方式描写相邻成分中心词之间的连接关系。链语法的生成能力等价于上下文无关语法。但是，上下文无关语法只强调成分之间的层次与前后关系，而链语法的重点是成分中心词之间的关系。两词要合法地连接，它们必须带有同一类链，且一词带链头，另一词带链座。一个句子的合法分析链结构称之为链系统，语法分析图由一些链系统组成。链语法是由一组词汇及相应词的定义所组成。词的定义是由一系列“逻辑

辑或'组成。

句子的链系统必须满足以下四个条件：

(1) 连接性；(2) 平面性；(3) 有序性；(4) 排他性。

链语法是基于词的语法系统，很容易得到词间关系的统计数据。

如果通过大规模真实文本的语料库来获取每种链所链接的单词对的出现频度，在句法系统遇到链接的歧义时，便可以凭借这样的统计数据作出判断，从而为句子的分析选出概率意义上的最佳结果。链语法的语法信息主要记录在词典中的单词上，依靠单词的语法信息来进行句法语义分析，单词中记录的信息是整个链语法运算的基本资源，这种语法特别强调词汇的作用。所以，链语法是当代计算语言学中词汇主义倾向的具有代表性的一种语法理论（冯志伟，1999）。

1.4 自然语言理解与 Montague 语义理论

1.4.1 国外语义学流派

语义学的研究从指称开始，随后又涌出各种语义流派。

结构语义学 (constructural semantics) 主张将结构主义语言学的研究原则和研究方法用于语义研究。其主要贡献是进行了语义场、义素分析及词汇语义结构的研究。把语言的意义看作系统，这无论在理论上还是在方法上，对传统语义学都是一个突破。代表人物和主要著作有：莱昂斯 (J. Lyons) 《结构语义学》(1963) 和《语义学》(1972) 乌尔曼 (S. Ullmann) 《语义学》(1962) 利奇 (G. Leech) 《语义学》(1987)。

解释语义学 (interpretational semantics) 60 年代初，接受 TG (transformational-generative grammar) 基本框架的卡兹 (J. Katz)、福德 (J. Fodor) 等人认为 TG 中应该包括语义部分，并论述了它的性质与作用。乔姆斯基接受了这个建议，在《深层结构、表层结构和语义解释》一文中，提出了一套解决语义问题的理论。语义部分的语义

规则对句子的深层结构作出语义解释。解释语义学的其他著作还有贾肯多夫 R. Jackendoff 的《生成语法中的语义解释》(1972), 卡兹、福德的《语义理论的结构》(1963), 卡兹的《语义理论》(1972)。

格语法 (case grammar) 菲尔默 (C. J. Fillmore) 在《格辩》(The Case for Case, 1968) 中提出一套用来说明句子中动词与名词之间语义关系的语法理论。他认为, 每个句子都由情态 (modality) 和命题 (proposition) 构成。其中的命题 (P), 可以扩展为一个动词和一个或一个以上的语义格 (semantic case)。语义格是任何语言都有的, 通过规则, 可以把深层结构转变为句子的表层形式。

切夫语法 (Chafe's grammar) 切夫 (W. Chafe) 在《意义和语言结构》(Meaning and Structure of Language, 1971) 一书中提出一套说明句子语义的语法理论。他认为, 任何句子都是围绕谓语部分构成的, 谓语一般是动词。句中的动词总带着一定数量的名词。句子里名词的语义作用是由动词决定的。他认为对语言的分析不能停留在句子这个层面上, 应该扩大到话语或篇章, 进行话语分析 (discourse analyses)。切夫和菲尔默都是把句法结构和语义结合起来分析句子, 但切夫是以动词的语义特征作为分析的依据的, 有人把他的理论称为动词中心说 (verb-centrality), 而把菲尔默的格语法称做名词中心说 (noun-centrality)。

情景语义学 (situation semantics) 情景语义学是 80 年代初美国斯坦福大学 J. Barwise & J. Perry 提出的。他们认为情景语义学可以克服传统的真值条件语义学遇到的一些困难。他们提出了六个重要问题, 即语言的外部意义问题、语言的能产性问题、语言的效率问题、语言角度相对性问题、语言的歧义问题、语言的心理学意义问题。情景语义学给自己定了较高的目标。他们希望能用情景语义学的理论处理各种各样的语言问题。他们构造的一种自然语言模型理论比较适合人工智能和自然语言处理的需要。

概念依存理论 (concepts dependency) 概念依存理论是美国人

香克 (R. C. Schank) 和他的同事在 70 年代提出的 (Schank, 1975, Schank and Abelson, 1977)。该理论有三个层次: 第一个层次是概念依存层次, 它规定了一组动作基元, 其他动作则是由这些动作基元组合而成的。第二个层次是剧本。它主要用来描写平时在遇到一些常见场景或场合时所采取的一些基本固定的成套动作。当剧本中每个动作都用基本动作来组合时, 我们就可以在相应场合中用基本动作来刻画各种套路行为。第三个层次是计划。为完成某项任务或达到某个目的, 往往需要有个计划。计划中的每一步都是一个剧本。这样, 从小到大, 我们可以用最简单的一组动作来表示复杂的行为。

这套 CD 理论, 正是希望对常识进行系统、具体的描写, 并利用那些基本动作来推理, 从而达到对语言的自动理解。从另一方面看, CD 对常识描写是相当刻板的。香克和他的同事在以后的工作中对此做了很多改进, 包括对剧本等引入各种组织和记忆方式、学习机制等。这套理论对范围有限的应用领域是非常有用的。这个方法在抽象化、规范化方面很有代表性, 后来发展成框架—平面描述法。CD 理论是针对理解的, 输出的是回答, 而不是等价的句义。CD 理论是逻辑语义学的形式化方法。在自然语言理解技术特别是人工智能技术当中有一定的影响。CD 理论的基本点之一是将句子中的隐含语义表面化, 这实际上相当于背景知识。在 CD 理论中只有两类背景知识, 即原因和状态。CD 理论提供了一种描述背景知识的实验模型, 而不是使用方法。实现 CD 理论的条件之一是要有成功的语法分析器。

优选语义学 (preferential semantics) 维尔克氏 (Wilks, 1973, 1975) 提出了优选语义学。它的核心是放松词与词之间的语义限制, 视这些限制为上优选择, 同时也允许有其他选择。动词可以根据其主、宾位置上词的语义特征, 赋予不同的数值。但没有说明如何来获得各词之间的优选数值, 它们的合理性以及随着上下文的变化这些数值应该如何改变。近年来, 他的学生 (Slator and Wilks, 1990) 直接

从英语词典中提取这些数值，并为建立大型而又实用的分析系统提供词汇资源。另外一些研究人员开始用统计的方法来获取这类优选数值，并已取得一些结果。

1.4.2 Montague 语义理论

蒙太古理论是美国数理逻辑学家蒙太古 (R. Montague) 所创立的一种用数理逻辑方法研究自然语言的形式语义学理论。本书的很多想法都借鉴于该理论，所以这里需要对蒙太古理论作较多的介绍。

熟悉转换产生式文法的语言学家将语义学看成是不可驯服的、几乎不能深究的语言部分。在讲到语言理论的框图时，往往用一空框框表示，语义成分究竟是什么，没有表示出来。

逻辑学家认为语义学是可以驯服的，但是与清晰严格的谓词演算、时态和模态逻辑相比，自然语言充满着歧义和无规律的例外，因此仍旧让人有一种不认同感。然而蒙太古对此提出了挑战，他认为形式语言和自然语言之间不存在重要的理论差别。他建立了一个框架，是一种清晰的实质的语义理论框架。他在 1970 年连续发表了三篇论文，其代表作是《日常英语中量化的特定处理》(“The Proper Treatment of Quantification in Ordinary English” 提出了著名的 PTQ 系统。他认为，自然语言与人工形式语言在本质上没有什么区别，因此都可以作精确的数学描述。他提出这样一个命题：每个句法规则都有一个相应的语义规则。据此，他认为，句法描述不仅要定义结构式的合法与否，而且要为语义解释提供基础。在该文中，他用数理逻辑方法解释自然语言的语义，从而开辟了用严格数学方法研究自然语言语义的新方向。

在自然语言理解的研究中，Montague 语义理论有特殊的贡献。在 GPSG 和 LFG 中都已采用了类似 Montague 语法的模型理论语义学。目前已有有人在设法使生成语法的语义研究与 Montague 的模型理论靠拢。能适用于计算机处理自然语言的语义理论中 Montague