

逻辑与语言研究会编

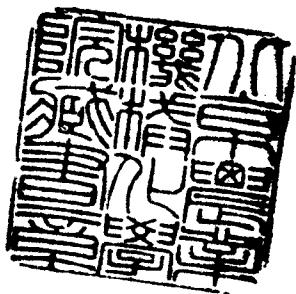
逻辑与语言研究



# 逻辑与语言研究

(3)

逻辑与语言研究会 编



中国社会科学出版社

**逻辑与语言研究**

(3)

\*

中国社会科学出版社出版  
新华书店北京发行所发行  
房山县印刷厂印刷

---

787×1092毫米 32开本 8印张 173千字

1983年12月第1版 1983年12月第1次印刷

印数1—15,000册

统一书号：2190·071 定价：0.76元

# 目 录

汉语人机对话的语法分析问题·····	范继淹 ( 1 )
从歧义句谈起, 试说转换-生成语法	
·····	徐思益 边兴昌 ( 21 )
论单独指称·····	江天骥 ( 44 )
浅析混淆概念·····	徐庆凯 ( 61 )
写作中概念关系处置不当问题的分析·····	江显芸 ( 73 )
谈谈语词定义·····	王耀堃 ( 83 )
“不”和“没有”的某些逻辑差别·····	苑玉成 ( 93 )
关于问句逻辑的几个问题·····	张尚水 ( 100 )
略谈问题(问句)逻辑·····	张家龙 ( 117 )
关于问题的逻辑特征·····	刘鸿钧 ( 123 )
<b>问题讨论</b> 横看成岭侧成峰	
——也谈概念和词义的关系·····	陈宗明 ( 137 )
也谈必要条件假言判断的	
联结词·····	倪鼎夫 ( 152 )
有适用于语言学家的逻辑学吗? ·····	[瑞典]
奥尔伍德等著 王维贤译 李先焜校	( 165 )
A、E、I、O的逻辑意义·····	周礼全 ( 183 )
逻辑代数简介·····	宋文淦 ( 207 )
亚里士多德区分多义词的方法·····	王 路 ( 227 )
<b>书评</b> 介绍《说话写文章的逻辑》·····	郑功伦 ( 237 )
<b>读者来信</b> 也谈复杂概念及其语词表现·····	吴 平 ( 242 )

# 汉语人机对话的语法分析问题\*

(中国社会科学院语言研究所) 范继淹

本文根据“80型汉语人机对话模型”的设计和实验写成。<sup>①</sup>“汉语人机对话”指人和电子计算机用汉语普通话直接“交谈”。交谈可以采取口语或书面两种方式。口语交谈涉及语音识别、语音合成等问题，困难尚多。这里讲的是书面交谈，即人用打字机输入语句，计算机“阅读”后把它的对答语句也打印出来，或者在荧光屏上显示出来。由于汉字识别还有待研究，目前输入、输出都用汉语拼音。

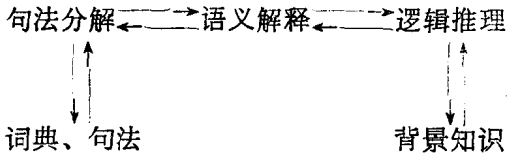
“人机对话”是通俗的说法，正式文献里称为“自然语言理解”(understanding natural language)，即研究计算机怎样才能理解人类社会的自然语言，如英语、日语、汉语等等。“自然语言理解”的探索还只有十几年的历史，然而已经成为当前人工智能研究的核心课题，在实用上和理论上都具有重大的意义。七十年代国外建立了一些英语理解的计算机模型，采用的理论和方法各不相同，都取得了一定的成果，同时也都面临着许多困难问题。国内关于汉语理解

\* 本文部分内容曾于1981年8至4月在北京大学中文系讲授。

① “80型汉语人机对话模型”由中国社会科学院语言研究所范继淹、徐志敏于1980年10月设计，11月开始在PDP11/03微处理机上进行实验，12月下旬初步试通。1981年扩大词汇和句型继续实验，现已获得成果。详细内容将另文专题报告。

的研究刚刚开始，一切尚待摸索。国外的经验可以借鉴，但是主要的是如何根据汉语的特点寻求我们自己的途径<sup>①</sup>。

自然语言理解的研究涉及语言学、心理学、逻辑学、计算机科学中的一系列重要问题，包括句法、语义、概念、联想、判断、推理、模式识别、知识表达等等，其中句法、语义和推理是三个主要部分。一个语句输入，按词查阅词典进行句法分解，语义部分配合句法进行解释，同时指令推理系统进行演绎推论，而推理又必须根据一定的背景知识。语义和推理随时反过来检验句法结构，以求得出正确的理解。三者交互作用，它们之间的关系可以图示如下：



本文只是从语言学的角度简略地谈一谈汉语人机对话中有关句法、语义和推理三个方面的一些基本概念和存在的问题，总称为语法分析问题。语法这个术语的含义，已经远远超出了传统的概念。传统的语法上包括词法和句法。乔姆斯基的转换生成语法认为，语法 (grammar) 由三大部分组成：句法部分 (syntactic component)，语义部分 (semantic component) 和语音部分 (phonological component)。此后的一些新兴的语言学派如语义生成学派、格变语法、切夫语法、功能语法等等，也都沿用了这一概念。据此，我们

① 请参阅范继淹、徐志敏：《人工智能和语言学》（《中国语文》1980年第4期）、《自然语言理解的理论和方法》（《国外语言学》1980年第5期）、《关于汉语理解的若干句法、语义问题》（《中国语文》1981年第1期）。

不妨把语法这个术语理解为“语言的法则”，词法并入词典，这样传统语言学所指的语音、词汇、句法、语义就都属于语法的内容了。

## 一、句法形式

语言是社会交际的工具。从信息论的角度来看，一种语言就是一个民族的成员在共同的社会生活中，传输信息使用的一个约定俗成的符号系统。语言形式就是信码形式，语义内容就是信息内容。句法形式是语言这个符号系统中信码形式的一个子集(和语音形式、词汇形式组成全部信码形式)。<sup>①</sup>

汉语的句法形式可以用短语结构 (phrase structure) 来描写。按乔姆斯基的重写规则，基本句型的短语结构是：句子由名词短语和动词短语构成，名词短语又可由限定语和名词构成，动词短语又可由动词和名词短语构成，如此等等。

设：S代表句子(sentence)

NP 代表名词短语 (noun phrase)

VP 代表动词短语 (verb phrase)

DET 代表限定语 (determiner)

DEM 代表指示词 (demonstrative)

NUM 代表数词 (numeral)

CLF 代表量词 (classifier) <sup>②</sup>

ADV 代表副词 (adverb)

V 代表动词 (verb)

<sup>①</sup> 请参阅范继淹：《语言的信息》（《中国语文》1979年第2期）。

<sup>②</sup> 这里是句法上的量词，例如“一个人”、“两本书”、“三支钢笔”里的“个、本、支”等等，不同于逻辑上的量词 (quantifier)。量词是汉语句法的一大特点，印欧诸语言中没有这个词类范畴。

N 代表名词 (noun)

下面的例句就可以用重写规则说明它的结构。

(1) 那两个老头儿在下象棋。

短语结构如下：

$S \rightarrow NP + VP$

$NP \rightarrow DET + N$

$NP \rightarrow N$

$VP \rightarrow V + NP$

$DET \rightarrow DEM + NUM-CLF$

$DEM \rightarrow$  那

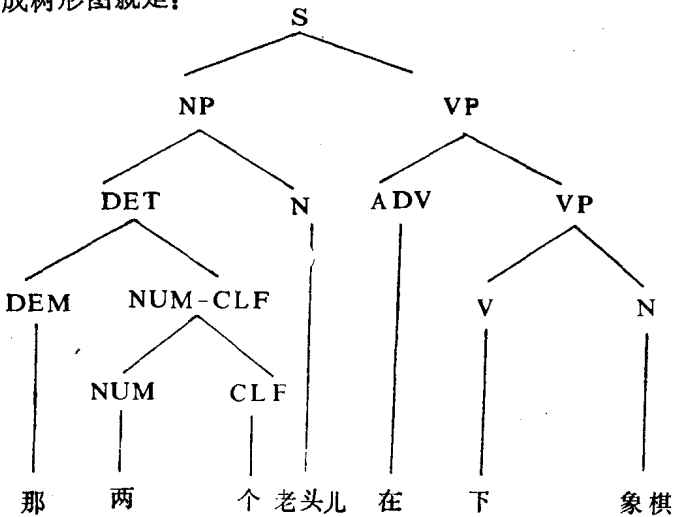
$CLF \rightarrow$  个

$N \rightarrow$  老头儿, 象棋

$ADV \rightarrow$  在

$V \rightarrow$  下

画成树形图就是：





以上只是输入句的表层结构 (surface structure)

表层结构千变万化，计算机往往很难据以作出语义解释，尤其是表层结构的形式不同而意思相同的句子。因此还必须根据表层结构求出深层结构 (deep structure)，才便于计算机理解语义。现代语法理论对深层结构有不同的解释：乔姆斯基的标准理论指的是句法上的深层结构，语义学派指的是语义结构。这里采用乔姆斯基的标准理论，为的是便于说明汉语句法上的一些变化。所谓句法上的深层结构，简单地说，就是一个句子的句法基础结构，包括词汇形式和句法范畴在内，凭以解释语义。下面举一个实例来说明。

(3) 你看这个电影吗

(4) 我看过这个电影了

一个语句输入，首先必须判明它是疑问句、陈述句还是祈使句。是疑问句就组织回答，是陈述句就贮存信息，是祈使句就执行指令。书面上用标点符号来表示疑问、陈述等等，但是标点符号是外加的，不是语言本身固有的。我们要从句法特征上来判断这些不同的语气，所以本节的例句一概不加标点（上机实验时也不用标点）。（3）是问话，句末有“吗”字；（4）是答话，句末有“了”字。我们可以根据“吗”来判断疑问句，根据“了”来判断陈述句。可是表示疑问、陈述、祈使等语气的句法手段不止一种，句式变化更是多种多样。（3）至少有以下几种表达方式：

(3) a. 你看这个电影吗

b. 你看不看这个电影

c. 你看这个电影不看

d. 你看这个电影不

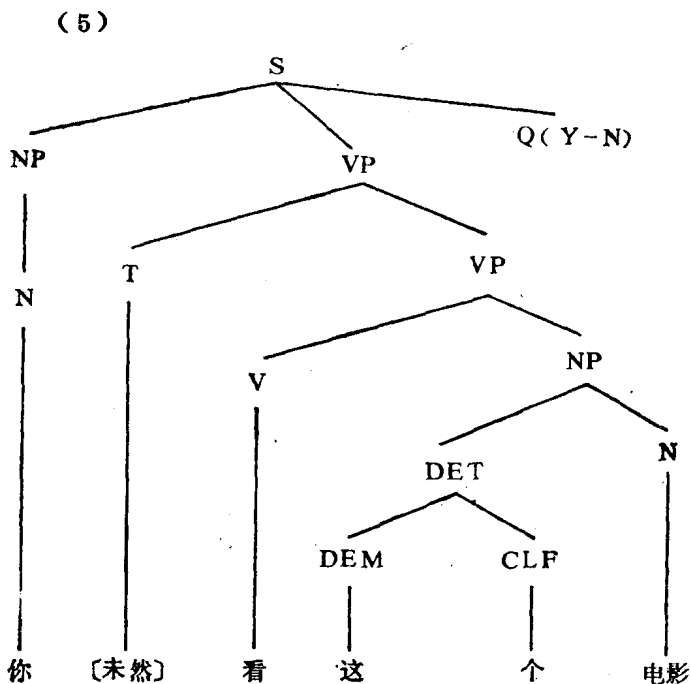
(4) 也有多种表达方式：

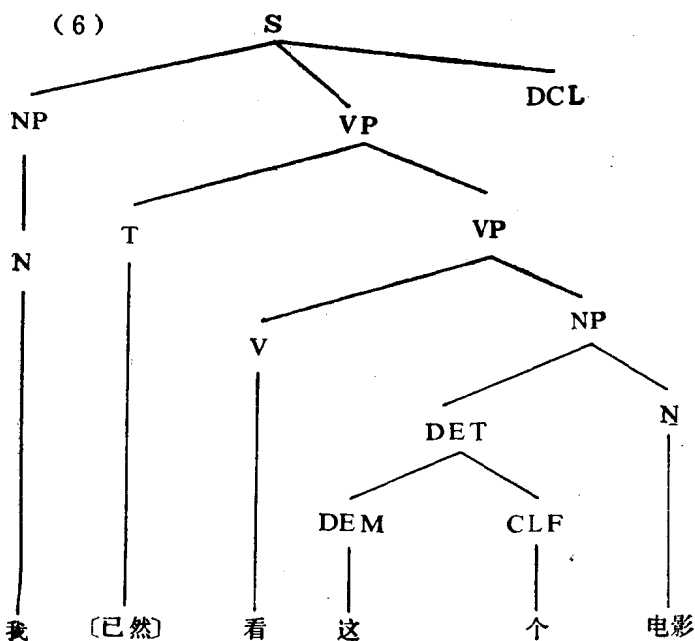
(4) a. 我看过这个电影了

b. 我看过这个电影的

c. 我看过这个电影

(8) 有五种不同的表层结构，然而意思相同，  
(4) 有三种不同的表层结构，意思也一样。这说明这些不同的表层结构有一个共同的句法基础，这个基础就是深层结构。(8)，(4) 的深层结构可以分别用 (5)，(6) 来粗略地表示：





Q (Y-N) 是是非问句的疑问标记(yes or no question), DCL是陈述标记 (declarative), T是时态标记(tense)。

〔未然〕、〔已然〕是时态标记的具体内容,属于句法范畴,不是词汇形式,加方括号表示。(3)的五种表层结构都由(5)生成,所以意思相同。(4)的三种表层结构都由(6)产生,所以意思也一样。

用这样的深层结构来解释语义就比较方便。(5)是是非问句,VP的时态是未然。“你”指听话人。“看”的义项很多,可以表示观看、阅读、看望、诊治、认为……等等,但是跟“电影”组合,根据语义选择规则只能表示观看。“这个”表示定指。全句的语义可以解释为“询问对方是否将要

观看某个定指的电影，要求对方作出肯定或否定的答复。”

(6) 是陈述句，VP 的时态是已然，“我”指说话人，其他成分同(5)。全句的语义可以解释为“说话人过去曾经观看对方刚才说到的、定指的电影。”在本例的情况下，这就是一个否定的答复。但是全句并无否定的词语，否定的语义从何而来我们在下一节里再谈。

一个句子的句法形式是从深层结构到表层结构的衍生过程中，一系列句法选择的结果。同一个深层结构选用不同的转换规则，就可以衍生出不同的表层结构来。(5)的Q如果取值为“吗”，就生成(3a)；取值为“V不V”(前后两个V必须相同)，就生成(3b)。(6)的DCL取值“了”，就生成(4a)；取值“的”就生成(4b)；取值为零，就生成(4c)，等等。转换生成语法研究从深层结构到表层结构的转换规则，是一个编码的过程。计算机理解输入句，要从表层结构的形式特征求出原来的深层结构，是一个译码的过程。据上述各例，我们可以建立以下的句型译码规则。

(7) 是非问句句型的译码规则

NP + VP + 吗		⇒Q(Y - N)
NP + V + NEG + DUV + (NP)		
NP + VP + NEG + DUV		
NP + VP + NEG		

竖线左边是输入句的各种形式特征，凡符合其中之一者就是是非问句，在深层结构中就要建立标记Q(Y - N)。(NP)表示NP可以不出现。NEG代表否定词(negative)“不”或“没”。DUV代表跟前面动词重复的动词(duplicated verb)。

(8) 陈述句句型的译码规则:

NP + VP + 了

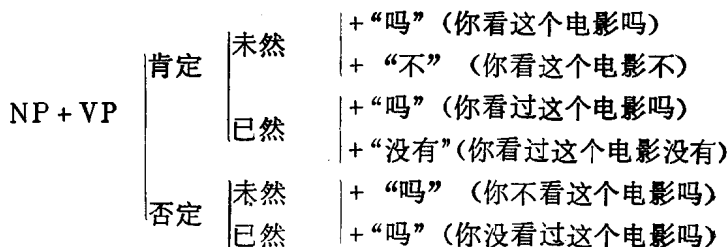
NP + VP + 的  $\xrightarrow{\text{DCL}}$  CL

NP + VP +  $\Phi$

∅表示零，即句末没有语气助词。符合以上形式特征之一者，就是陈述句，在深层结构中就要建立标记DCL。(7)和(8)只是一个很粗略的基本规则，某些细节尚待增添。还有一些语气助词如“吧”、“呢”既可表示疑问，又可表示陈述，尚待深入考察它们的用法规律。

(7)的最末一条形式特征 NP + VP + NEG 就否定词“不”或“没有”居于句末(你看这个电影吗 | 你看过这个电影没有)，是省略了后面重复的动词。但是从计算机处理的角度看，不妨认为句末的“不”和“没有”已经虚化为表示疑问的语气助词，句法功能相当于“吗”。这样，“吗、不、没有”就构成一个表示是非问句的疑问助词系统：

“吗”既可用在肯定式句末，又可用在否定式句末；既可用于未然，又可用于已然。“不”和“没有”只能用在肯定式句末，“不”只用于未然，“没有”只用于已然。换言之，“吗”是一个全面的疑问助词，“不”和“没有”各自相当于“吗”在肯定式句末的一种用法。图示如下：



图中竖线右边是一组互斥的语义成分或词汇形式。

## 二、语义内容

一个句子的语义内容,就是它所传输的全部信息,包括句内信息和句外信息。句内信息就是本句内部构成成分的词汇意义、词义之间的组合关系以及时态、语气等句法范畴。句外信息指来自本句之外、非本句所固有的信息。句子不是孤立的,而是在整段语流中出现或者在特定环境下说出来的。上下文、现场场景往往附加某种信息于句子之上,甚至改变句子本身的某些信息。

1. 句内信息。上一节我们谈了乔姆斯基的句法深层结构,用它来解释语义。一方面是便于说明汉语的表层形式变化多端,一方面也确实可以解释一些简单的句子。但是,如果句子的语义稍为复杂一点,问题就产生了。首先,短语结构以及由深层转化到表层的转换规则都只是一些词类的标记,如NP、VP等等,并未表明句中哪一个NP是主语、哪一个为宾语、哪一个是谓语,句法关系不明。其次,即使设法把主语、谓语、宾语等标示出来,也无法表示多种多样的语义关系。例如:

(9) a. 门开了。

b. 我开了门。

c. 这把钥匙开了门。

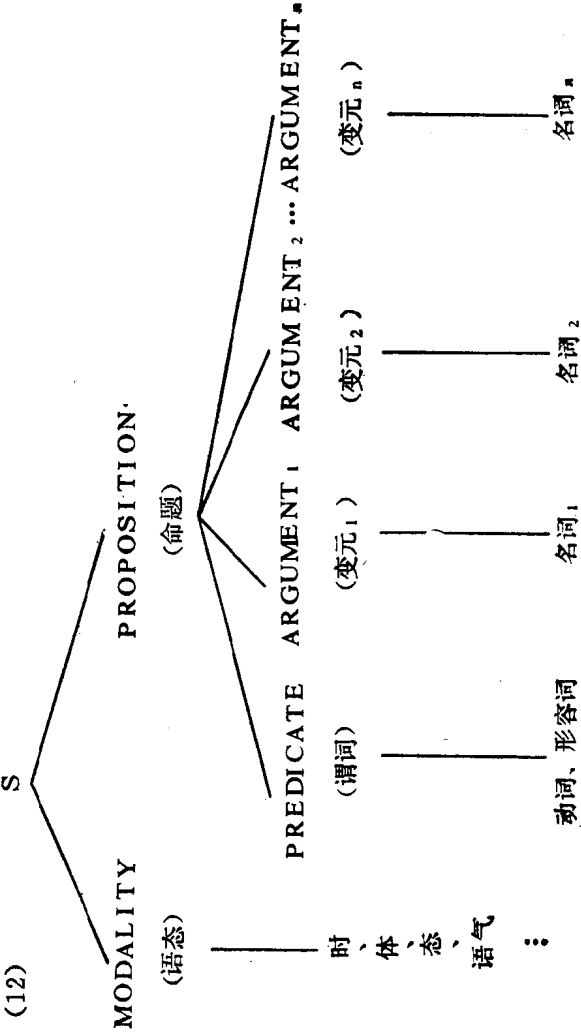
a的主语NP“门”是描述对象,b的主语NP“我”是施事,c的主语NP“这把钥匙”是工具。同是主语,含义不同。而且“门”在a,b,c三例中都是同一过程的经受者(由关闭状态到开启状态),然而在a中是主语,在b,c中却是宾语。语义所指相同,句法职能不同。这说明动词和名词之间的语义关系不能靠主语、宾语等句法术语来解释。特别是汉

语，动词和名词之间的语义关系极其复杂：

- (10) a. 他走了 (施事 + 行为)  
b. 问题解决了 (受事 + 行为)  
c. 香山去过了 (处所 + 行为)  
d. 牛皮纸包书 (工具 + 行为)  
e. 名字写上了 (结果 + 行为)  
.....
- (11) a. 看了一篇文章 (行为 + 受事)  
b. 写了一篇文章 (行为 + 结果)  
c. 住了四个人 (行为 + 施事)  
d. 住老王家 (行为 + 处所)  
e. 坐飞机 (行为 + 工具)  
f. 唱女高音 (行为 + 类型)  
.....

(10) 是主语和谓语的例子，(11) 是动词和宾语的例子，都有多种多样的语义关系，句法上的深层结构对此很难作出解释。于是语义学派认为，一个句子的深层结构，应该是它的语义结构。

语义很抽象，各家的分析不同。我们不拘泥于一家之言，凡是便于说明汉语的，便于计算机处理的，我们都采用，并结合汉语的特点，加以修改补充。人机对话最困难的部分就是语义，目前尚未很好解决，还处于摸索阶段。图(12)是一个句子的基本语义结构。语态包括时(现在、过去、将来)，体(未然、已然、持续等)，态(主动、被动等)、语气(陈述、疑问、祈使等)以及其他范畴。谓词包括动词和形容词或由它们构成的短语，变元由名词或名词短语构成。一个句子的语义由语态和命题两部分构成。语态就是各种句法范畴。命题是语义

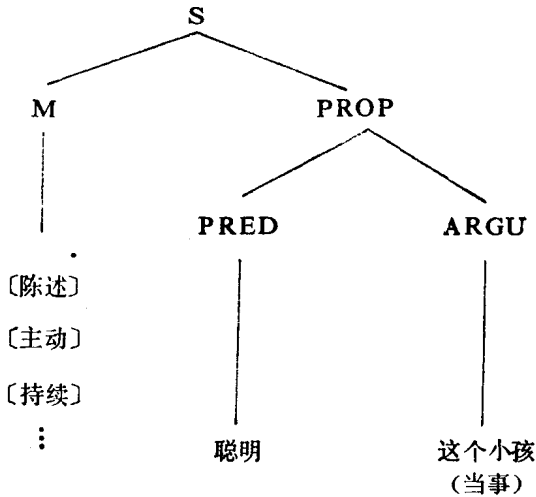




的核心部分,由一个谓词带上若干个变元构成。谓词表述一种行为或状态,变元是这个行为或状态的参与者。一个谓词能带几个变元,由它的语义决定。只能带一个变元的称为“一元谓词”,能带两个变元的,称为“二元谓词”;以此类推,“三元谓词”……直到“n元谓词”。

汉语的一元谓词以形容词最典型,例如“聪明”,只能跟一个变元发生关系:“这个小孩聪明”。二元谓词以动词为常,例如“看”,可以跟两个变元发生关系:“我在看小说”。三元谓词也是动词,但是较少,例如“送”,可以跟三个变元发生关系:“他送了我一本书”。汉语有没有四元谓词、五元谓词等等,尚待进一步考察。现在把以上三个例子的语义结构图示如下:

(13)这个小孩聪明



M是语态的英文缩写, PROP是命题, PRED是谓词, 下