

HANYU YUFA DE YIHE WANGLUO

鲁川 著

# 汉语语法 的 意合网络



商务印书馆

# 汉语语法的意合网络

鲁 川 著

商 籍 印 书 馆

2001 年·北京

**图书在版编目(CIP)数据**

汉语语法的意合网络/鲁川著. —北京:商务印书馆,  
2001

ISBN 7-100-03347-0

I. 汉… II. 鲁… III. 汉语-语法 IV. H14

中国版本图书馆 CIP 数据核字(2001)第 042117 号

**所有权利保留。**

**未经许可,不得以任何方式使用。**

HÀNYǔ YÚFǎ DE YÌHÉ WǎNGLUÒ

**汉语语法的意合网络**

鲁 川 著

---

**商 务 印 书 馆 出 版**

(北京王府井大街36号 邮政编码100710)

**商 务 印 书 馆 发 行**

**北 京 民 族 印 刷 厂 印 刷**

ISBN 7-100-03347-0/H · 847

---

2001年10月第1版

开本 850 × 1168 1/32

2001年10月北京第1次印刷

印张 9 1/2

**定价: 16.00 元**

# 前 言

汉语语法研究面临来自两方面的挑战:一是对外汉语教学,一是中文信息处理。前者迫切要求对于种种的语言现象,特别是对于语言对比中相同和相异的语言现象做出理论上的解释;后者迫切要求对于人类的语言机制,特别是对于通过语言来进行知识处理的大脑机制做出认知理论的探索和电脑软件的模拟。

本书的作者毕业于哈尔滨工业大学计算机系,长期从事中文信息处理的科研和教学工作。1987年当选中国中文信息学会首届计算语言学专业委员会主任。在陈力为院士领导的“八五”国家重点科研项目“中文信息处理应用平台工程(905工程)”中担任《汉语谓词语意组合词典》课题组负责人。

本书根据上述科研成果写成,是作者在实际工作中长期坚持学习汉语语法的心得笔记,是为电脑编制汉语语法规则集的经验总结,也是从中文信息处理的角度撰写的一本汉语语法专著。可供关心中文信息处理事业的读者以及语言专业(中文系、外语系)和信息专业(计算机系、自动化系、通信工程系)本科生和研究生阅读,也可供语言学家、语言教师和广大的汉语学习者参考。

信息时代已经到来,电脑正在从数据处理过渡到知识处理。电脑对自然语言的理解就是对语言中显性和隐性的知识的发掘。

语言是知识的载体。语法是语言的结构规则。适用于电脑的语法就是电脑通过语言来进行知识处理时所遵循的结构规则。

本书汲取了百年来国内外语言学者的研究成果,重点参考了吕叔湘先生《中国文法要略》和转换生成语法、格语法、依存语法、角色参照语法、认知语法的理论。

### 一、贯穿本书的基本理论主要有列六个论点

#### 1. 谓词是句子结构的中枢

《中国文法要略》和“依存语法”都提出了“动词中心论”。汉语动词和形容词都是谓词,本书的论点是“谓词中枢论”。

句法平面上:在一个单句中,一个由谓词充当的述语支配着主语、宾语、状语、补语等句子成分。

语意平面上:在一个事件中,一个由谓词充当的“中枢事元”支配着若干个相关的“周边事元”。

#### 2. 语意平面是汉语语法的基础

汉语语法研究的三个平面:语意平面研究语言的意义内容,句法平面研究语言的结构形式,语用平面研究语言的交际效果。说话首先从传递“语意内容”的目的出发,然后找到符合句法结构的多种形式,最后选定一种语用效果最佳的表达方式。这体现了“语意的先决性,句法的强制性,语用的选定性”。

英语有形态,是“形合法”语言,注重句法平面。

汉语无形态,是“意合法”语言,注重语意平面。本书以“意合网络”的形式全面讨论了汉语语意平面上的各种问题。

#### 3. 语块是三个平面同构分析的单位

“语块”是心理学的术语。美国心理学家米勒(G. Miller)指出:人脑短时记忆限度是 $7 \pm 2$ 个语块。本书将“语块”移植到语法学中以取代“短语”。短语(phrase)种类繁多,在不同的层次中长

度不定,不适于“同构分析”。语块(chunk)总共只有两种(中枢语块、周边语块),长度确定,边界清晰,特别适于句法平面、语意平面和语用平面互相对照的“同构分析”。

#### 4. 分清语言中的客观信息和主观信息

用语言传递客观信息的同时,不可避免地要附加说话者的主观信息。主观信息表达说话者在观察客观事物时的主观视角和参照系,以及说话者在叙述客观事物时的主观情绪、态度和交际意图等。主观信息表现为主观因素和主观标记。

实词主要表达客观信息。隐含在实词内的主观信息叫做主观因素,如表达说话者主观态度的褒贬因素等。

虚词主要表达主观信息。虚词本身就是显露在客观信息(实词)之外的主观标记,如表达说话者主观情绪和态度的“情态标记”(情态辅词和情态助词)等。

#### 5. 汉语语序的理据是临事性

汉语的编码机制是“援物、取象、比类、尽意”。“援物”就是汉语的编码以客观物质世界为依据。汉语语序的理据就是对客观有序世界的“临摹性”(iconicity)。本书在海内外的语言学者研究成果的基础上,归纳出下列三条基本语序规律:

① 时空事理先后律:时空早的先说,时空晚的后说。

原因条件先说,结果推理后说。

② 时空地位大小律:时空大的先说,时空小的后说。

地位高的先说,地位低的后说。

③ 信息旧新轻重律:旧知信息先说,新传信息后说。

次要信息先说,重要信息后说。

## 6. 里层网络和表层序列相互转换

转换生成语法提出语言有“深层结构”和“表层结构”。我国学者范继淹先生和邢福义先生指出“深层”应为“里层”。

本书提出:语言的“里层结构”是多维的“网络”。

“里层网络”是由“结点”和“联线”所组成的图式。每个结点中是一个实词,作为标记的虚词依附在结点之上,结点间的联线中断处注释各种“语意关系”和“语意依附”。

语言的表层结构是一维的序列。从语用的角度来看,主要是安排话题和焦点的顺序以达到最佳的交际效果。

里层网络体现语言信息的“完备性”和知识的“显露性”。

表层序列体现语言信息的“简洁性”和知识的“隐含性”。

电脑对人类语言的理解表现为能把表层序列转换成里层网络。

电脑对人类语言的生成表现为能把里层网络转换成表层序列。

## 二、本书以大规模语料库的成果作为研究的依据

“语法”来源于全部词汇的“用法”。本书的研究基于大规模的语料库,所依据的“八千常用谓词”和“八百常用虚词”是语料库中的高频词。本书对这些语料进行“逐个分析、归纳分类、探求解释、揭示规律”,所列出的“128个基本句模”是从语料库中大量现代汉语句子中抽象出来的实用模式。

## 三、本书以汉语的特点作为研究的基点

在本书的概论中,论述了现代汉语语法的下列六个特点:

- ①最小语法单位是字,
- ②语法组合是意合法,
- ③语法理据是临摹性,
- ④语言节奏的和谐律,
- ⑤语言表达的简缩化,
- ⑥语言铺排的话题链。

汉语以“话题—说明”为框架,英语以“主语—谓语”为框架。

本书只借鉴英语语法的理论和方法而不生搬硬套其语法体系。在书中各有关章节中还进行了汉语跟英语的对比。

在本书写作过程中得到北京信息工程学院软件工程研究中心周锡令教授和各位同事的支持和帮助,在此一并致谢。

在本书出版的时候,我衷心地感谢直接领导我们科研攻关的陈力为院士,感谢国内外语言学界和计算语言学界的良师益友们。最后要特别感谢商务印书馆的张万起先生和周洪波先生,他们为本书的审订和出版倾注了大量的心血。

由于作者水平所限,书中多有不当之处,敬请批评指正。

鲁 川

1998年11月于北京信息工程学院软件工程研究中心



# 目 录

|                   |     |
|-------------------|-----|
| 前言                | 1   |
| 第一章 概论            | 1   |
| 第一节 信息和语言         | 1   |
| 第二节 汉语的语法研究       | 13  |
| 第三节 汉语信息语法的句法平面   | 27  |
| 第二章 意合网络的内容和形式    | 39  |
| 第一节 人脑联结机制和电脑知识网络 | 40  |
| 第二节 意合网络中的语意单位    | 45  |
| 第三节 意合网络中的语意组合    | 59  |
| 第四节 意合网络的图式       | 69  |
| 第三章 配合关系的中枢角色     | 76  |
| 第一节 中枢角色的分类       | 77  |
| 第二节 中枢角色的定义       | 80  |
| 第三节 中枢角色的区别       | 87  |
| 第四节 谓词的配元和名词的配系   | 95  |
| 第四章 配合关系的周边角色     | 108 |
| 第一节 周边角色的分类       | 109 |
| 第二节 周边角色的定义       | 112 |
| 第三节 周边角色的区别       | 121 |

## 2 目录

|                           |     |
|---------------------------|-----|
| 第四节    周边角色的标记·····       | 135 |
| 第五章    加合关系和接合关系·····     | 144 |
| 第一节    加合关系·····          | 144 |
| 第二节    接合关系·····          | 154 |
| 第六章    语言模型和汉语句模·····     | 168 |
| 第一节    自然语言的语言模型·····     | 168 |
| 第二节    现代汉语的基本句模·····     | 177 |
| 第七章    语意依附及其标记·····      | 208 |
| 第一节    客观信息和主观信息·····     | 208 |
| 第二节    事件的依附及其标记·····     | 217 |
| 第三节    述谓意结的依附及其标记·····   | 226 |
| 第四节    指称意结的依附及其标记·····   | 236 |
| 第八章    从网络到表层序列的生成·····   | 240 |
| 第一节    信息结构和话语结构·····     | 240 |
| 第二节    汉语的语序及其基本规律·····   | 245 |
| 第三节    意合网络在电脑翻译中的应用····· | 271 |
| 第四节    联想网络在语言理解中的应用····· | 277 |
| 附录    语意注释的代码总表·····      | 284 |
| 主要参考书目·····               | 287 |

# 第一章 概 论

## 第一节 信息和语言

### I . 信息和知识

#### 一、信息(information)

信息是当代最流行的词语之一。翻开书报充满眼帘的是“信息时代”、“信息科学”、“信息技术”、“信息产业”、“信息就是财富”等等。但“信息”的定义却是众说纷纭。

《现代汉语词典》(1996年修订本)对“信息”的释义只是：“①音信；消息。②信息论中指用符号传送的报道，报道的内容是接收符号者预先不知道的。”

参考了国内外许多学者的论著，我们建议信息的定义如下：

**信息是物质的基本属性之一，是物质间接存在性的自身显示。  
信息的作用是消除不确定性。**

这个定义正确地阐明了“信息”的本质和作用。

1、物质(matter)有三个基本属性：“材质”(material)、“能量”(energy)、“信息”(information)。

材质取决于物质结构。“氢”跟“氧”材质不同是因为“氢”原子核只有一个“质子”而“氧”原子核有八个“质子”。

能量取决于运动状况。同一个“花盆”摆在二楼阳台上“势能”

较小,摆在九楼阳台上则“势能”较大。

材质和能量反映的是物质的直接存在性。它们是“守恒”的,质量守恒定律和能量守恒定律是我们都知道的。

信息反映的却是物质的“间接存在性”。它是不守恒的。有一种儿童学英语的字母积木,是做成英语字母形状的 26 个塑料块。我们取出三块字母“E”、“N”、“T”进行排列。从材质上看,无论怎么排列这三块字母的总重量都是不变的,即总质量是守恒的。从能量上看,如果它们每次都静止地摆在同一地点,无论怎么排列这三块字母的总能量也是守恒的。但是,这三块字母如排列成“TEN”,则显示的信息是“十”;如排列成“NET”,则显示的信息是“网”。

“控制论”奠基人维纳(N. Wiener)说:“信息既不是物质,也不是能量,信息就是信息。”维纳所说的“物质”,指的就是“材质”。他指出:信息既不是物质第一基本属性“材质”,也不是物质第二基本属性“能量”,信息是物质的第三基本属性。

2、有人说“信息是被反映的物质属性”。似乎有了反映才有信息。其实,我们能看见(反映)某个星体,是因为接收到了它发出的以光为载体的信息。在我们看到这个星体之前,这颗星的材质早已直接存在,它的信息也早已间接存在。可见,是先有信息而后才有反映。就是说,信息是物质自己发出来的。也就是说:信息是物质的“自身显示”。

3、“信息论”奠基人申农(C. Shannon)指出:“信息是用以消除不确定性的东西”。这个定义正确地指明了信息对人类的重要性,就是信息能够消除人类在认识和改造客观世界中必然要遇到的大量的不确定性。

从人们传递信息(说话)的过程来考察一下。人在说话时,总是首先说出一个话题(topic),然后再对这个话题加以说明(comment)。这就是说,说话者先用话题来提出一个不确定性,以引起听话者的悬念,然后再用说明去消除这个不确定性,从而解除听话者的悬念。

## 二、知识(knowledge)

知识是人类所积累的系統化的优化信息之总和。

1、知识是高级的信息。从本质上看:知识就是信息。但是知识比一般信息的级别高。所以知识处理是高级的信息处理。

2、知识是优化的信息。宇宙中充满了各种各样的信息:有些是对人们有用的,有些是对人们无用的或暂时无用的。因此,人们对于通过多种途径所获得的大量信息要通过实践的检验,并进行“去伪存真、去劣存优”的优化(optimization)。

3、知识是系统化的信息。人类所获得的有用信息非常之多,如果是杂乱无章的,则不便于利用。所以在对所获信息不断优化的同时还要分类整理,加以系统化(systematization)。

4、知识是全人类长期积累的有用信息的总和。人类的认识总是要受到历史的局限。随着时间的推移,知识也不断地更新。随着社会的发展,生产力中的知识含量也不断地增加。当生产力中的知识含量超过一定的阈值时就会出现知识经济。

## 三、人类文明史和信息时代

人类文明史是按照生产力的水平来分期的。并按照依次开发材质资源、能量资源和信息资源的顺序。

### 1、农业时代

农业时代的特点是人类对材质资源的第一次开发。就是对野生植物进行种植而发展为农业,对野生动物进行驯养而发展为畜牧业。

随着农业的发展,人和牲畜的力量太小限制了生产力的提高,发展生产力的需求呼唤着开发新的强有力的动力机械。

### 2、工业时代

18 世纪 60 年代英国的瓦特(J. Watt)发明了蒸汽机。从而奏响了工业时代的序曲。

工业时代的特点是人类对能量资源的第一次开发。就是利用自然能源来发动机器,逐步代替人类的繁重的体力劳动。

随着工业的发展,跟市场需求、原料供应、商品价格、股市行情有关的大量信息亟待处理。人类信息处理的小记忆量和低速度限制了生产力的提高,发展生产力的需求呼唤着开发新的高速的和记忆量的信息处理机器。

### 3、信息时代

1946 年美国第一台电子计算机 ENIAC 的问世和 1957 年苏联第一颗人造卫星 SPUTNIK 的升空,揭开了信息时代的序幕。

信息时代的特点是人类对信息资源的第一次开发。就是开发高速度大容量的信息处理机器和信息传输设备,以逐步地辅助和代替人类的脑力劳动。

在信息时代,经济的发展不再只取决于物质资源(材质资源和能量资源)的开发和劳动力的多少,而越来越取决于信息资源的开发深度和劳动者的知识素质和创新能力,即取决于全社会对信息和知识的重视和开发以及科学技术迅速转化为生产力的综合实

力。就是说,在信息时代,信息比材质和能量更为重要。

20 世纪 80 年代,由“超大规模集成电路”引起的计算机微型化,使之获得了“电脑”的美名。一方面是开始研制以知识处理为特征的智能计算机,一方面是电脑的大幅度降价和大面积普及。以信息技术为龙头的多种高新技术的发展,出现了“知识给产品增添了高额附加值”的现象。软件的飞速发展,网络的无所不在,虚拟现实的广泛应用,知识经济的初露端倪,这说明逐步实现人类脑力劳动机械化和知识就是财富的信息时代已经来到了我们的身边。

## 四、语言信息处理

### 1、人类的语言信息处理

- ① 编码:把意义内容按照语法规则编成一定的代码形式。
- ② 发送:通过说话写字印刷把编好的代码形式发送出去。
- ③ 传递:语言的声音图像通过各种媒体的信道向外传播。
- ④ 接收:通过听觉器官和视觉器官接收语言的代码形式。
- ⑤ 解码:对语言代码形式按语法规则去理解其意义内容。

### 2、电脑的语言信息处理

电脑是模仿人类信息处理功能的机器。

电脑(electronic brain)是一个通俗的名称,其正式名称是“信息处理机”(information processor)。

第一阶段是数据处理阶段。电脑所进行的处理主要是对于数值和字符的数学运算和逻辑运算。即使是进行文字处理,也是不理解文字所表示的意义,只是把文字也看成一种数据而对其进行运算。所以这一阶段的电脑被名符其实地称为“计算机”(computer)。

第二阶段是知识处理阶段。电脑进行的处理主要是对语言文字的识别、理解、问答、摘要、扩充、转换或翻译。电脑把语言文字作为知识的载体,通过对语言文字的意义的理解而进行知识处理。这时仍称为计算机就名不符实了。要么称为“知识处理机”,要么名符其实地称为“电脑”。

电脑能否顺利地进入知识处理的阶段,不仅取决于硬件的发展水平,更重要的是取决于软件的发展水平,特别是取决于语言研究的发展水平。也就是说,为了让电脑能够高水平地进行知识处理,必须有适于电脑信息处理的高水平的语言研究。

## II. 语言和语法

### 一、语言(language)

原先,语言有个定义:“语言是人类最重要的交际工具”。

1946 年出现了电脑,1948 年出现了信息论、控制论、系统论。这使人类改变了语言观。在信息时代,语言有了新的定义:

**语言是人类和电脑用来传递信息、实施控制的符号系统。**

这种新的语言观反映了下列的新观点:

#### 1、语言是人类表达知识的编码体系

人类在认识和改造客观世界中形成了一定的知识,把这种知识编成一种“码”,就形成了“语言”。于是,语言就成了知识的载体。人们用语言进行知识的表示、存储、传播、交流、继承、创新。每个民族都有自己的编码体系,反映了不同民族在思维方式上既有相似性也有差异性。因此,既要研究人类语言的共性,也要研究



民族语言的个性。

## 2、语言是探索人类思维奥秘的窗口

思维(thinking)是人类的一种精神活动,是在表象、概念的基础上进行分析、综合、判断、推理等认知活动的过程。

思维和语言如形影相随,不可分离。当人们不说话而仅仅是动脑筋时,也在用语言来进行思维。可见,语言也是思维的工具。人们对表象进行概括而形成的概念,总要落实到词语的形式上;人们进行判断和推理时,总要把表示概念的词语组成句子。因此,要想对看不见、摸不着的思维进行研究,只有通过研究语言来实现。

## 3、语言是人和电脑共有的交际工具

进入信息时代以后,运用语言的主体不再仅仅是人类,也包括电脑。从此进入了人机对话和人机互助的时代。

电脑已经成为帮人类进行语言交际的助手,正在成为向人类学习语言交际的学生,将要成为跟人类进行语言交际的主体。

目前电脑还不能完全理解人类的语言。为了跟电脑实际使用的“程序设计语言”(PASCAL, PROLOG, JAVA 等)相区别,把不受限制的人类语言称为“自然语言”(natural language)。

## 4、语言是传递信息的最主要的载体

传递信息可以有各种载体:从古代边疆的烽火台到现代街道的红绿灯,真是多种多样。当前风靡全球的“多媒体”,是声音、图像、语言三种媒体的融合。但是从传播知识信息的角度来看,语言是最主要的媒体。音乐家写讲如何作曲的书,不可能全书都是乐谱,书的大部分篇幅还是得用语言文字。所以,语言是传递信息和知识的最主要的载体。