

XIANDAI HANYU  
JUZI SHENGCHENG  
WENTI YANJIU

现代汉语句  
子生成问题研究

——一个以语序为样本的探索

刘鑫民 著



华东师范大学出版社

## 图书在版编目(CIP)数据

现代汉语句子生成问题研究:一个以语序为样本的探索/刘鑫民著. —上海:华东师范大学出版社,2004. 12  
ISBN 7-5617-4025-5

I. 现... II. 刘... III. 汉语-句法-研究-现代  
IV. H146.3

中国版本图书馆 CIP 数据核字(2004)第 125994 号

## 现代汉语句子生成问题研究

著 者 刘鑫民  
特约编辑 黄曙辉  
封面设计 黄惠敏  
版式设计 蒋 克

出版发行 华东师范大学出版社  
市场部 电话 021-62865537  
门市(邮购)电话:021-62869887  
门市地址 华东师大校内先锋路口

业务电话 上海地区 021-62232873  
华东 中南地区 021-62458734  
华北 东北地区 021-62571961  
西南 西北地区 021-62232893

业务传真 021-62860410 62602316

http://www.ecnupress.com.cn

社 址 上海市中山北路 3663 号  
邮编 200062

印 刷 者 上海长阳印刷厂  
开 本 890×1240 32 开  
印 张 10.125  
字 数 273 千字  
版 次 2004 年 12 月第一版  
印 次 2004 年 12 月第一次  
印 数 001-3 100  
书 号 ISBN 7-5617-4025-5/H·279  
定 价 16.00 元

出 版 人 朱杰人

(如发现本版图书有印订质量问题,请寄回本社市场部调换或电话 021-62865537 联系)



## 作者简介

刘鑫民,毕业于复旦大学汉语言文字学专业,获博士学位。现为复旦大学国际文化交流学院副教授。主要从事汉语语法研究及对外汉语教学研究,发表论文40多篇,著有《语言文字探索集》等,参编教材《现代汉语概论》、《标准汉语教程》。

Apr 30 / 10

## 序

句子的生成问题,是计算机处理自然语言的难点,也是目前国内  
外语言学界关注的热点。人们根据不同的语言学理论,提出了不少  
的句子生成模式,真可以说是“百花齐放”。刘鑫民《现代汉语句子  
生成问题研究》一书以三个平面的理论作为指导思想,结合汉语的  
句子事实,提出了一个独创的句子生成的新模式,即以语义为基础的  
串行整合句子的生成模式,并在这个框架内系统地刻画了汉语句子  
生成的过程。这不仅有重大的理论意义,而且对人工智能的研究和  
机器自动翻译的实践也是很有实用价值的。

本书作者认为:思想的产生是句子生成的前提,因此语义是句子  
生成的基础,句子生成的过程就是将语义信息和特定语境中建构起  
来的语用信息通过句法编码、语音编码转换成一个表层可以感知的  
线性序列的过程,即以语义为基础串行整合加工的过程。作者把这  
个过程分为语义建构、语用建构、句法编码和语音编码四个阶段。在  
这个理论框架的基础上,较详细地说明了汉语句子语义网络的构成  
形式、对语义网络压模形成句模的规则以及句模向句法平面投射的  
操作过程,并以语序为重点考察并论述了制约汉语句子生成的种种  
因素和这种因素之间相互制约、相互作用的互动关系以及制约句子  
生成形式的规则之间的相互整合方式。作者从宏观上提出了一个全  
新的、独特的句子生成模式。这个模式比较符合句子生成的客观实  
际,在实践上应该是可行的。

作者对国内外相关文献掌握得比较全面,有较深的理论功底,勤

于探索,并善于运用先进的语言理论和研究方法来研究。此书具有创造性,概念清晰,论证严密,分析细致,条理通顺,是一本高质量的优秀的学术著作。

句子的生成问题涉及的问题很多,书中对某些问题的论述还不够清楚,有的还没有很好地展开。理论是否得当,还需要在实践中进一步地检验。此书的出版,是一个好的开头,希望作者在此基础上继续研究,争取获得更大的成绩。

范 晓

2004 年春于复旦大学

# 序

人类有一个复杂的语言系统,正是这一点把人和动物区别开来。因此,人类一直在试图弄清楚思想是如何转换成语言表达出来的。早在1922年,法国的布鲁诺(Brunot)就指出,探索由思想到表述的途径是语言学的一项基本任务。但是,从外部刺激引起言语动机到言语的产生是一个复杂的过程,在这个过程中,进行运算的装置是人的大脑,由于这是一个不能打开直接观察的黑箱装置,这就造成了对这个问题的研究有很大的难度。刘鑫民博士知难而进,撰写了这本《现代汉语句子生成问题研究》。这是沿着思想到表述的途径来探讨句子生成问题的一个研究成果,也是迄今我所见的在现代汉语句子生成问题研究上最具系统性和全面性的一部专著。纵观全书,其特点表现在以下诸方面:

## 一、理论探索与切合应用相结合

句子生成研究是一个具有重大的理论意义的研究课题,作者在对现代汉语句子生成过程进行深入的理论探索的同时,又力求设计出一个可以为计算机语言处理所应用的句子生成模式。

作者认为,思想的产生是句子生成的前提,因此语义是句子的生成基础,从整体上看,句子生成经过了语义建构、语用建构和句法编码、语音编码几个阶段,这几个阶段是前后相继的,因此,句子生成呈现出串行(线性)加工的特点;但在每个阶段内的操作过程中可能会有多种因素同时作用,又呈现出并行加工的特点。因此句子生成是一个串行整合加工过程。基于这样一种基本的理论认识,作者提出

了一个新的句子生成模式:串行整合句子生成模式。作者认为,句子的生成就是在相关知识的指导下将所要表达的信息进行编码转换成特定语境中的句子的操作过程。为了反映这样一个过程,作者提出的串行整合加工模式整体上由三个模块构成:即知识库模块、操作模块和控制整合器模块。知识库包括多个子库,它负责提供句子生成时所需要的相关知识,操作模块是句子生成时进行编码、转换等运算的模块,控制整合器的作用是在各个模块之间起协调作用和对各种制约因素、制约条件进行临时整合。系统内部的基本运作方式是:操作是以线性(串行)的方式进行的,在进行每一步操作时,通过控制器的作用,指导操作从哪些知识库中获取相关的知识,同时负责将新获取的知识和前边模块操作的结果进行整合,这样使生成形式实现从上一层次到下一层次的转换,从而完成这次操作。这样的操作从语义建构开始,依次经过语用建构、句法编码、语音编码,最后在叠加了制约句子生成形式的各种因素之后,输出一个确定的解,即表层的线性序列。这样的句子生成模式,是比较科学的句子生成模式,它具有线性句子生成模式的优点,即比较容易处理各个模块之间的关系;同时又兼有整合模式的长处,即可以在句子生成过程中附加多种约束条件从而获得一个更确定的解,以此提高生成系统的可信度。

理论是为实践服务的。作者在书中考虑得比较多的是如何用自己建立起来的理论体系为计算机自然语言处理服务。实际上,在汉语作为第二语言教学(即对外汉语教学)领域中,句子生成问题的研究同样具有很强的实践意义。汉语作为第二语言的语法教学的目的,不在于让学生简单地掌握一些静态的语法规则,而在于让学生在学静态的语法规则的基础上习得汉语句子的生成机制。因此,汉语句子的生成研究的结论,实际上就决定了我们怎样去设计一个科学、合理的对外汉语教学语法系统的基本框架,同时也决定了对外汉语语法教学的基本思路。在对外汉语语法教学中,我们往往会发现,学生学了语法,但并不能避免生成一些不合格的句子,这说明我们对汉语句子的生成的制约因素的认识还是很全面的,还有待系统深入地

进一步探讨。

## 二、宏观研究和微观研究相结合

本书的主要内容分为两大部分,两部分实际上是从宏观和微观两个不同的角度来考察汉语句子的生成过程。

第一部分从宏观的角度考察了汉语句子的生成过程,目的是建立一个科学的句子生成模式。作者构拟的句子生成的宏观过程可以大致描述如下:句子生成的第一个阶段是语义建构阶段。语义建构阶段在语义知识库中语义网络框架知识的指导下,通过语义网络框架的同化与赋值操作形成特定句子的语义网络,这是句子生成的基础。然后在动核结构合并规则、关系化规则、序列化规则等知识的指导下,将语义网络压模形成句模。语义建构操作输出的形式是句模。句子生成的第二个操作阶段是语用建构。语用建构操作主要包括语体注册、表达目的的选择、主题选择、焦点选择、口气选择、语式的确定、时变项的赋值、体变项的赋值、动貌特征的赋值几种具体的操作。语用建构操作的输出形式是一个语用信息集。句子生成的第三个阶段是句法编码阶段。句法编码的任务是将语义建构输出的句模、语用建构输出的语用信息集和句法平面的一些规则进行整合最终将句模转换成一个附加了语用信息的并句法化了的线性语符列。句法编码包括四种操作,即词汇代换和标记词添加、句模向句干的投射、语用校验、句法完形。句子生成的最后一个阶段是语音编码。

第二部分从微观的角度考察了句子生成过程的几个主要环节。在这一部分,作者以语序的生成过程作为样本重点考察了句子生成过程中决定句子生成形式的一些因素和这些因素之间的互动关系。这些微观的考察包括:

(一) 影响句子生成方式的一些语义因素。在这方面,比较详细地考察了动词的事件分类和力发散方式分类以及动词的事件类型和力发散类型对句子生成形式的影响;重点考察了如何对语义网络框架进行压模形成句模以及影响句模中成分配位方式的因素;着重



考察了动核结构合并的方式以及和动核结构合并有关的语义问题。

(二) 影响句子生成方式的一些句法因素。在这方面,比较详细地论述了句模向句法平面直接投射过程中各种语义角色句法身份的指定和句法配位方式,考察了复杂句模向句法平面直接投射和进行句法编码的方式,分析了造成句法移位的原因。

(三) 影响句子生成方式的一些语用因素。分析了在语用校验操作中由于复杂语用加工造成的句模向句法平面的变异投射,即语用移位。

(四) 影响句子生成方式的各种因素之间的关系。说明了汉语句子的生成过程中决定语序的诸因素及这些因素之间的协同作用与相互矛盾,以此来说明汉语句子生成过程中决定句子结构形式的诸因素之间的互动关系。

应该说,这种宏观和微观相结合的研究,将会使我们对汉语句子生成过程有一个全面而且具体的了解。

### 三、吸纳借鉴与开拓创新相结合

正像作者所说的,本书的理论基础是三个平面的语法理论,但在此基础上,也吸纳借鉴了转换生成语法、格语法、配价语法、功能语法以及心理语言学、神经语言学的一些理论成果。而在吸纳借鉴的同时,作者立意创新,勇于开拓。这种开拓创新的精神体现在以下几个方面:

(一) 作者创造性地把三个平面语法理论应用于句子生成研究上,力求科学地处理语义、句法、语用三个方面之间的关系,这就对于三个平面语法理论作出了新的发展。

(二) 对汉语句子生成过程进行了新的阐释。其中,深层语义结构的心理表征、语义网络框架的构成形式、语义网络进行压模形成句模的方式和规则、力发散模式对语义成分句法身份的指定和配位方式的影响、从语义平面到句法平面的变异投射、句法编码中的语用校验和语用移位等问题的论说都颇有新意。

(三) 考察了句子生成时不同平面之间的转换、投射方式,并放

在这样的一个过程中,说明了制约句子生成形式的各种因素之间的相互制约、相互作用的互动关系。

(四) 对句子生成中涉及到的语序规则进行了详细的描写,并探讨了在汉语句生成模式的哪个层面或者说哪个阶段来处理这些不同的规则,同时说明了这些规则之间相互整合的方式。

汉语句生成问题,是一个意义重大而内容复杂、难度极大的研究课题。鑫民博士对这个问题的研究提出了具有创意的理念、思路和方法,作出了积极的贡献。但是,要完全说清楚汉语句生成的机理,并能切合于计算机自然语言处理和汉语教学应用的需要,还有许许多多的问题要探索。因此,希望鑫民博士继续努力对这个问题作出进一步的研究。同时希望以后能够更多结合汉语教学进行研究,探究一下如何把句子生成研究的成果有效地运用于汉语教学特别是汉语作为第二语言的教学。我们期待着他早日贡献出新的成果。

是为序。

陈光磊

2004 年春于复旦大学

# 目 录

引言 .....	(1)
一、句子生成研究的重要性 .....	(1)
二、本书的基本思路 .....	(3)
第一章 句子生成研究回顾 .....	(5)
第一节 句子生成研究中存在的主要争议 .....	(5)
一、和生成模型模拟的目标有关的争议 .....	(5)
二、和模式的构成有关的争议 .....	(6)
第二节 几种不同的句子生成模式 .....	(10)
一、有限状态语法和自左向右模型 .....	(10)
二、转换生成语法 .....	(12)
三、配价语法模式 .....	(14)
四、基于三个平面语法理论的句子生成模式 .....	(16)
第三节 句子生成中的语序问题 .....	(18)
一、句子生成模式中对语序问题的处理方式 .....	(18)
二、关于句子生成中制约语序的因素 .....	(22)
第二章 以语义为基础的串行整合句子生成模式 .....	(28)
第一节 句子生成研究要解决的基本问题 .....	(28)

一、关于句子生成的起始点和终结点 .....	(29)
二、关于从输入物到输出物的操作过程 .....	(35)
第二节 以语义为基础的线性整合模式 .....	(40)
一、模式的基本框架 .....	(40)
二、几个问题的说明 .....	(41)
第三节 串行整合模式句子生成的操作过程 .....	(54)
一、语义建构阶段 .....	(54)
二、语用建构阶段 .....	(84)
三、句法编码阶段 .....	(90)
四、语音编码 .....	(99)
第三章 语义网络框架的压模 .....	(100)
第一节 动词的分类 .....	(101)
一、以价为基础的分类和以事件类型为基础的分类 ..	(101)
二、动词的力发散(force transmission)类型 .....	(118)
第二节 动词与动元、状元的组配关系 .....	(123)
一、不同类别动词的语义特征及其与动元、状元的组配 关系 .....	(123)
二、前景角色和背景角色 .....	(125)
三、动词支配角色的隐现 .....	(126)
第三节 简单语义网络的压模及序列化规则 .....	(126)
一、相邻原则——补充项围绕其核心项配位 .....	(127)
二、语义特征对成分配位的影响 .....	(128)
三、时序原则 .....	(133)
四、小结 .....	(145)
第四节 复杂语义网络的压模 .....	(148)
一、动核结构的合并与事件结构临摹 .....	(148)
二、动核结构的区分 .....	(150)
三、动核结构合并的方式 .....	(158)

四、和动核结构合并有关的语义问题 .....	(166)
五、小结 .....	(169)
<b>第四章 句模向句法平面的投射及句法成分的配位规则</b> ...	(170)
<b>第一节 句模向句法平面的投射和语义成分句法成分的指定</b> .....	(170)
一、力发散类型和动元句法身份的指定 .....	(171)
二、状元句法身份的指定 .....	(178)
三、复杂句模向句法平面的投射 .....	(179)
<b>第二节 句干中决定成分配位方式的句法规则</b> .....	(183)
一、联系词中置原则 .....	(184)
二、句法身份对成分位次的影响 .....	(192)
三、句法复杂度对成分位次的影响 .....	(193)
四、词的句法个性对成分配位的影响 .....	(195)
<b>第三节 句模向句法平面的变异投射</b> .....	(196)
一、由于成分复杂度限制造成的句法移位 .....	(197)
二、由于动核结构合并而造成的句法移位 .....	(198)
三、由于词语的句法个性而造成的句法移位 .....	(202)
四、其它具体的句法规则造成的句法移位 .....	(202)
<b>第五章 句法编码中的语用校验:复杂语用加工和移位</b> .....	(204)
<b>第一节 语用移位的动因</b> .....	(205)
一、什么是语用移位 .....	(205)
二、语用建构和语用移位 .....	(206)
<b>第二节 主题化和移位</b> .....	(208)
一、主题的性质 .....	(208)
二、主题的位置 .....	(216)
三、主题化及主题化的方式 .....	(218)
<b>第三节 焦点化和移位</b> .....	(225)

一、焦点的表达对词序的制约作用 .....	(225)
二、焦点的性质和焦点在序列中的分布 .....	(226)
三、焦点化和语序 .....	(236)
四、分裂句——一种表达焦点的句式 .....	(239)
第四节 移位所受到的限制 .....	(245)
一、语义对移位的限制 .....	(246)
二、界标原则对移位的限制 .....	(253)
三、移位限制和主题化提升移位 .....	(255)
<b>第六章 从语序看句子生成中决定句子结构的诸因素及其相互</b>	
<b>    关系 .....</b>	<b>(258)</b>
第一节 决定语序的因素及相互关系 .....	(258)
一、汉语语序的功能 .....	(259)
二、汉语语序的制约因素 .....	(260)
三、制约汉语语序诸因素的协同作用和相互矛盾 .....	(273)
第二节 从动词后成分的语序看制约汉语语序诸因素及其	
相互关系 .....	(277)
一、动词后可以出现的成分及其组配方式考察 .....	(278)
二、决定动词后成分配位方式的因素及其相互作用 ...	(282)
三、决定动词后成分语序的各种因素的优先序列 .....	(287)
四、制约汉语语序的各种因素之间的互动关系 .....	(289)
<b>结语 .....</b>	<b>(291)</b>
<b>主要参考文献 .....</b>	<b>(294)</b>
<b>后记 .....</b>	<b>(306)</b>

# 引 言

## 一、句子生成研究的重要性

语言生成研究根据研究的目标可以分为句子生成研究和篇章生成研究,语言生成研究的最终目标是说明篇章的生成方式,但是要说明篇章的生成方式,句子生成研究是基础,如果句子生成的问题还没有搞清楚的话,根本就没有办法解决篇章生成的问题。因此,句子生成研究在语言生成研究中占有基础地位。

任何一种自然语言,都是由一个句子无穷集构成的,也就是说,每一种语言都包含了无限多的句子,这样的以句子为基本构成成分的无穷集构成了一种语言。对于这样的自然语言,我们可以有几种不同的描写方法。一是采用列举法,将一个语言所有的句子列举出来,但很明显,由于一种语言所包含的句子是无穷无尽的,用列举法来描写语言系统是不可行的。二是采用一套原则或规则系统来说明这些句子都是遵循一定的“语法”构成的,并通过说明一种语言中的“语法”来描写这种语言。显然这种方式是比较经济和可行的。当然,不同学派对这种“语法”的认识区别颇大。有的认为一种语言中的“语法”就是这种语言中的句子构成的规则系统,因此要描写一种语言的语法就是通过对这种语言中实际的句子进行抽象找出其背后隐藏的规则系统。传统语法和结构主义语法属于这种类型。有的认为一种语言中的“语法”可以看成是生成这种语言中的

句子的一套装置,对一种语言进行研究不仅要描写这套装置,而且还要通过描写这套装置解释句子是如何生成的。在这一派的许多学者看来,这套装置模拟了句子是如何从某种深层形式经过转换形成表层句子的。这种语法我们称之为生成语法。从这个意义上讲,本书的目的就是构划出这样一种生成语法的框架并通过这套语法系统来说明汉语句子的生成。

句子生成研究是语言研究中的一个重大问题,它不但具有重大的理论价值,而且在现代科学技术中还具有重要的实践意义。之所以这样说是因为:语言处理是决定计算机能不能真正实现智能化的一个关键,而语言处理的主要过程就是话语的理解和生成。正像我们刚才指出的,句子生成是话语、篇章生成的基础,因此句子生成研究就成了语言学、心理学、认知科学、思维科学共同关心的一个重大的理论问题,同时,怎样设计一个科学实用的句子生成模式,也是人工智能、计算机自然语言处理亟待解决的一个具有重要实践意义的问题。正像范晓先生指出的:“研究句子的生成机制,是语法研究中的重大课题。这个问题的研究对机器处理自然语言、对实现不同语言间的自动翻译(即“机器翻译”)以及对语言教学都是至关重要的。”(范晓(1996b))正因为句子生成研究具有重要的理论和实践价值,在最近几十年中,它吸引了许多学者投身于这项研究,不管是在语言学还是在心理学及计算机自然语言处理领域,学者们都提出了一些不同的生成模式。然而和自然语言理解的研究比较起来,语言生成的研究相对还很薄弱,特别是由于语言生成实验的难控制性,这方面的试验性研究还很不充分,因此,我们对于自然语言的生成问题还缺乏充分的认识。正如沈政、林庶芝指出的那样,“言语产出的信息加工过程,实际上是怎样从思维转化并生成语言的过程。应该承认,我们对这一过程了解得甚少,除了已知少数外显的心理语言学过程外,还有大量的内隐过程有待于我们去探讨。”(沈政、林庶芝(1995:79))“目前对于言语产出过程的细节了解得实在太少,值得探讨的问题太多。”(沈政、林庶芝(1995:80))最近几十年,计算机的硬件技术得到了很大的发展,但是语言研究却明



显滞后了,正像刘涌泉(1991:7)指出的:“实现机器翻译的技术条件目前已基本具备,存在的主要问题是语言研究跟不上。”

## 二、本书的基本思路

本书基于三个平面的语法理论来研究汉语的句子生成问题。

三个平面的语法理论为我们全面地认识汉语的句子生成问题提供了一个理论框架。我们知道,句子是在多种因素的制约下生成的,这些因素包括语言系统本身的因素,也涉及语言系统以外的一些因素。制约句子生成形式的语言系统内的因素又包括语法系统内的因素和语法系统外的因素。语法系统内的因素主要包括语义、句法、语用几个方面的因素,语法系统外的因素包括修辞等方面的因素。语言系统外的因素包括文化的、认知的等许多方面的因素。这些因素对语言结构的制约作用,都必然在句子生成的句法编码过程中表现出来。那么句子生成的最终结构形式是如何在这些因素的相互制约、相互作用的互动过程中形成的?本书主要就是在三个平面语法理论的框架内,通过语义结构、语用结构、句法结构的相互制约与转换,说明决定句子生成形式的各种因素之间的互动关系,并以此作为我们的理论基础来设计一个句子生成的模式。同时,为更全面地说明制约汉语句子生成形式的因素,我们在本书中有时也可能会涉及语义、语用、句法之外的其它方面的因素。

思想的产生是句子生成的前提,因此语义是句子的生成基础,句子的生成过程就是将语义信息和特定语境中建构起来的语用信息通过句法编码、语音编码转换成一个表层的可以感知的线性序列的过程。句子生成的过程是串行加工和并行加工交替进行的过程,从整体上看,句子生成经过了语义建构、语用建构和句法编码、语音编码几个阶段,呈现出串行加工的特点;但在每个阶段内的操作过程中可能会有多种因素同时作用,呈现出并行加工的特点。因此,句子生成过程从宏观上看是一个串行加工过程,但在这个串行加工过程中的