

本书由上海市马克思主义出版基金资助

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 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》。

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《现代汉语句子的生成问题研究》。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的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的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。这样的一个句子生成模式,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也具有强烈的实践意义。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的基本过程,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,并探讨了在汉语句式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。

汉语句式生成问题,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式生成的机理,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)
第四章 句模向句法平面的投射及句法成分的配位规则 ...	(170)
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	(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)
第二节 句干中决定成分配位方式的句法规则	(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)
第三节 句模向句法平面的变异投射	(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)
第五章 句法编码中的语用校验:复杂语用加工和移位	(204)
第一节 语用移位的动因	(205)
一、什么是语用移位	(205)
二、语用建构和语用移位	(206)
第二节 主题化和移位	(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)
第三节 焦点化和移位	(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)
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关系	(258)
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)
结语	(291)
主要参考文献	(294)
后记	(306)

第四章 句模向句法平面的投射 及句法成分的配位规则

在语义规则的作用下,语义网络中的语义成分被序列化为句模。句模是属于语义平面的,在句子生成的过程中,句模还需要经过句法编码投射到句法平面形成一个句法结构。本章考察在这种投射过程中涉及的一些问题。

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

如果说语义是句子生成的基础,那么我们就必须说明语义结构是如何转换成句法结构的,比如语义结构在转换为句法结构的时候,语义角色是如何获得句法身份的呢?然而,在这方面,语法研究一直面临着许多棘手的问题。首先,语义身份的指定具有很大的不确定性。语义身份的获得和事件的性质及参与成分之间的相互关系有关。同样一个实体在不同的事件中,根据成分之间的关系不同,可能获得不同的语义身份。由于作为句法身份指定基础的语义身份的指定具有不确定性,那么如何说明在不同的事件中怎样指定不同的语义角色,以及不同的语义角色是如何获得句法身份的,是语义为基础的句子生成模式揭示句子生成机制所面临的一个棘手问题。其次,一个语义结构在表层可能体现为多种不同的句法结构,同样一个语义角色在表层句子中可能投射为多种不同的句法成分,这样就使得一个语义角色在表层可能表现为多种不同的句法角色。如:

- [1] a. 我跌断了腿。 b. 我腿跌断了。

- [2] a. 我简直要累死了。 b. 简直要累死我了。
- [3] a. 一个犯人从监狱里跑了。 b. 监狱里跑了一个犯人。
- [4] a. 他用这把菜刀切菜。 b. 这把菜刀他切菜。

在上面这些例子中,加线的部分在语义层面其语义身份是一样的,但是在表层的句子中,所充当的句法身份却不一样。这些例子向我们表明,由于语义角色句法功能的多样性,这种语义角色的区分和语义关系的描写似乎对于预测语义角色之间的句法关系没有太大的帮助,因为同样一种语义角色可以在多种句法位置出现,同一种句法位置也可以出现多种不同的语义角色。这样,我们可能会得出如下结论:语义身份和句法身份之间的关系是非常纷繁复杂的,没有严格的对应关系。这样从语义结构转换成句法结构时,没有了语义身份和句法身份之间的对应关系,从语义角色到句法角色的转换没有严格的规则可循,那么从语义结构到句法结构的转换就不具有可操作性。然而,如果我们分核心句和变换句两个层次来看的话,我们就会发现情况并不是这么糟。语义结构是先转换成核心句(大致相当于本书所说的句干),^①然后在核心句的基础上又转换成变换句的。语义结构转换成核心句时,语义身份和句法身份之间具有相对严格的对应关系,语义身份和句法身份之间具有高度相关性。本书正是循着这样的思路试图找到问题的答案。

一、力发散类型和动元句法身份的指定

以语义为基础来考察句子的生成,我们面临的另一个难题是:在一种语言中,语义角色具有多功能性,即同一种语义角色能够充当的句法成分是多种多样的。由于语义角色在句法上的这种多功能性,就造成了在配位方式上的多样性和复杂性。那么如何解决这个难题呢?看来比较可行的办法还是从事件类型出发,来说明在不同事件

^① 不同的学者对核心句的认识不一样,但有一点是相同的,即核心句是派生出转换句的基本的句式,本书用句干指称这种基本的句式。

中会涉及到的那些类型的语义角色和语义关系,以及这些不同的语义角色在中性的序列(静态句)中典型的配位方式,进而说明这种典型的配位方式在某种句法、语用因素的作用下产生的变异。我们相信这种方式,可以比较有效地说明语义角色和句法角色之间错综复杂的关系。由于事件类型和动词的类型密切相关,所以事件的类型在语言中体现为动词的语义类型。

动元的原型句法身份的指定,和动元之间的力发散方式密切相关。我们在上一章中指出,语义角色在力发散关系中的地位决定成分在句模中的配位方式,但是在那里我们还没有来得及对这个问题进行详细的讨论。实际上,这种力发散关系不仅决定了成分在句模中的配位方式,也决定了句模直接投射为句干时成分在句干中的句法身份。因此我们有必要进一步详细探讨一下力发散模式在从句模到句干投射过程中以及在句法配位中所起的决定作用。

句法身份的获得是以语义为基础的,成分在核心句中将获得什么样的句法身份,至少和以下三种语义因素有关:1. 事件类型,2. 成分的语义特征,3. 成分间的语义关系。其中前两点是形成第三点的基础,因此,在对语义成分进行句法身份指定的时候,起关键作用的是第三点。

(一) 力发散模式

典型的事件中,往往在参与成分之间存在一种力传递的过程,一个典型的事件就是一个主体通过对动作的控制作用于另一个客体的过程,在这个过程中形成了如下的力发散链:

[5] $A \rightarrow V \rightarrow T$ (我榨苹果)

这样一个力发散链可以解释为:

主体(主事)A 通过动作 V 作用于客体(客事)T。

这是一个简单事件内部成分与成分之间的力发散方式。当然这个力发散链可能还会有另外的环节,这样就形成了事件与事件之间的有机联系。在客观现实中,事件往往并不是孤立存在的,此一事件可能是致使彼一事件的原因,彼一事件可能是此一事件致使的结果。