

# 语言模式化：概念、方法和表达

傅兴尚

(黑龙江大学俄语语言文学研究中心, 黑龙江 哈尔滨 150080)

**摘要:** 语言模式化是计算语言学中的核心概念, 它不仅是语言研究中强有力的认知工具, 也是面向信息处理的语言形式化、定性化描写的科学方法和表达手段。本文围绕以下几个问题对语言模式化问题展开研究: 1) 模式·模式化·语言模式化; 2) 语言模式的类型; 3) 语言模式化的方法; 4) 语言模式化表达形式和手段。

**关键词:** 语言; 模式; 模式化; 计算语言学

**中图分类号:** H085

**文献标识码:** A

语言模式化 (языковое моделирование) 是计算语言学 (компьютерная лингвистика, 简称 КЛ) 中的核心概念, 这可反映在以下阐述中。“计算语言学是应用语言学的一个分支, 指利用计算机这一工具——程序、信息组织技术以及信息处理技术对某种条件、场合、领域语言行使的功能进行模式化, 也指语言学和其交叉学科中对语言可计算模式的诸多应用。”(энциклопедия: “кругосвет” 在线查询系统)

不难发现, 计算语言学与其它语言学分支学科的不同主要体现在 1) 研究目的上, КЛ 要实现各种语言功能的模拟, 即语言信息的自动化处理 (如, 机器翻译、信息查询、语言知识的自动获取和表达等), 同时, 也要关注如何利用语言信息处理技术促进理论语言学研究 (在这方面但不限于这方面表现明显的, 如语料库语言学[корпусная лингвистика]); 2) 研究手段上, КЛ 研究手段主要为计算机技术、信息技术、控制论、人工智能等, 体现为自动化、快速化、准确化和实证化; 3) 研究方法上, 可归结为模式化, 包括语言知识的模式化和实现过程的模式化。

计算语言学研究不仅跨越各类语言单位 (词素、词、句子、篇章等), 而且覆盖语言的各个侧面 (语音、语法、语义、语用等), 这和只把某类语言单位作为研究对象的或只把某一语言侧面作为研究对象的语言学分支学科 (前者如词汇学; 后者如语义学) 也是不同的。

一般计算语言学围绕文法和算法两大方面以及二者的契合对接展开研究。文法是语言范畴、语言规则的总称, 体现为静态的语言知识库, 同时包括近几年逐步重视的基于语料库的语言统计模式; 算法是以文法为基础实现特定语言处理目标的指令集, 体现为动态的操作过程。然而, 不论是文法, 还是算法, 都需要模式化。语言模式化主要指文法的模式化。本文从以下几个方面阐释语言模式化问题: 1) 模式·模式化·语言模式化; 2) 语言模式的类型; 3) 语言模式化的方法; 4) 语言模式化表达形式和手段。

## 1 模式·模式化·语言模式化

不论是自然科学还是社会科学，建构模式、运用模式都是强有力的认知工具。这是因为现实的客体或过程太复杂，太多样，具有无法尽览的多侧面性。此外有的现实对象具有宏观性、微观性或无法直接观察的抽象性。所以，最好的研究方法是建构所要研究的对象物（изучаемый объект），即本体（оригинал）的模式。

所谓模式是模拟本体某些属性或功能的人造认知装置（искусственно созданное мысленное устройство），是便于研究和认知的本体近似替代物（аналог-заместитель）。不同的模式模拟本体不同的属性，它可能是研究和描写本体内部结构的，即结构模式（структурные модели）；也可能是描写功能表现的，即功能模式（функциональные модели）；可能是模拟本体稳定的常规属性的，即静态模式（статистические модели）；也可能是模拟本体的动态发展的，即动态模式（динамические модели）。

在俄罗斯语言信息处理专家比奥特罗夫斯基（Пиотровский Р. Г.）看来，模式具有不同于本体的一系列特点：1）简化性——模式作为客观现实的表达手段，是简化的近似物，决不是本体的镜像复制；2）简单性——模式比本体要简单，同时运用模式也比运用其它研究本体的方法方便和简单；3）逻辑一致性——建构模式应该避免矛盾，保持逻辑的同一性。实践中这比详尽性和简单化准则更为重要，而详尽化准则比简单化准则更为重要；4）通用性——模式应该具有通用性，运用它可以描写不同的相关本体，比如，英俄词典模式，还可以表示其它语言偶对之间的关系，如汉俄词汇；5）强解释力——模式不仅能够解释已发现和未发现的本体属性，而且挖掘本体潜在的属性；6）衍生性（эвристические возможности）——模式含有可作为发现新事实源泉的知识，以此为基础能够建构新的理论。

模式作为认知和规范外部世界的手段不应该是虚假的人造结构，而应该是反映本体重要特征和内部本质属性的类似物，换言之，模式以较为简化的形式反映现实对象的某些侧面，而不是所有方面。所谓模式化正是指上述建构模式、运用模式的方法。人类的科学发展史证明，这种方法是非常有效的。

模式化的方法在人类各个活动领域的广泛应用，并解决各种实际问题，使其成为科学研究中的通用手段。这在语言学研究中尤其具有重要的意义。首先，语言作为交际工具和思维外壳是社会的产物，较之自然属性的实体，如，房子、汽车等具有无法直接观察性，正如梅尔丘克（Мельчук И. А.）所言，“由于语言学家不可能研究人在说话和听话的时候在大脑里发生了什么，所以语言转换器就作为‘黑匣子’（черный ящик）模拟这种机制，语言学家只能观察到语言功能的结果，但不能直接观察到语言行使功能的过程——语言本身。因此对于语言学家最感兴趣的正是这种语言，在研究言语的时候他们试图解释和描写言语背后的东西和生成言语的机制，即语言。”（И. А. Мельчук 1999: 15）这说明在梅尔丘克看来，语言是一种转换过程或是一种机制，这种机制就是无法直接观察的称作“黑匣子”的东西。

其次，语言本身就是反映世界的图景，即模式；语言具有任意性、离散性、层级性、非单元性、模糊性、递归性等，在行使其功能的时候更为复杂，具有变体性、个体性等。这决定了语言模式化研究的必然性，同时决定了语言模式化研究的特殊性。

“模式”一词源于拉丁语“modulus”，意思是“度量（мера）”。“在语言学中指语言学家人工建构的（искусственно созданное）具体或抽象的装置（реальное и мысленное устройство），以此复现语言本体的种种表现。”（ЛЭС）

语言学中可分为语言模式和言语模式，生成模式和综合模式。各类模式有不同的本质属性。作为语言本体的类似物可以是：1）现实存在的客体，比如，为了检查文本自动分析和自动综合系统，人作为计算机的类似物在检查算法时模拟的是自动机的工作过程；2）是一种人造的物质实体，比如，借助于语音合成器实现了具体语言的语言发声机制，语音合成器

就是一种模式；3）公理化的数学结构或逻辑数学结构，如公式、图形，模拟的或者是语言系统结构，如形态结构；或者是语言功能，如生成语法。狭义的语言模式指第三种模式。

## 2 语言模式的类型

语言模式化概念最早出现在结构主义语言学家（如哈里斯等）的有关理论中，直至 20 世纪 70 年代出现了数理语言学（математическая лингвистика），而后发展为计算语言学，这一术语才真正被纳入科学的轨道，同时控制论的方法也引入到计算语言学中。根据莫斯科语义学派代表人物 Ю. Д. Апресян 的分类，语言学中应分 3 种模式类型，它们所模拟的本体各不相同：1）人类言语活动模式（модели речевой деятельности человека），模拟具体的言语过程和现象；2）语言研究模式（модели лингвистического исследования），模拟语言学家发现某一种语言现象的研究过程；3）元模式（метамодели），通过这类模拟可实现对言语活动模式或语言研究模式在理论上和实验中进行评测。

2.1 因为言语活动是最复杂的行为，所以言语活动模式具有最为复杂、多样的结构。根据语言持有者（носитель языка）具备哪方面的语言能力并以这方面的能力为模式化对象，把言语活动模式分为语法正确性模式（модели грамматической правильности）和功能模式（функциональные модели）。前者模拟的是区分语言正确与否的能力，例如，根据这种模式建造出语言自动纠错系统，实现了文本中词汇、语法、甚至修辞错误的自动检查；后者模拟把言语内容（内容平面）和言语形式（表达平面）联系起来的能力，例如，基于这种模式的文字（字符）自动识别系统，语音自动识别系统等。

2.1.1 根据语言处理过程中始端（на входе）和终端（на выходе）信息类型的不同，语法正确性模式又可分为识别模式（распознающие модели）和生成模式（порождающие модели）。识别模式的始端是自然语言文本或者抽象的人工语言表达式，终端要回答的是该段字符串语法上是正确的还是异常的，类似这样的模式如：《范畴语法》。

生成模式与识别模式正好相反，在始端是一套规则和一些范畴单位，由此在终端转换为结构化的语句。这种模式如：N. 乔姆斯基的生成语法（генеративная грамматика）。

2.1.2 根据模式化针对的不同言语活动方面，功能模式又分为分析模式（аналитические модели）和综合模式（синтетические модели）。分析模式在信息始端是文本片断，在终端则获得相应的语义记录或语义表示，这种语义记录或语义表示是用语义元语言表达的。综合模式则恰好与分析模式相反，在始端是语义描述，在终端才获得表达这个意思的文本。

典型的分析模式就是现代句法学中关于简单句结构类型的有关理论，它们都建立在语言模式化基础上，这种方法的应用体现在简单句的结构模式被确定为抽象的句法样板，根据这些样板可以建构最小的相对完整的句子，结构模式根据下列特征进行区分和归类，模式的词形构成及结构，包括词形表现，这些词形之间的关系；模式的语义；用该模式建构句子的聚合性能，常规表现系统和扩展规则。比如，前苏联科学院 1980《俄语语法》中的简单句句子模式：

$N_1 - Vf$  (Лес шумит; Отец работает ; Дети радуются);

$Inf Vf_{3s}$  (Курить запрещается);

$Adv_{quant} N_2$  (Много дел; Мало времени);

$N_1$  (Ночь; Тишина);

Vf<sub>3pl</sub> (Звонят);

Inf cop Inf (Руководить значит проверять)

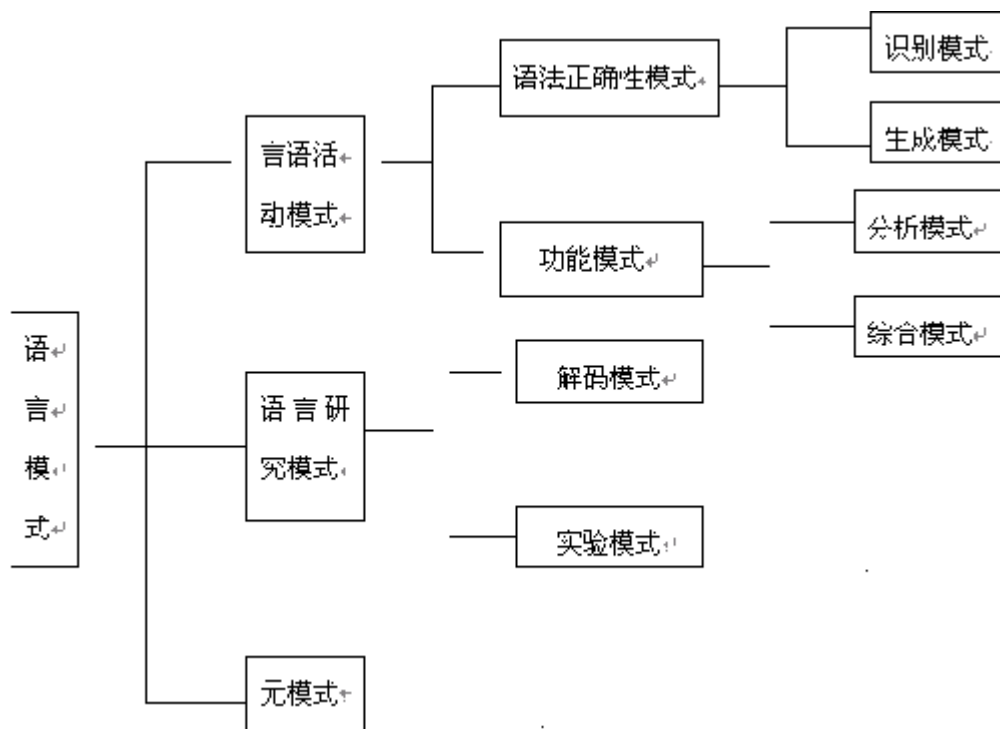
典型的综合模式是功能语法以及梅尔丘克 (И. А. Мельчук) 的《意思↔文本》模式。在该模式中 И. А. Мельчук 从寻找意义与话语(смысл и текст)对应关系的规则出发, 以一套元语言为工具离散地再现了言语生成的过程。他认为, “语言学家应尽力提出一套表达意义的形式规则或一定形式的解释规则, 即建立意义(смысл)与话语(текст)之间的对应关系。” (И. А. Мельчук 1974: 17) 其操作模式表明, 音义之间的对应不能直接建立起来, 而要借助其间多个层次的逐层转换(переход)得以实现。按从“义”到“音”的方向顺序, 要经过下列流程的层次及层次转换: ①语义表象(семантическое представление)→②深层句法表象(глубинно-синтаксическое представление)→③表层句法表象(поверхностно-синтаксическое представление)→④深层形态表象(глубинно-морфологическое представление)→⑤表层形态表象(поверхностно-морфологическое представление)→⑥音位表象(фонемная транскрипция фразы)→⑦语音表象(фонетическая транскрипция)。其中①属语义层, ②+③为句法层, ④+⑤为形态层, ⑥为音位层, ⑦为语音层。这一生成模式与生成语义学派坚持的言语生成模式(从语义到语音的单向性多层过渡)是雷同的。言语生成是一个复杂的过程, 简言之, 就是由无序的语义关系实现为有序的句法语义关系, 进而以语音形式输出, 生成合格句子的过程。

分析模式和综合模式构成机器翻译, 包括一般翻译的必需部分, 也是人工智能系统中必须面对和解决的问题。分析模式为分析算法提供语言知识, 分析算法由最表层逐渐向深层过渡; 综合模型为综合算法提供语言知识, 综合算法的方向是由深层逐渐向表层过渡。载有算法的模式转化为令类性质的模式, 这种模式称作语言处理器, 语言处理器成为人工智能系统的重要功能模块。

2.2 言语活动模式是重要的语言模式类型, 与该类模式不同, 语言研究模式和元模式起到的是辅助作用。语言学家为了描述言语活动模式, 要选择和确定一些概念和范畴。研究模式就是为这一活动提供客观依据, 最大限度地避免主观因素的干扰和影响。根据原始信息的大小, 研究模式分为解码模式(дешифровочные модели)和试验模式(экспериментальные модели)。解码模式的信息源是受限的文本数据库, 所有有关语言的信息通过模式从文本数据中提取出来, 例如, 借助该类模式可以建造语言知识的获取和挖掘系统, 作家语言风格研究系统等; 试验模式的信息源不仅是文本数据库, 而且应该确认这是一个该语言正确文本的集合, 在进行试验时, 语言学家求助于信息载体(语言持有者), 如利用这类模式可以建造语言知识的检测和实证系统。

2.3 元模式是一个标准系统或称为理论证实系统, 借助于这类系统选择好的语言模式。元模式是数理语言学研究的一种数学模式, 其对象物不是独立的语言概念, 而是整个的语言模式。如, 机器翻译的评测系统就是该类模式的应用。

上述语言模式可概括地表示如下:



### 3 语言模式化的方法

语言模式化不仅是一种认知工具，也是语言研究，尤其是计算语言学的主要研究方法。将语言模式化视为过程，则涉及本节的研究内容——语言模式化的方法。

计算机对自然语言的研究和处理，一般经过 3 个必需的过程：1) 把需要研究的问题在语言学中加以模式化，并以高度形式化的数学形式，严密而规整地表达出来；2) 把这种严密而规整的数学形式表示为算法，使之在计算机上模式化和形式化；根据算法编写计算机程序，使之在计算机上加以实现。这和信息学解决实际问题所经过的工艺流程是一致的：

确立目标 (постановка) → 模式(модель) → 程序(программа) → 结果(результат)

在这个流程中，模式起到非常重要的作用，是解决任务的必经阶段。在此，模式是所研究事物的约定性“表现”，反映所研究事物的重要属性。(С. В. Кириченко 2001: 26)

我们认为，语言模式化的方法，在研究内容、应用目的、实用手段等方面必需遵循以下原则：1) 高度的概括性。模式化的方法所表达的内容是语言的语法、语音、语义、词汇等系统的一般的、概括的、抽象的原则与规则，而不是上述系统中个别现象的具体描述。换言之，模式化的方法是从具体语言现象中抽象出一般的规律。2) 表达的高度形式化。形式化表达的手段是形式语言，可以是数学语言、逻辑语言，即用形式语言作元语言，而不是用自然语言作元语言。3) 过程的客观性和严密性。模式化的过程必须具有数学和逻辑的严密性，以保证结果的客观性和较强的解释力。

基于上述原则，模式化是一个复杂的过程，其本质是抽象概括，舍弃具体的非本质的特征，保留本质的共性的特征，模式化的结果是构拟出层次分明的“类”体系。模式化的基础是对所研究的对象（本体）的结构化 (структурирование)。

结构是实体间稳定关系的总合，实体在发生内外部变化的时候，能保证自身的完整性

(целостность) 和同一性 (тождественность), 即保持基本的性能。根据被研究的是实体、过程、还是现象, 结构化的方式是不同的。

首先, 把研究的客体或过程切分为彼此联系的成素, 每一个成素应具有独立性 (уникальность), 有相应的命名。成素命名是信息模式化的基本过程, 也是其基础。成素本身也是最简单的信息模式。

第二步就是选择特征和属性参数, 通过它们定义成素, 使模式简单化。特征和参数应该是概括的、通用的。各个成素的不同体现在属性值上。

第三步是建立成素之间相互关系的系统, 这种关系可能是时间关系、空间关系、因果关系、部分—整体关系、形式—内容关系、外部—内部关系等。

最后一步是反复修改、完善和调试。可通过人工和机助两种方式进行。

语言模式化属于信息模式化 (информационное моделирование, 与 предметное моделирование 相对) 的范畴, 信息模式有两种类型: 1) 分类模式 (классификационный модель), 该类模式描写对象是由什么组成的, 各个组成部分如何关联。其功能是为了识别, 即如何区分 A 和 B。主要是各种信息库、语言知识库和专家系统等。2) 动态(操作)模式 (динамический модель), 模拟的是过程, 其核心问题是如何由当前状态转向下个状态, 哪些状态是可能实现的。该类模式的典型应用就是语言自动机 (языковой автомат)。

#### 4 语言模式化的表达形式和手段

把语言模式化看成一个结果, 则体现为语言知识的表达形式和表达手段。有关语言知识的元语言表达手段, 许多学者进行过系统的研究, 提出了一些理论和方法。较有影响的有谓词逻辑表示法、产生式系统表示法、框架表示法、网络表示法、矩阵表示法、树形图表示法公式表示法等。

谓词逻辑是一种形式语言系统, 研究的是问题假设与结论之间的蕴含关系, 即用逻辑方法研究推理的规律。它可以看作是自然语言的一种简化形式, 由于它精确、无异义性, 故十分容易转换为计算机语言, 为计算机所理解。从另一方面看, 这种语言与自然语言也很相似, 也可以用来表示句子所具备的命题意义, 当然谓词逻辑作为形式语言远远不能表示自然语言所能表达的全部思想和知识。从思维推理的角度看, 谓词逻辑所体现的逻辑推理规律也不能模拟人类思维的全部。谓词逻辑这种元语言表达手段在《蒙太古语法》中是主要描写工具。

产生式系统 (production system) 是美国人工智能专家塞蒙 (H. A. Simon) 和列维尔 (A. Newell) 提出来的。他们认为产生式系统与智能活动有关, 而智能活动可分解为一系列最基本的单位, 这种单位有两种: 1) 根据某种环境采取某种行动; 2) 根据某些前提做出某种结论。所谓人们有智能就意味着能够根据某种特定的环境产生某种行动或者根据某一特定的前提产生某种结论, 而这种基本活动的智能单位就叫做产生式系统。产生式系统实质上是规则系统, 它采用规则对符号进行替换运算, 可表示为  $\alpha \rightarrow \beta$ 。不难看出, 在产生式系统中, 知识分成两部分: 产生式左边为事实, 它表示静态的知识; 产生式右边则是该事实的动作和行为的过成, 是动态的知识, 隐含了对知识的理解和推理。由于这类系统的知识主要用规则表征, 所以又称基于规则的系统。建立一个产生式系统, 其核心问题是构造一组知识规则及实现规则之间的匹配, 进而实现知识的演绎推理。这种表达方式和表达手段在乔姆斯基转换生成语法的短语规则中得以体现。如以下产生式规则:

S	→O <sub>в</sub> S
S	→O <sub>м</sub> S
S	→VN <sub>им</sub>
O <sub>в</sub>	→P <sub>вин</sub> N <sub>вин</sub>
O <sub>м</sub>	→P <sub>тв</sub> C <sub>тв</sub>
N <sub>вин</sub>	→П <sub>вин</sub> C <sub>вин</sub>
C <sub>им</sub>	→шум, спор...
C <sub>вин</sub>	→время, дни...
C <sub>тв</sub>	→окном, кулисами...
V	→раздался, шел...
Н	→очень, неоправданно...
N <sub>им</sub>	→A <sub>им</sub> C <sub>им</sub>
A <sub>им</sub>	→НП <sub>им</sub>
P <sub>вин</sub>	→в...
P <sub>тв</sub>	→за...
П <sub>им</sub>	→громкий, бурный...
П <sub>вин</sub>	→это...

(其中 S ——句子, O<sub>в</sub>——时间状语, O<sub>м</sub> ——地点状语, N<sub>вин</sub> ——一格名词词组, N<sub>им</sub> ——四格名词词组, A<sub>им</sub> ——一格形容词词组, P<sub>вин</sub> ——支配第四格的前置词, P<sub>тв</sub> ——支配第五格的前置词, П<sub>им</sub>——一格形容词, П<sub>вин</sub> ——四格形容词, C<sub>им</sub> ——一格名词, C<sub>вин</sub> ——四格名词, C<sub>тв</sub>——五格名词, V ——动词人称形式, Н ——副词)

以上这套规则由 S 出发依照操作顺序可以生成一串语符链, 最后可生成俄语句子 **В это время за окном раздался очень громкий шум.**

框架表示法是美国人工智能专家明斯基 (M. L. Minsky) 提出的一种知识表示方法, 亦称为框架理论 (frame theory)。这是一种基于认知过程的模型。人们对事物的认识往往是一种类似框架的结构存储在大脑中。世界上的万事万物以及它们的状态、属性、变化和相互关系往往有一种内在的联系。当面临一个新事物时, 这种联系就要和头脑中某一个框架进行匹配, 如果匹配成功, 就得到了对该事物的认识, 如果匹配不成功, 再新取一个与新事物相近的框架, 或者修改、补充刚才匹配的不太成功的框架形成新的认识, 即以新的框架存储在头脑中。可见, 框架提供了一种可能的知识结构, 人们面临新情境时能从过去的知识经验中进行预测, 引起有关事物的注意、回忆和推理。语义框架实际上表示的是事物或概念相关的数据结构, 表达了人对客观世界有关事物和事件之间的关系。由框架名和一组槽 (slot) 构成。框架名位于最顶层, 用于指称某个概念、对象或事件; 其下层的槽由槽名和槽值两部分组成。槽值可以是逻辑的、数字的, 也可以是一个子框架。如果我们把一个词相关的语言知识看作槽, 把词看作框架名, 则一个词可视为一个框架。框架表示法在菲尔默的格语法中得

以体现。如，格语法中，每一条动词词条下用缩写字母标出该动词可以插入的格的框架，这种标记称为框架特征（frame feature）。框架特征置于方括号内。例如，动词 плакать（哭）、бегать（跑）可以带有一个表示施事的 NP，如，Он плачет（他哭），Я бегаю（我跑）。因而它们可以插入（——A）的框架。这样，在词库中就标明它们的框架特征是（——A）。记作：（方框前面的符号“+”表示该动词具有方框内的格特征。方框内的横线表示可以插入该动词的位置。）

a. плакать: + {——A}

b. бегать: + {——A}

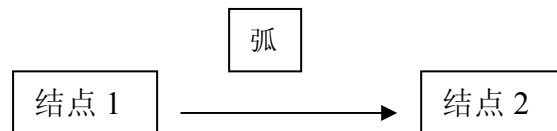
当一个动词可以出现几种不同的格框架中时，词库中就必须在其中区分出必选格特征和可选格特征。例如动词 открыть（打开）可以出现在以下 3 种格的环境中：

a. {——O+A}，如，Он открыл дверь。（他打开了门。）

b. {——O+I}，如，Открыло дверь ветром。（风把门吹开了。）

c. {——O+A+I}，如，Он открыл дверь ключом。（他用钥匙开了门。）

语义网络表示法最早由奎廉（R. Quillian）作为记忆联想网络于 1968 年提出。经过西蒙斯（R. Simmons）和亨德里克斯（G. Hendrix）等学者的改进，比较成功地用于人工智能中的知识表示和计算机对语言的理解。这里所谓的语义是指符号表达式同它所描述的对象（含义）之间的关系。人类知识一般可分两种基本形态，一是事实性的知识，一是这些事实之间的联系，即能够从一些事实找到另一些事实的信息。语义网络很好地表达了事实和事实之间的联系，其基本特点就是由节点和连接点的弧组成，其中节点表示物体、概念和态势，而弧则表示它们之间的联系。如下图所示。



图中结点表示概念，弧是有方向的、有标记的。在三元组中，弧由结点 1 指向结点 2，结点 1 为主，结点 2 为辅，弧的方向体现了主次，弧上的标记表示结点 1 的属性或结点 1 与结点 2 之间的关系。

上图中结点与结点之间的关系是多种多样的，主要有隶属关系（即弧=ISA，如马和动物的关系）；整体与部分关系（即弧=PART-OF，如手和身体的关系）；同一关系（即弧=IS，如莫斯科和俄罗斯首都的关系）等。

语义网络的推理则是以网络结构的匹配为基础。先根据待求的问题，构造一个网络片断，然后搜索语义网络知识库。在匹配过程中，只有当片断中的节点能使匹配完全成功时，它们的值才能被约束。语义推理也在匹配的过程中进行，当需要在一对概念之间进行推理时，则要寻找代表这两个概念的节点之间的联系，即从这两个节点开始，然后利用启发式知识，激活与之相联的部分节点，之后又依次激活同它们相联结的全体节点，从而在每一个初始的概念周围形成一个扩充的激活区，推理沿网络传播。当某一个概念同时从两个方向上被激活时，也就找到了一种期望联系，这种联想式知识推理过程使语义网络可以具有一定的创造性。

此外还有矩阵表示法、树形图表示法、公式表示法等大家比较熟悉，此不赘言。

以上概述的几种知识表示法各有特色，有的适合表示这类语言模式，有的适于表征另一



种语言模式。但是从设计目的、操作原理、与算法连接的可行性等各个方面来看，它们都为我们计算语言学中语言模式化研究提供了实践基础，起到启迪、参考作用，但都不能照搬照用。语言是非常复杂的，就模式化的不同侧面而言，至少包括语音模式化、词汇模式化、形态模式化、句法模式化、语义模式化、篇章（连贯和照应等属性）的模式化。

模式化的结果是建造认知的中介对象——模式。在认知语言事实的过程中模式具有一系列的功能：替代本体、信息功能、认知功能、形式化功能、算法功能、实证列举功能。所以模式化不论作为方法还是手段，不论作为过程还是结果都具有重要的意义，只有在此基础上才能实现语言学的对象物——文本的分析，揭示和模拟语言的各种功能，才能使运用数学机制和计算机技术研究语言得以实现。

### 参考文献

- [1]Герд А.С. 1998 Структурная и прикладная лингвистика [М]. С.-Петербургский университет.
- [2]Герд А.С. 1996 Прикладное языкознание [М]. С.-Петербургский университет.
- [3]Григорьев Н. В. 1999 Восходящий алгоритм построения дерева зависимостей для системы ЭТАПЗ // Труды Международного семинара Диалог' 1999 по компьютерной лингвистике и ее предложениям, том 2.
- [4]Золотова Г. А. 1988 Синтаксический словарь [М]. М.
- [5]Кириченко С. В.2001 О технологии информационного моделирования // информационные модели в лингвистике. Омск.
- [6]Максименко О. И. 2002 Формальные методы в современной прикладной лингвистике [М]. М.
- [7]Марчук Ю.Н. 1983 Проблемы машинного перевода [М]. М.
- [8]Марчук Ю.Н. 2000 Основы компьютерной лингвистики [М]. М.
- [9]Мельчук И. А. 1999 Опыт теории лингвистических моделей «смысл ↔ текст» [М]. М.
- [10]Пиотровский Р. Г. 1999 Лингвистический автомат и его речемыслительное обоснование [М]. Минск.
- [11]冯志伟 1996 《自然语言的计算机处理》 [M]，上海：上海外语教育出版社。
- [12]傅兴尚 1999 《现代俄语事格语法》 [M]，北京：军事谊文出版社。
- [13]傅承德 2000 《自然语言理解的方法与策略》 [M]，郑州：河南人民出版社。
- [14]侯敏 1999 《计算语言学与汉语自动分析》 [M]，北京：北京广播学院出版社。
- [15]石纯一、黄昌宁等 1993 《人工智能原理》 [M]，北京：清华大学出版社。
- [16]赵铁军等 2000 《机器翻译原理》 [M]，哈尔滨：哈尔滨工业大学出版社。

## Language Modelization: Concept, Method and Representation

FU Xing-shang

(Center of Russian Language and Literature Studies of Heilongjiang University, Harbin 150080, China)

**Abstract:** Language modelization is the kernel concept in computational linguistics. It is not only the powerful cognition tool in lingual research, but also the scientific method and representing means for information-processing oriented language formal and qualitative descriptions. This paper mainly discusses the following aspects: 1) model, modelization, and language modelization; 2)types of language models; 3)methods for language modelization; 4)relative representation and means.

**Key Words:** language; model; modelization; computational linguistics

**收稿日期:** 2003-07-05

**作者简介:** 傅兴尚(1966-)男,河北青龙人。黑龙江大学博士后,教授,硕士生导师。研究方向为计算语言学、语义学、词典学等。主持国家级课题3项(现代俄语事格语法;面向信息处理的俄语语言知识库;俄罗斯计算语言学与机器翻译);省级课题2项。出版专著3部,发表论文近30篇。

**[责任编辑: 靳铭吉]**