

文本生成与理解的语言学模拟*

——伊戈尔·梅里丘克«意思⇔文本»模型评介

易绵竹¹, 南振兴², 李绍哲³, 薛恩奎⁴

1,4. 黑龙江大学俄语语言文学研究中心, 黑龙江 哈尔滨 150080;

1,2,3. 解放军外国语学院国防语言文化研究所, 河南 洛阳 471003

E-mail: yimianzhu@sohu.com

摘要: 20世纪60年代中期, 由前苏联学者伊戈尔·梅里丘克等创立的«意思⇔文本»模型是一种研究意思及其表达形式(文本)之间双向对应关系的语言学理论, 其目标是模拟人的语言能力——文本的生成与理解机制, 其基本思想成为随后兴起的莫斯科语义学派的方法论纲领。近年来, 该模型受到国际学术界的广泛关注, 并成功地应用于词典编纂和机器翻译设计。在本文我们将着重介绍该模型的理论框架及其主要构件, 这对于我国计算语言学的基础研究必有助益。

关键词: «意思⇔文本»模型, 莫斯科语义学派, 文本生成, 文本理解, 计算语言学

Linguistic Modeling for Text Generation and Understanding —— A Review of Meaning ⇔ Text Model by Igor A. Mel'čuk

Yi Mianzhu¹, Nan Zhenxing², Li Shaozhe³, Xue Enkui⁴

1,4. Center for Russian Language and Literature Studies of Heilongjiang University, Harbin, Heilongjiang Prov., 150080, China; 1,2,3. Institute of Language and Culture Studies for National Defence, PLA University of Foreign Languages, Luoyang, Henan Prov., 471003, China

E-mail: yimianzhu@sohu.com

Abstract: The Meaning ⇔ Text model (MTM), created by the former USSR scholar Igor A. Mel'čuk et al. in the mid-1960s of XX century, is a linguistic theory that researches the bidirectional correspondences between sense or meaning and its expression form or text. The goal of MTM is the modeling of human language abilities — text generation and understanding mechanisms. Its basic ideas have become the methodological program of the Moscow Semantic School, which emerged afterwards. In recent years MTM attracted widespread attention from people in the international academic circles and has been successfully applied to dictionary compilation and machine translation design. The present article serves as a brief commentary on the theoretical framework of MTM and its main components. The work we introduce here most probably is helpful to fundamental studies on computational linguistics in China.

Keywords: Meaning ⇔ Text model (MTM), Moscow Semantic School, text generation, text understanding, computational linguistics.

* 本文承教育部人文社会科学重点研究基地重大项目《俄罗斯计算语言学与机器翻译》和本部科研项目《面向俄文信息处理的通用语义码模式》的资助。

1 MTM 背景介绍

根据伊戈尔·梅里丘克的自述,在 20 世纪 60 年代初由于阅读了一群法国数学家的学术论文,深受其学术思想的影响,认为他们为整个现代数学建立了统一的概念系统和形式语言。鉴于当时前苏联现代语义学尚不成型,句法学还处于萌芽状态,只有词法学的研究相当成熟,已具备整齐化和形式化的条件,因此梅氏撰著了一本《普通词法学纲要》,但是由于权威语言学家指责该书缺乏理论引导而未能出版。可以说,MTM 模型正是脱胎于此书。

阿·卓尔科夫斯基(A.K.Žolkowskij)是 MTM 的主要合作者,其学术兴趣是语义学。他们曾联名发表了四篇论述语义综合的论文^{[10], [11], [12], [13]},形成了语句的语义表达式、详解组配词典、词汇函项等理念。由此,MTM 的基本理论框架得以确立^[5]。作者决定以两卷本来阐述自己的学术思想,其中第一卷本构思就是要建立能模拟说话人言语活动的语言学模型,极力申论一种简明而自然的语言观。后来由于政治原因,MTM 主创者于 1977 年移居加拿大,在蒙特利尔大学翻译 & 语言学系组建“意思—文本语言学观察站”(Observatoire de Linguistique Sens—Texte),潜心于 MTM 的理论研究和实际应用,著述颇丰,他本人被西方语言学界誉为“了不起的局外人”(great outsider)。经过近 40 年的发展,MTM 在国际计算语言学界和理论语言学界的影响不断扩大,在 COLING、ACL、ANLP、DIALOG 等大型国际学术会议论文集中常能见到讨论 MTM 的文章。2003 年 6 月,在法国巴黎将召开首届意思—文本理论国际会议(MTT 2003: First International Conference on Meaning—Text Theory)。

以尤·阿普列祥(Ju.Apresjan)为代表的莫斯科语义学派秉承 MTM 的核心思想,逐渐发展为语言的整合描写与系统词典学理论^[1]。阿氏率领俄罗斯科学院信息传输问题研究所计算语言学实验室同仁,将基础研究确定为研制一种基于 MTM 的全效形式语言模型,并主持研发了名为 ETAP 的系列机器翻译实验系统(<http://proling.iitp.ru/>)。我国俄语学者张家骅先生对莫斯科语义学派的理论背景和学术成就进行了全面系统的评述,请参看^[14]。可见,MTM 是集集体学术智慧的结晶,梅氏、卓氏和阿氏则是构建该模型的“三驾马车”。

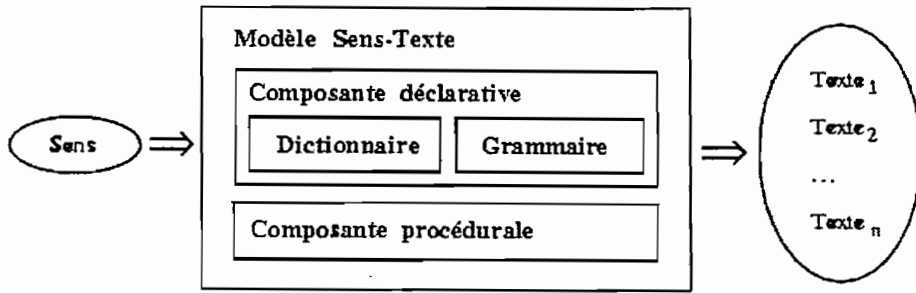
2 MTM 理论述要

梅氏援引美国语言学家列·布龙菲尔德(L.Bloomfield)的一段话作为其专著^[5]的卷首语:“在人类言语中,不同的声音具有不同的意义。研究一定的声音与一定的意义之协调关系(coordination)就是研究语言。”显然,作者从这段话中获得了构思 MTM 的灵感。

2.1 理论假设

梅氏提出一个鲜明的论题(理论假设):自然语言是一个特种的转换器,它将给定的意思加工转换成相应的文本,并将给定的文本加工转换成相应的意思。换言之,语言被看作是意思和文本之间确定的对应关系再加上某种机制,以便于用具体程序实现这种对应关系,即完成意思与文本之间的双向转换。作者对“文本”和“意思”作了进一步解释:前者相当于索绪尔的 *signifiant*,指言语的离散表达式,它是一种特殊的构造体,既有书面形式也有语音形式;后者相当于弗雷格的 *Sinn*、莫里斯的 *Significatum*、卡尔纳普的 *Intensional* 和索绪尔的 *Signifié*,它是语句所有同义转换的常体。记录意思需要采用一套专门的语义语言,这就是语义表达式。我们改用更通俗的语言学术语来说,MTM 反映了语言符号的形式和内容两个侧面,实际上涉及文本的生成与理解两种过程。为了更直观地显示意思与文本之间的对应关系,

我们从阿·珀尔谷埃尔(A. Polguère)绘制的图形中就可见一斑，它清晰地勾画出意思与表达这些意思的各种文本集之间的转换关系^[8]：



(意思↔文本对应关系图)

2.2 理论渊源

首先，梅氏从罗·雅各布森 (R. Jakobson) 的言语通讯模型^[4]中找到理论依据，指出构成该模型的六个要件中有三个可用于解释 MTM：属于传达和感知范围的信息就是意思；负荷这些信息的物理信号就是文本；表现信息与信号之间对应关系的代码则反映意思与文本之间的对应关系。作者引用心理学家米勒 (G. Miller) 等人的观点：“尝试在计算机上模拟这样或那样的过程，在很大程度上是由于期望检验或向他人展示模型建立者对其所坚持的理论有多深的理解……建造模型就是对理解清晰度的明证。”^[8]由此强调：现代语言学的中心任务之一乃是建造语言的功能模型，即类似«意思↔文本»的运作模型。作者借用乔姆斯基 (N.Chomsky) 的术语指出，MTM 模拟的是说话人的语言能力 (linguistic competence)，并非语言行为 (linguistic performance)。作者详细阐述了构建功能运作模型 MTM 的理论基础，包括乔姆斯基的转换-生成语法、兰姆 (S. Lamb) 的层次语法、单克 (R. Schank) 的概念依存理论、西蒙 (R. F. Simmons) 的语义网络理论，以及菲尔默 (Ch. Fillmore)、莱昂斯 (J. Lyons)、雷柯夫 (G. Lakoff)、维日比茨卡 (A. Wierzbicka) 等人的结构语义学思想。作者指出，MTM 是乔姆斯基语言学说的自然发展，二者既有交叉点，又有本质的差异。总之，MTM 理论观念的形成与前苏联早期的机器翻译实验密不可分，作者曾言机器翻译应该是“没有翻译、没有机器、没有算法的翻译”；“没有任何事物比一个叫得响的理论更实用。”(刘海涛先生提示)

2.3 理论特色

梅氏对 MTM 的建造提出三项要求，这可看作是对该模型理论特色的概述：

1) MTM 应当成为一个运作模型，它是一种具有输入和输出并能将输入物加工转换成输出物的装置。这就是该模型的动态性特色，它与传统的分类描写模型有着本质的差别。为了描述意思与文本的对应关系，作者认为有三大要素不可或缺：a) 意思基本单位 (义素) 的清单以及由义素组合成更复杂意思单位的规则 (语义表达式)；b) 文本基本单位 (形素) 的清单以及由形素组合成更复杂文本单位的规则；c) 任何意思单位与文本单位的双向对应规则。这种对应规则可形式化地表示为：‘A’⇒[A] (‘A’⇒A) 或 [A]⇒‘A’ (A⇒‘A’) 其中‘A’是某种意思，[A]是负荷这种意思的语音文本，A 是书面文本，箭头⇒表示“替换为”。作者指出，语言学工作者的任务是研究意思与文本的相互转换器，MTM 的语言学部分实际上是一个确定意思与文本之间对应关系的规则系统，这些规则尽管不涉及实现转换的具体程序，但仍然是动态

的。这样，作者把 MTM 划分为两个部分：(1) 在语言学部分收集该语言的数据资料(确定意思与文本的双向对应关系)，它只是一种运算，即一套允许或禁止命令(可能或不能做什么)，而不是一套算法或指令(应当做什么)；(2) 在算法部分对这些语言数据资料的使用机制或程序进行描述，它具有实际操作的算法性质。算法部分和语言学部分是彼此独立的。

2) MTM 应当成为能够完全形式化的模型，它只需借助单值而合乎逻辑的准确表述，不要求添加辅助信息。其控制准则是，该模型及其任何部件在原则上的计算机可实现性。作者认为，数理语言学所关注的主要是形式方面，即模型是如何构造的——模型的基本部件及其相互联系方式；而纯语言学则主要关注内容方面，即模型反映的是什么。当然，模型无论如何应当形式化，因为‘形式的’等于‘科学的’(«формальное» = «научное»)^[3]。

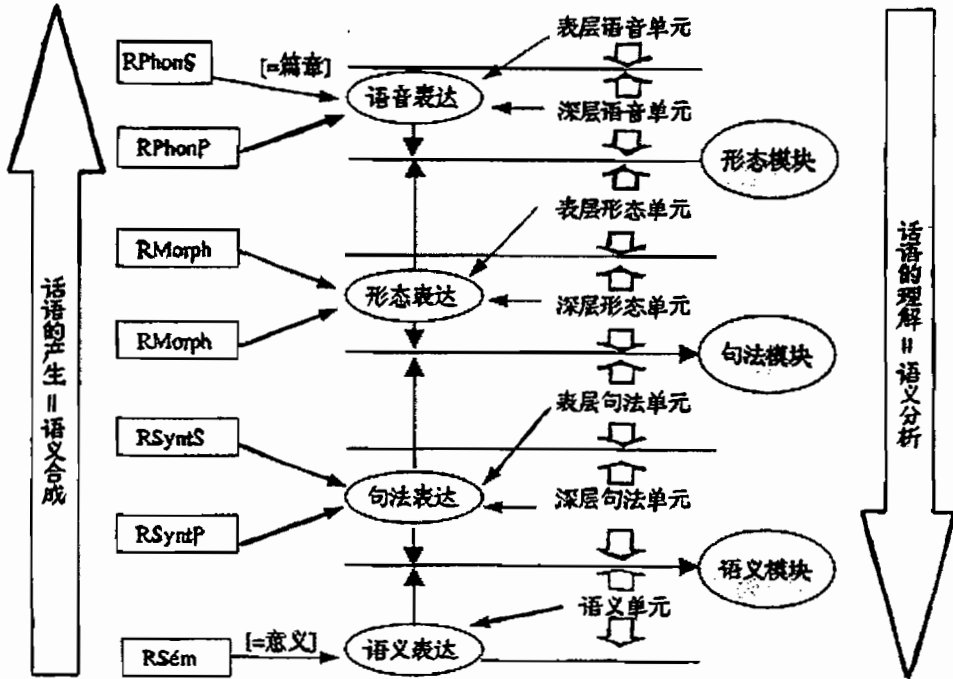
3) MTM 不是一种生成装置，而是一种转换装置，它仅仅是意思与文本的双向转换器。作者只要求该模型能够把给定的形式化的意思加工转换成相应的文本，并从给定的文本中提取意思。作者强调，MTM 作为语言模型而非思维模型，它应当能够对文本的(语法)不合格性作出反应，但是未必要求它对给定的意思或从文本中提取的意思的异常性(abnormality)作出反应。以俄语词组 *круглый квадрат* (圆形的正方体) 为例，其异常性并非出现在词汇组配层次上，而是出现在意思组配层次上。再以两个俄语句子为例：*Иван улетел в Рим.* (伊万飞往了罗马) *Иван улетел на Венеру.* (伊万飞往了金星) 如果一百年前说出这两个句子，语义上都异常，现在两句语义都正常，但前一句可能为真，后一句显然为假。在作者看来，要鉴别语句的语义正常性/异常性和真/假值，有赖于说话人对句子所描述的情景的了解、事实知识和逻辑推理能力，而这些因素均与具体语言无关。若用语言描述某种情景，则需要建立«现实 ⇔ 意思»模型。因此，设计文本信息处理或逻辑处理用的自动系统(如信息检索、自动问答等)，往往需要两套完全不同的机制：其一是语言模型，它将文本转化为某种专门的写入，便于意思内容的单值记录，并将这种写入转化为文本；其二是逻辑思维模型，它对写入本身实行必要的逻辑处理。

墨西哥籍俄裔学者博尔沙科夫与格尔布赫(I. A. Bolshakov & A. F. Gelbukh) 将 MTM 置于现代西方计算语言学的学术背景下加以审视，指出 MTM 与其他语言模型的相同之处在于：模型的功能性、语言的文本/语音形式与其语义表达式的对立性、语言的概括性、模型的动态性、模型的形式性、模型的非生成性、模型与转换方向的独立性、算法与数据的独立性以及对详解词典的高度重视；而其个性特征则表现在：面向语言综合过程、模型的多层次性、结构与形式化方法的多样性、区分深层与表层句法表达式、句子中词汇组成及词序排列的独立性、考虑文本的交际结构、面向不同于英语的其他语种、提供同义变异和词汇函项手段、标注词汇之间的句法关系、建立词汇支配模式及保留传统和经典语言学学术语^[2]。我国词典学研究者章宜华先生尤其看重 MTM 的另外三种特色：释义性、完整性和全面性、普遍性^[15]。在法国巴黎即将召开的首届意思 ⇔ 文本理论国际会议的网页上(<http://mtt2003.linguist.jussieu.fr>)，MTM 的理论特色被概括为：(1) the modeling of language as a correspondence between meaning and texts/sounds; (2) the central position of the lexicon; (3) the privileged position of semantics over syntax; (4) the use of a dependency representation in syntax.

3 MTM 框架结构及基本构件

3.1 框架结构

章宜华先生根据梅氏有关图式改编的 MTM 综合结构图^[15]直观明晰, 我们不妨摘录于此。

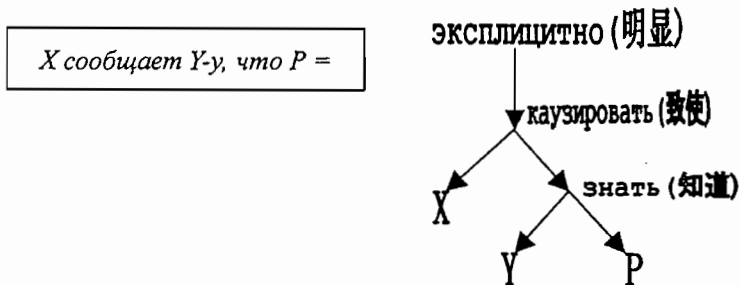


(MTM 综合结构图)

梅氏在初版 MTM 中设置了语句表达式的五个层次, 后又增设两个层次^[6]。简言之, 语义层次是对语句意思的写入, 形成语义表达式, 它由语义图和意思的交际组织两个部分组成; 句法层次的基本单元分别是深层句法结构和表层句法结构, 句法表达式为依存树形图; 形态层次包括深层与表层单元, 形态表达式对语句的所有词形式进行标注; 音位层次对语句进行音位标注; 语音层次对语句进行语音标注, 语音表达式就是由后两个层次构成。

3.2 基本构件

从 MTM 综合结构图中可见, 其基本构件就是语义、句法和形态三大模块。鉴于 MTM 的理论特色之一是语义优先于句法 (the privileged position of semantics over syntax), 在此我们只察看其语义构件。梅氏指出, 语义构件的作用是确定语义表达式与句子深层句法表达式之间的对应关系。句子语义表达式表示为语义图, 其顶端是义素符号, 其弧则标明谓词—义素与论元之间的联系。如俄语句子 *X сообщает Y-y, что P.* (*X 通知 Y-与格, 什么 P*) 可图示为:



作者认为, 语义学研究的首要任务是建立通用语义语言 (universal semantic language), 并制订了三条用于研制语义表达式的原则: 1) 意思原子化 (义素); 2) 所指 (意思原子复杂组合) 高度结构化; 3) 区分意思原子的不同逻辑类型 (谓词 vs. 体词)。在语义构件部分, 作者还详细论述了词汇函项 (Lexical Functions)、详解组配词典 (Explanatory Combinatorial Dictionary)、词汇的配价及支配模式、深层句法结构与句子同义转换系统等方面十分庞杂的内容, 并讨论了语义分析和语义综合问题, 演示了语义构件在 MTM 中的运行机制和程序。顺便指出, MTM 的句法构件同样重要, 它合理地借鉴了依存语法的思想^[7]。由于篇幅受限, 本文不再赘述, 我们将在另文探讨。

4 结语

梅氏创立的 MTM 被认为是前苏联第一个具有控制论特色的语言学理论, 其实用目的是建立双向的语言处理器。近几十年来, 在俄罗斯、加拿大、西班牙、法国、德国、日本和韩国等国家, 一批志同道合的研究者纷纷开展了基于 MTM 模型设计机器翻译系统或尝试以本国语为目标语编纂双语词典等方面的工作, 取得了许多应用性成果。本文对 MTM 模型产生背景、理论框架及其基本构件的介绍, 应该说还很不全面。我们下一步工作是: 从本体论、认识论和方法论角度挖掘该理论模型的深邃思想, 汲取其合理内核, 密切跟踪其发展动态, 为建立面向俄文信息处理的通用语义码模式和研制特殊用途的俄汉机器翻译系统奠定基础。

致谢: 北京广播学院应用语言学系教授刘海涛先生对本文提供了颇有参考价值的意见和资料, 并对 MTM 的理论特色发表了自己的独到见解, 特此深表谢忱。

参 考 文 献

- [1] Apresjan Ju. D. Integral description of language and systemic lexicography [in Russian]. Moscow: Shkola «Jazyki russoj kul'tury», 1995.
- [2] Bolshakov I. A. & Gelbukh A. F. The Meaning Text \leftrightarrow Model: Thirty years after [also published in Russian]. J. International Forum on Information and Documentation, FID 519, 2000, (1).
- [3] Gladkij A. V. & Mel'čuk I. A. The elements of mathematic linguistics [in Russian]. Moscow, 1969.
- [4] Jakobson R. Linguistics and poetics. Style in language. Cambridge (Mass.), 1960.
- [5] Mel'čuk I. A. An experiment of theory of linguistic «Meaning \leftrightarrow Text» models [in Russian]. Moscow: Nauka, 1974; Moscow: Shkola «Jazyki russoj kul'tury», 1999.
- [6] Mel'čuk I. A. & Polguère A. A formal lexicon in the meaning — text theory (or how to do lexica with words). Computational linguistics, 1987, Vol.13, (3—4).
- [7] Mel'čuk I. A. Dependency syntax: theory and practice. Albany, New York: The SUNY Press, 1988.
- [8] Miller G. & Gallanter E. & Pribram K. Plans and the structure of behaviour. New York, 1960.
- [9] Polguère A. La Théorie Sens—Texte. Dialangue, 1998, (8—9).
- [10] Žolkowskij A. K. & Mel'čuk I. A. On potential method and tools of semantic synthesis [in Russian]. Naučno—tečničskaja informacija, 1965, (6).
- [11] Žolkowskij A. K. & Mel'čuk I. A. On system of semantic synthesis. I. Construction of dictionary [in Russian]. Naučno—tečničskaja informacija, 1966, (11).
- [12] Žolkowskij A. K. & Mel'čuk I. A. On system of semantic synthesis. II. Rules of Transformation [in Russian]. Naučno—Techničskaja Informacija, 1967, (2).
- [13] Žolkowskij A. K. & Mel'čuk I. A. On semantic synthesis [in Russian]. Problemy Kibernetiki, 1967, (19).
- [14] 张家骅: “莫斯科语义学派”, 《外语研究》, 2001年第4期。
- [15] 章宜华: “意义—篇章模型及其释义功能”, 《现代外语》, 2002年第1期。