

俄语前置词的机用语义词典描述研究

崔晓菊

(解放军外国语学院, 洛阳 471003)

提 要: 前置词的机用语义词典描述可为俄语文本语义分析提供必要的计算资源。有前置词联系是俄语句法联系的重要类型, 前置词词典描述的参数选择应当满足有前置词联系的形式化描述机制和语义分析算法的需要, 这样才能保障前置词语义词典描述的可操作性和适用性。

关键词: 前置词; 语义词典; 语义分析算法

中图分类号: H356

文献标识码: A

0 引言

俄语中的前置词没有形态变化, 一般不能单独使用, 只有同其它实词构成前置词短语才能用作句子成分。前置词对俄语文本的构成具有特殊的作用, “据粗略统计, 俄语句子里平均10—12个词中就有1个前置词。有的句子中前置词的使用频率很高, 可以占到句中单词总数的1/3。”(穆慧春等 2006: 64) 俄语中的前置词通常作为实词与实词之间的连接词, 用以表示实词所指涉的事物、行为、状态和特征之间的相互关系意义。(Н.Ю. Шведова 1980: 706) 因此, 构建面向俄文信息处理的机用语义词典, 不仅要关注实词的语义描述, 更要重视对前置词的语义描述。

传统的详解词典在对前置词的描述中, 通常尽可能详细列举每个前置词的各种语义内涵, 这种做法常常会遇到意义繁杂的困扰, 且其描述结果往往缺乏一定的结构性, 因而很难将其形式化和算法化。根据语言哲学的观点, “词只在句子里才有意义”。(转引自陈嘉映 2006: 349) 我们认为, 须将俄语前置词的语义描述纳入句子的框架, 因为“前置词本身所表示的关系是很抽象的, 具有很大的不确定性, 只有在组配过程中才能得到具体化, 它的意义是在动态的组配过程中产生的”。(薛恩奎 1999: 48)

在俄语简单句的句法构成中, 前置词的句法构式通常为:

$$W1 \rightarrow P \rightarrow W2$$

其中, W1、W2 代表实词, 分别为句法支配成分和句法从属成分。P 代表前置词, 用于标识 W1 和 W2 之间的联系, 这一联系称为有前置词联系 (предложная связь), 有前置词联系是俄语句法联系的重要类型之一, 是语义分析程序处理的直接对象。前置词词典描述参数的选择应当依据有前置词联系的形式化描述和语义分析算法的需要, 这样才能保障前置词词典描述的可操作性和适用性, 因此本文尝试通过对有前置词联系的逻辑描述和分析程序的研究来探讨前置词的机用语义词典描述。这里需要强调的是, 本文中所探讨的有前置词联系不是前置词和实词间的联系, 而是前置词所连接的实词和实词间的联系。

1 有前置词联系的形式化描述语言

实现计算机对自然语言理解的关键是将自然语言转换为形式化的知识表示语言,这种语言对文本形式化规则和推理规则都应有准确的描述。逻辑语言具有较强的表达能力,是广泛使用的知识表示语言。其它的知识表示语言(如语义网络表示法,框架表示法,产生式规则表示法,关系数据库表示法等)都可以转换为逻辑语言中的某一受限公式类。(马少平等2004: 125)虽然逻辑语言中没有自然语言前置词系统的直接对应物,但由于有前置词联系是简单句的重要句法语义成分,而简单句在逻辑语言中的对应物被认为是形如 $P(x_1, x_2, \dots, x_n)$ 的逻辑公式(其中 P 为谓词, x_i 为个体词, $i=1, 2, \dots, n$),因此对前置词的语义描述应当与该公式的元素构成联系起来。

在简单句的逻辑表示式 $P(x_1, x_2, \dots, x_n)$ 中,个体词的排列顺序可以表示个体词相对于谓词 P 的角色意义,如 x_1 是谓词 P 的第一对象, x_2 是谓词 P 的第二对象……等等,但却不能表示角色本身的具体内涵,如:谓词 *покупать* (购买)的第一个角色位(ролевая позиция)是 *покупатель* (кто купил) (买者),第二个角色位是 *продавец* (у кого купил) (卖者),第三个角色位是 *товар* (что куплено) (商品)。此外,逻辑语言中用于表达角色间关系的手段贫乏,它只能表示出角色间的等值关系(如表达式 $a=b$)和不等值关系($a>b$ 或 $a<b$),而不能表示出角色间关系的内涵意义。由此就需要在上面的逻辑表示式中引入一种新的标记,以标识角色位的抽象语义内涵,为此需要进行以下两个方面的工作:

1) 预先确定好逻辑表示语言的角色清单如 Agent, Experiencer, Source, Goal, Instrument 等,并规定相应的角色标示符: $\rho_1, \rho_2, \rho_3, \dots$;

2) 将角色标示符引入逻辑公式:

$$P(\rho_1: x_1, \rho_2: x_2, \dots, \rho_n: x_n)$$

这种逻辑表示式亦称增强型谓词公式。

2 有前置词联系的逻辑描述机制

有前置词联系可分为强句法联系和弱句法联系两种。强句法联系指在从属词空缺时原词组或句子就丧失语义完整性的从属联系(подчинительная связь),如 *войти в комнату* (走进房间), *выйти из кабинета* (走出办公室), *оторваться от семьи* (脱离家庭)等。弱句法联系指在从属词空缺时原词组或句子仍具有语义完整性的从属联系,如 *книга на столе* (桌子上的书), *день без солнышка* (没有阳光的白天), *человек в шляпе* (戴着帽子的人)等。(Русская грамматика, том II 1980: 19—20)强句法联系和弱句法联系的句法语义表现不同,因而其描述机制也应不同,下面我们将分别探讨上述两类有前置词联系的逻辑描述机制。

强句法联系通常指谓词(情境词,句法联系的支配成分)和论元(一定谓词情境角色位的填充词,句法联系的从属成分)之间的联系。在强句法联系中前置词和被前置词支配的名词的语法格是论元角色位的指示符。例如在 *прибыл в Москву* (来到莫斯科)和 *прибыл из Москвы* (来自莫斯科)两个句子中,正是由于前置词及名词语法格的不同导致了句子意义的根本改变。在增强型谓词公式中, ρ_i 为论元角色位的形式标识符,由此可以得出,前置词及其后名词的语法格对应于增强型谓词公式中的 ρ_i ,这样我们就可以把有前置词的强句法联系描述为:

$$\text{描述机制 1: } P(\dots\rho_i: x_i \dots)$$

其中,支配前置词的词对应谓词 P ,被前置词支配的词对应论元 x_i 。例如在我们上面的例句中,前置词 *в* 和名词四格对应角色指示符 Goal,前置词 *из* 和名词二格对应角色指示符 Source。

显然,有关前置词所标识的角色位的信息应当记录在语义词典中。这些信息应当包括诸

如以下的内容：俄语前置词 ИЗ, ОТ, C_{gen} (C+名词二格) 标识角色位 Source；俄语前置词 К, В_{acc} (В+名词四格), НА_{acc} (НА+名词四格) 标识角色位 Goal；俄语前置词 НА_{loc} (НА+名词六格), В_{loc} (В+名词六格), ПОСРЕДСТВОМ 标识角色位 Instrument 等等。不能用上述描述机制描写的前置词如 ДЛЯ, У, ПРИ, ПО, ПОД, ЧЕРЕЗ 等, 在语义词典中将不记录其角色位信息。

对于弱句法联系来说则需要另外一种描述方法, 将前置词所指示的句法联系形式化。基于该方法的描述机制如下:

描述机制 2: R(x, y)

其中, x 对应句法联系的支配成分, y 对应句法联系的从属成分, R 表示 x, y 所指涉实体间的关系, 这一关系的确定可以有很多依据。对于有前置词的弱句法联系来说, 前置词本身就指示了实体间的关系, 因而可以作为其确定依据。下面就来看一些示例词组:

рукопись на столе (桌子上的手稿) → рукопись находится на столе;
рукопись в столе (桌子的手稿) → рукопись находится внутри столе;
рукопись под столом (桌子底下的手稿) → рукопись находится под столом;

箭头右侧的划线部分是左侧词组中实体间关系的词汇化。由上述示例可以看出, 箭头左侧的三个词组具有相同的实体组成(即 рукопись, стол), 但由于其中前置词及其后名词语格的不同, 导致了三个词组中实体间关系的不同。因此, 在语义词典中还应记录前置词在一定文本环境下所能指示的实体间关系的信息。

上述描述机制适用于句法联系的双方具有不同指涉的情况, 当句法联系的双方具有相同指涉时就需要另外一种描述机制:

描述机制 3: A(x)&B(x)

其中, A、B 对应句法联系的双方, x 对应句法联系的双方所共同指涉的实体。例如: банка из темного стекла (用深色玻璃制作的罐子) → БАНКА(x) & СТЕКЛЯННЫЙ(x)。

上面示例中前置词 из 指示的是名词及其参与者间的同指关系。这类句法联系在俄语中通常为无前置词句法联系。例如, 上面的示例词组常常具有同义表达形式: банка темного стекла。但是并非所有的前置词都可指示同指关系, 例如, 在词组 банка за темным стеклом (深色玻璃后面的罐子), банка под темным стеклом (深色玻璃下面的罐子) 中句法联系的双方显然具有不同的指涉。因此, 在对前置词的词典描述中还应引入相应的“可表同指关系”标记。

此外, 还应特别注意前置词在有前置词的数量词组中的作用。只有在词典中记录相关前置词所能指示及如何指示的数量意义的信息, 才能实现对该类有前置词的数量词组的准确描述。这里至少应区分如下情况:

- 1) 指示确切数量意义的前置词, 如: вода кипит при 100 градусах (水在 100 度时会沸腾);
- 2) 指示取值上限的前置词, 如: набить цену до пятидесяти рублей (把价格抬到 50 卢布);
- 3) 指示取值下限的前置词, 如: дети от восьми лет (8 岁以上的孩子);
- 4) 指示近似数量意义的前置词, 如: на расстоянии около 100км (大约 100 公里的距离);
- 5) 指示增减倍数的前置词, 如: мощность выросла в 2 раза (功率增加了一倍);
- 6) 指示增减量的前置词, 如: вес уменьшен на 20кг (重量减少了 20 公斤)。

综上所述, 如果要对文本中的有前置词联系进行准确的描述, 那么在前置词的词典描述

中就应记录以下信息：

- 1) 该前置词可以标识谓项的哪些角色位；
- 2) 该前置词可以标识哪些实体间关系；
- 3) 该前置词是否可以指示同指关系；
- 4) 该前置词可以指示哪些数量意义。

3 前置词词典描述信息在文本语义分析算法中的应用

本节将简要探讨前置词的语义描述信息在语义分析器中的应用情况。

语义分析器的输入是经过句法剖析后的文本，其中包含了剖析器所提取的文本中句法联系的所有变体。语义分析器的作用是通过对上述变体进行遍历分析，从而得出一个最优的选择，并输出其相应的语义表示，其中涉及词汇消歧和句法消歧的问题。语义分析器的分析过程依赖于包含有词汇句法语义信息的机器词典描述。（冯志伟等 2005：344—346）

有前置词联系的语义分析算法不仅要基于前置词本身的描述信息，而且要基于前置词所连接的主从双方的词典描述信息。语义分析器利用这两类信息进行推理和验证，从而确定有前置词联系的语义描述机制。具体来看，有前置词联系 $W1 \rightarrow P \rightarrow W2$ 的语义分析过程如下：

1) 如果 $W2$ 为数字，且在前置词的词典描述中记录有其能指示某种数量意义的信息，那么就确定该有前置词联系表示前置词指示的数量意义。

2) 如果在 $W1$ 的词典描述中记录有语义支配模型，且在前置词的词典描述中记录有其能指示角色位 ρ_i 的信息，那么就检验 $W2$ 是否满足谓项 $W1$ 对其角色位 ρ_i 的语义约束条件。如果所有的条件都被满足，那么就确定该有前置词联系为强句法联系，从而应当选择描述机制 1。

3) 如果 $W1$ 和 $W2$ 的语义类别均为“объект（事物）”，且在前置词的词典描述中记录有其能指示实体间关系 R_{prep} 的信息，那么就检验 $W1$ 和 $W2$ 之间是否可以构成关系 R_{prep} ，如果可以构成，那么就确定该有前置词联系表示实体间关系 R_{prep} ，从而应当选择描述机制 2。

4) 如果 $W1$ 和 $W2$ 的语义类别均为“объект（事物）”，且在前置词的词典描述中记录有其能指示同指关系的信息，那么就检验 $W1$ 和 $W2$ 是否可以兼容，如果可以兼容，那么就确定该有前置词联系表示同指关系，从而应当选择描述机制 3。

5) 如果上述 4 种情况都不符合，那么语义分析失败，并确定该有前置词联系为不合法的句法语义变体。

从以上对有前置词联系的语义分析过程的描述可以得出，在语义词典中应引入以下前置词函项：

$PrepNumValTyp(P)$ ，函项返回值为前置词 P 能够指示的数量意义类型；

$PrepRole(P, Rf)$ ，函项返回值为前置词 P 和其后名词的语法格 Rf 可以标识的角色位；

$PrepCoref_IsPossible(P)$ ，函项返回值为布尔值 {Yes/No}（即是/否可以标识同指关系）；

$PrepRs(P, Rf)$ ，函项返回值为前置词 P 和其后名词的语法格 Rf 可以标识的所有实体间关系；

4 结束语

机用语义词典是文本语义分析的重要支撑，其在机器翻译系统和信息检索系统中发挥着重要作用。与人用语义词典不同，机用语义词典的编纂要面向机器理解和应用程序的需要，将分析算法中所需的语义信息形式化和代码化。（傅兴尚等 2009：153—155）现今的很多研究主要集中于实词的词典描述，而对前置词、连接词和小品词的关注较少。有前置词联系是俄语句法联系的重要类型，本文通过对有前置词联系的逻辑描述和分析算法的探讨，提出了对俄语前置词进行机用语义词典描述的参数。由于这些参数是基于语义信息形式化和分析程

序的需要提出的，因而可以保障采用这种参数设计的前置词词典描述的可操作性和适用性。

参考文献

- [1]陈嘉映. 语言哲学[M]. 北京: 北京大学出版社, 2006.
- [2]傅兴尚、许汉成、易绵竹、李向东. 俄罗斯计算语言学与机器翻译[M]. 北京: 语文出版社, 2009.
- [3]冀刚. 俄语前置词用法词典[M]. 上海: 上海译文出版社, 1992.
- [4]马少平、朱小燕. 人工智能[M]. 北京: 清华大学出版社, 2004.
- [5]穆慧春、曲雅静. 俄语前置词与汉语介词差异研究[J]. 吉林省教育学院学报, 2006年第6期.
- [6]薛恩奎. 俄语空间前置词语义情景结构[J]. 解放军外国语学院学报, 1999年第5期.
- [7]Н. Ю. Шведова (Гл. ред). Русская грамматика[M]. Москва: 1980.
- [8]Daniel Jurafsky, James H. Martin. 自然语言处理综论(冯志伟、孙乐译)[M]. 北京: 电子工业出版社, 2005.

On the Description of Russian Preposition in the Machine-readable Semantic Dictionary

CUI Xiao-ju

(Eurasian Department of PLA Foreign Languages University, Luoyang 471003, China)

Abstract: The formalization of prepositional meaning can offer necessary calculation resources for the automatic Russian text analysis. This paper analyses the formalization and semantic processing of prepositional structure, in the view of which proposes several parameters for the description of preposition in the machine-readable semantic dictionary.

Key words: preposition; semantic dictionary; semantic processing

基金项目: 国家社会科学基金项目“俄罗斯应用语言学研究新景观”(项目编号: 05BYY008)。

作者简介: 崔晓菊(1984—), 女, 河南许昌人, 解放军外国语学院博士研究生。主要研究方向: 计算语言学。

收稿日期: 2010-12-06

[责任编辑: 张春新]