

俄罗斯词汇语义资源工程概览 (1)

——《俄语通用语义词典》解析

易绵竹¹, 王京武², 南振兴³, 何英玉⁴, 武 斌⁵

(1. 黑龙江大学俄语语言文学研究中心, 黑龙江 哈尔滨 150080; 2. 解放军外国语学院训练部研究生处, 河南 洛阳 471003; 3. 解放军外国语学院一系, 河南 洛阳 471003; 4,5. 解放军外国语学院国防语言文化研究所, 河南 洛阳 471003)

摘 要: 俄罗斯学术界历来十分重视词汇语义学研究和词汇语义资源工程建设, 取得了世人瞩目的理论成果和应用成果。全面梳理俄罗斯学者进行的词汇语义资源建设工程项目, 必将有助于我们充分利用这些资源开展俄汉词汇语义对比研究和俄汉机器翻译系统研制工作。我们准备以系列论文介绍几个主要的词汇语义资源建设工程项目, 首先拟对尼·尼·列昂婕娃 (Н.Н.Леонтьева) 主持开发的《俄语通用语义词典》加以解析和评述。

关键词: 莫斯科语义学派; 词汇语义资源; 俄语通用语义词典

中图分类号: H353

文献标识码: A

1 引论

莫斯科语义学派倡导的词汇语义学研究方法, 在俄罗斯及国外理论语言学界和计算语言学界均产生了广泛而深远的影响。特别值得指出的是, 在继承 И.А.Мельчук (1974/1999) 等创立的«意思 文本»语言模型理论之基础上, 莫斯科语义学派的代表人物 Ю.Д.Апресян (1967, 1974/1995a, 1986, 1991, 1995b), 通过对俄语词汇语义学的实验研究, 于 20 世纪 80 年代初期形成了自然语言的整合描写思想 (идея интегрального описания естественного языка), 其实质就是词典和语法的整合描写。这种学术思想是语言学理论俄罗斯化 (русизация лингвистической теории) 的一项重要成果, Ю.Н.Караулов (1994: 6) 将其确定为当代俄语学第一个方法论纲领 (методологическая программа)。应该看到, 莫斯科语义学派的学术影响已超出俄罗斯国界, 该学派代表人物 Ю.Д.Апресян 院士的学术著作被译成英、法、德、西班牙、葡萄牙、匈牙利、波兰文流行于世。我国著名俄语学者华劭先生(2003: 191) 在其新著《语言经纬》中正确指出: “莫斯科语义学派把重点放在词汇的研究上。”在纪念 Ю.Д.Апресян 院士诞辰 70 周年的文集中, 国内外学者讨论的中心议题就是“词在文本中与在词典中”(Слово в тексте и в словаре)。 (Л.Л.Иомдин, Л.П.Крысин 2000) 在第二届汉语词汇语义学研讨会上, 本文第一作者应邀作了题为“词汇语义学研究在俄罗斯 — 阿普列相现象简评”的学术报告 (易绵竹 2001), 引起了两岸三地学者的浓厚兴趣。由我国著名俄语学者张家骅先生主持完成的教育部人文社会科学重点研究基地重大项目成果《俄罗斯当代语义学》, 对莫斯科语义学派的理论背景和学术成就进行了更为详尽的介绍与评述, (参见: www.rlls.edu.cn/download/2003-05-24.txt; 张家骅 2001) 此项成果必

将极大地推动中国语义学研究，拓宽其研究范围。

无疑，本文主角 Н.Н.Леонтьева 受到了莫斯科语义学派思想的浸染和影响。她现为莫斯科国立大学科研计算中心高级研究员，最早曾在前苏联第一个机器翻译实验室工作 (Н.Н.Леонтьева 在致本文第一作者的电子邮件中写道：“Я ведь была первым сотрудником первой Лаборатории машинного перевода, созданной в 1959 году в МГПИИЯ.”)，国际知名学者 И.А.Мельчук 和 О.С.Кулагина 曾担任该实验室领导。Н.Н.Леонтьева 的学术兴趣是机器翻译、文本的信息分析和文本知识库的自动构造，她的副博士论文题为“文本的语义分析与意思不完备性”(Семантический анализ и смысловая неполнота текста)，博士论文题为“文本自动理解信息模式中语义要素的构造”(Строение семантического компонента в информационной модели автоматического понимания текста)，其主要科研成果和学术论文目录详见 <http://www.srcc.msu.ru/nivc/about/inf/lnn.html>。Н.Н.Леонтьева 曾在俄罗斯科学院美国和加拿大研究所率领课题组成功研制了政治文本分析系统 ПОЛИТЕКСТ 和法—俄自动翻译系统 ФРАП，从 1998 年开始研制商用俄—英机器翻译系统 ДИАЛИНГ。她是最先主张需要利用语义信息进行文本机器处理的学者之一，她提出的语义关系、语义范畴和语义特征等基本概念已成为应用机器 (计算) 语义学的基础。尽管其可能导致某种学理性的成就得到了普遍承认，但她依旧对应用系统感兴趣，并随时准备改变和完善自己的学术观点。本文介绍的《俄语通用语义词典》(Русский общесемантический словарь, 简称 РОСС) 是其理论的核心，该词典以对俄语实体词位 (полнозначные лексемы) 进行语义描述为己任。其方法论基础是将文本分析划成不同层次，每个层次就是输入文本的某种表征。在每个层次上都有特定的表征语言，而不是某种限定语 (атрибуты)。这种分析的最终结果就是构造语义表达式，后者由文本情景事实构成(текстовые факты-ситуации)。情景与其他情景和其他行动元(актанты)之间的关系被称作语义关系。(http://www.aot.ru/history.html)

2 РОСС 的设计理念

Н.Н.Леонтьева 指出，词典既是整个语言学的核心构件，又是任何文本处理系统，即语言转换器 (лингвистический транслятор) 的重要模块。在词典中能够反映出系统的哲学，根据词典可以揭示语言的层次和系统所采用的描述语言，词典信息有助于判定系统的语义解释力。Ю.Д.Апресян 提出的语言整合描写思想已成为应用系统建造的必要原则，它使语言理论和实践彼此协调。Н.Н.Леонтьева 认为，每一部词典中都隐含着编纂者心仪的动机，比如 И.А.Мельчук 设计“意思—文本”模型的词典之主要动机就是系统地描述词汇的组配性能：在语法层次上详细描述支配模式(модели управления)，在词汇层次上引入词汇项(лексические функции)，这就在很大程度上改善了传统词典释义的方法，从而降低了词典描述的主观性。(И.А.Мельчук 1974/1999; ТКС 1976-1996)

在 РОСС 设计过程中，Н.Н.Леонтьева 的基本理念就是要保证词汇语义的描述能够同时通向文本和事物领域 (“выход в текст” и “выход к предметной области”)。前一部分提供文本配价描述的语言和方式，后一部分使该词典与领域词典统合起来，从而实现语言信息和领域信息在文本语义结构形成过程中的柔性对接。(Н.Н.Леонтьева 1997: 6) 作为 ПОЛИТЕКСТ, ФРАП 和 ДИАЛИНГ 系统的主要语义模块，РОСС 建造的基石是 Н.Н.Леонтьева 创立的信息—语言学模型 (информационно-лингвистическая модель)，其实质乃是调和语言学与事物领域之间的关系，模拟自然语言文本的柔性理解 (“мягкого понимания”) 机制，从而形成所分析文本的信息空间。简言之，所谓文本的柔性理解是指，由于理解过程的条件和步骤不同，同一文本允许有不同的理解结果。(Н.Н.Леонтьева 2000: 4-6)

3 РОСС 的总体构架

3.1 层级结构

РОСС 的结构具有层级性,其底层是表示具体词义的场(поля),其顶层则由众多场名—域(зоны) 所构成,这些场和域的数据就是词典信息的表征方式。该词典新版包括 10 个域和 50 多个场。(Н.Н.Леонтьева 1997: 7; Leontyeva 2003)

这 10 个域分别用下列字母符号表示:

ВХОД: 输入词汇单位辨识 (词义名称及编号);

ОБЩ: 词的一般信息;

МОРФ: 用以证明赋予该词意义合理性的词法限制、以及派生词的信息;

СИН: 词的句法类别及其在句子中的典型角色;

СЕМ: 词的语义特征,包括一整套意思 (语义) 配价;

СИТ: 受词义制约并在该事物领域或然率最大的情景之结构;

ИНФ: 词的领域归属属性标注、变体 (准同义词)、在语义和信息表征中该词的信息权重、百科性功能、由该词构成的术语、以及该词在某领域中的义类联系;

ПРАГМ: 该词在给定领域中的语用信息,即事件构造方式、推论和预设、事件的评价;

ЛЕКС: 该词与其他词的词汇搭配、固定词组、成语、词汇函项;

ЭКВ: 该词的外语 (如英语) 对应词及与其相关的术语;

КОММ: 该词条编写者的备注、词义的例证、编写者姓名等信息。

(注: **МОРФ** 和 **СИН** 统称词的语法特征,可合并为一个域。)

由于场的种类和名称十分繁杂,我们仅举语义域 **СЕМ** 为例以窥 РОСС 之一斑。该域包括以下若干场: **СХ** (语义特征), **ВАЛ** (配价), **СХ1-7** (语义特征), **АЛФ1-7** (行动元词汇函项), **ВОПР1-7** (问题), **ГХ1-7** (语法特征), **ИЛЛ1-7** (例证), **ПОЗ** (词汇及其行动元的相互位置), **ДОП** (行动元之间补充的意思联系), **НЕСОВМ** (配价实现方式/语法特征的不兼容性), **СОВМ** (语法特征和/或语义特征的兼容性), **КОРР** (配价结构的校正规则), **ТРАН (СХ)** (语义特征的转换), **ТРАН (ВАЛ)** (配价的转换) 等。

如下列名词的语义特征 **СХ** 表示为:

август

СХ=ВРЕМЯ, ИНТЕРВЛ/время, интервал/

автобус

СХ=УСТР, ДВИЖ/устройство, движеник/

又如,下列形容词和动词的配价 **ВАЛ** 表示为:

враждебный

ВАЛ=СУБЪЕКТ (А1,С); АДРЕСАТ (А2,С); АСПЕКТ (А3,С)

компенсировать

ВАЛ=СУБ, А1, С; ПРИЧ, А2, С; АДР, А3, С; СОДЕРЖ, А4, С

ДОП=1. АКТАНТ, А1, А4; 2. ПАЦИЕН, А3, А2; 3. АДР, А3, А4.

3.2 语义关系表

语义关系是 РОСС 最基本的语义工具,它被界定为“操某种语言的人在文本中所发现的共性联系”(некая универсальная связь, усматриваемая носителем языка в тексте). 这种联系是二元的,它标示一个语义节点与另一个语义节点之间的联系。语义关系的写入格式为: **R (A, B)** 这里 **R** 表示语义关系, **A** 表示关系的从属成分, **B** 表示关系的支配成分。如短语 **роман Толстого** 可表示为: **АВТОР (Толстой, роман)**,意思是“Толстой является АВТОРОм романа”。

在 ДИАЛИНГ 系统中，总共罗列出 35 种语义关系，现以表格形式引录于此：

语义关系代码	示例	结构
АВТОР/作者	Роман Толстого Указ Президента	АВТОР (ТОЛСТОЙ, РОМАН) АВТОР (ПРЕЗИДЕНТ, УКАЗ)
АГЕНТ/施事	Мы сократили отставание	АГЕНТ (МЫ, СОКРАТИТЬ)
АДР/接收者	Я отдал стул отцу.	АДР (ОТЕЦ, ОТДАВАТЬ)
В-НАПР/方向	указатель на Монино	В-НАПР (МОНИНО,УКАЗАТЕЛЬ)
ВРЕМЯ/时间	Это произошло вчера.	ВРЕМЯ (ВЧЕРА, ПРОИЗОЙТИ)
ЗНАЧ/值	Высота дома – 20 метров.	ЗНАЧ (20 МЕТРОВ, ВЫСОТА)
ИДЕНТ/标识符	Дом N 20	ИДЕНТ (N 20, ДОМ)
ИМЯ/名字	Дворник Степанов	ИМЯ (СТЕПАНОВ, ДВОРНИК)
ИНСТР/工具	резать ножом	ИНСТР (НОЖ, РЕЗАТЬ)
ИСХ-Т/起点	яблоки из Молдавии	ИСХ-Т (МОЛДАВИЯ, ЯБЛОКИ)
К-АГЕНТ/配动者	купил у старьевщика	К-АГЕНТ (СТАРЬЕВЩИК, КУПИТЬ)
КОЛИЧ/数量	два яблока	КОЛИЧ (ДВА, ЯБЛОКО)
КОН-Т/终点	уехать в Москву	КОН-Т (МОСКВА, УЕХАТЬ)
ЛОК/处所	жить в глуши	ЛОК (ГЛУШЬ, ЖИТЬ)
МАСШТ/规模	Банк России	МАСШТ (РОССИЯ, БАНК)
МАТЕР/材料	сумка из кожи	МАТЕР (КОЖА,СУМКА)
НАЗН/用途	книга для детей	НАЗН (ДЕТИ, КНИГА)
ОБ/客体	уничтожить мост	ОБ (МОСТ, УНИЧТОЖИТЬ)
ОГРН/限制	выделять по возрасту	ОГРН (ВОЗРАСТ, ВЫДЕЛЕНИЕ)
ОЦЕНКА/评价	хорошо относиться	ОЦЕНКА (ХОРОШО, ОТНОСИТЬСЯ)
ПАРАМ/参数	высота дома	ПАРАМ (ВЫСОТА, ДОМ)
ПАЦИЕН/受事	арест преступника	ПАЦИЕН (ПРЕСТУПНИК, АРЕСТ)
ПОСРЕД/凭借	закончить доклад анекдотом	ПОСРЕД (АНЕКДОТ, ЗАКОНЧИТЬ)
ПРИЗН/特征	красивый шар	ПРИЗН (КРАСИВЫЙ, ШАР)
ПРИНАДЛ/所属	дом отца	ПРИНАДЛ (ДОМ, ОТЕЦ)
ПРИЧ/原因	деревья повалены ураганом	ПРИЧ (УРАГАН, ПОВАЛИТЬ)
РЕЗЛТ/结果	испечь пирог	РЕЗЛТ (ПИРОГ, ИСПЕЧЬ)
СОДЕРЖ/内容	рассказать о весне	СОДЕРЖ (ВЕСНА, РАССКАЗАТЬ)
СПОСОБ/方式	идти босиком	СПОСОБ (БОСИКОМ, ИДТИ)
СРЕДСТВО/手段	красить белилами	СРЕДСТВО (БЕЛИЛО, КРАСИТЬ)
СТЕПЕНЬ/程度	весьма преуспеть	СТЕПЕНЬ (ВЕСЬМА, ПРЕУСПЕТЬ)
СУБ/主体	любовь отца	СУБ (ОТЕЦ, ЛЮБОВЬ)
ТЕМА/主题	говорить о Москве	ТЕМА (МОСКВА, ГОВОРИТЬ)
ЦЕЛЬ/目的	забастовка в целях повышения	ЦЕЛЬ (ПОВЫШЕНИЕ, ЗАБАСТОВКА)

	зарплаты	
ЧАСТЬ/部分	ножка стула	ЧАСТЬ (НОЖКА, СТУЛ)

顺便指出，上列各类语义关系同样具有层级性。Н.Н.Леонтьева 对以下几组语义关系 АДР, В-НАПР, КОН-Т//СПОСОБ, ИНСТР, СРЕДСТВ, ПОСРЕД//ИДЕНТ, ИМЯ// СУБ, АГЕНТ//ПРИЗН, ОЦЕНКА, ПАРАМ//ЦЕЛЬ, НАЗН//ОБ, СОДЕРЖ. 进行了详细的比较分析，充分揭示了它们之间的区别与联系。如 *I talk to you => АДР (you, talk) ; I go to the town => КОН-Т (town, go) ; The vane points to the north => В-НАПР (north, point) .*

为了证明自己提出的语义关系列表之合理性，Н.Н.Леонтьева 将该表与 Е.В.Падучева 的语义角色清单和 Ю.Д.Апресян 的语义配价清单进行了对比。我们不妨照录于此，以加深对各家学说的理解。

(1) Е.В. Падучева 的语义角色清单

语义角色名称	示例	ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ
Агент	Иван приехал	АГЕНТ (Иван, поехать)
Конечный пункт	приехал в Варшаву	КОН-Т (Варшава, приехать)
Адресат	сказать отцу	АДР (отец, сказать)
Степень	наполнить полностью	СТЕПЕНЬ (полностью, наполнить)
Направление движения	поехать направо	В-НАПР (направо, поехать)
Исходный пункт	приехать из Москвы	ИСХ-Т (Москва, приехать)
Пациент	съесть яблоко	ПАЦИЕНТ (яблоко, съесть)
Место	банка кишела	ЛОК (банка, кишеть)
Содержание	чтение нравится ему	СОДЕРЖ (чтение, нравится)
Посессор	борода Ивана	ПРИНАДЛ (борода, Петр)
Инструмент	открывать ключом	ИНСТР (ключ, открывать)
Результат	резать на куски	РЕЗЛТ (куски, резать)

(2) Ю.Д.Апресян 的语义配价清单

语义配价名称	示例	ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ
Sub (субъект)	поезд движется	СУБ (Поезд, двигаться)
Contrag (контрагент)	покупать у старьевщика	К-АГЕНТ (старьевщик, покупать)
Сар (глава)	вина перед коллективом	
Obj (объект)	гладить руку, стрелять в мишень	ОБ (рука, гладить) КОН-Т (мишень, стрелять)
Content (содержание)	знать об отъезде	СОДЕРЖ (отъезд, знать)
Adr (адресат)	сообщать президенту	АДР (президент, сообщать)
Recip (получатель)	давать детям, дарить людям	АДР (дети, давать) АДР (люди, дарить)
Via (посредник)	передовать через секретаря	ПОСРЕД (секретарь, передавать)
Is (источник)	брать в кассе	ИСХ-Т (касса, брать)
Loc (место)	находиться в лесу	ЛОК (лес, находиться)
Ab (начальная)	вывести из леса	ИСХ-Т (лес, вывести)

точка)		
Ad (конечная точка)	везти в город	КОН-Т (город, везти)
Itin (маршрут)	идти по дороге	
Med (средство)	прибивать гвоздями	СРЕДСТВО (гвозди, прибивать)
Instr (инструмент)	резать ножом	ИНСТР (нож, резать)
Mod (способ)	обращаться плохо	АСПЕКТ (плохо, обращаться)
Cond (условие)	если P, то Q	УСЛ (P,Q)
Motiv (мотивировка)	награждать за храбрость	ПРИЧ (храбрость, награждать)
Caus (причина)	радоваться подарку, проистекать из-за Q	ПРИЧ (подарок, радоваться) ПРИЧ (Q, проистекать)
Result (результат)	превращать в воду	РЕЗЛТ (вода, превращать)
Dest (цель)	стремиться к общему благу	ЦЕЛЬ (благо, стремиться)
Asp (аспект)	превосходить по качеству	ОГРН (качество, превосходить)
Quant (количество)	пять человек	КОЛИЧ (пять, человек)
Period (срок)	отпуск на два месяца	ВРЕМЯ (на два месяца, отпуск)
Temp (время)	начаться в полночь	ВРЕМЯ (полночь, начаться)

3.3 语义特征集

РОСС 的语义特征在词汇语义描述中起着关键作用, 这类特征共计约 40 个。这样, 每个词都被写入由语义特征所组成的逻辑式子(用逻辑连接词 **и**, **или** 标示)。每个词的配价结构记录为 $\langle A_1, \dots, A_7 \rangle$, 其中 A_i 是对行动元(актант) 的描述, 它采用偶对 $\langle GX_i, CX_i \rangle$ 来表示, GXi 是对行动元语法表达的描述, CXi 是对行动元的语义描述。并且, 词的配价结构是由算法来实现的。下面仍以表格形式罗列 РОСС 的语义特征集。

语义特征 CX 的代码	注释	带有该语义特征 CX 的例词	带有该语义特征 CX _i 的例词
АБСТР/ 抽象物	Любое абстрактное существительное или прилагательное	модель, план, тенденция, обстоятельство	внедрять технологию; использовать план
АРТ/ 人造物	Артефакт. Все, что сделано человеком.	машина, хлеб, памятник	выпускать машины; требование закона
ВЕЛИЧ/ 度量	Прилагательные, образованные от параметрических существительных и от существительных, обозначающих какое-либо значение на параметрической шкале	<i>высокий,</i> <i>мощный</i> <i>большой</i>	—
ВЕЩВО/ 物质	Любое название химического вещества или того, что можно как-либо дозировать, отмерять, продавать по весу или объему.	аммиак, бензин, ядохимикат	вес натрия; дом из бетона; месторождение нефти
ВЛАСТЬ/ 政权	Высшие государственные и военные должности и	генерал президент	Министр назначил;

	учреждения. Любые должности, связанные с непосредственным управлением людьми.	руководитель	Дума ратифицировала; секретарь директора
ВМЕСТЛ/ 容器	Все, что предназначено для содержания чего-либо другого.	мешок, сейф, гараж	—
ВОСПР/ 感知	все глаголы и существительные восприятия	слушать видеть чувствовать	—
ВРЕД/ 有害事物	Все, к чему человек обычно относится как к нежелательному.	катастрофа, война, перегрузка	устранить недостаток
ГЕОГР/ 地理客体	Любой географический объект	остров, река	набережная реки впадать море стоять на холме
ГОС/ 国家名	Любое название государства или тип государства	республика, Россия	компартия России парламент Украины политика США представитель ЮАР
ДВИЖ/ 运动	Глаголы движения (в том числе декаузативы)	идти, ронять	—
ДОЛЖ/ 职务	Должность, профессия, социальный статус	повар, парработник	—
Д-УСТР/ 装置零件	Деталь устройства	карбюратор валик	—
ИЗМ/ 促变	Действия-изменения	наращивать, реформировать, копить	—
ИНТЕЛ/ 智能行为	все действия, непосредственно связанные с мыслительной деятельностью	надеяться, изучать, решать	—
ИНТРВЛ/ 时段	Временной интервал	день, неделя	продлить на два дня; бюджет на 2000 год
ИНФ/ 信息	Слова, обозначающие информацию	знание, команда, утверждение, новость	—
КОММУНИК/ 交际	Глаголы речи	выражать, выступать репортаж	—
НОСИНФ/ 信息载体	Носители информации. Это можно прочесть, а потом сжечь.	книга, газета, заметка	заголовок книги публикация заметки написать статью

Н-ТРЕБ/ 要求或指令	Набор требований (разновидность НОСИНФа)	закон инструкция	нарушать закон соблюдать инструкции отменять указ
ОДЕЯТ/ 活动领域	Область деятельности	физика балет быт	министр культуры
ОДУШ/ 有生命的客体	Семантически одушевленный объект	папа, президент	Отец рассказывает любовь Ивана
ОРГ/ 组织	Любая организация	колхоз, школа	актив банка директор института облигации банка
ПРЕДМ/ 物品	Любой предмет (объект, который меньше по размерам среднего человека) . Часто является АРТ. В отличие от УСТР устроен просто.	марка, бинокль	—
ПРОТЯЖ/ 延伸的地理客体	Протяженные географические объекты	дорога, граница	река впадает в море
СОБИР/ 集合体	все, что обозначает множество однотипных объектов	библиотека, молодежь	—
СОЦ/ 社会情景	Любые ситуации, выходящие за пределы одной семьи	митинг, коррупция, олимпиада	—
УСТР/ 装置	Любое устройство	компьютер, лифт	чинить компьютер разобрать лифт
ФИН/ 财经	Все, что связано с финансами	деньги, банк, долг	продавать за рубли индексировать пенсию
ХОР/ 肯定评价	Все, что оценивается как положительное	взаимопомощь, мужество	—
ЭМОЦ/ 情感	Обычно прилагательные, которые выражают эмоции	мизерный, могучий несчастный	—
ЯВЛЕН/ 现象	Ситуация, для которой трудно найти причину	смерч, терроризм	—

应当指出, 上述某些特征不能单独使用, 如集合特征 СОБИР 需要与其他特征配合使用: СХ (библиотека) = СОБИР, НОСИНФ; СХ (молодежь) = СОБИР, ОДУШ. 此外, 某些特征是合成的: НОСИНФ = ИНФ, ПРЕДМ; Н-ТРЕБ= ВЛАСТЬ, НОСИНФ. 还有一些特征是反义的, 禁止合用: ВРЕД, ХОР; АБСТР, ПРЕДМ. 也有一些特征是另一些特征的变种: ОРГ – разновидность ОДУШ; ДОЛЖ – разновидность ОДУШ; Д-УСТР, УСТР – разновидности АРТ; СОЦ, ФИН – разновидности ОДЕЯТ.

3.4 语义分类体系

在 РОСС 的 ОБЩ 域下设置了一个重要的范畴场 КАТ (категория), 以确定词汇的基本语义类别。根据词汇语义表现(семантическое поведение)的差异性, 划分出以下 5 大语义类别。

(1)标签词 (слова-этикетки, ЭТК) — 这是一个最大的开放词类, 它只包括实体词。标签词构成事物领域的词汇核心, 在词典中其信息权重 (информационный вес) 最大, 并且随着领域的变换, 其词汇组成、词义系统和领域词典的检索系统发生变化, 自动索引系统正是依据这类词研制的。标签词又可分为 3 个子类: ①客体词 (ЭТК-ОБ, 如 *президент, статья, законопроект*); ②特征词 (ЭТК-ПРИЗН, 指所有形容词和副词); ③ 情景词 (ЭТК-СИТ, 如: *война, обсуждение, утечка*)。对标签词还可进行更为细致的语义区分, 它主要由语义域 СЕМ 下设置的语义特征场 СХ 和配价场 ВАЛ 来确定。

(2)指代词 (местоименные слова, МЕСТ) — 它是标签词的一个变种, 但语义为空 (семантически пусты), 需要由指称语义关系 (референт, РЕФ) 或特征语义关系 ПРИЗНАК 来加以确切。如 АКТАНТ (*они, В*), РЕФ (? , *они*); ПРИЗНАК (*такой, В*), РЕФ (? , *такой*)。

(3)关系词 (слова-отношения, ОТН) 亦称意思关系(смысловые отношения, СО) — 它在逻辑式 R (A, B) 中占据 R 位置, 如 РАВНО (A, B); АДРЕСАТ (A, B)。这类词主要包括虚词 (前置词、连接词和标点符号), 名词的格范畴意义和动词的时—体范畴意义也归入该类词。如表人名词第三格的语义表达式 *Ивановскому А.П.* → АДРЕСАТ (*Ивановский А.П., ?*)。在问号“?”位置可以插入表示给予义的动词 (*дать, передать, выдать...*)。

(4)角度词 (аспектные слова, АСП) — 从语法和语义角度来看, 这类词各式各样, 但主要包括形容词、名词、动词和副词。它们在语义表达式中具有共同的结构属性: 在逻辑式 R (A, B) 中占据第一个位置 A, 并给定关系名 (种概念)。如 СТАДИЯ (*начинать, строить*) МОДАЛЬНОСТЬ (*должен, строить*), ОЦЕНКА (*первостепенный, важность*), СПОСОБ (*быстро, передвигаться*), ПАРАМЕТР (*размер, В*), ЧАСТЬ (*член, В*), ЧАСТЬ (*центр, В*)。

(5)算子词 (слова-операторы, ОПЕР) — 在初级语义表达式中, 这类词占据逻辑式 ОПЕР (C, B) 的第一个位置 C, 并给出第二个位置 B 的语义主导词。算子词包括语气词 (*не, даже, только*), 插入语 (*кстати, в частности, действительно, на самом деле*) 和其他词 (*особенно, наоборот, соответственно*)。在语义特征场 СХ 和配价场 ВАЛ 中, 算子词可表示为: ФУНКЦИЯ (C, B), МОДАЛЬНОСТЬ (C, B), КВАНТОР (C, B)。

3.5 词条格式

РОСС 的词条格式 (формат словарных статей) 严密规范, 具有一套自定义的字母、数字和标点符号写入方式。输入词条 ВХОД 采用偶对<слово/词, номер значения/词义号>, 如 ЗН=2 (3) 表示描写的是第二个义项(总共 3 个义项)。每个输入词条都有一套偶对<название поля/场名, значение поля/场义>, 在场名和场义之间打上等号‘=’。РОСС 中最基本的场名和场义如下表所示:

场名代码	解释	场义示例
КАТ	词位的范畴	ЭТК.ОБ, ЭТК, ЭТК.СИТ...
СХ	词的语义特征	ОДУШ, ФИН...
СХ1,...,СХ7	行动元的语义特征 (1,...,7)	ОДУШ, ФИН
ГХ	词的语法特征	СУЩ: ИГ ГЛ: ГГ
ГХ1,...,ГХ7	行动元的语法特征 (1,...,7)	к_доп: Р ОБСТ_ГР: y+P

ВАЛ	词的配价结构	СУБ, А1, С
ЛФ	词汇函项	Магн: жгучий

为了登录场义，在 РОСС 中设定了众多域 (常项/константы)，由此可以构成复杂的逻辑式。一个要素“элемент”通常由几个域的常项所组成，如要素 “к_доп: Р” 包括 3 个域 (домен синтаксических отношений/句法关系域, домен двоеточия/冒号域 домен падежей /格域)。下表列出 РОСС 中常用的域 (Д):

域名代码	解释	示例
Д_КАТ	范畴	ЭТК.ОБ, ЭТК.СИТ,...
Д_ГРУППЫ	短语	ОБСТ_ГР, ГГ, ИГ,...
Д_СИН_О	句法关系	к_доп, с_опр, п_доп, отпредик...
Д_ЛФ	词汇函项	Магн, Опер, Func,...
Д_ПАДЕЖ	格	И,Р,Д,В,Т,П
Д_ПО	事物领域	фин, информ,...
Д_ПРЕДЛОГ	前置词	в,на,у,...
Д_СЕМ_О	语义关系	СУБ, ИДЕНТ,...
Д_СХ	语义特征	ОДУШ,ОРГ,...
Д_ЧАСТЬ_РЕЧИ	词类	СУЩ, ПРИЛ, НАР,...
Д_ЧИСЛО	数	ед, мн
Д_АКТАНТЫ	词条中行动元和主导词的标识	С,А1,...,А7

从形式上来看，场义的登录方式主要有下面 3 种类型：

(1) 单项式 <значение поля> = элемент 如: КАТ = ЭТК.ОБ.

(2) 多项式 <значение поля> = элемент

....

элемент,

如: ВАЛ = СУБ (А1, С)

ОБ (А2, С)

(3) 复杂式 <значение поля> = 1 элемент

....

элемент,

2 элемент

....

элемент,

....

n элемент

....

элемент,

如: СХ = 1 ДОЛЖ

ФИН

2 ДОЛЖ

СОЦ

А.В.Сокирко (1997: 20-23) 在讨论 РОСС 的程序实现时, 对词条进行了全面的形式化描述。为了更直观地显示 РОСС 的词条写入格式, 我们从该词典数据库中摘录 2 个词条作为样例。(Н.Н.леонтьева 1997: 19)

<i>акционер</i>			1 (1)
ТИП	ОБЩ	1	ЛЕ
КАТ	ОБЩ	1	ЭТК. ОБ
ГХ	СИН	1	СУЩ; ИГ
СХ	СЕМ	1	ДОЛЖ
ВАЛ	СЕМ	1	ПРИНАДЛ, С, А1
СХ1	СЕМ	1	ОРГ
		2	К_ДОП: Р
ГХ1	СЕМ	1	С_ОПР: ПРИТ_МЕСТ
АНГ	ЭКВ	1	shareholder
		2	stockholder

<i>активный</i>			1 (1)
ТИП	ОБЩ	1	ЛЕ
КАЧ	МОРФ	1	+
КР	МОРФ	1	+
ПРОИЗВ	СИН	1	СУЩ (актив)
ГХ	СЕМ	1	ПРИЛ: С_ОПР
СХ	СЕМ	1	ОЦЕНКА, ДЕЙСТВ
ВАЛ	СЕМ	1	ОЦЕНКА, С, А1
СХ1	СЕМ	1	ДЕЙСИВ
		2	ПРОЦЕС
		3	ОДЕЯТ
		4	ОДУШ
ГХ1	СЕМ	1	Х!: ИГ
ЛФ	ЛЕКС	1	SO: активность
ПО	ИНФ	1	общ
ВЕС	ИНФ	1	3
ВАР	ИНФ	1	действующий
		2	деятельный
АНГ	ЭКВ	1	active

4 结束语

20 世纪 90 年代以来, Н.Н.Леонтьева 主持开发的 РОСС 曾先后得到多项基金的资助, 其中包括俄罗斯人文科学基金 (Российский гуманитарный научный фонд, РГНФ) 和俄罗斯基础研究基金 (Российский Фонд фундаментальных исследований, РФФИ). 进入新千年后 Н.Н.Леонтьева 申请的科研课题“用于文本自动分析的俄语语义词典系统开发”(Развитие системы русских семантических словарей для автоматического анализа текстов) 起了一个颇有诗意的名称“РУСЛАН-1”, 获得了 2001-2002 年度俄罗斯人文科学基金的后续资助。(http://www.rfh.ru/) 根据 Н.Н.Леонтьева 自述, 其主要学术思

想集中体现在莫斯科国立大学出版社新近出版的三本书中。(Н.Н.Леонтьева 2000, 2001, 2002)

毋庸置疑,以 РОСС 为基础的俄语语义词典系统是一项重要的词汇语义资源工程,可视其为莫斯科语义学派大旗下一个具有相当竞争力的信息产品,我国俄语学和计算语言学研究均能从中获益匪浅。我们下一步工作是,更深入地挖掘和借鉴 РОСС 所提供的词汇语义资源,以尽快完善我们的在研项目“面向俄文信息处理的通用语义码模式”。

附注

①此次会议由北京大学计算语言学研究所、台湾中央研究院语言研究所和香港城市大学中文、双语及翻译学系共同主办。

参考文献

- [1] Апресян Ю.Д. 1967 Экспериментальное исследование семантики русского глагола [M]. М.
- [2] Апресян Ю.Д. 1986 Интегральное описание языка и толковый словарь [J]. ВЯ, №2.
- [3] Апресян Ю.Д. 1991 Об интегральном словаре русского языка [A]. Семиотика и информатика [C]. Вып.32. М.
- [4] Апресян Ю.Д. 1974/1995a Избранные труды. Том I: Лексическая семантика. Синонимические средства языка [M]. М.
- [5] Апресян Ю.Д. 1995b Избранные труды. Том II: Интегральное описание языка и системная лексикография [M]. М.
- [6] Иомдин Л.Л., Крысин Л.П. (Отв. ред.) 2000 Слово в тексте и в словаре: Сб. ст. к семидесятилетию академика Ю.Д. Апресяна [C]. М.
- [7] Караулов Ю.Н. 1994 Фундаментальные феномены русистики [J]. Русистика сегодня, №1.
- [8] Леонтьева Н.Н. 1997 Русский общесемантический словарь (РОСС): структура, наполнение [J]. НТИ, Сер.2, №12.
- [9] Леонтьева Н.Н. 2000 К теории автоматического понимания естественных текстов. Часть 1. Моделирование системы "мягкого понимания" текста: Информационно-лингвистическая модель [M]. М.
- [10] Леонтьева Н.Н. 2001 К теории автоматического понимания естественных текстов. Часть 2. Семантические ловушки: состав, структура, методика создания [M]. М.
- [11] Леонтьева Н.Н. 2002 Препринт 3: Первичный семантический анализ текста [M]. М.
- [12] Leontyeva N. N. 2003 Semantic Dictionary for Text Understanding and Summarization [J]. International Journal of Translation, Vol. 15, № 1. Bahri Publications.
- [13] Мельчук И.А. 1974/1999 Опыт теории лингвистических моделей «Смысл Текст» [M]. М.
- [14] Сокирко А.В. 1997 Программная реализация русского общесемантического словаря [J]. НТИ, Сер. 2, № 12.
- [15] ТКС 1976-1996 Толково-комбинаторный словарь [Z]. Апресян Ю.Д., Мельчук И.А. и др. Серия публикаций ПГЭПЛ [C]. Институт русского языка, М.
- [16] 华 劭 2003 语言经纬[M], 北京: 商务印书馆。
- [17] 易绵竹 2001 词汇语义学研究在俄罗斯 — 阿普列相现象简评[A]//第二届汉语词汇语义学研讨会论文摘要集[C]. 北京大学计算语言学研究所。
- [18] 张家骅 2001 莫斯科语义学派[J]. 外语研究, 第 4 期。

An Overview of Lexical Semantic Resources Engineering in Russia (1)

— A Parsing of «Russian General-purpose Semantic Dictionary»

Yi Mian-zhu¹, Wang Jing-wu², Nan Zhen-xing³, He Ying-yu⁴ Wu Bin⁵

(1. Center of Russian Language and Literature Studies of Heilongjiang University, Harbin 150080, China;

2. Graduate students' office of Training Department, PLA University of Foreign Languages, Luoyang

471003, China; 4. Department One, PLA University of Foreign Languages, Luoyang 471003, China; 1, 3,

5. Institute of Language and Culture Studies for National Defence, PLA University of Foreign Languages, Luoyang 471003, China)

Abstract: Lexical semantic studies and the construction of related resource projects have always caught the attention of Russian linguists, and as a result, many achievements have been reached in terms of both theoretical researches and their applications. A comprehensive review of the lexical semantic resource projects initiated by Russian linguists will no doubt shed light on the comparative studies on Chinese and Russian lexical semantics and the development of Russian-Chinese MT systems. Therefore, we are planning to introduce the most important lexical semantic resource projects in Russia in a series of papers, the first of which is a parsing and review of the Russian General-purpose Semantic Dictionary directed and developed by N. N. Leontyeva.

Key words: Moscow Semantic School; lexical semantic resources; Russian General-purpose Semantic Dictionary

收稿日期: 2003-08-11

第一作者简介: 易绵竹 (1964--), 男, 四川营山人, 解放军外国语学院教授, 博士生导师; 中国中文信息学会理事, 黑龙江大学俄语语言文学研究中心专职研究员。主要研究方向: 计算语言学、心理语言学。

[责任编辑: 惠秀梅]