

基于俄语国家语料库的同义词差异研究

孙敏庆

(黑龙江大学俄罗斯语言文学与文化研究中心, 哈尔滨 150080)

提 要: 语料库语言学的蓬勃发展为语言研究提供了新的视角和方法。本文借助俄语国家语料库, 通过对同义词在不同语境、不同时期的频次进行统计, 揭示其词频分布差异; 通过对同义词搭配关系提取、分析, 揭示其语义韵差异。研究表明, 基于语料库的研究对传统的定性分析是一种有效的补充, 拓宽了同义词的研究范围, 其研究成果有助于对同义词的全方位认识, 为外语教学及辞书编撰提供一定借鉴及启示。

关键词: 语料库; 同义词; 词频; 语义韵

中图分类号: H03

文献标识码: A

1 引言

俄语中存在着大量同义现象。“词的同义关系(синонимия)是同一词类的不同词在意义上全部或大部分重合的关系, 词与词在这种情况下可以相互替换而不改变语句的内容。具有同义关系的词叫做同义词(синонимы)。”(张家骅 2006: 72)这里的“意义重合”, 是指概念意义层面的重合, 可能体现在不同词的全部义位之间, 也可能体现在不同词的一个或若干个义位之间。根据词与词之间同义程度的大小, 可将同义词分为两类: 绝对同义词和相对同义词。前者在概念意义、搭配领域、表情色彩、语体属性、使用范围、使用频率等方面没有差别, 如 языкознание(语言学)/лингвистика(语言学), здесь(在这里)/тут(在这里); 后者在上述各个方面和某个方面存在着差异, 如 посещать(访问未知的客体)/навещать(访问已知的客体), аплодировать(鼓掌)/хлопать(<口>拍巴掌)。(张家骅 2006: 71, 76—78)事实上, 真正意义上的绝对同义词是极少的, 由于受到语言发展规律的影响, 绝对相同的符号即使存在, 也是暂时的。随着语言的不断发展, 要么区别特征逐渐显化, 要么语义仍然等值, 但一个成为高频的积极词汇, 另一个成为低频的消极词汇, 甚至退出语言系统。因此, 语言中更多的同义词都属于相对同义词, 而这部分词汇的辨析问题始终是外语教学中的一大难题。

传统的同义词辨析, 无论是研究者还是学习者。主要借助各类详解词典、辨析词典, 采用释义比较、语例体悟等方法, 把握同义词在概念意义、搭配结构、语体范围、表情色彩方面的差异。这样的辨析主要是基于规则的定性分析。但是, 同义词之间的差别, 有些是通过上述方法无法解决的。就详解词典而言, 不同词典所给出的词条信息并不完全一致, 且存在许多循环释义的情况, 查阅结果往往使得学习者更为混乱; 就辨析词典而言, 收录范围很难达到理想的程度。更为重要的是, 同义词之间的有些差异特征, 单纯基于规则的定性分析是无法发现的, 需要对大量语料进行量化分析后才能呈现出来。因此我们认为, 同义词辨析方法研究还有待深入。

随着语料库语言学的蓬勃发展,各种大规模标注语料库不断建立和完善,为语言学中的许多问题提供了新的研究方法和研究视角。正如国内语料库学家杨慧中所指出的,“语料库语言学为语言学研究提供了一种全新的研究思路,它以真实的语言数据为研究对象,从宏观的角度对大量的语言事实进行分析,从中寻找语言使用的规律。”(杨慧中 2002: 4)同义词研究,作为语言研究的一个重要课题,也在语料库语言学的推动下得到了进一步的拓展。目前国内英、汉语学界利用语料库在同义词比较方面取得了丰硕成果,主要在词频、类联接、搭配强度、语义韵等方面探讨了同义词之间的差异。英、汉语学界所取得的相关成果对俄语同义词研究具有重要的参考价值和启示意义,但是由于语言本身的特点,所使用的语料库在规模大小、标注体系、检索功能等均有差别,即使采用同样的研究视角和方法,具体的操作步骤也不尽一致,分析过程也会因语言而异。本文尝试以俄语国家语料库为基础,辅以语料检索软件 Antconc 3.2.1,对俄语同义词差异特征进行调查研究,以期对俄语同义词研究提供一个新的研究思路。

2 基于语料库的研究方法及优势

语料库(corpus)是指“按照一定的语言学原则,运用随机抽样方法,收集自然出现的连续的语言运用文本或话语片段而建成的具有一定容量的大型电子文库。”(杨慧中 2002: 33)概括来讲,语料库就是以电子文本形式存放语言材料的数据库,是代表自然语言使用情况的一个样本。

2.1 基于语料库的研究方法

利用语料库进行语言研究一般分为三种方法:语料库例证法(corpus-illustrated)、基于语料库法(corpus-based)和语料库驱动法(corpus-driven)。第一种方法,一般不涉及概率统计问题,先提出某一假设或理论,然后从语料库中提取适合例句对该假设或理论进行佐证,属于自上而下的方法,不是严格意义上的语料库语言学研究方法。后两种方法是基于概率的研究方法,通过对大量语言现象进行半自动或自动的统计量化、观察、概括、分析,提出观点,属于自下而上的方法,是语料库语言学的典型方法。后两种方法在依赖数据程度、具体操作以及研究任务上,又有所差别。基于语料库的方法,并不排斥研究者的语言学知识,同时主张使用“标注文本”,语料库的标注愈完善愈有利于语言研究,确切地说,这种研究是在利用传统知识的前提下,注重量化分析,通过所呈现数据挖掘语言隐在的现象,揭示一些单纯靠直觉所不能发现的规律。语料库驱动的研究方法则要求尽可能摆脱已有的语言描写体系和方法,偏重“干净文本”,完全通过数据源挖掘语言现象,得出结论。

本文采用第二种研究方法,即基于语料库的方法。作为语言研究的依托,语料库的选择至关重要。正如俄罗斯语料库语言学家 В.П. Захаров 和 С.Ю. Богданова 所指出的,“语料库的规模要足够大,才能保证数据的代表性和语言现象的覆盖率;语料库的文本类型要丰富,才能使研究全面、客观;所建数据库要能够服务于为研究者的各种语言研究目的,只有这样,语料库的创建才是合理的、基于语料库的语言研究才是有意义的。”(Захаров, Богданова 2011: 8)。目前,俄罗斯已经建成并投入使用的语料库有“俄语国家语料库”(Национальный корпус русского языка,简称 НКРЯ)、“图宾根语料库”(Тюбингенский корпус русского языка)、“20 世纪末俄语报纸文本计算机语料库”(Компьютерный корпус текстов русских газет конца XX века)三大语料库。其中,《俄语国家语料库》始建于 2003 年,目前上网词次已达 3 亿多,无论从语料规模、信息标注、文本类型以及检索功能看,都属世界领先行列,基于这样的语料库进行语言研究是可信的。

2.2 基于语料库的研究优势

作为一种方法,语料库在语言研究中的一个重要作用就是“可以为研究者提供更一般的、经验的语言数据,这些经验数据可以使语言学家作出的结论更客观”。(黄昌宁、李娟子,2002:

154) 其实, 关于语言现象统计数据对语言意义研究的积极作用, 俄罗斯语言学家 В.А. Москович 早在 1969 年就有过阐述: “统计数据可以揭示语言意义之间结构关系的某些本质特征, 只有考虑到这些数据才能实现对语义的充分描写。”(Москович, 1969: 5) 但是在电子文本语料库诞生之前, 研究者对各类语言现象的数据统计, 主要是先搜集相关语言现象, 并逐一记录在卡片上, 其整理、分类及分析过程也都是手工操作, 极其耗时费力。而随着大规模语料库的创建与完善、相关检索软件统计功能的日趋强大, 许多语言现象的量化信息可以借助计算机自动或半自动地提取, 使语言研究在深度、广度和效度上都有了新的突破。

基于语料库进行语言研究的最大优势在于语料真实、充足、均衡, 检索快捷, 提取出的索引行可以展示翔实、丰富的用法, 并提供有力数据, 使研究者能够在客观描写基础上做出更为准确的结论。以同义词辨析为例, 通过对语料库各项参数的设定, 获取相应的检索语料及统计数据, 这些信息能够反映语言系统以及语言使用的重要特征, 可以对同义词的使用频率、搭配强度、语义韵等方面的描写提供有力支撑, 这是对传统分析一种重要补充。

3 同义词的词频差异分析

词频 (word frequency) 是指 “各词形 (word form) 在文本中出现的频数, 是语料库分析中最基本的统计量之一, 可以通过分析软件自动提取”。(杨慧中 2002: 341) 同义词的词频差异既可能体现在语言系统的时间分布上, 即不同历史时期的使用频率不同, 也可能是语言系统的空间分布上, 即不同语域的使用频率不同。下面我们以两组同义词为例, 对其词频进行比较。

3.1 词频历时分布差异

词语在不同历史时期的词频分布变化能够反映语言发展的动态特征。通过对同义词的词频历时分布可以从一个侧面揭示出二者之间的差异。这一点在俄语外来词与固有词构成的同义词列中表现得尤为鲜明。外来词的借入势必会影响固有词在语言系统中的地位。一部分外来词由于借入程度还不深, 使用频率还很低, 还有一部分外来词则使用频率日趋提高, 甚至超过相应的固有词, 极大地撼动了固有词的地位, 大有取而代之的趋势。对于这些动态特征的掌握, 有助于我们更准确、更合理地选词。下面以同义词 **вывоз/экспорт** (**export**) (出口) 为例, 对同义词的使用频率进行历时考察。

具体操作分两步: 1) 依据历史分期, 建立子库: 将 1700—2000 年, 每 50 年划分为一段, 2000—2012 年为一段, 共计 7 个时间段, 分别建立子库。2) 分别在建好的各个子库的词汇—语法检索框中输入 **вывоз** 与 **экспорт** 进行检索, 记录出现频次。由于每个子库总词次不一致, 因此单纯的频次之间无可比性, 为科学统计, 我们需要人工辅助计算出每百万词频中检索词出现的频次, 即频率。统计结果如下:

表 2 **вывоз/экспорт** 的词频历时分布

子库 (按年代划分)	子库词次	вывоз		экспорт	
		频次	频率	频次	频率
1700 — 1749 年	541431	14	26	0	0
1750 — 1799 年	2779147	33	19	0	0
1800 — 1849 年	9512684	39	4	0	0
1850 — 1899 年	36605808	135	4	9	0.25
1900 — 1949 年	57901364	390	7	167	3
1950 — 1999 年	65263610	270	4	584	9
2000 — 2012 年	56858446	610	11	2549	49

从表 2 可知, 固有词 **вывоз** 在 1700—1749 年间已广泛使用, 但使用频率在历时分布上总体呈递减趋势。外来词 **экспорт** 是在 1850—1899 年间出现在俄语中, 此后与固有词 **вывоз**

并行使用,且使用频率逐渐增大,至 21 世纪初其使用频率突增,频率升为 49,显著高于 **вывоз** 的使用频率,大有取代 **вывоз** 的趋势。研究表明,尽管固有词 **вывоз** 与外来词 **экспорт** 的全部意义成分相同,在表情色彩、语体属性、适用范围方面也无差别,但 **экспорт** 的使用频率却高于 **вывоз**,且呈上升趋势,为使表达更地道,选词时建议优先使用 **экспорт**。基于语料库的量化研究,为我们做出更为准确结论提供了有力的客观依据,这往往是传统词语辨析方法无法企及的。

3.2 词频语域分布差异

语域 (register) 是人们在实际的语言活动中,出于交际的需要,或因其所从事的职业和兴趣相异,亦或因其话语发生的情景、说话的对象、地点和话题的不同而产生的一种言语变体,体现为语言中的不同语体风格、用语格调等。(张继东,刘萍 2005: 49)“同义词之间除了有定性差异外,在使用频率以及不同语域词频分布上都有明显的差异,这种差异也可成为判断的依据之一,而且更适合作判断的前提,判断应优先考虑哪一个”。(潘璠,冯跃进 2000: 8)借助俄语国家语料库自有的检索功能,可以实现对同义词词频语域分布的量化统计。我们以俄语同义词 **глаза/очи/гляделки** (眼睛) 为例,分析三者在词频语域分布上的差异。

具体操作分两步: 1) 依据语域类别,分别自建子库。在自建子库 (задать подкорпус) 页面,分别选取基础库 (основной корпус) 中的文学库及非文学库的 8 个功能领域 (сфера функционирования),统计时间均设定为默认值,即 1700 年到 2012 年,建成含有 9 个语域类别的书面语子库,与俄语国家语料库中的独立子库“口语库” (устный корпус),构成存有不同语域的 10 个子库。2) 分别在每个子库的词汇—语法检索框内输入关键词 **глаз, око, гляделка** 进行检索,可以得到三个词在各个子库中出现的频次,由于各子库总词次不同,因此如同前文,需要计算出频率 (每百万频次中检索词出现的频次)。统计结果如下:

表 1 **глаза/очи/гляделки** 的词频语域分布

子库 (按语域划分)			子库词次	глаза		очи		гляделки	
				频次	频率	频次	频率	频次	频率
书面语	文学		101800661	181267	1780	4218	41	9	0.09
	非文学	政论	85441370	42999	503	1342	16	1	0.01
		公文事务	3942577	177	45	0	0	0	0
		生产技术	1367781	179	131	0	0	0	0
		教学科研	27992504	7651	273	740	26	0	0
		教会神学	3218238	928	288	782	243	0	0
		广告	604693	121	200	0	0	0	0
		日常生活	4849097	1627	335	82	17	0	0
		电子交际	2860072	1218	426	0	0	0	0
口语		10361479	3 343	323	95	9	1	0.01	

对每个词在 10 个语域中的频率进行纵向比较可知, **глаза** 在文学作品中的频率最高,每百万词次中出现 1780 次,在公文事务语篇中频率最低,每百万词次中仅有 45 例,而在其他语域中,均为几百例,差异不明显。而 **очи** 则在教会神学著述中有很高的使用频率,每百万词次中有 243 例,在文学作品、口语、政论语篇中也会出现,但在公文事务、生产技术、广告、电子交际基本不出现; **гляделки** 的出现频次极少,仅出现在文学语篇、政论语篇及口语中,且频率均不到 1。

对各语域中 3 个词的使用频率进行横向比较,除教会神学语篇外,其他语域中 глаза 的使用频率均远高于 очи 和 гляделки。如书面语中,每百万词次中 глаза 出现 1780 例,而 очи 仅为 41 例,гляделки 则为 0.09 例;政论语篇中,每百万词次中 глаза 出现 503 例,而 очи 仅为 16 例,гляделки 为 0.01 例。在教会神学语篇中,每百万词次中 глаза 出现 288 例,而 очи 为 243 例,频率差异无几,гляделки 未出现。

俄语详解词典对三个词的差异描写主要体现在修辞色彩上,对 очи 的标注为<旧、雅或诗>,对 гляделки 的标注为<俗、方>。量化数据的引入,使我们发现,глаза 使用语域广泛,且使用频率很高,属于典型的积极词汇,而 очи 的语域有限,在教会神学作品中使用频率很高,гляделки 则每百万词次中出现不到 1 例,属于典型的消极词汇。可见,三个词的词频语域分布差异信息从一个新的视角审视三者之间的差别,使往往被忽略的信息得到挖掘,使认识更为全面。

4 同义词的差异分析

语义韵或称语义韵律 (semantic prosody),作为语料库语言学研究的一个重要术语,从新的维度反映了词语搭配特点。可以说,揭示同义词的语义韵差异是辨析它们的一个有效途径。

4.1 术语阐释

“语义韵”这一术语是由 John Sinclair 于 20 世纪 80 年代引入到英语词汇研究中。Sinclair 发现,词语具有在特定语义环境反复出现的趋势,如英语动词 happen 通常和指称不愉快事件的名词搭配。此后,许多学者借助语料库对词的搭配行为进行了更为深入的观察、分析,推进了语义韵的研究。他们指出,有些词项的搭配行为显示出一种特殊的趋向:它们习惯地吸引某一类具有相同或相似语义特点的词项,与之构成搭配。由于这些具有相同或相似语义特点的词项习惯地、反复地与节点词在文本中共现,其语义相互感染、相互渗透,节点词也就被“传染”上了有关的语义特点,它的整个语境内也就形成一种特殊的语义氛围。(Sinclair 1991: 74—75; Louw 1993: 156—159; 杨慧中 2002: 266) 这种语言现象即为语义韵。相较而言,俄罗斯学者 В.Б. Кашкин, К.М. Шилихина 对语义韵的定义,即“词语在语篇中同积极联想意义或消极联想意义词汇搭配的趋势被称为语义韵”(Кашкин, Шилихина 2010: 26),虽然更为简练,但却也道出了语义韵的实质。

M. Stubbs 将语义韵划分为三大类:消极语义韵 (negative prosody)、积极语义韵 (positive prosody)、中性语义韵 (neutral prosody) (Stubbs 1996: 76)。与传统语言学中修辞色彩中的褒义、贬义、中性义有所不同,后者是针对词语本身的属性而言,而语义韵是针对某词与其搭配词之间的关系而言,揭示的是搭配关系的语义倾向性。具体而言,消极语义韵表现为节点词所吸引的词语绝大多数都具有鲜明的消极语义特点,使整个语境弥漫着一种强烈的消极语义氛围;积极语义韵的情况则与之相反,吸引的几乎都是具有积极语义特点的词,由此形成一种积极的语义氛围;中性语义韵又称为错综语义韵,与节点词搭配的词既有消极涵义的、也有积极涵义或中性涵义的,所占比例无显著差异,因而形成一种错综混合的语义氛围。语言中大多数的词的搭配行为会呈现出中性语义韵,但是对呈现消极和积极语义韵的词的搭配行为,一般的外语学习者是无法直接感知的,对此可以借助语料库进行确定。

与语义韵研究密切相关的术语有节点词、类联接、跨距、搭配词。节点词 (node word) 就是指研究人员要在语料库中分析其搭配行为的关键词,通常为输入语料库的检索词。类联接 (colligation) 就是词语搭配出现于其中的语法结构和框架。如 N+V 就是一个类联接。跨距 (span) 是指节点词左右的语境。如跨距定为-4/+4,表示的是在节点词左右各取 4 个词为其语境。所有落入跨距的词项都视作节点词的搭配词 (collocates) 或共现词。(杨慧中 2002: 336—338) 按所处距位,搭配词又可分为左搭配词和右搭配词。搭配词与节点词所构成的搭

配有些有语法限制关系,有些无语法限制关系。根据不同的研究目的,可自行设定跨距范围,以及通过建立类联接等手段限定搭配词与节点词的关系类型。“节点词与搭配词之间的逻辑关系是语义韵研究的重要依据。”(Partington2004: 154—155, 转引自陆军 2010: 277)也就是说,在语义韵特征分析之前,对节点词与搭配词之间的语义、语法、位置等关系的限定是必要的环节。

4.2 操作方法

卫乃兴曾提出语义韵的三种研究方法:1)建立并参照类联接,用基于数据的方法(data-based approach)研究;2)计算节点词的搭配词,用数据驱动的方法(data-driven approach)研究;3)用基于数据与数据驱动相结合的折中方法研究。(卫乃兴 2002: 300)本文将采用第一种方法,以同义词 **вызвать/пробудить**(引起、产生)为例,尝试利用俄语国家语料库对同义词语义韵进行分析。

具体操作为:1)在基础库页面,建立类联接。通过查阅词典,可知两词在用于“引起、产生”之意时,均支配非动物名词第四格形式,因此,在基础库检索页面,分别输入关键词的同时,将其支配的词语法特征中的“**существительные**”,“**винительные**”,“**неодушевленное**”项勾选,这样,得到相应的类联接即为“**V+N**”,且N被限定为非动物名词第四格形式。为便于数据提取,我们将“**расстояние**”项设置为“**от 1 до 1**”,即要求节点词后边的第一个名词要符合上述语法限制条件;2)提取索引语料,进行数据统计:将检索到的语例,分别以 Excel 格式下载(得到的索引行都是以节点词为中心排列,一大好处就是可以快速锁定节点词),观察索引行,人工剔除无效数据:以 **вызвать** 为例,剔除如含有 **вызвать помощь**, **вызвать полицию** 等索引行。3)借助检索软件 Antconc 3.2.1 对搭配词进行统计:将语例从 Excel 格式转换成 TXT 文本形式,并导入 Antconc3.2.1 中,在 search term 中勾选 words 框,输入相应的检索词,同时将右跨距设置为 1,左跨距为 0,点击 collocates 和 start,便可自动统计出动词所支配名词的频次。

4.3 语义韵比较

鉴于统计方便,我们分别从索引行中随机抽取 200 例作为分析对象。

4.3.1 **вызвать** 的语义韵

通过对 200 个含有 **вызвать** 的索引行进行观察,统计其所支配名词有 114 个,按语义类别划分为以下 5 类:

1)表示冲突、疾病、灾害、困难类: **протест**(抗议), **арест**(查封), **вспышка**(爆发), **взрыв**(爆发), **тревога**(骚动), **спор**(争论), **приступ**(迸发), **шум**(喧哗), **возражение**(反对), **неприятие**(拒绝), **обострение**(激化), **разногласие**(分歧), **развал**(瓦解), **раскол**(分裂), **нарекание**(责备), **критика**(批评), **колебание**(动摇), **отказ**(拒绝), **отповедь**(驳斥), **несоединимость**(排斥性), **нападение**(攻击), **обеспокоенность**(妨碍), **ограничение**(制约性); **аллергия**(过敏), **заболевание**(疾病), **пандемия**(流行病), **эпидемия**(流行病), **осложнение**(并发症), **слюноотделение**(流涎症), **симптомы**(症状), **рвота**(呕吐), **выкидыш**(流产), **ожог**(烧伤), **остеохондроз**(截骨术), **шок**(休克); **пожар**(火灾), **землетрясение**(地震), **шквал**(暴风), **коррозия**(腐蚀), **нарушение**(破坏), **катастрофа**(灾难), **обвал**(塌陷); **обвисание**(下坠), **буря**(爆发), **затруднение**(困难), **вопрос**(问题), **проблема**(问题), **тщетность**(无效), **замечание**(责备), **дефект**(故障), **поражение**(失败), **инфляция**(通货膨胀), **пауза**(停顿), **отсутствие**(缺乏), **падение**(下跌)。搭配名词为 55 个,合计 90 例。

2)表示负面情感、状态类: **беспокойство**(不安), **недовольство**(不满), **недоумение**(困惑), **раздражение**(气愤), **возмущение**(愤怒), **удивление**(惊讶), **страх**(恐惧), **подозрение**(猜忌), **подозрительность**(多疑), **сомнение**(怀疑), **паника**(恐慌), **сочувствие**

(同情), ярость (暴怒), фурор (狂怒), смятение (焦急), фибрилляция (颤动), волнение (激动), зависть (妒忌), ненависть (憎恨), недоверие (不信任), напряжение (紧张), неудовольствие (不满), негодование (愤慨), гнев (愤怒), галлюцинация (错觉)。搭配名词为 25 个, 合计 51 例。

3) 表示正面情感、状态类: интерес (兴趣), восторг (兴奋), улыбка (微笑), желание (愿望), восхищение (赞赏), энтузиазм (热情), оптимизм (乐观), почтение (敬重)。搭配名词为 8 个, 合计 17 例。

4) 表示数量变化类: рост (增长), сокращение (缩减), повышение (提高), снижение (下降), увеличение (扩大), уменьшение (减少), подъем (上升), тенденция (趋势), переход (转折), изменение (改变), потепление (变暖)。搭配名词为 11 个, 合计 16 例。

5) 其他: необходимость (必要性), появление (出现), эффект (效果), отклик (反响), чувство (情感), реакция (反应), явление (现象), ощущение (感觉), подражание (模仿), резонанс (反响), действие (行为), дискуссия (讨论), мутация (运动), перераспределение (重新分配); потребность (需求)。搭配名词为 15 个, 合计 26 例。

不难发现, вызвать 所支配的词语中, 1) 类和 2) 类词汇都含有明显的消极、否定意味, 共计 80 词, 约占总词数的 70%, 出现 141 例, 约占总例数 71%。尽管有些看似不具有积极或消极涵义词汇, 如 4) 和 5) 类, 但是通过对下文的进一步分析, 便会发现它所传递的信息大部分都是具有消极意味的: масса нареканий (诸多责难), подъём заболеваемости (发病率上升), повышение артериального давления (动脉压升高), ряд критических замечаний (一系列批评意见), эффект «бутылочного горлышка» (“瓶颈”效应) 等。相较而言, 与第 3) 类表示正面情感、状态类的词汇搭配的数量就非常有限, 共计 8 词, 约占总词数的 7%, 出现 17 例, 约占总例数 9%。综上, 我们认为 вызвать 用于“引起”意义时, 出现的大多为不良现象, 因此, 其语义韵律为消极的。

4.3.2 пробудить 的语义韵

通过对 200 个含有 пробудить 的索引行进行观察, 统计其所支配名词有 83 个, 按语义类别划分为以下 4 类:

1) 表示冲突、疾病、灾害、困难类: война (战争), хаос (混乱); боль (疼痛); вопрос (问题), гарь (烧焦); удар (打击)。搭配名词为 6 个, 合计 9 例。

2) 表示负面情感、状态类: жалость (惋惜), ярость (狂怒), угрызение (内疚), тоску (苦恼), суровости (严厉), стыд (羞愧), страх (恐惧), страдание (痛苦), ревность (醋意), раскаяние (后悔), подозрения (猜疑), ненависть (愤恨), презрение (鄙视)。搭配名词为 13 个, 合计 19 例。

3) 表示正面情感、状态类, 如 интерес (兴趣), активность (积极性), вера (信心), вдохновение (灵感), влюбленность (爱恋), сочувствие (同情), воля (意志), гордость (自豪感), любопытство (好奇心), любовь (爱情), мысль (思想), надежда (希望), совесть (良心), сознание (意识), стремление (渴望), страсть (激情), силы (力量), симпатия (好感), самолюбие (自尊心), склонность (爱好), смелость (勇气), самостоятельность (主动性), аппетит (胃口), говорливость (滔滔不绝), уважение (敬意), преданность (忠诚), улыбка (微笑), радость (快乐), искренность (真诚), мужество (勇气), разум (理智), любознательность (求知欲), чистота (纯真), честолюбие (虚荣心), жажда (渴望), дух (勇气), доверие (信任), инициатива (首创性), честь (敬意), энтузиазм (热情), расположение (好感), желание (愿望)。搭配名词为 42 个, 合计 125 例。

4) 其他: внимание (注意), воображение (想象), воспоминание (回忆), деятельность (活动), чувство (感情), явь (现实), эхо (回声), чтение (阅读), состояние (状态), создание (创建), прошедшее (过去), интуиция (直觉), инстинкты (本能), звук (声

音), звон (音响), замысел (构思), образ (形象), ощущение (感觉), память (记忆), потребность (需求), правосознание (法制观念), движение (运动)。搭配名词为 22 个, 合计 47 例。

пробудить 所支配的词语中, 1) 类和 2) 类词汇都含有明显的消极、否定意味, 共计 19 词, 约占总词数的 23 %, 出现 28 例, 约占总例数 14%。3) 类表示正面情感、状态类的词汇搭配的数量共计 42 词, 约占总词数的 51 %, 出现 125 例, 约占总例数 63%。4) 类基本为中性涵义的词汇。由此, 我们认为 пробудить 用于“引起”意义时, 大多支配表情感类词汇, 其中积极情感类占大多数, 因此, 其偏于积极语义韵。

通过对两个动词的搭配词进行分析, 我们揭示出了二者的语义韵特征, вызвать 主要和具有消极意义的词汇搭配, 包括冲突、疾病、灾害、困难、负面情感及状态类。相较而言, пробудить 即使也会与表消极、中性涵义的词汇搭配, 但更多的是与表示正面情感、状态类词汇搭配。

“语义韵律沉淀在人们的语言运用习惯中, 本族人用母语说话大多会自然而然地按照语义韵律的要求, 当然也会有违背韵律的现象。由于缺乏母语一样的语感, 语义韵律问题对于外语学习和使用来说更值得注意。”(王泽鹏, 张燕春 2005: 91) 基于语料库研究同义词, 使过去仅靠直觉和少量例证较难发现的语义韵呈现出来, 这对缺少语言直觉的外语学习者在同义词掌握上, 是极具指导价值的。

5 结束语

本文借助俄语国家语料库对同义词的频率及语义韵进行了实证研究。研究表明, 语料库以其自有的优势, 使得其成为进行同义词差异比较的一种有效工具。语料库丰富的资源, 可提供详实的证据, 使研究者能够在对关键词的搭配情况和特点客观、扎实描述基础上, 通过引入量化数据, 做出单纯依靠直觉和少数例句无法得出的结论。需要注意的是, 基于语料库的词语辨析方法并不是要取代传统的辨析方法, 而是对后者的有益的和必要的补充, 从而推动词语辨析研究朝着更科学、更客观方向发展。我们相信, 随着语料库的日臻完善, 语言研究也会更加深入、细致、全面。

参考文献

- [1] Захаров В.П., Богданова С.Ю. Корпусная лингвистика [М]. Иркутск: ИГЛУ, 2011.
- [2] Кашкин В.Б., Шилихина К.М. Метапрагматическая рефлексия по данным национального корпуса русского языка [А]//Обыденное метаязыковое сознание: онтологические и гносеологические аспекты[С]. отв. ред. Голев Н. Д. Часть 3. Кемерово: Изд-во Кемеровского гос. ун-та, 2010.
- [3] Москович В.А. Статистика и семантика[М]. Москва: Наука, 1969.
- [4] Louw B. Irony in the Text or Insincerity in the Writer? The diagnostic potential of semantic prosodies[A]// Baker M., Francis G. & Tognini-Bonelli E. (eds.) Text and Technology: in honor of John Sinclair [C]. Amsterdam: John Benjamins, 1993.
- [5] Sinclair J. Corpus, Concordance, Collocation[M]. Oxford: Oxford University Press, 1991.
- [6] Stubbs M. Text and Corpus Analysis [M]. Oxford: Blackwell Publishers, 1996.
- [7] 黄昌宁, 李娟子. 语料库语言学[M]. 北京: 商务印书馆, 2002.
- [8] 陆 军. 基于语料库的学习者英语近义词搭配行为与语义韵研究[J]. 现代外语, 2010(3).
- [9] 潘 璠, 冯跃进. 基于语料库的同义词差异性特征调查[J]. 山东外语教学, 2000(4).
- [10] 王泽鹏, 张燕春. 语义韵律理论[J]. 同济大学学报 (社会科学版), 2005(4).
- [11] 卫乃兴. 语义韵研究的一般方法[J]. 外语教学与研究, 2002(4).

- [12]杨慧中. 语料库语言学导论[M]. 上海：上海外语教育出版社，2002.
- [13]张继东，刘 萍. 基于语料库同义词辨析的一般方法[J]. 解放军外国语学院学报，2005(6).
- [14]张家骅. 新时代俄语通论[M]. 北京：商务印书馆，2006.

Russian National Corpus-Based Research on Synonym Differences

SUN Min-qing

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

Abstract: The rapid development of corpus linguistics provides a new perspective and method for language study. By using Russian national corpus, frequency statistics based on synonyms in different language domain and different periods can reveal the distribution of words. Through collocation extraction and analysis of synonyms, semantic prosody differences can be revealed. Studies have shown that research based on corpus is an effective supplement for the traditional qualitative analysis, broadening the research scope of synonyms. The research results help to understand synonyms comprehensively and provide reference and inspiration for foreign language teaching and dictionaries compilation.

Key words: corpus; synonyms; frequency; semantic prosody

基金项目：本文系黑龙江省高等学校哲学社会科学学术创新团队“俄语语言学创新研究”(TD201201)的阶段性成果。

作者简介：孙敏庆(1985—)，女，黑龙江大庆人，黑龙江大学俄罗斯语言文学与文化研究中心在读博士生。主要研究方向：语义学、语用学。

收稿日期：2013-06-03

[责任编辑：叶其松]