文章编号：1003-0077(2010)06-0003-07

基于动词的汉语复合名词短语释义研究

王 萌[1,2]，黄居仁[2]，俞士汶[1]，李 斌[3]

（1. 北京大学 计算语言学教育部重点实验室，北京 100871；
2. 香港理工大学 中文及双语学系，香港；
3. 南京师范大学 文学院，江苏 南京 210097）

摘 要：复合名词短语的语义解释的主要目的是恢复修饰语和中心词之间隐含的语义关系。该文针对汉语复合名词短语的语义解释，首次采用动态的策略，提出了“基于动词的短语释义”的方法，利用语料库及 Web 数据，自动获取复合名词短语的释义短语，实验结果表明，该方法不仅可以为复合名词短语提供多种可能的语义解释，而且能够反应相似的复合名词短语之间细微的语义差别。此外，该文的研究结果可以服务于问答系统、信息检索、词典编纂等多个应用领域。

关键词：汉语复合名词短语；语义解释；释义短语；释义动词

中图分类号：TP391 文献标识码：A

Chinese Noun Compound Interpretation Based on Paraphrasing Verbs

WANG Meng[1,2], HUANG Chu-ren[2], YU Shiwen[1], LI Bin[3]

（1. Key Laboratory of Computational Linguistics, Peking University, Beijing 100871, China;
2. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China;
3. School of Chinese Language and Literature, Nanjing Normal University, Nanjing, Jiangsu 210097, China）

Abstract: Noun compound interpretation is to recover the implicit semantic relation between the head and modifier. In this paper, we present a dynamic approach to use paraphrasing verbs to interpret the meaning of Chinese noun compounds automatically for the first time in the literature. The experimental results show that this approach not only provides the possible interpretations for one noun compound, but also reflects the subtle semantic differences of similar noun compounds. In addition, our research can be applied in some other fields such as question answering, information retrieval and lexicography.

Key words: Chinese noun compounds; interpretation; paraphrase; paraphrasing verbs

1 引言

复合名词短语（noun compounds）是一种特定类型的短语，它由相邻的名词序列组成，其功能整体上相当于一个名词[1]，如“电子 警察”、“电脑 公司”和“空气 质量 问题”。通常把复合名词短语中的最后一个名词称为中心词（head），前面的成分称为修饰词（modifier）。从语法角度来说，复合名词短语和词比较相似，整个复合名词短语的功能相当于中心名词的功能。

复合名词短语广泛存在于各种语言，经常出现在各种文体中，衍生能力很强，组成方式简单但是歧义性高。这些特点使得复合名词短语在语言学和计算语言学领域成为一个热点研究课题，其涉及到的研究范围也越来越广泛，包括复合名词短语的自动获取、句法分析、语义解释、翻译以及语义焦点分析等等。

收稿日期：2010-05-06 定稿日期：2010-07-12

基金项目：国家 973 计划资助项目（2004CB318102）；国家社科基金资助项目（07BYY050）

作者简介：王萌（1977—），女，博士生，主要研究方向为计算语言学；黄居仁（1958—），男，讲座教授，主要研究方向为词汇语义学，知识本体，中文语言资源；俞士汶（1938—），男，教授，主要研究方向为计算语言学。

复合名词短语的语义解释(noun compound interpretation)的主要目的是自动获取修饰语和中心词之间隐含的语义关系。这种语义信息的显性化对信息检索、问答系统、机器翻译等诸多自然语言处理任务有所帮助。例如,在问答系统中,若用户的问题是"the causes of headaches",如果已知"caffeine headache"的语义解释是"headache caused by caffeine",就可以给出正确的回答。再如,在信息检索系统中,用户输入查询"headache pill",系统可以提供"pill causes the headache"或者"pill prevents the headache"等不同的语义解释来帮助用户改进查询。

本文主要研究了汉语复合名词短语的语义解释,首次采用动态的策略,提出了"基于动词的短语释义"的方法,自动获取复合名词短语的语义解释。本文组织结构如下:第2节介绍国内外相关研究,第3节介绍汉语复合名词短语释义方法的步骤,第4节至第6节对该方法中的每一步进行详细描述,第7节报告了评价方法及实验结果,最后,第8节对本文工作和下一步研究方向进行总结和展望。

2 国内外相关工作

目前,英文复合名词短语的语义解释研究得较为充分。总的说来,主要有两大处理策略,一种是自上而下的策略(top-down strategy),这种方法首先要求有一组已经定义好的、明确的关系集合,然后根据这个关系集合,为每个复合名词短语分配适当的语义关系,这实际上就是一个分类问题,也有文献将这种方法称之为"清单为本法(Inventory-based method)"。

不同的研究者提出的语义关系集合各不相同,文献[2]最早提出了九种"可恢复的删除谓词(recoverably deleteable predicates)",它们表示的语义关系可以用介词短语或者关系从句的方式予以表达,如CAUSE(exam anxiety),HAVE(vegetable soup),MAKE(electricity station),USE(laser printer),BE(player coach),FOR(concert hall),IN(morning class),FROM(peanut butter),ABOUT(computer expert)。文献[3]提出了一个语义关系的分类体系,上层有6个主要的语义关系类(CONSTITUTE,POSSESSION,LOCATION,PURPOSE,ACTIVITY-ACTOR,RESEMBLANCE),每个类下面包含若干子类。文献[4]定义了13种语义关系,提出了判断复合名词短语语义关系的准则,即通过wh-questions(who,what,when,whom,where,whose,how)对复合名词短语进行提问,根据修饰语能否回答这些问题对短语进行语义归类。文献[5]用8个介词(of,for,in,at,on,from,with,about)来定义语义关系,如baby car(car for the baby)。2007年SemEval组织了一项评测"Classification of Sematic Relations between Nominals"[6],定义了七种语义关系(Cause-Effect,Content-Container,Instrument-Agency,Origin-Entity,Part-Whole,Product-Producer,Theme-Tool)。

第二种是自下而上的策略(bottom-up strategy),研究者认为第一种方法存在一些缺陷:首先复合名词短语存在的语义关系是不能由一组固定的集合穷尽的[1],无论根据何种关系定义,总存在一些短语不能被正确归类;其次,固定的关系集合难以反映复合名词短语的多义性;最后,一个复合名词短语根据不同的解释可以属于多个语义类,如"lab printer",按照介词的语义关系分类,即可以是"printer in the lab",也可以是"printer for the lab"。因此研究者采用一种非受限的、开放式的方法,不事先定义语义关系集合,而是通过大规模的语料去发现词语组合时隐含的语义关系,并通过某种模式进行释义(paraphrase)。

在这种思路下,很多研究者尝试用动词来解释复合名词短语的语义关系[7-10],寻找能够连接中心词和修饰词的"事件框架(event frame)"。例如,"butter knife"和"kitchen knife",它们的语义解释分别为"knife for cutting the butter"和"knife used in the kitchen",其中"cut"和"use"就是释义动词。2010年的SemEval有一项英文的评测任务"Noun Compound Interpretation Using Paraphrasing Verbs and Prepositions"[11],要求参赛者为每一个复合名词短语提供释义动词集合,同时给出这些动词的排名。

在汉语方面,相关研究成果较少。文献[12]研究了具有名物化现象(nominalization)的汉语复合名词短语的语义分类问题,如"鸟类 迁徙","迁徙"是名动词(具有名词功能的动词),作者参照动词的语义角色(semantic roles)定义了四种粗粒度语义关系(Proto-Agent,Proto-Patient,Range和Manner),对300个复合名词短语进行了实验。该方法是属于第一种策略的,按照定义好的语义类别对复合名词短语进行分类,迄今为止,还未见汉语中采用第二种策略处理复合名词短语语义关系的相关报

道。因此本文则尝试第二种策略, 利用语料库及 Web 数据, 自动获取"基于动词的释义短语"对复合名词短语进行语义解释。

3　汉语复合名词短语的释义方法

文献[13] 对汉语谓词隐含(implying predicate)进行过详细论述, 从句法和心理实现性等方面对谓词隐含现象进行了验证。动词在复合名词短语的语义解释中起着相当关键的作用, 对于一个复合名词短语"n1 n2", 人们首先根据 n1 和 n2 之间的语义联系去激活被隐含的动词, 进而获得正确的语义解释。在一定的语境中, 这个隐含的动词是可以复原的。例如, "红木 家具", 解释为"红木 制造 的 家具", 其中"制造"就是隐含谓词。"爱情 故事", 解释为"描写 爱情 的 故事", "描写"是隐含谓词。隐含谓词可以有多个, 例如, "水果 价格", 可以解释为"买/卖/销售 水果 的 价格"等。我们将上述包含动词以及目标名词的短语称之为"基于动词的释义短语", 本文的目的就是自动发现这些释义短语, 并按照释义的可能性给出排名, 即越恰当的释义排名越靠前。该过程分为以下三步:

(1) 动词获取。对一个复合名词短语"n1 n2", 找到与 n1 和 n2 概念相关的动词。

(2) 释义短语生成。将 n1、n2 以及第一步中获取的动词放入已定义的释义模板中, 生成所有可能的释义短语。

(3) 释义短语过滤。将第二步中生成的模板作为查询(query), 送入搜索引擎, 得到命中次数, 并按照命中次数的降序对释义短语排序。

4　动词获取

名词通常指涉概念, 概念有不同的特征, 名词与名词组合构成复合名词短语时, 某一方面的特征会凸显出来。例如, "钻石"作为一种坚硬的材质可以被切割和打磨, 并可以镶嵌在戒指上作为装饰物, "钻石 锯片"和"钻石 戒指"分别凸显了"钻石"的两种不同特征, 而这种特征可以通过不同的"动词"进行解释, 即"切割 钻石 的 锯片"和"镶嵌 钻石 的 戒指"。因此, 动词获取的主要目的就是获取所有可能的与名词概念相关的动词, 在两个名词组合时, 与被凸显特征相关的"动词"就是合理的语义解释。

在自然语言中, 任何形式和结构都是为了表达一定的意义, 而任何意义及其关联都要通过一定的形式和结构表现出来。从句法层次上看, 连接动词与名词的最为直接的语法关系就是"述宾(verb-object)"和"主谓(subject-verb)", 即名词充当动词主语或者宾语, 名词与动词之间存在语义关联。因此, 本文从形式上可以把握的线索——表层的句法结构入手, 利用"述宾"和"主谓"两种语法关系, 获取与名词概念相关的动词。但是, 这就要求语料是经过深层加工并标记了句法结构的, 而目前可以利用的中文短语结构树库资源十分有限, 这会直接影响获取动词的数量(覆盖率)。因此本文采用一种回退的策略, 获取与名词在指定语法关系下具有搭配意义的动词, 并不要求语料经过深层次的句法加工和标注。中文词汇特征素描系统(Chinese Sketch Engine)[①]即可以胜任此项任务。

4.1　中文词汇特征速描系统

Sketch Engine 是一个大规模语料处理系统[14-15], 该系统除了提供一般的关键词及语境查询外, 还提供了词汇特征素描(word sketch)、语法关系(grammatical relation)以及同近义词分析(thesaurus)等自动产生的语法知识。目前这个系统已经应用在英语、汉语、法语、德语、日语等多国语言, 产生了广泛的影响。中文词汇特征素描系统(CSE, Chinese Sketch Engine)是 Sketch Engine 系统与十四亿字的 Chinese Gigaword 语料相结合的产物[16], 提供了绝大部分中文词汇实际使用的描述, 可以服务于诸多自然语言处理任务。

Word Sketch 描述了词语在某些语法关系下与其他词语的搭配情况。根据词类的不同, 其对应搭配词的语法关系也不同。例如, CSE 中名词的搭配关系有述宾关系(object_of)、主谓关系(subject_of)、领属关系(possession/possessor)、修饰关系(A_modifier/ N_ modifier/modifies)及并列关系(and/or)等 9 种。所有的搭配关系可以用一个三元组(triple)表示, 即(word1, relation, word2), 其中 word1 是查询的关键词, relation 是语法关系, word2 是在这种语法关系下的搭配词。

4.2　获取步骤

利用 CSE 中的 Word Sketch 功能可以方便地

① 该系统是一个网络在线系统, 访问地址: http://word-sketch.ling.sinica.edu.tw/.

获取某个名词的在各种语法关系下的特征素描，本文只使用“subject_of”和“object_of”两种关系。以复合名词短语“n1 n2”为例，经过两步获取释义动词。

第一步，将n1和n2作为查询关键词，分别获取它们在“subject_of”和“object_of”两种语法关系下的搭配词，本文只为每个名词抽取前200个显著性最高的搭配词，这样分别得到名词n1和n2的相关动词集合，记为VerbSetn1和VerbSetn2。

第二步，求VerbSetn1和VerbSetn2的交集，得到名词n1和n2共有的动词，作为最终的释义动词获取结果。

以“爱情 故事”为例，表1给出了两个名词在subject_of和object_of语法关系下的搭配动词样例，以及它们求交集的结果。表2给出了其他两个复合名词短语“水果 价格、网球 场地”的释义动词获取样例，同时给出了获取动词的个数。

表1 “爱情 故事”的释义动词获取过程

爱情（VerbSetn1）	背叛 追求 象征 遇见 追寻 表达 描述 描写 相信 昭告 见证 演绎 万岁 冲昏 酿 加温 滋润 不渝 献给 挡 追求 降 超越
故事（VerbSetn2）	讲述 诉说 讲 叙述 取材 描述 演绎 述说 冒险 编造 描写 听 编 流传 讲起 改编 搬上 传遍 精选 创作 浓缩 说明 娓娓道來 铺陈 讲完 重演 编排
爱情 故事（集合交集）	叙述 围绕 有关 经历 表示 发生 追寻 表达 充满 看 追求 演出 分享 演绎 诠释 描写 描述 寻找 诉说 向往 了解 展现 写 牺牲 留下

表2 释义动词样例

复合名词短语	动词个数	释义动词样例
水果 价格	40	促销 导致 批发 提高 普遍 上涨 含 买到 卖出 购买 贩售 持续 抵消 下挫 进口 销 收购 出口
网球 场地	23	开展 赞助 管理 进入 举行 打 看到 增加 给 参加 选择 离开 展示 练习 举办

5 释义短语生成

文献[13]提出了典型的谓词隐含的句法模式，本文借鉴其研究成果，采用四种句法模板来生成释义短语。见表3，其中，“n1 n2”是复合名词短语，“v”是获取的动词。以“爱情 故事”为例，根据释义模板产生了152个基于动词的释义短语，表4给出了部分样例。这里按照释义模板生成释义短语时，采用的是一种穷尽的方式，产生所有可能的释义短语，这些释义短语中除了包含正确的解释之外，也必然带来很多噪音。因此需要对这些短语进行过滤和排序，尽量把最恰当的释义短语排在前面。

表3 释义模板

编号	模 板	编号	模 板
P1	n1 + v + 的 + n2	P3	n2 + v + n1
P2	n1 + v + n2	P4	v + n1 + 的 + n2

表4 “爱情故事”的释义短语样例

动词	释义短语	动词	释义短语
叙述	爱情叙述的故事	发生	爱情发生的故事
	爱情叙述故事		爱情发生故事
	故事叙述爱情		故事发生爱情
	叙述爱情的故事		发生爱情的故事
围绕	爱情围绕的故事	追寻	爱情追寻的故事
	爱情围绕故事		爱情追寻故事
	故事围绕爱情		故事追寻爱情
	围绕爱情的故事		追寻爱情的故事

6 释义短语过滤

释义短语过滤的目的是去除噪音（即不合理的解释），保留合理的解释，并将最恰当的解释给予较高的排名。为此，最为直观的解决方法就是在语料中为每个释义短语寻找“证据”，如果该释义短语经常在语料中出现，那么就认为它是常用的、合理的解释，可信度较高；如果出现频次很低，就认为该解释并不可信。因此，释义短语过滤的基本假设就是：正确的释义短语应该出现在语料中，并且随着其出现次数增加，释义的可信度随之增加。

然而，在自然语言处理中，数据稀疏（Data Sparseness）是基于语料库的统计方法面临的一大难题。随着网络技术的发展和普及，网上文本资源越来越丰富，研究者提出把互联网（World Wide Web）看做一个巨大的语料库，利用Web数据构建语言资源[17-18]，或者为某些自然语言处理任务提供参数平滑（smoothing）以缓解数据稀疏问题[19-20]。

本文利用海量的Web数据，对释义短语进行验

证, 过滤掉不合理的短语。做法是: 将生成的释义短语作为查询送入搜索引擎, 得到命中次数, 按照命中次数的降序进行排序。命中次数越高的短语, 就越可能是最恰当的释义。目前, Google(www.google.com)和 Baidu(www.baidu.com)是使用最为普遍的中文搜索引擎, 本文利用这两大搜索引擎进行了实验, 查询的方式都是精确匹配(exact match)。

表 5　基于 **Baidu** 和 **Google** 结果的释义短语排名样例

排名	爱情 故事	
	Baidu	Google
1	描绘爱情的故事 11 400	爱情有关的故事 2 920 000
2	追求爱情的故事 10 100	爱情经历的故事 1 280 000
3	寻找爱情的故事 7 900	爱情发生的故事 892 000
4	有关爱情的故事 4 070	爱情有关故事 272 000
5	爱情有关的故事 3 860	故事写爱情 233 000
6	为爱情的故事 3 080	描绘爱情的故事 186 000
7	故事诠释爱情 1 390	爱情为故事 154 000
8	写爱情的故事 1 320	故事让爱情 133 000
9	属于爱情的故事 884	故事描写爱情 128 000
10	诠释爱情的故事 739	诉说爱情的故事 89 800

表 5 分别给出了复合名词短语“爱情 故事”基于 Baidu 和 Google 的结果样例, 表中显示的是前 10 个命中次数最高的释义短语, 短语后面的数字是命中次数。可以看出, 由于 Google 和 Baidu 对中文网页的索引规模不一样, 所以两者对于同一查询所返回的命中次数并不一致, 导致释义短语的排名有所差别。

7　实验结果及分析

本文共选择了 391 个汉语复合名词短语作为实验对象, 经过上述三个步骤, 为每个复合名词短语获取可能的释义短语, 并给出排名。

为了评测该方法的性能, 获取的释义短语都将经过人工检查, 作出二元判断(是或否), 即该释义短语与对应的复合名词短语是否意义相同或者相近。对每个复合名词短语, 返回排名最高的前 n 个(Top n)释义短语, 分别提供给 3 位标注者进行判断[①]。对每个释义短语, 如果有 2 个或以上标注者认为正确, 则判定为正确。然后对每个复合名词短语, 统计给出的候选释义短语中是否有正确的释义出现, 基于此就可以计算整个方法的准确率, 如公式(1)所示。

$$\text{Accuracy}=\frac{\text{the number of compounds with correct interpretation}}{\text{the total number of compounds}}\times 100\% \quad (1)$$

表 6　基于 **Google** 和 **Baidu** 结果的准确率

Top n		1	3	5	10
准确率 Accuracy/%	Google	68.79	90.28	93.35	96.41
	Baidu	71.09	89.25	92.83	96.67

本文给出了 n 取值为 1、3、5 或 10 时的四组评价结果, 见表 6。表中显示, 使用 Google 和 Baidu 所得结果的准确率非常接近。当只为每个复合名词返回排名最高的一个释义短语时, 它们的准确率在 70%左右, 随着返回的释义短语的个数增加, 准确率不断提高。当 n 等于 3 时, 大约 90%的复合名词短语都可以找到正确的释义短语, 比 n 等于 1 时提高了 20%多, 增幅显著。当 n 逐步增大到 5 和 10 时, 准确率依次提高了 3%～4%, 增幅不大。实验结果说明本方法可以有效地为大部分复合名词短语提供正确的释义短语, 当返回前三个排名最高的释义短语时, 准确率已经达到 90%。

本文分别列举了两组复合名词短语, 每组具有相同中心词, 并给出它们正确的语义解释, 见表 7(a)-(b)。可以看出, 对于同一个复合名词短语, 不同的释义短语给出了各种可能的语义解释, 例如“电影 公司”可以是“制作 电影 的 公司”或“发行 电影 的 公司”等不同职能的公司。对于同一组具有相同中心词的复合名词短语, 释义短语分别反映了

表 7(a)　中心词为“故事”的复合名词短语的释义

民间 故事	爱情 故事
民间 流传 的 故事	描绘 爱情 的 故事
流传 民间 的 故事	追求 爱情 的 故事
故事 源于 民间	寻找 爱情 的 故事
源于 民间 的 故事	有关 爱情 的 故事
故事 取材 民间	诠释 爱情 的 故事

① 三位标注者, 其中一位是语言学专业的博士生, 其余两位是计算语言学专业的硕士生。

表7(b) 中心词为"公司"的复合名词短语的释义

电影 公司	啤酒 公司
电影 发行 公司	公司 销售 啤酒
电影 制作 公司	啤酒 制造 公司
电影 投资 公司	啤酒 经销 公司
电影 进出口 公司	啤酒 代理 公司
电影 服务 公司	公司 经营 啤酒

它们进行语义关联的方式,例如,"民间 故事"用"流传、源于、取材"等动词进行释义,强调的是故事的来源或发生地,而"爱情 故事"用"描绘、诠释"等动词进行释义,强调的故事的内容。虽然两者在结构形式上是一样的,但是通过对比各自的释义短语,就可以发现语义联系是有差别的。

本文对没有找到正确释义的复合名词短语(当n等于10时没有找到)进行了分析,造成没有找到正确释义短语的原因主要有两个方面:

第一,没有获取到正确的释义动词,造成由这些动词生成的释义短语也不正确。例如,"农民 习气"获取到的释义动词只有两个:"摆脱"和"改变",都不能正确解释两个名词的语义关联(正确的应该是"来自"或"存在"类动词)。因此,提供的候选释义短语中没有包含正确答案。

第二,搜索引擎Google和Baidu索引的差异,导致释义短语排名不同,正确的释义短语可能排名靠后,因此某些复合名词短语可以在其中一个找到正确的释义短语,而在另外一个失效。例如,"奶油蛋糕"在Baidu提供的前十个结果中没有正确释义,而在Google提供的结果中,正确的释义短语"蛋糕包括 奶油"排名在前十位。

在分析的过程中,我们发现,汉语复合名词短语与英语相比,存在着不少差异。例如,把汉语复合名词短语翻译成英文时,一些名词的词性会发生变化,变成形容词。例如,复合名词短语"国际 标准"的英文是"international standard",名词"国际"变成了形容词"international"。在英文复合名词短语的释义任务中,这类形容词作为修饰语的复合名词短语是不包括在内的。而在汉语中,对于这部分名词实际上充当了形容词功能的复合名词短语,是很难找到合适的动词进行释义的。因此,本文在选词时候,并没有选择这部分复合名词短语进行释义。这样就使得中文和英文的释义任务有更多的共同点,可以相互借鉴和对比。

8 结语

本文首次在汉语中采用"基于动词的短语释义"的方法对复合名词短语进行语义解释,该方法不仅可以为复合名词短语提供多种可能的语义解释,而且能够反应组成相似的复合名词短语之间细微的语义差别。此外,本文的结果也可以服务于问答系统、信息检索、词典编纂等多个应用领域。

本文的方法优于以中心语或修饰语分类的方法,以表7中的两组复合词为例,中心语相同时并不表示其构成关系相同,本方法有效地解决了这个以复合名词短语成分展开无法解决的问题。下一步,我们将围绕动词获取和释义模板两个方向继续研究。获取动词的方法还可以进一步改进,本文利用Chinese Word Sketch获取在指定语法关系下具有搭配意义的动词,这样获取的动词还是有限的,可以借鉴英文的方法,定义一些模板在Web数据(如Google 5-gram web index)上进行扩充。此外,本文使用的释义模板还比较简单,需要改进和完善。例如,一些时间名词和地点名词在释义短语中实际上是作状语,"冬季 运动"解释成"(在)冬季 参加 运动"就比较合理,而目前的方法并没有将这种情况考虑在内。

参考文献

[1] Downing, Pamela. On the Creation and Use of English Compound Nouns[J]. Language, 1997, 53(4): 810-842.

[2] Levin, Judith. The Syntax and Semantics of Complex Nominals[M]. Academic Press, New York. 1978.

[3] Warren, Beatrice. Semantic patterns of noun-noun compounds[J]. In Gothenburg Studies in English 41, Goteburg, Acta Universtatis Gothoburgensis. 1978.

[4] Vanderwende, Lucy. Algorithm For Automatic Interpretation of Noun Sequences[C]//The 15th International Conference on Computational Linguistics (COLING). 1994.

[5] Lauer, Mark. Designing Statistical Language Learners: Experiments on Compound Nouns[D]. Ph. D. thesis, Macquarie University. Australia. 1995.

[6] Girju, Roxana, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney, and Deniz Yuret. Semeval-2007 task 04: Classification of semantic relations between nominals [C]//Proceedings of SemEval, Prague, Czech Republic. 2007: 13-18.

[7] Girju, Roxana, Dan Moldovan, Marta Tatu, and Daniel Antohe. On the semantics of noun compounds[J]. Journal of Computer Speech and Language - Special Issue on Multiword Expressions, 2005, 4(19): 479-496.

[8] Diarmuid O Seaghdha. Learning Compound Noun Semantics[D]. Ph. D. thesis, University of Cambridge. 2008.

[9] Nakov, Preslav. Noun compound interpretation using paraphrasing verbs: Feasibility study[C]//Proceedings of the 13th international conference on Artificial Intelligence: Methodology, Systems and Applications (AIMSA 2008), Springer. 2008: 103-117.

[10] Nakov, Preslav and Marti A. Hearst. Using verbs to characterize noun-noun relations[C]//Proceedings of the 12th international conference on Artificial Intelligence: Methodology, Systems and Applications (AIMSA 2006), Springer. 2006: 233-244.

[11] Butnariu, Cristina, Su Nam Kim and Preslav Nakov et al. SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions[C]//Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009). 2010.

[12] Zhao, Jinglei, Hui Liu and Ruzhan Lu. Semantic Labeling of Compound Nominalization in Chinese[C]//Proceedings of the Workshop on A Broader Perspective on Multiword Expressions, Prague, June 2007: 73-80.

[13] 袁毓林. 谓词隐含及其句法后果[J]. 中国语文. 1995 年, 第 4 期.

[14] Kilgarriff, Adam and David Tugwell. Sketching words. Lexicography and Natural Language Processing: A Festschrift in Honour of B. T. S. Atkins. Marie-Hélène Corréard (Ed.)[M]. EURALEX, 2002: 125-137.

[15] Kilgarriff, Adam, Pavel Rychly, Pavel Smrz and David Tugwell. The Sketch Engine[C]//Proc. Euralex. Lorient, France, July, 2004: 105-116.

[16] Huang, Chu-ren, Adam Kilgarriff, Yiching Wu et al. Chinese Sketch Engine and the Extraction of Grammatical Collocations[C]//Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing. 2005.

[17] Grefenstette, G. and J. Nioche. Estimation of English and non-English Language Use on the WWW[J]. Arxiv preprint cs. CL/0006032. 2000.

[18] Jones, R. and R. Ghani. Automatically building a corpus for a minority language from the web[C]//Proceedings of the Student Research Workshop at the 38th Annual Meeting of the Association for Computational Linguistics, 2000: 29-36.

[19] Grefenstette, Gregory. The World Wide Web as a resource for example-based machine translation tasks[C]//Proceedings of the ASLIB Conference on Translating and the Computer. London. 1998.

[20] Keller, Frank and Mirella Lapata. Using the web to obtain frequencies for unseen bigrams[J]. Computational Linguistics. 2003, 29(3): 459-484.