

# 怎样用物性结构知识解决“网球问题”？\*

袁毓林 李强

(北京大学 中文系/中国语言学研究中国中心/计算语言学教育部重点实验室, 北京 100871)

**摘要:** “网球问题”指怎样把 racquet (网球拍)、ball (网球) 和 net (球网) 之类具有情境联想关系的词汇概念联系起来、发现它们之间的语义和推理关系。这是一个自然语言处理和相关的语言知识资源建设的世界性难题。该文以求解“网球问题”为目标, 对目前比较主流的几种语言词汇和概念知识库系统(包括 WordNet、VerbNet、FrameNet、ConceptNet 等)进行检讨, 指出它们在解决“网球问题”上都存在一定的局限性, 着重分析它们为什么不能解决“网球问题”。进而指出基于生成词库论的名词物性结构知识描写体系可以解决“网球问题”, 主张用名词的物性结构知识和相关的句法组合知识来构建一种以名词(实体)为核心的词汇概念网络, 以弥补上述几种知识库系统的不足, 为自然语言处理提供一种可资参考的词汇概念知识库体系。

**关键词:** “网球问题”; WordNet; VerbNet; FrameNet; ConceptNet; 物性结构

## How to Use Qualia Structure to Solve “Tennis Problem”?

YUAN Yulin LI Qiang

(Dept. of Chinese Lang. & Lit., Peking University, Research Center of Chinese Linguistics,  
Ministry of Education Key Laboratory for Computational Linguistics, Beijing 100871, China)

**Abstract:** “Tennis Problem” tries to link racquet, ball and net that are in a situational association relationship, and find the semantic and reasoning relationship among them. This is a worldwide problem in natural language processing and in the construction of related language knowledge or resources. Aiming at solving the “tennis problem”, this paper reviews several mainstream systems of linguistic lexical and conceptual knowledge base (including WordNet, VerbNet, FrameNet, ConceptNet, etc.), illustrates their limitations on solving this problem, and focuses on why they cannot solve it. For this reason, this paper proposes that the descriptive system of knowledge based on the theory of generative lexicon, i.e., the qualia structure of nouns, can be adopted, and that the qualia structure and relevant syntactic combinations can be used to build a noun-based or entity-based lexical network. This conceptual network may make up for the inadequacy of the above-mentioned systems of knowledge base and provide a knowledge base of lexical concepts for natural language processing.

**Keywords:** “Tennis Problem”; WordNet; VerbNet; FrameNet; ConceptNet; Qualia Structure

### 1 什么是“网球问题”？

“网球问题”(Tennis Problem)最初是由 Roger Chaffin 在他跟 Christiane Fellbaum 的私人交流中提出来的, 后来由 Fellbaum (1998:10) 公诸于世, 并随着 WordNet 的普及而受到自然语言处理学界的广泛关注。简单地说, “网球问题”指的是如何将 racquet (网球拍)、

---

\* 收稿日期:

定稿日期:

基金项目: 国家社科基金重大项目《汉语国际教育背景下的汉语意合特征研究与大型知识库和语料库建设》(12&ZD175)

ball (网球)、net (球网) 等词语以一定的方式联系到一起的问题。从概念意义上看, 这些词语所指谓的都是 court game (球类运动项目) 的组成部分, 是由网球运动这一事件或场景激活 (activate) 出来的。与此类似的情况, 还包括: “医生” 和 “医院”、“蛋糕” 和 “蜡烛”、“粉笔” 和 “黑板”、“小狗” 和 “链子” 等。这些词语所表示的不同概念, 都需要通过一个事件或一个场景才能组织到一起。正因为这些不同概念之间缺乏相似性 (similarity)、组织性 (organization) 和推导性 (derivation), Fellbaum 认为: 对于此类 “网球问题”, 他们所研制的 WordNet 没能提供较好的解决办法。因为, WordNet 并不包含围绕一个特定话题组合起来的、篇章结构环境下的、各个语言成分之间的语义关系。那么, 对于这种需要通过事件或场景来维系的不同概念及其词汇, 有没有一种恰当而又充分的语言词汇知识体系, 能将它们关联起来, 并且把它们的语义关系描述清楚呢? 换句话说, 这些不同的概念及其词汇是否能够通过一种具有系统性 (systematicness) 和秩序性 (order) 的词汇知识库串联起来, 以弥补 WordNet 在解决这类复杂关系问题方面的不足呢?

本文以求解 “网球问题” 为目标, 对目前比较主流的几种语言和词汇概念知识库进行检讨, 说明为什么这几种知识库都无法解决 “网球问题”。进而指出, 生成词库理论中的物性结构知识描写体系, 对于解决 “网球问题” 具有重要的启发性意义。因为, 不同语言成分所代表的概念可以通过物性角色串联在一起, 从而形成一个网状的知识体系。这种意义的网络能够将概念与概念之间以及概念所具有的属性之间的关系清楚地揭示出来, 为下一步建立常识性推理知识库奠定一定的基础。

文章的结构组织大致安排如下: 第二节对 WordNet 进行介绍, 并说明为什么它无法解决 “网球问题”; 第三节对 VerbNet 进行介绍, 并说明它不能解决 “网球问题” 的原因; 第四节对 FrameNet 进行介绍, 并解释它在解决 “网球问题” 方面的局限性; 第五节介绍概念网 ConceptNet, 说明单纯的常识性推理知识不注重词语的组合方面知识, 不便于解决 “网球问题”; 第六节介绍名词的物性结构知识体系, 通过具体个案来展示它对 “网球问题” 的解释力; 最后对文章进行小结。

## 2 为什么 WordNet 无法解决 “网球问题” ?

WordNet 是目前自然语言处理领域非常重要且应用十分广泛的词汇知识库, 它是由美国 Princeton University (普林斯顿大学) 认知科学实验室的 G. A. Miller (米勒)、R. C. Beckwith (贝克威克)、C. Fellbaum (费尔鲍姆) 等人于 1985 年研制的, 至今将近有三十年的时间, 数据库版本从 1991 年的 version 1.0 到 2007 年的 version 3.0, 再到如今的 version 3.1, 收词规模和知识库维护正在逐步地扩大和完善 (Miller & Fellbaum 2007)。作为一种机读词典, WordNet 充分吸收了心理语言学的相关理论, 改变了传统以字母顺序组织词汇信息的方式, 而将在语义上紧密联系的相关概念聚合到一起形成同义词集 (synsets)。这样做的心理学依据是: 在人们的心理词库 (mental lexicon) 中, 词汇的组织 and 表征方式并不是随意和散漫的, 而是有组织性和系统性的。将这样一种词汇语义的分布模式移植并应用到计算机的词汇知识资源之中, 为计算机实现自然语言的语义理解提供了可能性。

WordNet 对于词条的组织方式基本上是属于聚合式的, 所以它看起来像是一部汇编式的同义词义类词典 (thesaurus)。在 WordNet 中, 主要包括名词、动词、形容词和副词, 所有的词都按照其义项以同义词集的形式组织在一起, 并标示相关的同义词集在语义上的联系。这些语义联系有上下位关系 (hyponymy/hypernymy)、整体 - 部分关系 (meronymy/holonymy)、反义关系 (antonymy)、蕴涵关系 (entailment) 等 (Fellbaum 1998: 23-104; Miller etc. 1990)。从本体知识 (Ontology) 的角度看, WordNet 通过各种语义关系将词汇联系在一起的方式, 展示了它构建语言本体知识的能力。因而, 它对于信息检索 (information retrieval)、词义消歧 (word sense disambiguation)、机器翻译 (machine

translation)、知识工程 (knowledge engineering) 等自然语言处理领域具有重要的推动作用, 并直接服务于下一代新网络——语义网 (semantic web) 的构建 (黄居仁 2005)。

然而, 作为一种计算词库 (computational lexicon), 袁毓林 (2008) 指出了它也存在若干缺陷。比如, 缺少动词句法配置方面的信息、缺少谓词论元的语义角色信息等等。其中, 最为重要的是, WordNet 没有能将词汇之间的组合关系揭示出来。组合关系和聚合关系是语言系统中的两种最根本关系, 索绪尔 (1980: 170) 分别称之为“句段关系”和“联想关系”。组合关系基于语言成分的线性排列, 而聚合关系基于语言成分的共同特点。作为义类词典的 WordNet 更多关注的是词语之间在意义上的相似性特征, 也即聚合关系, 而对那些不具有相似性特征的词语之间的可能的组合关系和语篇中的共现关系 (co-occurrence) 则关注不够。比如, “hospital” (医院) 和 “physician” (医生), 它们在 WordNet 中检索到的信息分别如下面的图 1 和图 2 所示。

## Noun

- (25){03545775} <noun.artifact>[06] S: (n) **hospital#1 (hospital%1:06:00::), infirmary#1 (infirmary%1:06:00::)** (a health facility where patients receive treatment)
- {08071169} <noun.group>[14] S: (n) **hospital#2 (hospital%1:14:00::)** (a medical institution where sick or injured people are given medical or surgical care)

图 1 “hospital” (医院) 在 WordNet 中的信息

## Noun

- (7){10040615} <noun.person>[18] S: (n) **doctor#1 (doctor%1:18:00::), doc#1 (doc%1:18:00::), physician#1 (physician%1:18:00::), MD#2 (md%1:18:00::), Dr.#2 (dr.%1:18:00::), medico#2 (medico%1:18:00::)** (a licensed medical practitioner) *"I felt so bad I went to see my doctor"*

图 2 “physician” (医生) 在 WordNet 中的信息

由图 1 和图 2 可见, 在 “hospital” (医院) 的语义描述信息中找不到 “physician” (医生); 相反, 在 “physician” (医生) 的语义描述信息中也找不到 “hospital” (医院)。因而, WordNet 没有办法在 “hospital” (医院) 和 “physician” (医生) 之间建立起有效的语义指针性关联。这对于自然语言处理, 尤其是对于文本理解来说是不利的。比如, 当某个人说 “我要去医院”, 隐含的意思是他要去看医生。如果不能在 “医院” 和 “医生” 之间建立起关联, 那么计算机在理解 “我要去医院” 这句话的隐含义时就会有困难。

此外, 心理学的研究成果显示, 在语义关系网络中, 所有的概念都不是孤立存在的, 总是跟其他概念相互联系。其中, 最重要的两种联系是类别关系 (taxonomic relation) 和主题关系 (thematic relation)。类别关系是基于相似性组织的层级关系, 对应于聚合关系; 而具有主题关系的两个概念之间通常并不共享相似的特征, 只是通过事件或场景组织起来, 对应于组合关系 (Warrington & Shallice 1984; Lin & Murphy 2001)。人们对于特定词汇的习得以及将其整合到语义网络的过程, 都离不开这两种学习方式。可见, 基于心理学研究成果之上的 WordNet, 它对词汇所进行的语义归纳和分类的做法只是体现了词汇学习中的类别关系, 而忽视了另外一种基于场景的联想式词汇组合关系和语篇中的共现关系, 也即上文第一节所说的 “网球问题”。要解决 “网球问题”, 就需要利用到组合性词汇知识资源, 通过语义关系指针将表示具有主题关系的不同概念的有关词语联系起来, 从而形成指针式的语义链接关系。而在一个场景或事件中, 这些不同的词汇概念往往是通过动词联系起来的, 比如, “小狗” 和 “链子” 就可以通过动词 “拴” 联系, 形成 “用链子拴小狗”, 或者其关系化的名词短语 “拴小狗的链子、拴着链子的小狗”; “母鸡” 和 “鸡蛋” 可以通过动词 “下/生” 联系,

形成“母鸡下/生鸡蛋”，或者其关系化的名词短语“下/生鸡蛋的母鸡、母鸡生/下的鸡蛋”。动词可以为不同的概念提供链接的中介，也可以同时激活这些不同的信息（Altmann 1999；Warren & McConnell 2007）。于是，接下来的挑战是：怎样找到这种桥梁性动词？什么样的语义学理论和语义知识资源可以解决这个问题？这要留到下文第六节再讨论。

### 3 为什么 VerbNet 无法解决“网球问题”？

上文提到，WordNet 只关注动词内部的语义组织关系，包括蕴涵关系（entailment）、对立关系（opposition）、致使关系（causation）等等，这些语义关系基本上是静态性的、聚合性的，至于动态性的、组合性的动词论旨角色及其句法配置等语义关系信息，WordNet 则没有涉及。在这一方面，University of Colorado at Boulder(科罗拉多大学波尔得分校)的 VerbNet，在一定程度上弥补了 WordNet 的不足。它对动词的句法框架、论旨角色（thematic roles）和选择限制（selectional restrictions）进行了细致的描写（Kipper etc. 2000）。VerbNet 的理论基础是 Levin（1993）的动词分类标准。该分类标准体现了动词句法和语义的互动关系，表现为：具有相同句法行为的动词往往属于同一个语义类，而不同语义类的动词则一定具有不同的句法行为。因此，可以通过句法形式上的变换（alternation）关系分析，达到划分动词语义小类的目的。VerbNet 为每一个动词设置了三类信息，包括：（1）Members，即同属于一个语义类的各个动词；（2）Thematic Roles，一类动词所能支配的不同论旨角色，并在括号中标明动词对于论元的语义选择限制；（3）Frames & Predicates，包括动词的句式类型、实例、句法配置和语义描述等（Kipper et al. 2000；袁毓林 2008）。比如，下面是动词 hit 在 VerbNet 中的信息：

MEMBERS	
BANG (fv 1, 2, 3; mv 1, 3; c 1, 4)	SQUELCH (mv 1)
BASH (fv 1; mv 1)	TAMP (mv 1)
CLICK (fv 1, 2, 3; mv 4; c 2)	THOMP (fv 1, 2; mv 3)
DASH (mv 3; c 2)	THRACK (fv 1; mv 1)
SMITE (mv 1, 3; c 1)	WHACK (mv 1; c 1)

ROLES
<ul style="list-style-type: none"> <li>• AGENT [+ENT_CONTROL]</li> <li>• PATIENT [+CONCRETE]</li> <li>• INSTRUMENT [+CONCRETE]</li> <li>• RESULT</li> </ul>

FRAMES	
<b>HP V HP</b> EXAMPLE "Paula hit the ball." SYNTAX AGENT V PATIENT SEMANTICS CAUSE(AGENT, E) MANNER(DURING(E), DIRECTEDMOTION, AGENT) NOT(CONTACT(DURING(E), AGENT, PATIENT)) MANNER(END(E), FORCEFUL, AGENT) CONTACT(END(E), AGENT, PATIENT)	REF
<b>HP V HP PP. INSTRUMENT</b> EXAMPLE "Paula hit the ball with a stick." SYNTAX AGENT V PATIENT [WITH] INSTRUMENT SEMANTICS CAUSE(AGENT, E) MANNER(DURING(E), DIRECTEDMOTION, INSTRUMENT) NOT(CONTACT(DURING(E), INSTRUMENT, PATIENT)) MANNER(END(E), FORCEFUL, INSTRUMENT) CONTACT(END(E), INSTRUMENT, PATIENT) USE(DURING(E), AGENT, INSTRUMENT)	
<b>HP V HP TOGETHER</b> EXAMPLE "Paula hit the sticks together." SYNTAX AGENT V PATIENT <+PLURAL> (TOGETHER) SEMANTICS CAUSE(AGENT, E) MANNER(DURING(E), DIRECTEDMOTION, PATIENT_i) MANNER(DURING(E), DIRECTEDMOTION, PATIENT_j) NOT(CONTACT(DURING(E), PATIENT_i, PATIENT_j)) MANNER(END(E), FORCEFUL, PATIENT_i) CONTACT(END(E), PATIENT_i, PATIENT_j)	
<b>HP V HP ADJP</b> EXAMPLE "Paul kicked the door open." SYNTAX AGENT V PATIENT RESULT SEMANTICS CAUSE(AGENT, E) MANNER(DURING(E), DIRECTEDMOTION, AGENT) NOT(CONTACT(DURING(E), AGENT, PATIENT)) MANNER(END(E), FORCEFUL, AGENT) CONTACT(END(E), AGENT, PATIENT) PRED(RESULT(E), PATIENT)	
<b>HP V HP ADJP PP. INSTRUMENT</b> EXAMPLE "Paul hit the door open with his foot." SYNTAX AGENT V PATIENT RESULT [WITH] INSTRUMENT SEMANTICS CAUSE(AGENT, E) MANNER(DURING(E), DIRECTEDMOTION, INSTRUMENT) NOT(CONTACT(DURING(E), INSTRUMENT, PATIENT)) MANNER(END(E), FORCEFUL, INSTRUMENT) CONTACT(END(E), INSTRUMENT, PATIENT) PRED(RESULT(E), PATIENT) USE(DURING(E), AGENT, INSTRUMENT)	

图 3 动词 hit 在 VerbNet 中的信息

从 VerbNet 对于动词的描写可以看到，动词相关的语义和句法信息都能清晰地展现出来，这离解决“网球问题”更进了一步。正如上一节所提到的，基于事件或场景的不同概念通常需要动词作为链接指针进行联系。那么，不言而喻的是，动词的句法语义信息对于情景中概念和词语之间的联系起了非常重要的作用。然而，我们发现 VerbNet 虽然对于动词的相关信息做了足够充分的刻画，但在解决“网球问题”方面还有以下三点不足：

（1）VerbNet 是以动词为中心的词汇知识库，以动词作为查询项，对名词所指事物的情景关系不能做出预判。比如，“tennis”（网球）和“racquet”（网球拍），它们在 VerbNet

中查询不到，只有先获知“‘racquet’是击打‘tennis’这个动作所凭借的工具”这层语义关系，进而查询动词 hit，才能得到[NP<sub>agent</sub> V NP<sub>patient</sub> PP<sub>instrument</sub>] (Paula hit the tennis with a racquet) 这种能将“tennis”和“racquet”联系起来的句法格式。

(2) 对于一些场景依赖性特别强的相关事物，VerbNet 提供不了一个能够将指称这些事物的名词囊括在特定句法结构之中的动词。比如，“生日”和“蛋糕”，能将它们联系起来的常识性推理应该是“为了庆祝生日而吃蛋糕”。但若要在 VerbNet 中查询动词“eat”（吃），它的句法框架只能容纳名词“蛋糕”，无法将“生日”和“蛋糕”关联起来并且都容纳进其中的某个句法框架。动词“eat”（吃）在 VerbNet 中的句法框架信息如下面图 4 所示。“celebrate”也是如此。虽然 VerbNet 中未收录表示“庆祝”义词条“celebrate”的句法框架信息，但根据语言学知识，它的句法框架是[NP<sub>agent</sub> V NP<sub>theme</sub>] (We celebrated Mary’s birthday)，也只能容纳名词“生日”，而不能容纳名词“蛋糕”。

FRAMES	
<b>NP V NP</b>	
EXAMPLE	"Cynthia ate the peach."
SYNTAX	AGENT V PATIENT
SEMANTICS	TAKE_IN(DURING(E), AGENT, PATIENT)
<b>NP V</b>	
EXAMPLE	"Cynthia ate."
SYNTAX	AGENT V
SEMANTICS	TAKE_IN(DURING(E), AGENT, ?PATIENT)
<b>NP V PP-CONATIVE</b>	
EXAMPLE	"Cynthia ate at the peach."
SYNTAX	AGENT V {AT} PATIENT
SEMANTICS	TAKE_IN(DURING(E), AGENT, PATIENT)
<b>NP V PP.SOURCE</b>	
EXAMPLE	"He ate off of the table."
SYNTAX	AGENT V {+SRC} SOURCE
SEMANTICS	TAKE_IN(DURING(E), AGENT, ?PATIENT) PREP(DURING(E), ?PATIENT, SOURCE)

图 4 动词 eat（吃）在 VerbNet 中的句法框架信息

(3) VerbNet 主要表现动词的论元结构和句法实现，动词为特定句法结构中的名词赋予论旨角色，可以将受动词支配的有关名词所表示的概念聚合在一起。但是，有些概念之间的联系依赖的是情景式联想。比如，“键盘”和“鼠标”，它们之间的联系并不依赖于某一个特定的动词，而是通过“电脑配件”这一概念域情景产生关联的。所以，在这种情况下，VerbNet 对不同概念之间的激活起不到应有的作用。也就是说，VerbNet 关注以动词为核心的词汇性组合关系，不能反映语篇中相关词语的常规性（routine）的共现关系。

综合以上三个方面来看，VerbNet 虽然能够在一定程度上反映动词对于名词的串联作用，即不同事物依靠特定动作关联在一起；但是，它仍然无法反映事物之间的情景关系，尤其是通过“扩散性激活”<sup>1</sup>（spreading activation）引起的不同事物之间的联想式语义网络。要想解决“网球问题”，还得求助于其他类型的组合性语义学理论和相关的词汇知识库。

#### 4 为什么 FrameNet 无法解决“网球问题”？

加州大学伯克利分校（University of California, Berkeley）的框架网项目（FrameNet），是基于框架语义学（frame semantics）、并以语料库为基础建立的在线英语词汇资源库。框架

<sup>1</sup> 语义的扩散性激活指，调用（即激活）一个词项的意义可以触发（trigger）知识网中相关的语义节点。比如，激活场景定义型名词“生日”，可以触发“生日蛋糕、生日蜡烛、生日歌”等概念；激活事件名词“火”，可以触发“红色、火光、消防车、水”等概念。



语义学是 Fillmore 提出的研究词语意义和句法结构意义的方法, 该理论主张对于词语意义的描述必须联系特定的语义框架, 因为框架是信仰、实践、制度、想象等概念结构或概念模式的图式表示, 是言语社团中人们相互交流的概念基础 (Fillmore et al. 2003; 袁毓林 2008)。Fillmore 把一个格框架看作是刻画一个小的抽象的“情景”(scene) 或“境况”(situation), 该“情景”或“境况”帮助理解一个动词的语义结构跟该动词的基本句法属性如何联系, 以及不同语言形成最小句子的不同方式。所以, 要理解动词的语义结构, 就必须首先理解这类图式化的情景 (Fillmore1982)。比如, 框架 heat 描述的是一个涉及烹调 (cook)、食物 (food) 和加热工具 (heating instrument) 的情境, 以及可能引发这一情境的一些词汇, 如 bake、blanch、boil、broil、brown、simmer、steam 等。出现在 heat 这一框架中的成分称为“框架元素”(frame element), 而能够引发激活这一框架的词语称为“词汇单元”(lexical unit)。

FrameNet 数据库主要由词汇库 (lexicon)、框架库 (FrameNet database) 和标注例句 (annotated example sentences) 三部分组成。词汇库主要包括词条传统的词典释义、语义结构和配价模式、与标注例句的链接以及跟其他机器可读资源 (如 WordNet/Comlex) 的链接。框架库主要描述框架基本的概念结构, 给出框架名称、框架元素以及框架之间的关系。标注例句呈现特定词项的语义和句法性质, 标注句中的框架元素、框架元素所在短语的短语类型, 以及框架元素的句法功能 (Baker et al.1998)。比如, 下面的图 5 是表示“击中”义的动词 hit 在 FrameNet 中的信息。

**Frame: Hit\_target**

**Definition:**

COD: strike (a target).

**Frame Elements and Their Syntactic Realizations**

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Agent	(2)	NP. Ext (2)
Subregion	(1)	PP[in]. Dep (1)
Target	(2)	NP. Obj (2)

**Valence Patterns:**

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns		
1 TOTAL	Agent	Subregion	Target
(1)	NP Ext	PP[in] Dep	NP Obj
1 TOTAL	Agent	Target	
(1)	NP Ext	NP Obj	

[Clear Sentences](#) [Turn Colors Off](#)

[X] He missed the snake and hit the leg of the end table that was on one side of the sofa .  
 [X] All the shots missed their target , but one hit a nearby tourist in the leg .

图 5 动词 hit 在 FrameNet 中的信息

由上图可见, 跟 VerbNet 以动词为核心、自下而上 (bottom-up) 的句法语义描写不同, FrameNet 是以框架为核心、自上而下 (top-down) 地对处于框架中的动词或事件名词的句法语义信息进行描写。框架和框架元素反映了事件和事件参与者之间的关系, 以及表示这种事件的动作跟其论元在句法配置上的关系。所以, FrameNet 可以看作是 VerbNet 的抽象版本, 在解决事物和概念的情景联想关系时, 它同样也会遇到和 VerbNet 一样的问题。比如:

(1) 对于某些表示事物概念的名词, 如“网球 (tennis)”、“键盘 (keyboard)”、“狗 (dog)”等, 基于情景联想关系, 它们可以分别与“网球拍”、“鼠标”、“链子”产生关联。但是, 这

些词语在 FrameNet 中都查询不到相关的信息, 这些事物之间的情景联想关系也就无法表示。

(2) 对于某些表示事物概念的名词, 如“生日 (birthday)”、“医院 (hospital)”等, 虽然它们在 FrameNet 中可以查询到相关的框架信息; 但是, 框架中所包含的元素并不能体现情景联想关系。比如, “hospital”所在的框架为“buildings”, 该框架的定义是“a building which houses an institution providing medical, surgical, and nursing care for sick or injured patients”, 出现在该框架中的框架元素包括“building、function、material、name、place、relative location、time of creation”, 这些框架元素中并不包含可由“医院”激活的、基于情景联想关系的“医生、护士、医疗设备”等概念。

(3) 正如袁毓林 (2008) 所指出的: “在 FrameNet 中, 做多少框架才能覆盖绝大多数的文本语料? 从哪些框架开始、止于哪些框架? 怎样确定情景的大小和覆盖面? 对于这些问题, FrameNet 都是无法给出明确答案的”。而它们将直接影响到 FrameNet 解决“网球问题”所能取得的实际效果。因为, 基于情景联想关系的不同事物或概念之间通常并不具有十分紧密的语义联系, 由一个事物激活另外一个事物的原因可能是由于它们经常性地共现于某一个场景, 或基于一般的生活经验性常识。那么, 可以预料到, 如果 FrameNet 所划分和定义的情景框架较小或覆盖面较窄, 它就一定不能触及由情景激活的事物。所以, FrameNet 在确定情景框架方面的主观随意性, 会给“网球问题”的解决带来一些困扰。

基于以上认识, 我们认为: FrameNet 设置情景框架并在此基础上描述词项的语义结构和句法配置信息, 这比 VerbNet 单纯对于动词相关信息的描写更进了一步。同时, 情景框架的设置也给“网球问题”的解决带来了一些新的启发和思考。但是, 因为 FrameNet 从根本上来说还是以动词为核心的、旨在反映事件和事件参与者之间的论旨角色关系; 并不能抓住相关词汇概念在语篇中的常规性的共现关系。而且, 情景框架的划分具有较大的任意性和主观性。所以, 它还不能为解决事物之间的情景联想关系这一问题提供一个理想的语言知识资源。

## 5 为什么 ConceptNet 无法解决“网球问题”?

人工智能领域的研究从 20 世纪 30、40 年代开始兴起, 早在计算机时代的曙光来临之前, 以图灵 (Alan Turing, 1912-1954) 为代表的一批学者就已经提出了关于计算机以及人工智能的设想, 他们希望未来可以研制出类似人脑那样具有判断、分析、推理、决策等高级认知能力的计算机, 而这也成为人工智能领域的终极目标。在人工智能的发展过程中, 很多学者都曾明确提出, 横亘在人工智能研究道路上的一块大顽石是现有的计算机程序缺乏必要的常识。换句话说, 在人类社会, 一个正常的 6 岁儿童就已经掌握了有关周围世界和环境的常识知识; 但是对于计算机来说, 它连最基本的常识都没有 (Henderson 2007)。人工智能专家为了解决这个问题, 开始为计算机研制添置常识知识库, 例如费根鲍姆 (Feigenbaum) 的“专家系统”和里南 (Lenat) 的“大百科全书”计划, 都旨在使计算机系统具有一个强大的知识库, 这是当代智能系统或智能代理所具备的一种常识性知识库。本节所介绍的 ConceptNet 也属于这样一种常识性推理知识库。

ConceptNet 是由麻省理工学院媒体实验室开发的一种开源工具, 它的开发者 Liu & Singh (2004) 指出, 基于关键词和数据统计的方法只能实现计算机语义理解的表层处理, 要想实现深层次的计算机对于文本的理解, 就必须添加各种不同的语义知识 (semantic knowledge), 使得计算机同时也拥有人所具备的常识性知识。ConceptNet 的开发建立在 OMCS (Open Mind Common Sense) 基础之上, 它是一个大型的常识知识库 (knowledgebase), 包含了来自于空间、物质、社会、时间和心理方面的日常生活知识。ConceptNet 从 OMCS 所收集的大约 70 万个句子中自动构造一个常识性语义网络, 用不同类型的链接描述物体、事件以及人物之间的关系。值得一提的是, 与传统的手工提取常识性知识不同, OMCS 通

过网络平台向普通大众寻求支持，体现了网络众包开发意识。在 ConceptNet 中，语义知识网络通过 160 万个箭头将超过 30 万个节点连接在一起，每一个箭头代表一种语义关系，类似这样的语义关系共有 20 种，构成了语义关系的本体知识系统。比如：

Is (A, B): A *dog* is an *animal*.

Location Of (A, B): *Books* are in the *library*.

Used For (A, B): *Forks* are used for *eating*.

Subevent Of (A, B): After *waking up in morning*, he *checked his email*.

在上面这些例子中，“Is”连接“dog”和“animal”这两个概念节点，这两个概念节点又可以通过其他类型链接与其他概念节点相连。其余情况亦是如此。通过词汇之间相互链接所形成的概念语义网络对于话题提取 (topic extraction)、情感标注 (emotion tagging)、词义消歧 (word sense disambiguation)、文本推理 (text inference) 等自然语言处理都有重要的作用。ConceptNet 常识语义网络如下面图 6 所示：

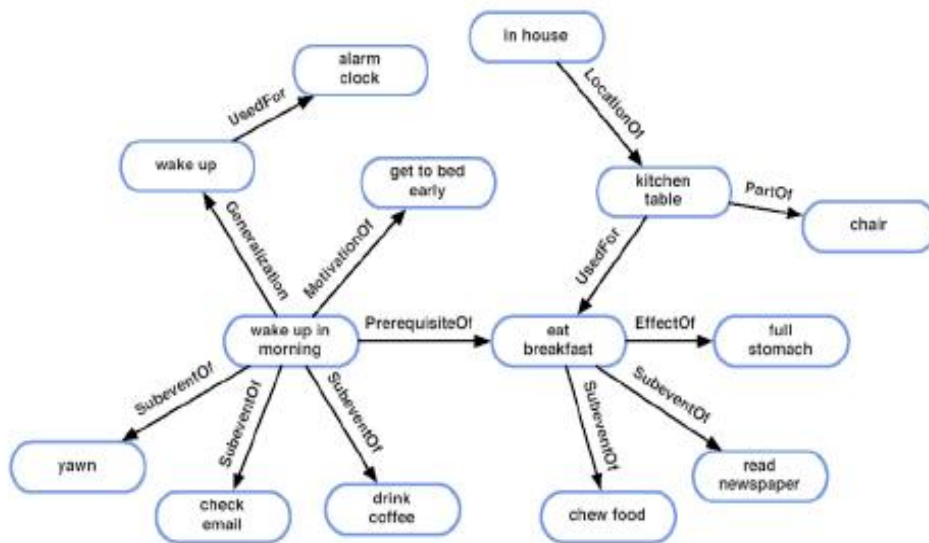


图 6 ConceptNet 常识语义网络片段

由上图可见，ConceptNet 实质上是一种基于常识和概念联想关系的词汇语义知识库。这种知识库通过认知上的扩散性激活机制 (spreading activation mechanism) 将日常生活中方方面面的知识都囊括在语义网络之中，并可通过指针进行追踪 (traversal)，从而为计算机建立了一个类似储存在人脑之中的概念系统，为相关的自然语言处理提供了强有力的支持。从理论上来说，这种常识知识库应该是解决“网球问题”的一种比较理想的资源。比如，“网球拍”、“网球场”、“网球服”等概念都可以通过“网球”来激活；“医生”、“护士”、“医疗仪器”等都可以以“医院”作为中心概念形成扩散性的辐射式网络；“蛋糕”、“蜡烛”等都是“生日”这一事件情景中经常出现的事物。但是，这种常识知识库主要侧重于概念和推理层面，在知识库中存储的信息是关于某一个概念可能与其他若干概念之间的生发 (generation) 关系以及这种关系的具体属性值，而对于词语之间的组合关系关注不够。这进一步限制了它在自然语言处理中的应用。比如，袁毓林(2002)就曾指出，在信息抽取 (information extraction) 领域中，中观层次的论元结构知识是非常重要的。以句子为考察对象，句子中的谓词和名词之间的论元结构关系这种低层次的语义关系可以通过一定的程序传递到高层次的关于事件的脚本和框架中。也就是说，只有弄清楚动词的论元结构信息，才能在特定的文本模板中抽取信息检索所需要的关键内容。虽然 ConceptNet 为不同概念节点之间设置了多达 20 种的语义关系链接，但这些语义链接呈现出来的只是概念之间的深层语义推导关系，而并没有说明表达这些不同概念的词语在句法表层是如何被组织在一起的。正因为缺乏句法层面的相关描



述,使得不同概念之间的关系链接不便于以一种规则化和形式化的标准确立,只能依赖于一般的常识性知识。此外,ConceptNet 主要关注短语结构所表达的复合概念(compound concept,如“开车”、“买食物”)之间的深层语义关系,大多数属于由事件范畴所激活的概念知识,对于由一般事物范畴所能引发生成的网络系统,ConceptNet 关注得还不够。因此,也就不便于进行以名词为查找项(入口)的相关词汇及其情境联想关系的检索。

总之,我们认为,与 VerbNet 和 FrameNet 相比,ConceptNet 建构了一种常识性知识网络系统,这对于“网球问题”的解决具有重要的作用。但是,ConceptNet 过分关注不同概念之间常识推理性的语义关系,而忽略了表示相关概念的不同词语在句法层面上的组合关系和语篇层面上的共现关系。如果有一种词汇知识库能同时涵盖跟有关词项相关的常识性知识和句法组合及语篇共现知识,那么,它对于解决“网球问题”应该会发挥更大的作用。

## 6 为什么物性结构知识可以解决“网球问题”?

围绕着自动构建词汇语义系统的尝试,在过去的一、二十年中一直在进行。其中,Pustejovsky (1991, 1995) 提出的生成词库论(Generative Lexicon, GL),确定了如下的总体目标:

开发生成词库的目的在于提供一种成分组合语义学,对源自现实语言使用中的意义进行语境调制(contextual modulation)。(Pustejovsky et al. 2013: 1)

生成词库论一方面关注词汇语义学(词汇的成分组合语义学),同时也关注语境语义学(意义的语境调制),并试图在它们之间进行调和<sup>2</sup>(吴国向、袁毓林 2013)。面对语言学知识和非语言学知识的表征接口,Pustejovsky 设计了一种语义框架,即物性结构,或称为物性角色描写体系。它其实是一套关于词汇本体知识(ontological lexical knowledge)的描述体系,是生成词库理论中最为核心的内容<sup>3</sup>。借助物性角色知识,我们可以对语言的创造性使用,尤其是对词项在不同的语言环境下浮现出(emerge)不同的意义这一动态性词义变化现象进行说明和解释。Pustejovsky (1995)谈到了下列四种物性角色<sup>4</sup>:

(1) 构成角色(constitutive role):描写一个物体与其组成部分之间的关系,包括材料(material)、重量(weight)、部分与组成成分等;也指物体在一个更大的范围内构成或组成其他物体;

(2) 形式角色(formal role):描写对象在更大的认知域内区别于其他对象的属性,包括大小(magnitude)、形状(shape)、维度(dimensionality)和颜色(color)等;

(3) 功用角色(telic role):描写对象的用途(purpose)和功能(function)。主要包括两种,一种是直接功用角色(direct telic),人可以与某物发生直接联系,如beer的功用角色是drink;另一种是间接功用角色(purpose telic),指某个事物可以用来协助完成某个活动,如knife的功用角色是cut;

(4) 施成角色(agentive role):描写对象怎样形成或产生的,如创造、因果关系等;比如,饺子是用皮和馅儿包出来的,那么饺子的施成角色就是“包”这个动作。

受 Pustejovsky 的这套物性角色的启发,袁毓林(2013, 2014)通过对大规模汉语真实文本中名词跟相关动词等的搭配关系和选择限制进行调查后发现,Pustejovsky (1991, 1995)

<sup>2</sup> 这实质上对应了“语言知识”与“非语言知识”(世界知识)的划分。而关于这一问题的讨论一直伴随着语义学的发展。早在词汇语义学发展的早期—历时语义学阶段,就有学者对如何协调这两者的关系这一问题进行了思考。比如,保罗在其《语言学原理》中就有关于语境和用法的论述;后来结构主义语义学阶段,主张语义识别完全属于语言内部层次的结构;到生成语义学阶段,非语言知识完全被排除在语义研究之外;再到新结构语义学阶段,如概念语义学、双层语义学等理论考虑如何将语言知识和非语言知识结合起来;随后的认知语义学甚至将非语言学知识全部纳入到语言知识的框架之中。关于这一问题详尽的讨论,可以参看 Dirk Geeraerts (2010)。

<sup>3</sup> 生成词库理论的原始内容及最新发展,可参看 Pustejovsky (1995, 2011)。

<sup>4</sup> 物性角色的定义源自 Moravcsik (1975)对亚里士多德“四因说”(aitia)的诠释。

的四种物性角色并不足以全面反映名词跟动词等的搭配限制，于是对物性角色框架进行了更深一步的扩展描述，将物性角色的总数增加至 10 个，以便适应汉语的语义描写和研究。除了上面的四个物性角色外，新增的六个分别是：单位、评价、行为、处置、材料、定位<sup>5</sup>。这些物性角色反映了我们人类对于名词所指事物的最为关切的内容，也即关于世界万事万物最基本、最朴素的哲学命题：是什么？为什么？怎么样？基于亚里士多德的“四因说”，物性角色体系将事物的客观属性特征与主观评价特征包装在一个描写框架之中，多维度地展示了名词自身所容纳的语义信息。这实际上是一种基于世间万物本体论的语言哲学思考，将事物跟相关事物、属性、动作、行为和评价等等之间的关系表现了出来。同时，在袁毓林（2013，2014）的物性结构体系中，相关词项的物性角色描述直接跟该词项所具备的基本句法格式挂钩，语义和句法组合的各种可能性被展示出来，为下一步自然语言处理应用提供了有关名词的句法-语义接口知识。这样一种不仅关注词项所代表事物的百科知识，同时也重视与其他语言成分相互组配的词汇知识库特别有利于“网球问题”的解决。

为了检验这种经过扩展的物性结构理论在解决“网球问题”方面的效力，我们以“网球、网球拍、网球场”和“生日、蛋糕、蜡烛”这两组基于情景联想关系的名词为例进行考察。首先，我们从北京大学已经编制的名词物性结构知识库中调出这些词条的描写信息，再检索北京大学汉语语言学研究中心 CCL 语料库，并且按照袁毓林（2013，2014）的物性结构描写体系，分别对“网球”和“生日”的词条信息（字形、语音、意义等）、物性结构和句法配置信息进行修正。首先来看“网球”的物性结构信息。

网球 02 wǎngqiú 〈名词，中性〉网球运动使用的球，圆形，有弹性。里面用橡皮，外面用毛织品等制成。

(1) 物性角色：

**形式 FOR:** 具体事物、运动器材；

**构成 CON:** 由橡皮、纺织材料（羊毛和尼龙）构成，圆形，颜色为白色或黄色，可以分为训练用球和比赛用球，等等；

**单位 UNI:** 个、只、种、堆、些、袋、桶、篮子，等等；

**评价 EVA:** 有弹性、软、硬、新、旧、重、轻，等等；

**施成 AGE:** 生产、制作、缝制，等等；

**功用 TEL:** 打、击打、拍，等等；

**行为 ACT:** 滚动、飞、弹动，等等；

**处置 HAN:** 买、卖、扔、捡、拿、踢、发现、掏出、看见、发、接，等等；

(2) 句法格式：

S1: CON+ (的+) - | -+有+CON

如：白色~ | 黄色~ | 圆形(的)~ | ~有橡皮 | ~有尼龙

S2: NUM+UNI+ -

如：一个~ | 一袋~ | 一些~ | 一桶~ | 一堆~ | 一篮子~ | 一只~ | 一种~

S3: EVA+的+ -

如：有弹性的~ | 软的~ | 硬的~ | 新的~ | 旧的~ | 重的~ | 轻的~

S4: AGE+ -

如：生产~ | 制作~ | 缝制~

S5: TEL+ -

如：打~ | 击打~ | 拍~

S6: -+ACT

如：~ (一直) 滚动 | ~飞 (起来) | ~ (上下) 弹动

S7: HAN+ -

如：买~ | 卖~ | 扔~ | 捡~ | 拿~ | 踢~ | 发现~ | 掏出~ | 看见~ | 发~ | 接~

在启发式 (heuristic) 规则 (让相关的目标概念通过名词或动词、形容词的有关语义角

<sup>5</sup> 受篇幅限制，本文未给出每一种物性角色的定义。详细内容可以参看袁毓林（2013，2014）。

色联系起来)的指引下,我们挖掘和会聚上面对“网球”物性结构和句法组合的有关描述信息,可以建立如下的概念关系网络,从而将“网球”、“网球拍”、“网球场”等概念串联在一起。

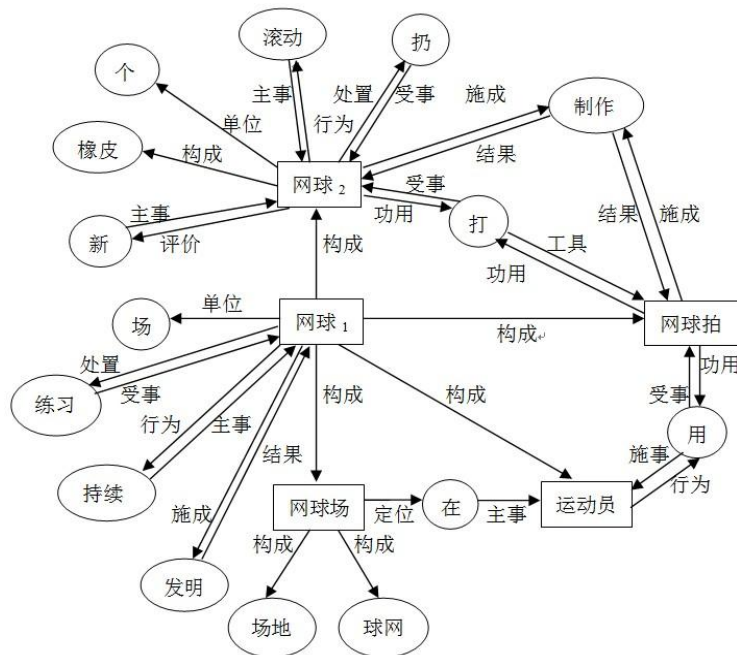


图7 “网球、网球拍、网球场”的概念关系网络

“网球”这类名词是“合成类名词”(complex nouns),既有事件义的义面(semantic facet),也有事物义的义面。在上图中,“网球<sub>1</sub>”和“网球<sub>2</sub>”分别表示“打网球”的事件和“圆形球体”的物质。基于“网球<sub>2</sub>”的功用角色“打”和施成角色“制作”,它可以与“网球拍”建立指针式的链接,因为“网球拍”也同时具有相同的功用和施成角色,从而这两个概念被牢牢地绑定在一起。同时,“网球”作为运动的事件义和作为实体的物质义之间的关系可以预先在知识库中绑定,设置“网球<sub>2</sub>”到“网球<sub>1</sub>”的自然链接,再通过利用“网球<sub>1</sub>”的构成角色,实现“网球<sub>2</sub>”与“网球场”、“运动员”等概念的相互关联。在此基础上,最终形成一个以“网球”(包括网球<sub>1</sub>和网球<sub>2</sub>)为检索核心的语义知识网络。此外,在上述不同概念的语义关联中,某些概念之间的联系由于可以通过不同途径得到链接从而得以强化。比如,“网球<sub>2</sub>”和“网球拍”,它们之间可以通过施成角色得到关联,也可以通过功用角色得到关联,还可以通过“网球<sub>1</sub>”的构成角色得到关联。随着这种关联次数和类型的不断增多,它们之间的概念联想关系也不断得到强化。由此,我们可以预测:“网球”和“网球拍”之间的语义紧密程度应该是非常高的,相对于其他事物来说,“网球拍”可能是“网球”最容易激活的事物。可见,这种语义学理论具有很强的心理现实性。

下面来看“生日”的词条信息、物性结构和句法配置信息。

生日 shēngrì 〈名词,积极〉(人)出生的日子。也指每年满周岁的那一天,即每年跟过生日相同的日子。

(1) 物性角色:

形式 FOR: 时间、节日、值得庆祝和纪念的日子;

构成 CON: 生日作为节日,隐含了庆祝这种事件。庆祝生日是一种场景定义型事件,主要由生日蛋糕、生日蜡烛、生日歌、生日礼物、生日卡片、生日聚会、生日舞会等场景要素构成;

单位 UNI: 次、个,等等;

评价 EVA: 难忘、特殊、普通、记忆深刻、低调,等等;

功用 TEL: 庆祝、纪念、过, 等等;

行为 ACT: 到来、来临, 等等;

处置 HAN: 记住、迎来, 等等;

(2) 句法格式:

S1: -+CON

如: ~蛋糕 | ~蜡烛 | ~礼物 | ~卡片 | ~歌 | ~聚会 | ~舞会

S2: NUM+UNI+ -

如: 一个~ | 一次~

S3: EVA+的+ -

如: 难忘的~ | 特殊的~ | 普通的~ | 记忆深刻的~ | 低调的~

S4: TEL+\_\_

如: 庆祝~ | 纪念~ | 过~

S5: \_\_+ACT

如: ~到来 | ~来临

S6: HAN+ -

如: 记住~ | 迎来~

同样, 在让相关的目标概念通过名词、动词和形容词等的语义角色会聚到一起这种启发式规则的指引下, 通过挖掘和组织上面对“生日”物性结构和句法组合的有关描述信息, 我们可以建立如下的概念关系网络, 将“生日”、“蛋糕”、“蜡烛”等概念串联在一起。

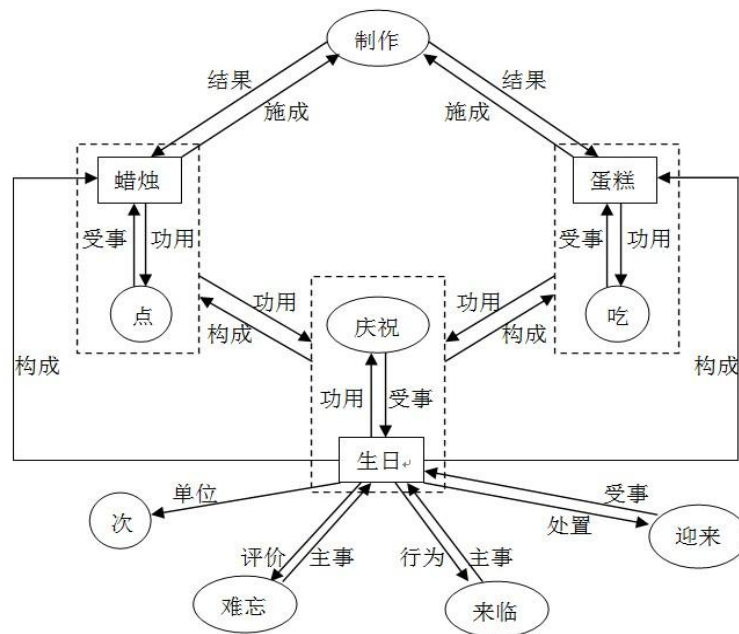


图8 “生日、蜡烛、蛋糕”的概念关系网络

上图中, “生日”和“蜡烛、蛋糕”可以通过两种途径获得关联。一方面, 作为场景定义型名词“生日”, 它的构成角色中包括“蛋糕”和“蜡烛”这类事物, 可以据此建立它们之间的直接联系, 比如“生日蜡烛”、“生日蛋糕”。另一方面, “生日”、“蜡烛”和“蛋糕”可以通过它们各自与相关动词的组合形成短语结构, 进而在短语结构的基础上建立语义关联, 实现间接的联系。比如, 以“蜡烛”(或“蛋糕”)的功用角色“点”(或“吃”)为基础形成短语结构“点蜡烛”(或“吃蛋糕”), 以“生日”的功用角色“庆祝”为基础形成短语结构“庆祝生日”; 随后进一步确认“点蜡烛”(或“吃蛋糕”)和“庆祝生日”之间的功用(或“目的”)语义关系, 即“点蜡烛”(或“吃蛋糕”)的目的是为了“庆祝生日”。同时, “庆祝生日”这一事件又是由“点蜡烛”、“吃蛋糕”等常规(routine)的子事件(sub-event)构成的。至于“蜡烛”和“蛋糕”, 一方面可以由它们各自的构成角色中都有“生日”作为

分类属性来关联；另一方面，也可以通过它们的共同施成角色“制作”获得关联。基于这样一种语义推导关系，可以为计算机理解“生日”和“蜡烛”、“蛋糕”之间的语义关系提供一种有效的知识表示和词汇语义资源。

利用物性结构知识，描绘诸如“网球、网球拍、网球场”和“生日、蛋糕、蜡烛”这种基于情景联想关系的名词的语义网络，相较于我们之前讨论过的其他词汇语义知识库有下面几个优点：

1. 物性结构知识库以名词为主要描写对象，通过对名词相关语义和句法信息的描写，构建属性、动作和事物之间相互关联的语义网络。这一做法不同于 VerbNet、FrameNet 等以动词为核心的知识库构建体系，在一定程度上是对动词句法语义信息描写的补充，对于“网球问题”的解决具有重要的作用；

2. 在物性结构知识库中，相同的语义信息是不同的词语相互关联的纽带，知识库网络的构建得益于这些具有链接作用的语义信息。在这种词汇本体知识库中，可以以特定词语为检索项寻找不同概念之间的语义关系。这弥补了作为单纯性情景联想关系网络的概念知识库 ConceptNet 不便于词汇检索和查找的不足；

3. 物性结构知识库一方面关注不同概念之间基于世界知识的语义联系，同时也注重词语之间的组合关系，尤其是各种不同物性角色与名词在句法配置上的组合信息。跟单纯的基于常识知识的 ConceptNet 相比，物性结构知识库在对语言事实的挖掘和呈现上更加深入，在不同概念之间的关联方面也更加精细；

4. 物性结构知识库为不同概念之间设置了动作指针链接，即特定的动词将两个名词关联在一起，形成了“谓词-论元”式的语义关系图式。这样做，一方面便于跟文本中相关的句子进行匹配；另一方面，也便于跟 VerbNet、FrameNet 等以动词为核心的词汇知识库进行整合。

5. 我们早先已经开发出了汉语谓词（动词、形容词）论元结构知识库，正好可以跟现在正在开发的汉语名词物性结构知识库整合；在有关目标词汇概念的驱动下，挖掘和发现相关词汇概念之间的语义联系和常识性推理关系。

## 7 小结

词汇语义知识库作为自然语言处理的一种特殊的知识资源，其建立的目的是要提供一个大规模的词汇语义及相关世界知识的知识库，这个知识库应该能够较好地揭示词汇概念与词汇概念之间以及概念与属性之间的关系，从而为文本语义的计算分析提供可靠的基础。不过，由于不同的语义资源是由不同的研究者为了不同的理论或应用目的开发出来的，在描写的语义内容、呈现格式等方面具有不同的类型；因而，就有了如何将这些不同的语义资源相互取长补短、统一并整合起来的问题（袁毓林 2008）。

目前比较主流的几种语义知识资源库，在设计和应用方面都有自己的特点。比如，WordNet 注重从聚合的角度对词语之间的语义关系进行描写，为我们提供了极为丰富的词汇语义信息，这些信息对于自然语言处理中的语义分析是非常有用的；VerbNet 在动词的句法和语义之间建立起一种有效和有用的链接，把句法-语义上具有相似表现的动词聚成一个类别，提供了更多的句法和语义的连贯性，并且建立了与其它词汇资源的映射（Loper 2007；Loper et al. 2006）；基于框架语义学的 FrameNet 在语义框架下验证相关的一批词在某种意义上语义和句法结构的各种表现形式，为使用者提供了一部内容丰富的语义和用法词典，同时在信息抽取、问题回答等自然语言理解领域也得到应用；建立在常识推理基础上的 ConceptNet，将日常生活中的百科知识纳入到其构建的基于情景联想的概念网络之中，为主题发现、文本分类、情感评价等提供了有用的知识库。但是，这几种语义知识库在解决“网球问题”方面，都存在一定程度上的局限性。本文建议，在谓词（动词、形容词）的论元结



构知识库的基础上,再以生成词库论中的物性结构知识为理论框架,并配置相关词项的句法组合信息,来建构名词的物性结构知识库,最后把谓词与名词的语义角色关系双向连接来解决“网球问题”。这样,通过利用不同词语(名词、动词、形容词)所共有的语义信息作为指针链接,构建相关词语之间的概念关系网络,把事物和跟事物相关的事件的有关世界知识及其语言表达形式表示出来,最终形成以名词(实体)为检索核心的、面向对象(object orientation)的语义知识库。

幸运的是,目前发展迅速的知识图谱(knowledge graph)技术正好为上述语义会聚(converge)和关联(connection)提供了知识表示和数据库技术支持。

### 参考文献:

- [1] Fellbaum, Christiane (ed.) WordNet: An Electronic Lexical Database [M]. Cambridge, Massachusetts: MIT Press, 1998.
- [2] Miller G. A., Fellbaum, Christiane. WordNet then and now [J]. Lang Resources & Evaluation, 2007, 41: 209 - 214.
- [3] Miller G. A., Beckwith Richard, Fellbaum Christiane, Gross Derek, Miller K. J. Introduction to WordNet: An On-line Lexical Database [J]. International Journal of Lexicography, 1990, 3(4): 236 - 244.
- [4] 黄居仁. 语意网与中文信息化前瞻: 知识本体与自然语言处理[C]//孙茂松、陈群秀. 自然语言理解和大规模内容计算. 北京: 清华大学出版社, 2005: 1 -10.
- [5] 袁毓林. 语义资源建设的最新趋势和长远目标——通过影射对比、走向统一联合、实现自动推理[J]. 中文信息学报, 2008, 22 (3): 3 - 15.
- [6] 索绪尔. 普通语言学教程[M]. 北京: 商务印书馆, 1980.
- [7] Warrington, E.K., Shallice, T. Category specific semantic impairments [J]. Brain, 1984, 107(3): 829 - 853.
- [8] Lin, E. L., Murphy, G. L. Thematic relations in adults' concept [J]. Journal of Experimental Psychology: General, 2001, 130(1): 3 - 28.
- [9] Altmann, G. Thematic role assignment in context [J]. Journal of Memory and Language, 1999, 41 (1): 124 - 145.
- [10] Warren, T., McConnel, K. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movement in reading [J]. Psychonomic bulletin & review, 2007, 14: 770 - 775.
- [11] Kipper Karin, Dang H. T., Palmer Martha. Class-Based Construction of a Verb Lexicon[C]//AAAI-2000 Seventeenth National Conference on Artificial Intelligence, Austin, TX, July 30 - August 3, 2000.
- [12] Fillmore C. J., Johnson C.R., Petruck M.R. Background to FrameNet[J]. International Journal of Lexicography, 2003, 16 (3): 236 - 250.
- [13] Fillmore C.J. Frame Semantics[C]//Linguistics in the Moring Calm. Seoul: Hanshin Publishing Co., 1982:111 - 137.
- [14] Baker, F. Collin, Charles J. Fillmore, John B. Lowe. The Berkeley FrameNet Project[C]//Proceedings of the 17th International Conference on Computational Linguistics [COLING '98] / the 36th Annual Meeting on Association for Computational Linguistics [ACL '98], Montreal, Canada, 1998: 86-90.
- [15] Henderson Harry. Milestones in Discovery and Invention: Artificial Intelligence [M].

- Shanghai: Shanghai Science & Technological Literature Publishing House, 2007.
- [16] Liu, H. & Singh P. ConceptNet – a practical commonsense reasoning toolkit [J]. *BT Technology Journal*, 2004, 22(4): 211 - 226.
- [17] 袁毓林. 信息抽取的语义知识资源研究[J]. *中文信息学报*, 2002, 16 (5): 8 – 14.
- [18] Pustejovsky, James. The Generative Lexicon [J]. *Computational linguistics*, 1991, 4: 409 - 441.
- [19] Pustejovsky, James. *The Generative Lexicon* [M]. Cambridge, Massachusetts: MIT Press, 1995.
- [20] Pustejovsky J., P. Bouillon, H. Isahara, K. Kanzaki, C. Lee (eds.) *Advances in Generative Lexicon Theory*[M]. Dordrecht: Springer, 2013.
- [21] 吴国向, 袁毓林. 词汇本体语义学的实证性研究[R]. 第五届当代语言学国际圆桌会议, 南京: 南京师范大学, 2013年10月.
- [22] Geeraerts, Dirk. *Theories of Lexical Semantics* [M]. New York: Oxford University Press Inc., 2010.
- [23] Pustejovsky, James. Coercion in a general theory of argument selection [J]. *Linguistics*, 2011, 49 (6): 1401-1431.
- [24] Moravcsik, J. M. Aitia as generative factor in Aristotle's philosophy [J]. *Dialogue*, 1975, 14: 622 - 636.
- [25] 袁毓林. 基于生成词库论和论元结构理论的语义知识体系研究[J]. *中文信息学报*, 2013, 27 (6): 23 – 30.
- [26] 袁毓林. 汉语名词物性结构的描写体系和运用案例[J]. *当代语言学*, 2014, 1: 31 – 48.
- [27] Loper, Edward. PropBank, VerbNet & SemLink[R]. Joint Meeting of the ACL/SIGSEM Working Group on Representation of Multimodal Semantic Information and the ISO Task Domain Group on Semantic Content Representation. Tilburg, the Netherlands. 2007.
- [28] Loper, Edward, Szu-ting Yi, Palmer Martha. Combining Lexical Resources: Mapping between PropBank and VerbNet[C]//Proceedings of the 7th International Workshop on Computational Linguistics. Tilburg, Netherlands. 2006.