

汉语动词资源馆 (Chinese Verb Library) 的构建*

汪梦翔^{1,2} 王厚峰¹ 刘杨¹ 饶琪³

(1.北京大学, 北京 邮编 100871; 2. 北京联合大学, 北京 邮编 100011; 3. 华中师范大学, 湖北武汉, 430079)

摘要: 本研究以动词的语义聚合层次为核心, 构建了含有四个语义层级的动词分类系统, 依托生成词库理论、语义格框理论和构式语法理论, 从事件结构、语义格框架、物性角色、句法格式四个角度来对动词的自身属性以及组合性等特征进行描述, 从而构造出能够解释并深入描述动动之间、动名之间甚至是超常搭配用法的汉语动词资源馆 Chinese Verb Library (CVL)。实验表明, 本研究可以为句法分析、语义角色标注、尤其是揭示隐含的谓间关系提供新的支持。

关键词: 汉语动词资源馆; 生成词库论; 事件结构

The Construction of Chinese Verb Library

Abstract: This research uses verbal semantic hierarchy as the core element to construct a verbal classification system which contains four semantic levels. With the theoretical supports from generative lexicon theory, semantic frame theory and construction grammar, we describe the attributes and combinatory features of verbs from event structure, semantic frame, qualia structure and syntactic format. The aim of research is to construct a Chinese Verb Library (CVL) which can be used to make an adequate explanation and in-depth description of “V+V” constructions, “V+N” constructions, and other creative collocations. Experiment results show that our research can support the syntactic analysis, semantic role labeling, especially can reveal the implicit predicate relationship.

Keywords: Chinese Verb Library; Generative Lexicon Theory; Event Structure

0 引言

我们先看两个句子:

(1) 我在买菜

(2) 我在赶论文

一般情况下, 目前计算机要理解和处理这两个句子, 必须理清其中谓词和其所支配成分间的语义关系, 而要准确的反映这些语义关系, 传统做法是在确定核心动词的前提下, 通过对其所支配的成分进行相关的标注, 来反映动词和论元的关系。到底如何才能准确标注, 这就需要建立一个以动词为核心的相应的语义资源, 陈列各种语义关系, 来为计算机提供选择的余地。传统的动词语义资源都是以动词为核心, 着眼于动词和其所支配成分间关系的刻画。但是一旦这种核心动词缺省或者隐含, 这种情况就无法通过计算机准确的表达出来。比如(1)句中的核心动词是“买”, 在相应的语义资源中“买”的支配对象可以用“受事”表示, 句中的“菜”直接标为“受事”就能够正确反映这种支配和被支配关系。但是(2)句中动词“赶”和“论文”没有直接语义关联, 其内含语义关系应该是“赶”和“写”, 以及“写”和“论文”, 但这个动词“写”隐含了, 自然这种关系就不能显现。

目前国内大部分语义资源都不能反映和再现谓词间隐含的语义关系, 而国外的Pustejovsky 等人所构建的 Brandeis Semantic Ontology (BSO) 虽然能够揭示诸如 “begin the novel” 这样的谓词隐含现象, 但这一数据资源库主要是面向英语, 而且缺乏句式的描述, 所以本文就是想在前人的基础上, 构建一种动词语义资源: 它能够集句法形式与语义分析于一体, 不仅能够反映汉语中常规的动动以及动名关系, 还能够找出非常规的、带有隐含的谓词逻辑关系。

* 收稿日期:

定稿日期:

基金项目: 本文为国家社科基金重大项目 (编号: 12&ZD227)、国家自然科学基金 (编号: 61370117, 61333018) 和教育部人文社科研究基金 (编号: 12YJC740085) 部分研究成果。

1 CVL 的框架搭建及层级系统

1.1 数据来源

本资源库综合了《动词大词典》、《现代汉语动词大词典》、《动词用法词典》的数据，包括拼音、释义、义项、语义角色、例句等，另外又额外补充了自己搜集的近 200 个常用动词，一共 2506 个动词，3299 个词项。在语义角色和句法格式的描述中，主要参考了《动词大词典》的例句，对于一些词典中没有搜录的词，我们在描述时，主要是通过 2002—2012 年电子版的人民日报和 CCL 语料库中的例句资源，还有 Chinese Giga Word 的 Xinhua News，此外还有部分数据来源于作者的语感，为自造例句。

1.2 基本框架

为凸显动词内部关联，多角度的描写动词间、动名间的常规或非常规关系，本文采用树形的动词语义表示框架，外部将借鉴 VerbNet 语义层次构建平台，聚合带有相同语义或句法功能的动词集，并以此作为父节点，内部将通过事件结构、物性结构、论元角色、句法格式的描写来反映集合内部成员的各方面特征，具体框架如图所示：

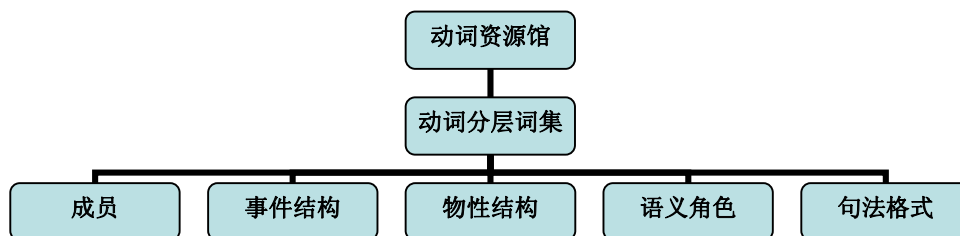


图 1：CVL 构建框架

1.3 CVL 的层级系统

我们把动词划分为四个层级。

第一层，我们主要依据动词的句法语义属性，把动词分为：行为动词、心理动词、交际动词、关系动词四类。

第二层，我们在依据语义聚合关系的同时还根据一些动词的句法特征的有无，分为创造、活动、动作、认知、言语交际等 20 类。这 20 类和第一个层次的继承关系如下表所示：

表 1：中级词集层次关系表

动词																			
行为动词								交际动词		心理动词					关系动词				
活动	现象	动作	移动	支配	作用	消亡	创造	变化	言语交际	行为交际	情感	情绪	认为	认知	态度	意愿	领属	系属	形式

在这 20 个类别中，有 9 种是属于行为动词的。

第三层，除了关系动词外，我们主要根据语义的近似度，来对上一层词集进行分类，将上一层的 20 类扩展为 646 个小类。其中行为动词最多，聚合了 493 类，其次是交际动词，聚合了 108 类，再次是心理动词，聚合了 38 类，关系动词最少是 7 类。

第四层，词的层次，含有 2506 个词，3299 个词项。

1.4 动词特征的知识表示

在对动词特征的发掘和规约上，我们采用了自顶向下和自底向上两种模式。比如对于交际动词，其语义角色的基本框架一般为“施事+对象”，那么如果自顶向下发掘的话，下属成员词集的语义角色基本框架都会有“施事+对象”，比如“学习类”和“帮助类”基本框架都含有“施事+对象”，这样它们就可以很容易和其他类动词进行区别，但是内部成员的差异性不容易凸显。如果是自底向上的描述比如“学习”和“帮助”这两个词，在最低级别的语义角色描述时，其语义角色基本框架分别为“施事+对象（受事）”和“施事+对象”，以这两个特征我们可以聚合一些词语，形成“学习类”和“帮助类”，“学习类”的词都是“施事+对象（受事）”，“帮助类”的词则都是“施事+对象”，如果再往上一层级聚合，那么其能作为基本特征保留的只有“施事+对象”，这一特征也是交际动词的基本框架。这样做虽然可以很好的凸显成员的差异性，但最底层词语的特征不易规约一致，且越往上丢失信息越多，且工程量也较大。

我们的做法是，先通过自顶向下规约基本特征，然后通过自底向上的描述补充差异性特征。具体思路下图所示：



图 2：动词特征表示思路

这样一方面保持了底层成员规约特征的一致性，凸显了成员间的共性特征，使得其和集合外成员进行有效区别，另一方面是也考虑到一些集合内部非典型性成员的个性化特征，并且让这些个性化特征处于某一基本特征的管辖。

2. 动词事件结构及事件动词描述

2.1 动词事件结构的描述

在动词事件特征知识库中，我们主要涉及动词的事件结构和是否为事件动词两个特征。

这里的事件结构 (Event Structure) 特征主要涉及事件内部经常会有事件的起始、持续、终结等以时间为衡量的特征。

我们知道动词内部是含有时间结构的，比如“开会”和“打”，一个是持续性动作，一个是瞬间动作，它们所隐含的时间特征有差别，所以这两个动词从时间的维度上或者是体结

构 (Tense) 维度上划分为两个不同的类别。同样, 事件内部也有时间结构, 而且事件结构主要就是指事件的起始、度量和界化 (delimitation) 等时间结构特性。因为事件主要靠动词表现, 所以事件结构和动词的体结构关系紧密。

关于动词事件结构的划分, 因为划分的角度不一样, 目的不一样, 自然划分的结果也不一样。我们结合前人的思想, 决定依据动词的内部时间情态来和动词的界性特征^①来对动词所表达的事件结构类型进行划分, 本人认为可分为三类: 状态 (state type)、过程 (process type) 和转变 (transition type):

状态: 这种事件结构类型, 一般是表示事情的起点或终点, 有时起点和终点合一。如果要细分的话, 还可以分为三种, 一种是起点状态, 表示事件的起点, 如“开始”, 一种是终点状态, 表示事件进行到终点的状态, 如“结婚”, 还有一种是瞬间状态, 表示没有明确的起始点, 也没有明确的终点, 或者说起点终点重合, 如“死”。一般这类事件结构所对应的动词只能出现一次 (死、结束等) 或者必须经历一段比较长的时间以后才可以再次出现 (结婚、离婚等), 一般可以带“了”, 也可不带 (如“发愁”), 但是一般不能带“着”。典型的是“死、喜欢、盼望、热衷 (于)、知道、主张、提倡、强调、放弃”等。

过程: 这种事件结构类型, 一般表示动作由发生到结束的中间过程, 一般不包括时间的起点和终点。表达这种事件结构的动词, 仍然是瞬间动词的另一个类别, 而且是无界的, 后面可以接结果性补语 (如“摔破”), 可以在短时间内连续出现, 如果是双音节形式都可以进入“对 NP2 的 V1”格式 (如“对疾病的预防”), 其有界性特征较前面两类要强, 前面可以接“(正) 在”, 后面可以接“着”或“了”。例如“跑、杀、学习、防止、尝试、预防”等。

转变: 这类动词的时间特征具有两面性, 一方面可以处于进行状态, 具有“无界”的特征, 因此可以前加“(正) 在”, 有时可以加“着”, 如“正在建房子/房子正建着”。但另一方面, 可以向终极时间点靠近, 即将达成一种完结状态, 一般不能跟“着”连用, 如“建了一栋房子”。这类词语主要为: 建、写、制订、安排等。

基本上, 每一个事件结构都有相应特征的动词来与之对应, 但是反过来对每一个动词来说不一定了。因为研究者发现, 事件结构除跟动词静态的语义特征或时间特征有关外, 在使用过程中, 还会受到一些外在因素的影响, 比如是否是动补结构, 是否有定, 是否带有时间助词等等, 因此我们可以说动词的事件结构特征是组合性的。提取过程中, 我们参考了《语法信息词典》^②中关于动词的一些特征 (主要是通过是否能添加一些表示时效性的成分), 如:

表 2: 动词事件结构特征表

事件结构	着	了	过	(正) 在	动结	后时量	后动量	例词
状态	-	±	±	-	-	+	-	死、喜欢
过程	±	+	±	+	±	-	+	跑、学习
转变	±	+	±	±	+	+	±	建、制定

需要指出的是, 汉语的形式表现和语义内涵并不是完全对应的关系, 这种形式上的提取标准如同时满足则比较严格, 而《语法信息词典》在描述这些信息时并不是完全正确, 容易造成一定的数据稀疏。因此事实上, 我们还要看动词的语义特征, 包括[持续性]、[动作性]、[完结性]等, 来判断动词所对应的事件结构。

另外需要注意的是, 虽然每一种事件结构都有相应的动词与之对应, 但是对于某一个动词来说, 它所传达的事件结构并不一定是特定的某一种, 也有可能可以传达两种事件结构。比如“屠杀”从形式上判断应该是“过程类”动词, 但是从语义上讲, 还包括“死”这个状态, 所以它的事件结构可以表达过程, 也可以表达状态。Pustejovsky 在描述“kill”的事件结构时就认为 kill 可以传达两种事件结构类型, 一个是 Process (过程), 一个是 State (状态)。

2.2 事件强迫与事件动词

^① 界性特征指的就是有界和无界的特征。有的动词界性特征强, 那么它的有界性就强, 界性特征弱就倾向于无界性。根据税昌锡 (2005) 的研究, 他依托动词的界性特征, 把行为动词分为七种: 起始动词 VP₁ (initiation verb)、持续动词 VP₂ (duration verb)、活动动词 VP₃ (activity verb)、跨界动词 VP₄ (transboundedness verb)、事件动词 VP₅ (event verb)、达成动词 VP₆ (achievement verb)、完结动词 VP₇ (Accomplishment verb)。界性特征 (boundedness feature) 从 VP₁ 到 VP₇ 依次增强。

^② 即使《语法信息词典》收录了近 2 万多个动词, 但是实际只覆盖了我们词库中 3192 条词项, 还有 107 条词项的语法信息它没有收录。

事件强迫是 Pustejovsky 生成词库理论的一种生成机制。Pustejovsky (2006) 认为词义单独来看是相对稳定的,但到了句子层面,一般会通过一些生成机制(分别是:纯粹类型选择 (pure selection)、类型调节 (type accommodation) 和类型强迫 (type coercion))获得延伸意义。事件强迫就是类型强迫的一种。例如,begin 要求其宾语是个事件论元,句法上通常表现为一个 VP。但有时出现“begin the novel”这种非 VP 作宾语的情况,因此 begin 就会强迫这个 NP 进行类型转换 (type shift),由一个物体名词变成临时的事件名词,这就是事件强迫机制。

事件强迫其实就是一种语义压缩形式,一些超常搭配往往是事件强迫的结果。比如“赶论文”就是“赶”和“写论文”两个事件的压缩。

事件强迫要发生,一般要满足几个条件,首先语义上,一般带有谓词隐含,且隐含的谓词多为虚义的轻动词,如“赶(做)论文”中“做”为隐含谓词,而且意义比较虚,可理解为“写”。其次是句法形式上,一般是出现在动宾结构中,比如“学钢琴”、“喜欢饺子”。另外,对动词也有要求,动词一般必须能够接 VP 作宾语,如“写”这类动词不能接 VP 作宾语,那么就不可能存在事件强迫现象。之所以要提到事件强迫,那是因为只有事件动词才具有事件强迫的功能。

不过对于事件动词 (eventive verb),学界的认识还比较混乱。一种外延比较大,认为事件动词是相对于性质动词 (property) 或关系动词来说的,比如陈平 (1988);一种外延相对较小,指的是一般行为动词内部一部分具有特殊时间性特征的动词,比如税昌锡 (2005) 就是把动作时间开始转向终止点的有界动词称为事件动词,可以表示一个完整的事件,具有“完结”义。宋作艳 (2011) 认为事件动词应该是不能独立表示一个完整事件,但可以引发事件,因此是不具有“完结”义的动词。我们觉得前者应该称为“事件性动词”,后者才是真正意义的“事件动词”。

一般情况下,事件动词因为可以后接事件,因此具备引起事件强迫的可能,但不是事件强迫发生的充分条件。比如“学英语”和“学钢琴”。这里“学”是事件动词,但是一个有事件强迫,一个没有。

对于事件动词的判定,我们不能根据是否有“完结”义来判定,因为宋作艳 (2011) 就将动词“完成”视为事件动词,只不过宋作艳认为“完成”是表示“达成”(achievement)。其实事件动词区别于其他动词的最典型特征是可以后接由 VP 构成的事件,并且具有事件强迫功能。因此本文认为:如果一个动词不能单独描述一个事件,必须借助其他动词或者依靠隐含动词的帮助才能描述一个完整事件,那么这样的动词就称为事件动词。这里要指出的是,谓词的隐含不一定直接放在动词之后,如“避免交通事故”我们可以说,“避免交通事故的(发生)”。

因此形式上,只要满足以下两个条件任一个,我们就可判定 V_1 为事件动词:

- A. $NP_1+V_1+(V_2)+NP_2$ B. $NP_1+V_1+NP_2$ 的 V_2

我们据此从 3299 个词项中,确定了 213 个动词为事件动词。提取事件动词的目的就是为了解释或描述一些带有省略的超常搭配组合,并且把它们内在的关联进行还原和补充。

3. 语义角色描述

我们根据林杏光 (1994) 的研究成果以及结合实际标注语料的需要,提取了一个含有三个层级,并带有 14 种语义角色的体系。在这个体系中,我们对林杏光语义角色体系进行了某种程度的规约,去除了“与事格”,提出了“对象格”^③,并对其他角色格和情景格进行了一些整合,从数量上减少了 8 个,这主要是从工程的角度来考虑。因为语义角色的设置要充分保证在实际语料标注过程中的可操作性。虽然从科学研究的角度来讲,我们应该把区分每一类的语义角色作为终极目标,但是对于目前的水平来说,过于精细的划分在增加工作量的同时,还不能保证标注的准确率。而向上规约自然可确定性就要高。需要指出的是,这不是此项研究的终点,随着研究的开展及人力、物力的跟进,以后还会进一步在此基础上细化,可以说这一体系为日后语义角色的深入研究留有了一定空间。

表 3: 本项目语义角色标注体系和林杏光语义角色标注体系比较

^③关于“对象格”的具体内涵可以参考汪梦翔 (2012)、(2014)

本人	角色											情景										
	主体			客体								凭借		环境			缘由					
	施事	领事	主事	受事	结果	对象			系事		工具	方式	时间	地点	范围	原因	目的					
林杏光	施事	当事	领事		受事	结果	客事	与事	同事	分事	系事	基准	工具	材料	方式	时间	方向	处所	范围	原因	依据	目的

另外，在动词语义角色的描述过程中，我们参照了鲁川《动词大词典》的框架，对动词的语义角色框架进行了基本式和扩展式的描述。基本式一般描述动词和其所接的必有论元的关系，而扩展式一般描述动词和非必有论元的关系。

需要指出的是，必有论元并不是简单指施事、受事、对象、结果等核心角色，非必有也不是单指时间、地点、原因等外围角色，只要这个语义角色是动词的一个“价”，而且足以描述某个动词的语义角色关系特征，那么就可以作为必有论元，我们就可以把它纳入到这个动词的基本式中进行描述。比如，对于动词“削”，它的基本式是“施事+V+受事”，如“我[施事]削了一个梨[受事]”；而对于动词“住”，它的基本式就可能是“施事+V+地点”，如“我[施事]住酒店[地点]”，这里的“酒店”虽然是“地点”，但是也是必有的论元，可以体现这个动词的语义角色特征，所以必须纳入到基本式中。

而扩展式虽然针对的是非必有论元，但也不一定只能由非核心论元充当，比如，“小李为张三买了一件衣服”中“张三”是“买”的“对象”，但不是必有论元，这就应该归入到“买”的扩展式中。所以无论是基本式还是扩展式在语义角色的选择上没有严格的界限。我们选择的标准参照了《动词大词典》，基本式或扩展式的例句的来源，部分来源于《动词大词典》，部分是为人为造句，然后再根据我们的语义角色体系，进行逐一描写。

4. 动词物性角色描述

物性角色的相关概念及描述主要来源于生成词库理论（Generative Lexicon Theory, GLT）。这一理论最初是为了回答为什么会出现类似“begin the novel”这类虽不和语法但可以为人们所运用和理解的句子。而这类现象传统语法不能解释，因为传统语法对词义的描写是静态的列举法、分义项，这样就阻止了词义的渗透性，反映不了创新性用法的变化轨迹，从而无法还原词义在上下文的变化。为了解释词的不同意义及其在上下文中的创新用法，生成词库理论为每一个词项设计了比较立体的语义表达平面，包括四个层面：论元结构、事件结构、物性结构和词汇继承结构。

需指出的是，生成词库理论中的四种语义表达层面中，一般可能认为论元结构和事件结构考察动词的语义特征显得较为合适，而物性结构比较适用于描述和考察名词，因为名词内部本身带有一定的继承性和物性。但事实上 Pustejovsky 所创立的物性结构描写体系是适用于所有词项的（不仅限于名词），而利用生成词库理论中的物性结构来描述汉语中的动词，目前在汉语学界做的还不多。

对于汉语动词的物性角色，我们参考了 Pustejovsky 的描述体系，对其形式角色、构成角色、功用角色、施成角色进行了描述，并额外增加了动词的情感角色描述。所谓情感角色，就是对动词所传达出来的感情倾向。一般分为积极、消极、中性。比如：“尊敬”的情感角色就是积极，而“勾结”的情感角色就是消极，“尝试”的情感角色就是中性。对于动词情感色彩的判定，我们主要依据人的感知。我们采取了人工双盲标注，对每一个词语进行推敲，然后再挑出有分歧的词语，从适用对象、实际句子语料、以及同义、反义对照的形式，共确定积极动词词项 330 个，消极动词词项 458 个，中性动词词项 2511 个。它们的分布比例如图所示：

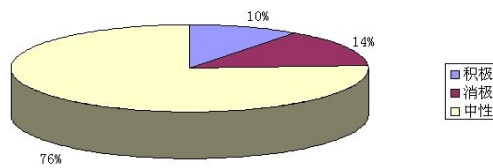


图 3: 动词情感角色分布图

而对于动词的形式角色、构成角色、功用角色、施成角色我们做了相应的规定，其内涵和 Pustejovsky 最初的定义有所出入。

在生成词库论的理论体系中，词项的形式角色一般指的是词项的本体属性（ontology）。比如 Pustejovsky 在描述“artifact”和“doctor”的形式角色时，主要突出的是它们的归属，分别是“实体（entity）”和“人（person）”

因之前我们在动词层级划分时就对动词的本体属性有某种程度的涉及，所以在这里，形式角色的内涵主要是指动词自主性特征。

对于功用和施成角色，我们主要从动词所接 NP 来进行表现。我们认为：动词的功用角色，可以理解为动词后所接的受事 NP；而动词的施成角色，可以理解为动词的结果 NP。比如：“熬”的功用角色可以是“骨头、树叶”等，而施成角色可以是“汤、稀饭”等。动词的功用角色和施成角色都是通过动词所接的名词性宾语来体现。只不过一个是支配，强调的是作用；一个是转变关系，强调的是结果。为避免数据的繁杂，我们在提取时，对名词性宾语做了相应的规约，比如，就施成角色来说，熬₍₁₎和熬₍₂₎所接的名词性宾语分别为“汁_粥_汤_羹_清汤_膏_稀饭”和“公婆_婆_精”，规约为“人工物”和“人”。如图所示：

ID	高层词集	中层词集	低层词集	动词	注音	释义	施成角色
31	行为动词	创造	生产类	熬	ao2 (1)	把粮食等放在水里，久煮使成糊状或有较浓的汁液。	汁_粥_汤_清汤_膏_稀饭_茶水_油_农药_烟_猪油
32	心理动词	态度	忍受类	熬	ao2 (2)	忍受(多指艰苦的生活)。	公婆_婆_精

↓

ID	高层词集	中层词集	低层词集	动词	注音	释义	施成角色
31	行为动词	创造	生产类	熬	ao2 (1)	把粮食等放在水里，久煮使成糊状或有较浓的汁液。	人工物
32	心理动词	态度	忍受类	熬	ao2 (2)	忍受(多指艰苦的生活)。	人

图 4: 动词“熬”的施成角色描述

对于构成角色的描述方式和角度，Pustejovsky 缺乏一定的规范，针对这点，我们就直接把动前出现的名词作为构成角色的数据源。所以本项目所描述的构成角色和生成词库中的构成角色内涵有所差别，其所指也不同。

我们可以把这几种物性角色的内涵简单归纳为：

表 4: CVL 的几种物性角色内涵表

动词物性角色	对应的内涵	举例
形式角色	自主性特征	跑（自主）
构成角色	动前 subject，为施事关系	（生物）跑
功用角色	动后 object，为受事关系	吃（生物）
施成角色	动后 object，为结果关系	建（人工物）
情感角色	词语的倾向性（积极、中性、消极）	尊敬（积极）

5. 动词句法格式描述

汉语的句法格式研究主要依托的理论是构式语法(Construction Grammar)的相关理论。本文在汉语动词句法格式的提取上,一方面基于一定的数据库语料,一方面参考了前人的研究成果,比如《动词大词典》的数据。

在句式的判定和划分上,我们首先以论元为依托。汉语中一般最多有三个论元参与到事件中,即主体、客体和邻体。需要注意的是,这种句法层次的划分和语义角色的划分不是一个层次。虽然有它们某种对应关系(主体一般由施事格充当,客体一般由受事格充当,邻体一般由对象格充当),但是也有可能一些外围角色成为主体或客体。比如“墙上挂了一幅画”、“飞机直达伦敦”,这里“墙上”和“伦敦”都是地点,但是一个是“挂”的主体,一个是“直达”的客体。而且邻体是从形式上来说的,专门针对第三个论元的情况,是在客体论元不够的时候,为区别客体而设置的,因此不可能出现“主体+邻体”的情况。

另外汉语在使用中,往往包括一些省略,我们判断一个动词的主体和客体还是要依据句法的最大自足性原则,也就是在尽量补足主体和客体的原型句式基础上,再考察一些自足的句式。比如:“钱包丢了”,这句虽然能说,但不是最完整的句法结构。而“我丢钱包了”这里主体和客体都有,因此作为判断主体客体的原型或基本句式,这样来看的话“钱包丢了”这个句式中,“钱包”就不能看做是“主体”,而应该看做是“客体”。

在汉语中,一个动词所联系的主体、客体,有时还有邻体,在表层结构中,可以有多个位置,因此构成了各种句子格式。比如“吃”的主体可以在句首或句中。如,“我吃了苹果、苹果我吃了,苹果被我吃了”。根据排列组合的原则,如果主体、客体、邻体和动词这四个成分组合的话有24种,但是实际上只有11种,很多情况不符合汉语的说话习惯,比如规则的汉语句子里,一般动词是不出现在句首的。

为彰显句法格式变换的系统性和区别性,我们是在三种常见句型的基础上,再最大限度的罗列了可能的变换句式:

I 无客式(不带客体):

001 主体+V 客人来了(常规)

002 V+主体 来客人了

II 带客式(带一个客体):

003 主体+V+客体:我吃了三个苹果(常规)

004 客体+V:苹果吃了

005 主体+客体+V:我把苹果吃了/我苹果已经吃了

006 客体+主体+V:苹果被我吃了/苹果我已经吃了

III 一客一邻式(带一个客体一个邻体):

007 主体+V+邻体+客体:我送他一支笔(常规)

008 主体+V+客体+邻体:我送一只笔(给)他

009 主体+客体+V+邻体:我把一支笔送给他

010 主体+邻体+V+客体:我向老师请教一个问题

011 客体+主体+V+邻体:书(被)我送给他了

这里我们参考了《动词大词典》的句法变换格式,但是我们也增加了一些《动词大词典》没有的句式,比如“客体+V”、“主体+邻体+V+客体”。

另外,《动词大词典》中把有无介词作为划分句型的条件之一,比如同是“主体+客体+V”,《动词大词典》就认为有“主体+客体+V”和“主体+介词+客体+V”两种。我们在这里没有采用,主要是因为正规的书面语中,一个论元如果不在常规位置,或者需要由句首或句末位置移至句中时,一般都会加入形式标记进行区别(比如“我吃了苹果→苹果被我吃了→我把苹果吃了”,而“苹果我吃了”这样的句式在口语中比较常见)。在我们的变换体系中,主要是凸显主、客、邻体的位置变换,而且这种变换已经把介词的因素考虑在内,因为有些论元移入到句中时必须加介词,我们出于人力物力的考虑,现阶段暂时不根据介词的有无再进行句式的细分。不过,在以后的研究中,如果有需要,我们还是有必要将某些句式根据介词的有无或介词的种类再进行进一步的划分。

6. CVL 的应用

6.1 在句法分析中的应用

在面对同一种组合形式的词组和句子时，我们可通过动词的一些特征来进行句法分析，因为动词的特征是需要依赖名词来体现的，所以我们必须要结合名词的相关语义知识表示。

我们来看两个形式一样，但句法构造不同的句子：

(1) 修理汽车的工人

(2) 搜查士兵的背包

这两句构造都是“VP+NP+的+NP”，但一个是偏正结构，一个是动宾结构。如何让计算机区别，我们可以借助这两动词在本资源库中的物性特征描述。

“修理”的功用角色为：人工物

构成角色为：人

“搜查”的功用角色为：具体物

构成角色为：人

然后我们需要结合名词的语义类属性来匹配，这里我们借助的是北大中文概念词典中名词的语义类属性：

工人[语义类:人] 汽车[语义类:人工物]

背包[语义类:具体物] 士兵[语义类:人]

这样，“修理”只能支配“汽车”，而不是“工人”，而“搜查”只能支配“背包”，而不是“士兵”。

具体分析过程如下图所示：

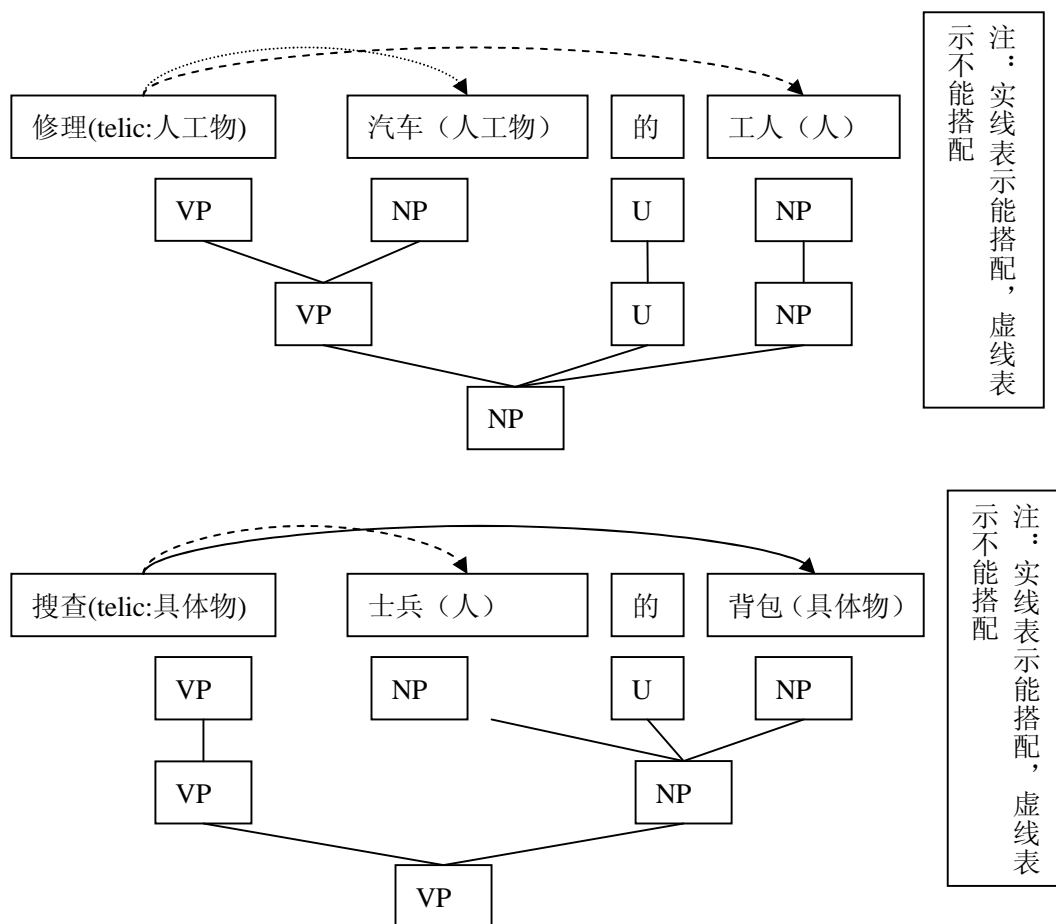


图 5：动词物性角色在句法分析中的应用

6.2 对动宾搭配中谓词省略现象的判定和还原

判定某些动宾搭配是否带有省略谓词现象，我们可以依据我们语义资源库的两个特征：一个是事件动词，一个是物性角色。其中是否为事件动词，是是否带有谓词省略的必要条件。而对于物性角色我们主要关注施成或功用角色是否对应。因为谓词的省略常常是事件强迫的结果，宋作艳（2011）指出“事件强迫结构中所隐含的动词通常是宾语名词的施成角色或功用角色”，所以要较好的确认或者还原省略成分，必须从动词和名词间的施成角色或功用角色是否对应开始。那么如何利用我们上面提到的两个特征来对谓词省略现象进行判别和还原呢？我们可以看两组搭配：

- (1) 赶论文 (2) 学英语
 看论文 学钢琴

在第一组中，动词不一样，名词一样，都是“V+论文”模式，但是一个有谓词省略，一个没有谓词省略。我们可以通过我们语义资源标注的一些动词特征(比如是否为事件动词)来判断是否有省略现象。在我们动词资源馆中，“赶”是事件动词，“看”不是事件动词，因此“看”就不具备事件强迫的可能，也就不存在省略现象。而对于“赶”来说，因为是事件动词，所以具备了省略谓词的可能。具体如何还原或补出谓词，还需要借助名词的物性角色特征。比如“赶论文”中“论文”的施成角色为“写”，我们要确认“写”是否为隐含动词，可以利用“写”的物性角色描述。比如我们资源库中描述到“写”的构成角色为“人”，而且只有施成角色“人工物”，而“论文”属于人工物，所以“写”可以和“论文”搭配，并且照应“论文”的施成角色。“赶”作为状态动词后可接状态、转变、过程类动词，且后接的动词必须是“do(做)”义的轻动词，这样作为过程动词和带有“do(做)”义的“写”就符合“赶”后接成分的条件，因此可以和“赶”搭配，同时在 CVL 中“写”的施成角色角色中也含有“论文”，这样相互照应，就可以判定“赶论文”是带有谓词省略“写”。

在第二组中，都是“学+NP”格式，因为“学”本身属于事件动词，因此它所接的宾语可能会有省略。事实上，“英语”的功能角色是“学”，而“钢琴”的功能角色是“弹奏”，那么对于“学英语”来说就没有省略，而对“学钢琴”来说，省略了谓词“弹奏”。

因此我们可以利用动词的事件特征和物性特征来帮助我们判断某些动宾搭配是否存在省略搭配现象，并可以加以适当的还原，这样省略的语义就得到补充，语义表达更加明确。

6.3 在语义角色标注中的应用

由于北大多视图汉语树库(PKT)的语义角色标注体系和本项目所描述的语义角色是一个体系，所以我们利用本语义资源库的语义角色体系，可以为北大多视图汉语树库(PKT)的语义角色标注提供部分支持。

北大多视图汉语树库(PKT)的语义角色标注采取的是机器标注和人工标注相结合的方法。在语义角色标注过程中，我们可以依据动词的类别和特征，然后查阅相关动词对应的语义角色框架，对实际语句中的动词和论元的语义关系进行语义角色的分配。比如：“感谢”在我们词典中的语义角色配对一般是“施事+对象”。那么只要是“感谢”的客体我们都可以标为“对象”，而“感谢”的主体我们都可以标为“施事”。这种方法就需要利用我们的动词语义资源中关于语义角色的描述来做参考。具体思路如下图所示：

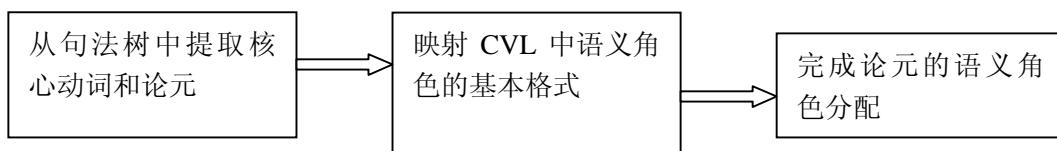


图 6: 语义角色标注思路

我们可以以实际语料中的标注例子说明：

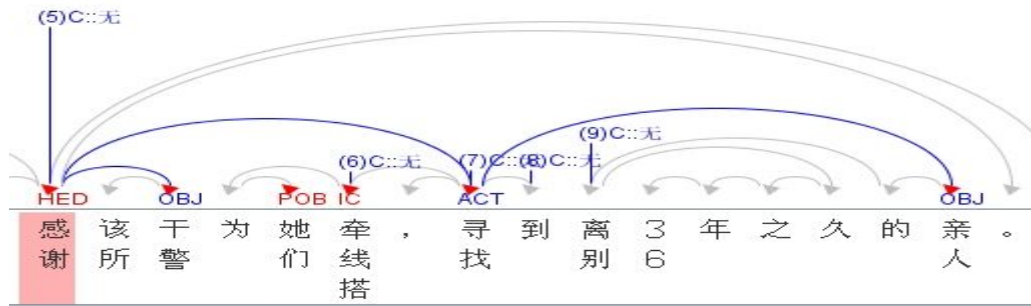


图 7: PKT 中句法分析实例

首先我们可以依据北大多视图汉语树库已有的句法树，提取出核心动词和论元，在这里有两个动词“感谢”、“寻找”，所对应的论元分别为“干警”和“亲人”。

然后查找动词语义资源中这两个动词的语义角色的基本格式，如下图所示：

	A	B	C	D	E	F	G
1	高层词集	中层词集	低层词集	动词	注音	释义	基本语义角色框架
1003	心理动词	态度	感激类	感谢	gan3xie4	好感，表示谢意。	施事(先生、读者、大夫、政治家)+感谢+对象(对方、师傅、医院):家长很感谢老师
2824	行为动词	活动行为	寻找类	寻找	xun2zhao3	找。	施事(老太太、父母、青年人、人们、全家)+寻找+受事(儿子、亲人、出路、真理、水

图 8: “感谢”和“寻找”的基本语义角色框架

紧接着可以依据语义资源库中的语义角色分配，直接给实际语料中“感谢”和“寻找”的客体分派语义角色。结果如图所示：

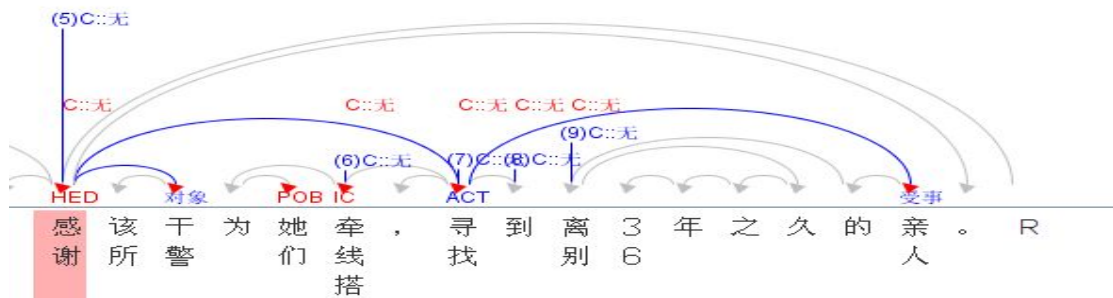


图 9: PKT 中语义角色标注实例

以上的标注方法可以说是一种静态的标注方法，它直接涉及到动词的分类，因为动词类别的不同，其与名词的语义关系就有所差别。但是在实际语料中，我们还要结合一些基于动态语义关系的语义角色标注。所谓动态语义关系，就是说对于有些动词来说，它们和名词间的语义关系有时并不是一层不变的，而是会随着所接名词自身的特征有所变化，这时我们主要根据实际句子中的动词与名词的语义关系来进行的标注。这个灵活性比较大，难度也比较大，比如动词“砍”，在我们的动词资源馆中基本语义角色是“施事+受事”，如果按照静态标注的话，后面的语义角色就都是受事了。而有时这些行为动词所接名词的语义角色随着句子不同，语义关系也不同。如同样是“砍”，所支配名词的语义角色却不同：如“他把树（受事）砍倒了”，“他把斧子(工具)砍坏了”，他砍了个洞（结果）。这时，我们可以利用我们动词资源馆中关于动词语义角色的非基本框架来作参考。

如下图所示：

1	动词	注音	基本语义角色框架	施事	主事	领事	受事	对象	结果	派事	工具	方式	原因	目的	时间	地点	范围
1509	砍	kan3	工、刽子手)+砍+受事	〈木匠〉砍木头			山民砍〈缆绳〉		房架砍好	这片小树	用〈斧子〉砍	砍坏了斧子			砍了〈半天〉	山里砍松树	助下李木匠
1510	看	kan1 (1)	人、放学娃、老母	〈瓜农〉看孩子			门卫看〈西瓜〉				人员〈用电子				妈在看孩子。	只看着这几	生路的情况

图 10: 标注实例动词的非基本语义框架样例

由于动词的非基本语义角色搭配涉及到十几种语义角色，这一步计算机是很难辨别的，所以只有人工来进行核对。

此外，一般情况下利用动词资源馆标注语料的最大问题就是语料中每一个动词语义角色都要有所收录，由于本树形动词词库收词不多，因此其资源库作用的发挥受到了限制。但我

们的做法是把这个动词向上规约，规约到我们比较大的层次中再去考虑语义角色的标注。比如“打乱”我们词典没有收录，但是我们可以把这些未收录的词上升到第二或者是第一层级进行分类，归为“行为动词”，那么行为动词的语义角色框架基本是“施事+受事”的语义框架，这样也可以实现语义角色的标注，当然最终还是由人工来帮助核对。

当然，CVL的应用不仅止于此，由于篇幅所限不进一步论述。

7. 总结与展望

本项目结合语义和句法对常用的 2506 个动词，3299 个词项，进行了四个层级的划分，并作了多维度多特征的描述，除了传统的静态语义描述、语义角色描述和句法格式描述外，还加入了事件结构特征和物性结构特征。对判断和还原一些带有谓词省略的超常动宾搭配现象有一定的积极作用，有利于句法分析的进一步开展。

只不过受到人力、物力和时间的限制，目前所构建的汉语动词资源馆收词不多，这极大地制约了该项目在语义角色标注、自动问答等领域的作用发挥。因此，在今后的工作中，应该进一步扩大收词规模、完善动词知识库的结构、继续核对数据的准确性和有效性，争取从更多的角度来揭示和描述动词的特征和属性。

参考文献：

- [1] Beth Levin. English Verbs Classes and Alternations [M]. Chicago: The University of Chicago Press, 1993.
- [2] Dowty, D. On Recent Analysis of the Studies of Control [J]. Linguistics and Philosophy, 1985(8): 291-331.
- [3] Fillmore, C.J. The Case for Case [M]. New York : Holt, Rinehart and Winston, 1968
- [4] Pustejovsky, James. Introduction to Generative Lexicon [M]. Manuscript, 2005.
- [5] Pustejovsky, James. Type Theory and Lexical Decomposition [J]. Journal of Cognitive Science 2006(6):39-76.
- [6] Vendler, Z. Linguistics in Philosophy [M]. Ithaca N. Y.: Cornell University Press. 1967
- [7] 陈平.论现代汉语时间系统的三元结构[J].中国语文, 1988,(6).
- [8] 戴耀晶.现代汉语时体系统研究[M].杭州:浙江教育出版社, 1997.
- [9] 宋作艳.现代汉语中的事件强迫现象研究[D].北京:北京大学博士论文,2009.
- [10] 宋作艳.类词缀与事件强迫[J].世界汉语教学, 2010,(4).
- [11] 宋作艳.轻动词、事件与汉语中的宾语强迫[J].中国语文, 2011,(3).
- [12] 林杏光等.现代汉语动词大词典[M].北京:北京语言学院出版社,1994.
- [13] 鲁川等.动词大词典[M].北京:中国物质出版社,1994.
- [14] 孟琮、郑怀德等.汉语动词用法词典[M].北京:商务印书馆,1999.
- [15] 税昌锡.VP 界性特征对时量短语的语义约束限制[J].语言科学, 2006, (5) .
- [16] 汪梦翔.对象格语义范畴及相关语法理论研究[D].武汉:华中师范大学博士论文, 2012
- [17] 汪梦翔, 王厚峰.现代汉语“对象格”自动识别研究[J].计算机工程与应用, 2014,(8)
- [18] 杨敏, 常宝宝.基于北京大学中文网库的语义角色分类[J].中文信息学报, 2011,(3).
- [19] 袁毓林.一套动词的论元角色的语法指标[J].世界汉语教学, 2003,(3).
- [20] 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,(6).

作者简介：汪梦翔（1984——），男，讲师，主要研究领域为汉语语法、中文信息处理，Email:wmx1984@pku.edu.cn；王厚峰（1965——），男，教授，主要研究领域为自然语言处理，Email:wanghf@pku.edu.cn；刘扬（1971——），男，副教授，主要研究领域为语义资源库建设。Email: liuyang@pku.edu.cn；饶琪（1982——），男，讲师，主要研究领域为汉语语法、中文信息处理，Email: 337245309@qq.com。

汪梦翔

王厚峰

刘扬

