

# 中文动词及分类研究：中文动词词汇语义网的构建及应用<sup>\*</sup>

刘美君 万明瑜

**摘要** 随着人工智能的日益发展,语言学成为“产”“业”“学”界寻求合作及突破的新契机。其中语言学语义资源的构建及标注问题成为了当前的一大热点及难点。文章针对中文动词语义分类问题,从理论研究、语义网构建及实践应用三方面进行了全面的探讨和分析。理论研究上,文章以“框架为本,构式为用”的研究方法为基石,依循框架语义和构式语法以区分动词和构式之间的“形-义”搭配,形成“格式塔”(Gestalt)般互补。语义网构建上,以语言学分析为基础,语料实证为依归,通过“框架元素”与“定义性构式”来定义动词属性,使语料兼有词汇表征、框架阶层及语义标注等信息。语义知识库目前包含“沟通”“认知”“感知”“情绪”“评价”“社会互动”“自动”和“致使移动”八大类框架动词,已有效运用于多种基于语义及事件框架的中文自然语言处理任务,包括中文语义自动消歧,自动语义角色标注,事件框架甄别及故事自动生成。

**关键词** 框架语义 构式语法 动词分类 语义标注 自然语言处理

DOI:10.16134/j.cnki.cn31-1997/g2.2019.02.004

## 一、导 论

大数据与深度学习技术的到来,使得人工智能技术飞速发展,Marcus(2018)指出,近六年来,人工智能在语音及图像的自动识别、机器翻译、信息自动检索等自然语言处理任务方面作用显著,几乎到达一个很难再超越的高峰期。研究者们开始关心语言学和人工智能如何进一步结合以求得突破,例如在词性标注上,从97%到100%的可能性。(Manning 2011)为此,语言学成为“产”“业”“学”界寻求再突破的新契机,究其原因有四:其一,语言学知识为人工智能的自然语言处理提供更接近语言本质的特征模型;其二,语言学规则为语言现象及行为差异提供更为合理的解释;其三,语言学特征编码为人工智能提供相对有效无噪的数据集;其四,语言学知识工程不依赖于大规模集成电路或复杂算法,大力节省软硬件成本。现阶段人工智能与语言学合作发展成为必然趋势,但合作的热点及难点在于语言学资源库的构建及语义标注问题。过去的自然语言处理任务在词语切分、词类标注、句法标注方面均已取得了良好成果。然而,语义标注的工作难度却更大,问题更为复杂。关于语义划分及标注的问题,国

<sup>\*</sup> 特别要感谢助理林琦铎及博士生薛时蓉同学协助整理初稿。

内外已经有许多相关研究和资源库(或线上知识库),例如国外有 The Proposition Bank (英文 PropBank Palmer *et al.* 2005), English FrameNet (英文框架语义网, Fillmore & Baker 2000), English VerbNet (英文动词语义网, Levin 1993; Schuler 2005); 国内有中文句结构树 (Chen *et al.* 1999; Huang *et al.* 2000), 汉语框架语义知识库/Chinese FrameNet (You & Liu 2005) 等。国内相关研究较国外起步晚,且现有语义研究及标注系统仍存在很大的不足和弊端,许多问题有待重新审视:如何系统而有效地区分语义?如何辨别形义搭配?形式与语义对应关系如何?什么样的语法表现可作为语义区分的标准?如何对动词进行分类?什么样的分析方法适合中文动词的语义区分?下面从国内外相关研究总结当前语义研究及标注的主要问题和难点,并有针对性地介绍中文动词词汇语义网的特色和优势。

## 二、国内外语义研究及标注系统综述

国外的语义研究起步较早,发展相对成熟,已有许多完整的理论构架及标注体系,代表性研究及体系有英文框架语义网、英文动词语义网、UCREL 语义分析系统(USAS)及英文 PropBank 等。相比之下,国内的语义研究起步较晚,发展相对滞后,但也有一些对应的体系和成果,比如汉语 PropBank、中文句结构树资源库 (Sinica Treebank)、构式义标注(刘洪超 & 詹卫东 2014)、汉语框架语义知识库 (Chinese FrameNet) 及中文动词词汇语义网 (Mandarin VerbNet) 等。中文的语义划分标准及理论研究大都沿袭西方的研究体系,缺乏对中文语法特殊性的处理。而且各体系基于不同语义关联分析语义及语法表现,侧重点不一,划分粗细度各异,各有特点,但也存在不足之处。

### 1. 英文框架语义网 (English FrameNet)

英文框架语义网<sup>[1]</sup>是根据“框架为本”(frame-based)的研究方法和分析方式而建立的一套专门分析英文语义的开源性知识库体系。(Fillmore 1982; Fillmore & Atkins 1992; Fillmore & Baker 2000)该系统基于 Fillmore(1982)提出的框架语义理论,以认知为驱动因素,以词汇(不仅限于动词)语义背后所涉及的“场景情境”(scene)来进行语义分类。这样的背景架构即为“语义框架”(semantic frame)。语义框架除了提供认知上的事件基模,也作为联系相关词汇的概念基准,以作为词汇语义分类的依据。传统上,语义框架是由一组主要的参与角色(即框架元素)来定义的,然后通过所需角色来呈现语义的依存关系。下面以动词“hit”为例,呈现该系统对于动词的核心的框架元素标注情况。

(1) [The massive metal foot/Agent] [HIT/Impact] [the ground/Patient] with [a huge thud/Result.]

如例(1)所示,动词“hit”在框架语义网中被归为 Impact 类动词,所牵涉的核心框架元素有“Agent”(施事者)、“Patient”(受事者)及非核心框架元素“Result”(结果)。系统对动词本身类别及其所关联的参与角色均进行了相应的语义标注,但显著的缺点是只关注框架参与角色,忽略词汇在句法上的特征表现;另外,词汇本身的分类也缺乏一个系统的准则,显得杂乱而无章法。

### 2. 英文动词语义网(English VerbNet)

英文动词语义网<sup>[2]</sup>是基于 Levin(1993)所提出的英语动词词汇分类方法(English Verb Classes and Alternations)而建立的关于英文动词分类及结构变化的标注体系。英文动词语

义网的构建是在 Levin(1993) 提出动词类别基础上进行次类的细化与添加,得到同类动词语义、语法的一致性。如 Carry(携带)类动词,其同类成员有 Carry, Drag, Draw, Haul, Heft, 这些动词具有共同的语义角色,即施事、客体、源位、目的地,且具有共同的语法表现,例如: NP + V + NP。下面的例子阐释了英文动词语义网对于 Carry 动词的一种最常见的句法结构及语义角色。

- (2) 例句: *Amanda carried the package.*  
 句法: Agent/施事 V Theme/客体  
 语义: motion [during( E0) , Theme]  
       equals ( E0 , E1)  
       motion [during( E1) , Agent]  
       cause ( Agent , E0)

尽管英文动词语义网兼顾了语义和语法因素,但其对动词分类缺乏标记、语义分类过于泛化、缺少对同类词元的列举且不适用于中文的特殊语法及语义表达。

### 3. UCREL 语义分析系统(USAS)

UCREL 语义分析系统<sup>[3]</sup>是一套对文本进行自动语义分析及标注的系统。系统的开发始于 20 世纪 90 年代,语义标记参考 McArthur(1981) 撰写的 *Longman Lexicon of Contemporary English* 《朗文英语当代辞典》。该系统的框架于 2013 年开始扩展,涵盖多国语言,包括:中文、荷兰语、意大利语、葡萄牙语、西班牙语和马来语。该系统对语义的划分基于一个多层结构,根据 21 个主要话语场进行细化分类,利用相同语义概念来聚集词意的语义场(标注见例 3)。

- (3) UCREL 语义分析系统标注示例“我是中国人”。

TOKEN	POSTAG	SEMTAG	MWE
我	PN	Z8mf	0
是	VC	A3 Z5	0
中国人	NN	S2mf	0

例(3)在 UCREL 语义分析系统自动标注体系中被自动分词且每个词语标有三层信息: POS 语法信息、语义关联类别信息及多字词(Multi-Word-Expression)信息。该系统的语义标注问题在于其语义关联并不能有效区分不同的语义类型。例如,基于 UCREL 语义分析系统的分析,例(3)中“我”不足以分辨出是历事(Experiencer)、施事(Agent)还是说话人(Speaker)。此外,UCREL 语义分析系统仅提供词类标注,而没有任何句法结构信息的标注。

### 4. 英文 PropBank( The Proposition Bank)

英文 PropBank<sup>[4]</sup>(Kingsbury & Palmer 2002)是一套基于 Treebank2 华尔街日报语料(WSJ)进行的语义标记资源库。语料库中每个动词都会被当作一个语义谓词,其周围的文本会被标注为该谓词的论元和附加角色,谓词本身也用细粒度的带有语义的方法进行标

记。语料库中的语义角色采用先决的通用标记, 过于泛化且灵活度低。以动词 *hit* 为例:

- (4) a. [The boy]施事 *hit* her hard.  
 b. ? [The wind]施事 *hit* her hard.  
 c. ? [His words]施事 *hit* her hard.

例(4)中的 *the boy the wind his words* 尽管语义特征不同, 分别为人类、非人类的物质实体、非人类非物质事物, 在英文 PropBank 中却不作区分, 统一分析为施事。更为完善的处理方法应是基于同一标记下的成分语义, 根据特征区别进一步细分语义角色, 如: 施事 (Agent)、致事 (Causer)、刺激物 (Stimulus) 等。

## 5. 汉语 PropBank

汉语 PropBank<sup>[5]</sup> 沿用英文 PropBank 的理论体系, 对中文的语料采用不同谓词对应不同基本语义命题的理论框架进行标注。且谓词与论元的关系也被添加到中文结构树资料库 (Chinese Treebank) 的句法树中。如动词“搬” (Move) 属于“致使移动” (Caused-to-Move) 类别, 其论元包括了移动者 (ARG0: mover)、被移动物体 (ARG1: moved)、移动起点 (ARG2: moved from) 及移动终点 (ARG3: moved to) 等。例(5) 为该系统的一个标注实例。

(5) [这个村庄]ARG0: mover [曾] ARGM-ADV 自力更生, 艰苦奋斗, [搬]REL [山] ARG1: moved 添沟造平原。

从例(5) 可知, 汉语 Propbank 明显的问题在于对谓词的简单罗列 (统一将动词标注为“REL”), 没有对动词进行系统性的分类。

## 6. 中文句结构树资源库 (Sinica Treebank)

中文句结构树资源库<sup>[6]</sup> (Chen *et al.* 1999; Huang *et al.* 2000) 由“中央研究院”词库小组 (Chinese Knowledge and Information Processing) 建立, 基于信息为本的格语法 (ICG) 对中文语料库 (Sinica Corpus) 的语料进行标注。该资源库的结构框架采用中心驱动原则 (Head-Driven Principle), 围绕句子或词组的核心中心对其论元或附加成分的语法构成进行标注。图 1 为一示范标注。

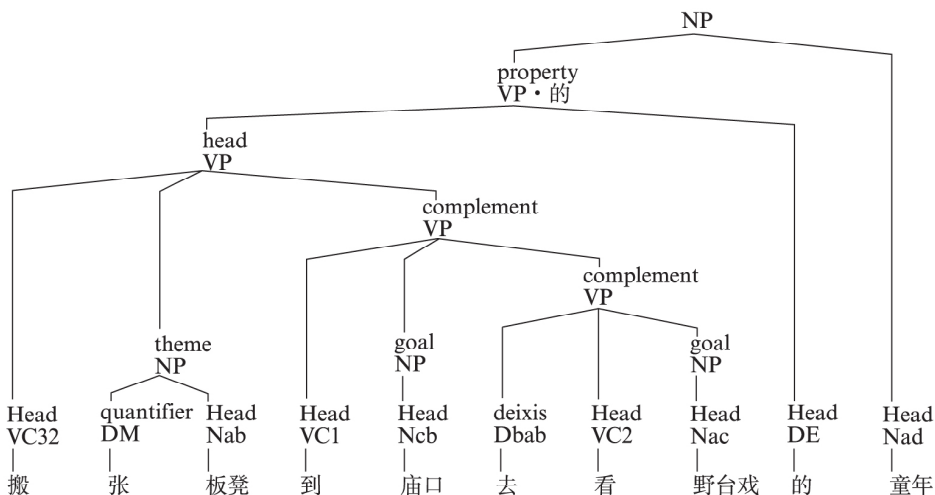


图 1 中文句结构树资源库示范标注

如图 1 所示, 尽管中文句结构树资源库标注了词的语法特征, 但未对动词语义进行分类, 且对论元、附加成分仅进行语法标注, 并没有凸显各自语义角色。

### 7. 北京大学构式义标注

北京大学构式义标注(刘洪超, 詹卫东 2014) 以构式为单位进行语义标注。构式指整体意义无法从其组成部分简单加合出来的语言单位。该语料库针对特定构式的认知机制进行分析, 尝试建立从表层语言到深层认知的映射。如:

- (6) a. 用一吨, 少一吨。  
b. 泡一次, 淡一次。

在构式义标注中, 例(6) 统一采用“A—X, B—X”的释义模板, 其语义为“越 A 越 B”, 表达因果倚变义。尽管这一语义标注包含构式信息, 但对动词的语义角色与特定构式的互动并未进行进一步的分析标注。

### 8. 汉语框架语义知识库

汉语框架语义知识库是由山西大学(You & Liu 2005) 开发, 以 Fillmore(1982) 的框架语义学为理论基础, 参照加州大学伯克利分校的 FrameNet 工程, 构建的以汉语真实语料为依据, 可供计算机使用的汉语词汇语义知识库。表 1 为一示范标注(以动词“到达”为例)。

表 1 汉语框架语义知识库示范标注

框架名	到达 Arriving
定义	指转移体到达目标的过程。目标可直接表达出来, 或从上下文中得到理解, 或者动词本身隐含目标之义
核心框架元素	目标(goal), 转移体(thm)
非核心框架元素	伴随体(thm_c), 形容(depic), 目标状态(goal_c), 修饰(manr), 方法(mns), 传送模式(mot), 路径(path), 源点(src), 时间(time)
词元	到达 v, 来到 v, 进入 v, 抵达 v, 返回 v, 走到 v, 走进 v, 赶到 v, 回来 v, 归来 v, 到 v, 回到 v

如表 1 所示, 汉语框架语义知识库对动词的语义框架、动词的核心框架元素, 非核心框架元素, 以及同类词元都进行了定义和整理, 但这一知识库的建立更像是英文框架语义网的中文译本, 缺乏对中文特殊语法现象的统一处理; 另外, 与英文框架语义网类似, 该系统也缺乏对构式语义信息的标注。

## 三、中文动词词汇语义网的特性

以上无论哪种语义分类方法, 虽各有所长, 但均有不足之处。对于语义的划分, 语言学家的共识就是动词的语义必然体现在语法表现上, 语法和语义是一体两面、相辅相成的。语义角色是可能进一步细化或延伸的语言范畴。框架中的角色本身也许不足以定义框架, 而需要与语法特征结合。换言之, 对语义角色, 要将其作为构式中的框架角色进行研究。Liu 和 Chang(2015) 指出, 动词与构式彼此互为表里, 有如“格式塔”(Gestalt) 般相互依存的紧密关系。构式与动词间存有彼此定义的搭配关系, 构式语义必须通过动词来呈现, 而动

词框架与不同的构式相联系。本文对于汉语动词词义分析和动词分类的问题,提出一种“框架为本,构式为用”(frame-based constructional approach)的分析方法,结合了“框架语义”理论中的事件框架(frames)和“构式语法”(Constructional Grammar)理论中构式的表征意义(Goldberg 1995, 2003),来分析和划分动词的语义类型。框架理论主要定义了“参与者”(participant roles),但缺乏语法形式上的明确性;而构式语法主张形式和意义间的表征和对应,正好作为在动词划分上显性的标准。也就是说,对于语料库的动词属性分析,首先确定所参与的语义框架,再定义每个框架下的“框架元素”(frame elements)与“定义性构式”(defining constructions)来共同描述框架的性质;根据实际应用中词义辨析中所需要的粗细程度,也需考虑与构式搭配的词汇和共现形式(lexical collocation)。

### (一) 研究方法 “框架为本 构式为用”

为了解决目前语义标注系统的问题及适应汉语本身特殊的语法特征,中文动词词汇语义网采取“框架为本,构式为用”的研究方法,这是一种以认知为驱动、以框架语义理论与构式语法为理论基石,以中文特色语法表现为补充的综合性研究方法(Liu 2003, 2005, 2016, 2018; Liu & Chang 2005; Liu & Chang 2015, 2018, 2019; Liu & Chiang 2008; Liu & Hu 2008, 2013; Liu *et al.* 2006, 2015)。

#### 1. 框架语义(Fillmore 1982, 1985)

本文根据 Fillmore(1982)的框架语义理论,从认知语义的角度,将动词依照其背后所涉及到的“场景情境”(scene)来加以分类,即“语义框架”(semantic frame)。语义框架提供认知上的事件基模和联系相关动词的概念基准,也可作为动词分类的依据。传统上,语义框架是由一组主要的参与角色,即框架元素(frame element)来定义的,然后通过所需角色来呈现语义的依存关系。举例来说,“买”这个动词,从认知经验来说,属于“商业交易动词”,涉及商业交易的认知框架。在这个框架中,主要的参与元素有买家、卖家、货品、金钱等,相关的动词包括买、卖、付、花(钱)、要价等;同一框架下的动词或者会搭配不同的框架元素,如表2与例7所示。

表2 买卖动词与框架元素

	买家	卖家	货品	金钱
买	+		+	
卖		+	+	
付	+	+		+

- (7) a. [我/买家]买了[一瓶水/货品]  
 b. [他/卖家]卖了[一瓶水/货品]  
 c. [我/买家]付了[他/卖家][五百块钱/金钱]

由此可以看出,以框架为本的分析方式可兼顾语义的共性与特性,提出符合实证经验的分类。此法已应用于英文动词的分析上,已得到广泛认可的结果(详见“英文框架语义网”)。

## 2. 构式语法( Goldberg 1995 ,1997 2010)

在词汇语义相关的研究中,最为广泛接受的前提是动词的语义属性会体现在其语法表现上( Levin 1993; Levin & Hovav 1996 2005) ,而具有语法辨识性的语义成分才是关键特征。基于这一原则,我们对框架与动词的关联做出进一步的语法定义,对原有的框架语义理论进行了扩展:借助“构式语法”中形义搭配的特点( Goldberg 1995 ,1997 2010) ,提出一种混合互补的方法,进一步将语义框架之标准优化为“形义搭配”的原则;即在框架的基础上,明确规定了框架元素,并列出“定义性构式”( defining construction) ,帮助界定语义要素的形式搭配。从宏观的角度,“构式”也如同词汇一般,是一种“形式和意义”的搭配对应关系,即构式自身就是具有语义的,并可通过语义上与其兼容互补的动词体现出来。构式和动词,两者结合,相辅相成,携手将框架的意义具体地表达出来。如同 Boas ( 2003) 所言:“动词的每一个义项( sense) 都形成一个微型构式( mini-construction) ,既包含框架语义也包含句法信息。”反过来说,就是一个框架的语义除了动词所带有的框架元素之外,也体现于动词所参与的词汇构式当中,而这些构式即为框架的“定义性构式”。

具体来说,构式的方法主要体现在语义网参考框架中的语义要素与动词的共现-构式特征( collo-constructural features) 。一方面,共现-构式模式可以清晰定义动词或框架之间的差异。如同为“致使移动类”动词的“搬”“放”,两者具有相似语义角色:移动者、被移动物、处所,在共现-构式模式方面存在显著差异,以施事性把字句与处所倒置式为例:

- (8) 施事性把字句: a. 我把书搬到/\* 在桌上。  
b. 我把书放到/在桌上。  
(9) 处所倒置式: a. \* 桌上搬着一本书。  
b. 桌上放着一本书

两种共现-构式模式区分了“搬移类”动词与“放置类”动词。就构式而言,根据 Liu 和 Chang( 2015) 的观点,处所处置式是放置类动词的区别性构式,而“搬移类”动词不能用于处所倒置式。

另一方面,就词汇共现模式( Morpho-collocation) 而言,基于中文十亿词语料库( Chinese Gigaword) 所做的“搬”“放”做施事性把字句谓语介词搭配的频率统计如表 3 所示:

表 3 搬和放做把字句谓语的词汇共现频率

	在	到	例 句
搬	2	3203	我把书搬在? * /到桌上。
放	24384	2704	我把书放在/到桌上。

其中,“搬”通常与表路径的“到”搭配使用,表示致使物体移动的路径;而“放”通常与引入处所的“在”共现,表示致使物体处于某位置。由词汇共现模式,还可归纳出同类动词如“搬移类”动词“投”“掷”,“放置类”动词“丢”“扔”。

基于这样的分析模式,动词与构式互为表里,互相定义。以这种“框架为本,构式为用”的方法来分析归类汉语动词,不但有语言学的理论基础,也具备操作上的实用性;无

论是在后续的分类研究中还是应用于汉语动词的教学,较以往的分类方法都更为合理适用。

## (二) 以真实语料为依归

本文中分析的文本材料,都是从各个语料库搜索得来的自然语言实例,旨在考察动词在实际使用中的情况。使用的语料库包括“‘中央研究院’平衡语料库”(Sinica Balanced Corpus)和“中文十亿词语料库”(Chinese Gigaword),利用“‘中央研究院’中文词汇速描系统”(Chinese Word Sketch Engine)进行查询。在语料缺乏的情况下,也有借用其他搜索引擎(例如谷歌)的搜索结果作为辅助查询。“中央研究院”的平衡语料库拥有涵盖六大主题(文学、生活、社会、科学、哲学、艺术)、超过一千万词目的语料;中文十亿词语料库囊括了来自两岸及东南亚的华文新闻出版媒体,如《新华社》《人民日报》《“中央”社》及《联合早报》等,包含了超过十亿字词的新闻语料。语料筛选方面,我们选定某一动词作为观察对象,以其为关键字在语料库中收集语例,经过整理和观察,得出该动词实际的句法表现、数量分布,以及最为显著的共现形式,而后进行语义上的分析和讨论。本文的最终目的为基于现有的语义标注系统的各类问题及汉语本身的语法特征,建立一个针对汉语动词语义的有效分类及标注体系及知识库,并能为中文自然语言处理所应用。

## 四、中文动词词汇语义网的构建与功能介绍

中文动词词汇语义网(Mandarin VerbNet)<sup>[7]</sup>是关于中文动词分类及语义标注的一个全面性词汇语义知识库。该知识库由语言学导向,基于语料库实证研究,且面向自然语言处理及教学等各种应用。该知识库的建立以“框架为本,构式为用”的研究方法为理论基础,为动词进行分类及语义标注,提供了有层级的动词框架信息和带有统计频率的动词语义标注信息。以下为语义网的具体介绍。

### (一) 框架信息及其层级结构

框架的语义范畴不同,具有层次性。(Liu & Chang 2005)按照范畴大小,可分为源框架(Archi-frame) > 初级框架(Primary-frame) > 基本框架(Basic-frame) > 微框架(Micro-frame)。框架间层层相关,低层框架可视为高层框架的子框架(sub-frame)。换言之,这种分层结构含有框架自上而下的继承和使用关系。

- 源框架是认知上广义的语义领域,作为框架上层是最为概括的大类,由框架最基本的语义原型与最基础的句式来定义,源框架为必须框架;

- 初级框架可以提供一个次高层次的语法语义搭配关系(如情绪类动词的基本框架至上还有五个初级框架,见表4);

- 基本框架是认知上显著的层次,有较明确清晰的语义成分及搭配构式,表达基本范畴所涵盖的语义特征。源框架和基本框架是必需的,而初级框架和微框架则不是必需的;

- 微框架用来描述基本框架下一些频繁出现的近义词集,用来区分近义词的细微语义差别,如“悬”“挂”“吊”为放置类(Placement)基本框架下面的三个微框架。

以下以情绪类动词为例,展示该类动词的层级框架信息:情绪被定义为源框架,在语义网中包括了五大初级框架及十个基本框架,具体框架信息如表4所示。



表 4 情绪类动词的框架层级

源框架	初级框架	基本框架
情绪 (EMOTION)	历事且不及物 (EXP-INTR)	满足 (CONTENT)
		高兴-难过 (HAPPY-SAD)
		后悔-遗憾 (REGRET-SORRY)
	历事且及物 (EXP-TRANS)	嫉妒-同情 (ENVY-PITY)
		爱-恨 (LOVE-HATE)
		担心-害怕 (WORRY-FEAR)
	刺激源且不及物 (STIM-INTR)	刺激源属性 (STIM-ATTRI)
	刺激源且及物 (STIM-TRANS)	刺激源且及物 (STIMULUS-TRANSITIVE)
	肇使者且及物 (AFFECTOR-TRANS)	吸引-刺激 (ATTRACT-STIM)
		打扰-激怒 (BOTHER-IRRITATE)

除了层级框架结构,语义网的每一层框架都会有框架定义及框架信息,包含框架定义、代表动词词元、定义性构式、框架元素/参与角色(包括核心与非核心框架元素)、构式标记(用星号来标记,用于区分框架元素标记)。以下以情绪类源框架下面的五个初级框架为例,介绍其相关框架信息。

表 5 情绪源框架下面的五个初级框架

a. 历事主语状态不及物动词	词汇定义: 历事的内心情绪状态 核心框架元素: 历事 定义性构式: 历事 + * 很 + 动词(例: 我很高兴。) 代表动词词元: 高兴、生气
b. 历事主语动态及物动词	词汇定义: 历事对对象的情绪活动 核心框架元素: 历事、对象 定义性构式: 历事 + * 很 + 动词 + 对象(例子: 我很爱他。) 代表动词词元: 羡慕、爱
c. 刺激源主语状态不及物动词	词汇定义: 刺激源的特性描述 核心框架元素: 刺激源 定义性构式: 刺激源 + * 很 + 动词(例: 电影很有趣。) 代表动词词元: 恐怖、有趣

(续表)

d. 刺激源主语动态及物动词	词汇定义: 刺激源对历事引发的情绪状态 核心框架元素: 刺激源、历事 定义性构式: 刺激源 + * 很 + 动词 + 历事(例: 这本书很吸引我。) 代表动词词元: 困扰、吸引
e. 肇使者主语动态及物动词	词汇定义: 情绪受到肇使者的影响 核心框架元素: 肇使者(Affector)、受制者(Affectee) 定义性构式: 肇使者 + * 把 + 受制者 + 动词(例: 他把我激怒了。) 代表动词词元: 激怒、打动

表5中五种初级框架,虽然都是描述情绪状态,却各自呈现明显的语法表征差异,语义网根据定义性构式来对动词加以分类。

## (二) 动词的相关语义信息

语义网包含经过标注的语料(用例来自十亿词库和中文平衡语料库),每个动词标记的内容包括动词在基本框架下的框架元素和构式标记。语义网还提供多义词的标注、两个动词的标注结果对比、半词汇化信息等。此外,语义网还提供每个动词在中文十亿词语料库中的出现频率,用来呈现动词的使用情况。

### 1. 动词的标注形式

以放置类动词“放”为例,语义网的动词语义标注主要通过两大类信息来实现:框架元素(frame element)和构式标记(construction marker)。框架元素是框架内动词特定的参与角色,并通过角色来呈现与语义的依存关系。构式标记是跟动词紧密连接的显著句法标记,用来说明动词类别和句法结构的关系,在语义网中用星号(\*)标记加以区分。

(10) 定义“放置类”动词“放”:

基本框架: 放置框架(Placement frame)

a. 核心框架元素: 放置者、放置物、放置点

b. 构式标记: \* 把、\* 处所标记、\* 体貌标记

c. 基本构式模式:

i. 及物的把字结构

如: [她/放置者][把/\* 把][玩具/放置物]放[在/\* 处所标记][房间里/放置点]

ii. 凸显放置物的不及物结构

如: [玩具/放置物]放[在/\* 处所标记][房间里/放置点]

iii. 凸显放置点的处所倒装结构

如: [房间里/放置点]放[着/\* 体貌标记][玩具/放置物]

除了动词最基本的语义,“框架为本,构式为用”的标注方法还可以标注动词语义的概念和隐喻延伸。以“放置”(Placement)类动词“放”为例:

(11) a. [我/放置者][把/\* 把][书/放置物]放[在/\* 处所标记][桌子上/放置点]。

b. [她/放置者][把/\* 把][老师的话/放置物]放[在/\* 处所标记][心上/放置点-认知延伸]

例(11)中的两个句子,具备相同的构式及句法结构,但是语义却不同。例(11)a为动词“放”的典型用法,表达的是普通放置处所的概念;而例(11)b由于搭配非处所的介宾短

语(在心上),使得“放”的语义从空间层面延伸至认知层面(记忆)。语义网对类似现象的处理使得词汇的基本语义与其语义的延展或转换得到明确的区分。

## 2. 多义词的处理

动词的多义性和义项划分是语言研究中值得关注的问题。语义网遵循“一个框架,一个语义”(one sense, one frame)的原则,根据所在语义框架不同,区分多义词的不同义项。以动词“放”为例,“放”在中文词汇网络(Chinese WordNet)的分析中包含 37 个不同的义项,可见其用法之广,语义之丰富。而在中文动词语义网中,根据“框架为本,构式为用”的分析,“放”初步分为两个空间动态的基础语义——“置放”类和“释放”类。(Liu & Chang 2015)。如:

(12) a. 放置义: [她/放置者][把/\*把][鸟/放置物]放[在/\*处所标记][笼子里/放置点]

b. 释放义: [她/施事][把/\*把][鸟/释放物]放[出/\*出标记][笼子/源点]

这两个基础义项,可以用“释出源点”(away from a source)到“置于终点”(land at an endpoint)的事件链来解释,又可经由位移路径衍生出“放”的其他语义和用法。在认知框架的基础下,“放盐、放书、放钱”等用法都和“置放”有关,构式上有处所终点出现;但“放人、放牛、放风筝”等则与“释放”较相关,构式上可带有源点出处(source)。但是当构式表现看起来相同时,框架元素的定义就成了区分的关键。根据语义网,可知“释放”义与“放置”义的“放”牵涉的框架元素不同,实际上是涉及了两个不一样的基本框架。“放置”框架包含放置者、放置物、处所等,而“释放”框架包含施事、“出”类标记、释放物等。

## 3. 两个动词之间的比较

中文动词词汇语义网还可以比较同一源框架下的任意两个动词之间的构式模式异同,主要用于区分近义词或反义词的语义特征,得出其模式上的异同,进而分析其语义异同。例如,“快乐”和“高兴”都同属于情绪源框架下的“快乐-悲伤”(Happy-Sad)基本框架,根据语义网的语料标注,通过自动比较的功能,可以得出两者具有共同的模式如:

[历事][快乐-悲伤]: 我高兴/快乐啊!

[历事][\*程度][快乐-悲伤]: 大家都很高兴/快乐。

[刺激源][\*致使][历事][快乐-悲伤]: 美丽的景色使她们高兴/快乐。

[历事][\*情绪感官][快乐-悲伤]: 她们觉得高兴/快乐。

[刺激源][\*程度][快乐-悲伤]: 有钱很高兴/快乐。

两者也具有不同的构式模式,如“快乐”具有大量名物化[\*快乐-悲伤+名物化]的用法,而“高兴”则几乎没有名物化;另外,“高兴”具有[历事][\*程度][快乐-悲伤][内容]的模式,比如“我很高兴你能过来”,而“快乐”没有。两者语法表现的异同说明他们语义上也存在细微差异,特别是“快乐”的高频率名物化用法,对比之下,“高兴”在语义上比“快乐”表现出更固化的句法特征。

## 4. 半词汇化信息

中文动词词汇语义网也包含部分动词与半词汇化成分的搭配,半词汇化结构是指是具有能产性(productive verbs)的词汇结构。网站现有的半词汇化结构包括“可+V”“V+人”“令人+V”等,其中V代表适应该结构的能产性动词,如例(3)，“可+V”的半词汇化结构,

适用于类似“可爱”“可恨”“可怜”等动词“V+人”的半词汇化结构,适用于类似“烦人”“气人”“感人”等动词“令人+V”的半词汇化结构,适用于类似“令人兴奋”“令人鼓舞”“令人沮丧”等动词;这三类半词汇化结构均属于“情绪类”框架下的子类“刺激源属性”的基本框架,且用“#”加以标注,以区别于其他标注信息。相关例句如下:

- (13) a. [女孩子/刺激源][比较/\*程度][可爱/#刺激源-属性]。  
 b. [这个画面/刺激源][十分/\*程度][感人/#刺激源-属性]。  
 c. [眼前的情景/刺激源][让人迷惑/#刺激源-属性]。

## 5. 语义网构建流程及进度

语义网的构建流程主要包括以下几个步骤(见图2):其一,语料收集与筛选。该环节旨在完成基本语料的准备,采用自动提取与手动筛选相结合的方式进行语料收集,在每个类别中定义最高频的动词列表,每个单词选取合格语料200句;其二,人工标注。该环节采用半自动辅助标记工具 Atom 结合人工方式标注动词语义,主要包含动词类别、框架元素、构式标记;其三,人工校对,为了确保标注准确性,初步标注结果由两位项目经理做交叉比对,进行一致性内部评分测试,以确保更高的标注准确度;其四,网页开发及管理,采用JS脚本语言开发网页,在网页后台对标注、定义完成的语料进行管理,利用 Atom 内嵌程序包自动统计并上传数据至城大伺服器,以公开网络用户使用;最后,基于用户回馈网站的体验评价及建议,更新标注以提高标注质量及改善网页设计以增加用户体验。

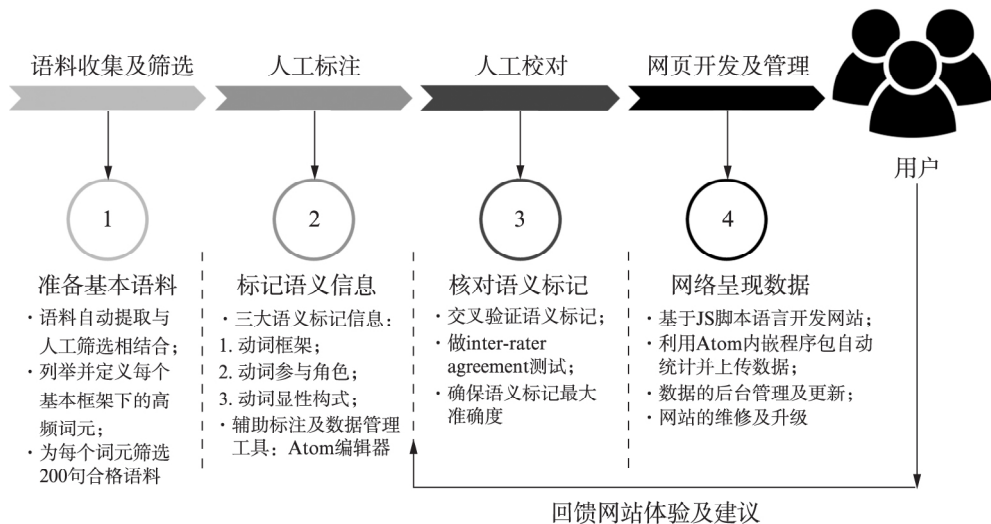


图2 中文动词词汇语义网的实现流程

目前,中文动词词汇语义网已建立了八个源框架,包括“情绪”“认知”“感知”“致使移动”“自动”“社交互动”“沟通”及“评判”类,涵盖大约80个基本框架。另外,语义网定义并分类了约800个动词,其中520个动词得到语义标注及构式统计。

在下一个阶段,语义网短期目标为完成中文高频的2000个动词的分类及标注,涵盖大概10个新的源框架,50个基本框架,并同步实行对“空间配置类”(Spatial Configuration)、“关系类”(Relation)、“存现类”(Existential)、“个人行为类”(Behavioral)等源框架的语义研究。

## 五、中文动词词汇语义网之 NLP 应用

中文动词词汇语义网构建的同步,产生了一个富含语言学信息的知识库。该知识库的建立对于语言学本体研究、中文教学及自然语言处理均有很好的启示及衍生作用。目前,我们已经成功将已有的中文动词词汇语义网资源库投入多种语义相关的中文自然语言处理任务,并已经取得不错的成果,以下三个应用为例。

### (一) 中文动词语义消歧

词汇歧义/多义指同一个词具有两个或以上不同但相关联的语义。词义消歧(Word Sense Disambiguation,简称 WSD)(Wan & Liu 2018)是指基于语境信息自动解决词义歧义的自然语言处理任务。近几十年来 WSD 仍是 NLP 研究中的一个热点话题,尤其在中文动词的一词多义问题上。例如,在例句“他很烦!”中,既可理解为“他”这个人很烦人;也可理解为“他”感到厌烦。针对情绪类动词的语义判别问题,Liu(2016)提出了一套适用于中文“情绪类”动词分类方法,全面且创新地归纳了五种“情绪类”动词的词汇化模式:

- (14) a. 刺激源主语 + 及物动词 “这个问题烦了我三天三夜。”  
 b. 刺激源主语 + 不及物动词 “明天的考试很烦。”  
 c. 历事主语 + 及物动词 “你在烦什么?”  
 d. 历事主语 “我好烦啊!”  
 e. 影响者主语 + 及物动词 + 历事 “那我就别再烦你了。”

以上五种词汇模式,能够很好地将多义词“烦”加以区分,其中包含三大主要语义:例(14) a、例(14) b 属于刺激源主语(属性类)语义,强调主语导致客体产生某种情绪的一种属性;例(14) c、例(14) d 属于历事主语(感受类)语义,强调主语的主观感受;而例(14) e 属于影响者主语(主动影响类)语义,强调主语的主观性及客体的受影响性。除了动词“烦”以外,很多其他情绪类动词也有类似歧义问题,例如“尴尬”“无聊”“讨厌”。这四个多义“情绪类”动词各自的歧义项及在语料中的分布如表 6 所示:

表 6 四个动词所有义项的分布频率

动词	框架	语义	用例	比例
烦	烦扰	S11: 主动烦人	50	10.0%
	刺激源-主语	S12: 是烦人的	162	32.4%
	历事-主语	S13: 感到烦	288	57.6%
尴尬	刺激源-主语	S21: 令人尴尬	188	37.6%
	历事-主语	S22: 感到尴尬	312	62.4%
无聊	刺激源-主语	S31: 令人无聊	184	36.8%
	历事-主语	S32: 感到无聊	316	63.2%
讨厌	刺激源-主语	S41: 令人厌恶	122	24.4%
	历事-主语	S42: 厌恶	378	75.6%

表 6 中,“框架”指的是语义网中各义项对应的语义框架。“语义”指四个多义动词的义项。“用例”指各义项在语料库中的频次。“比例”指基于每个动词找到的各 500 句中义项的分布比例。基于中文动词语义网中以上四个多义动词的参与角色及构式标注信息(FC),我们采用贝叶斯模型(NB)及优化支持向量机(SMO),进行了可监督的自动词义消歧任务,并比较了采用三种基本特征集——单语素(uni-gram)、双语素(bi-gram)、单名词性结构(uni-NP)进行同等消歧任务的准确度,发现以 FC 特征集的消歧准确度比其他三种特征集预测义项明显更高,F-score 最大差异达到 0.337,主要结果如图 3 所示。

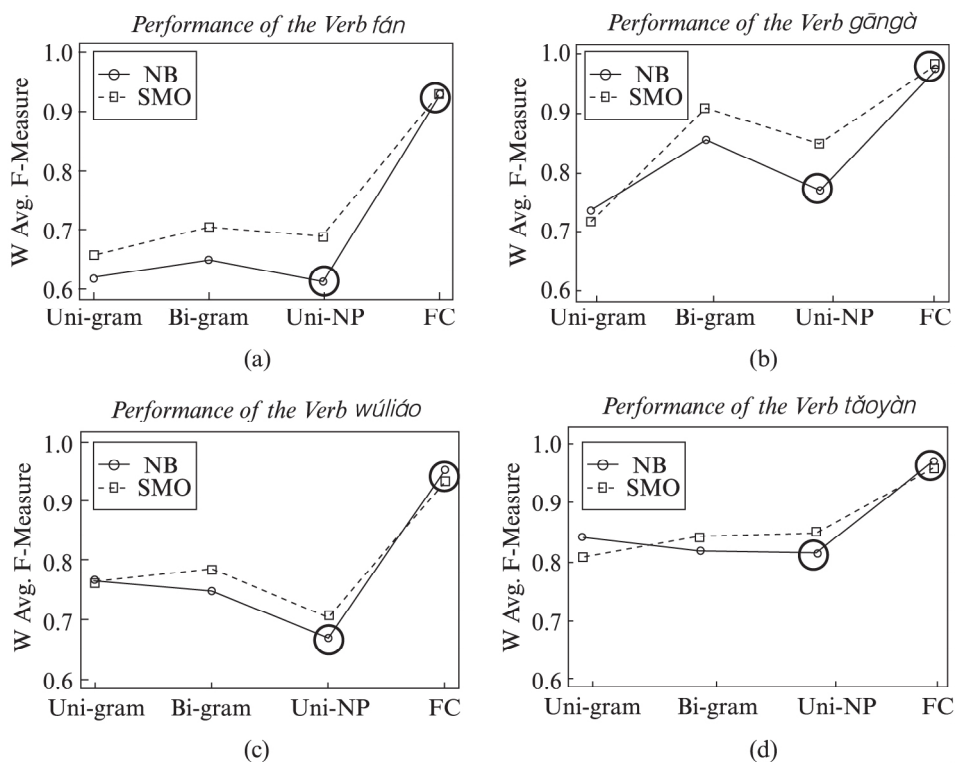


图 3 中文动词词义消歧的主要结果

另外,研究还发现,三个基本特征集中,名词性结构特征集的预测精确率最低,而双语素特征集最高,单语素特征集其次,但三者的总体表现差别不大;在四个动词的各义项中,历事主语的义项比刺激源主语的义项预测准确度更高,但刺激源主语的义项最易与其他义项发生混淆;机器学习曲线表明,FC 特征相比其他基本特征的预测优势随着样本空间的增加而呈现降低的趋势。但总体而言,使用 FC 用于词义消歧的功能非常显著且有效,印证了“一个框架,一个语义”(one sense, one frame)的理论正确性,也凸显了动词语义与其语法结构的有效互动。

## (二) 基于中文动词语义网的自动语义角色标注

中文动词词汇语义网现有的资源库的建立,由语言学导向,经过了实证分析,提供了一

个很好的语义角色自动标注训练集。目前数据库包含了 79707 条完成标注的句子及 265 种语义角色。在已有的资源库的基础之上,我们与北京大学孙薇薇教授合作了一个自动语义角色标注系统,并已经取得很好的标注成果(最好标注结果达到 88.6% 的准确度)初步实验结果如表 7 所示。其中“4l”和“6l”代表了神经网络模型的深度分别为四层及六层; Dev. S 表示词语输入的开发集自动标记准确性; Dev. F 表示具有语义标签输入的开发集自动标记准确性。

表 7 自动语义角色标注初步实验结果

	Dev. S	Test. S	Dev. F	Test. F
4l	88.0	86.7	84.4	83.4
6l	88.6	87.7	85.1	84.5

为实现语义角色的自动标注,该系统基于中文动词语义网前期标注好的语义数据集,采用了 N 层前馈神经网络模型(N-layer FNN)进行机器学习,如图 4 所示。

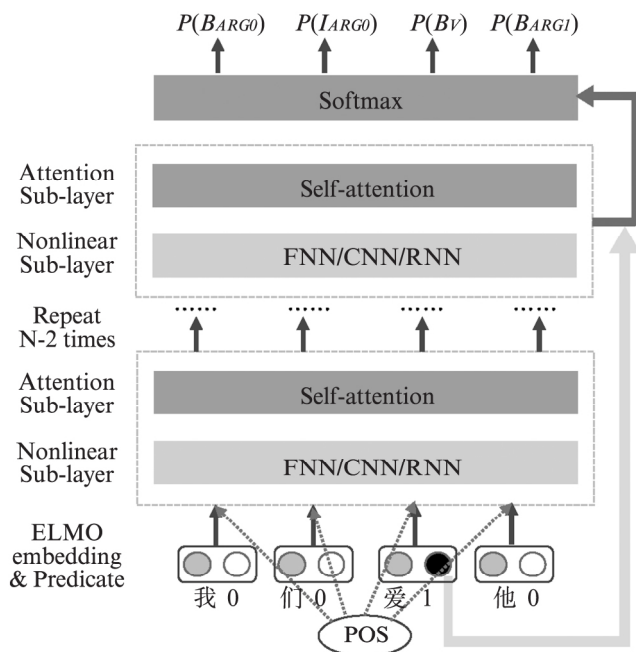


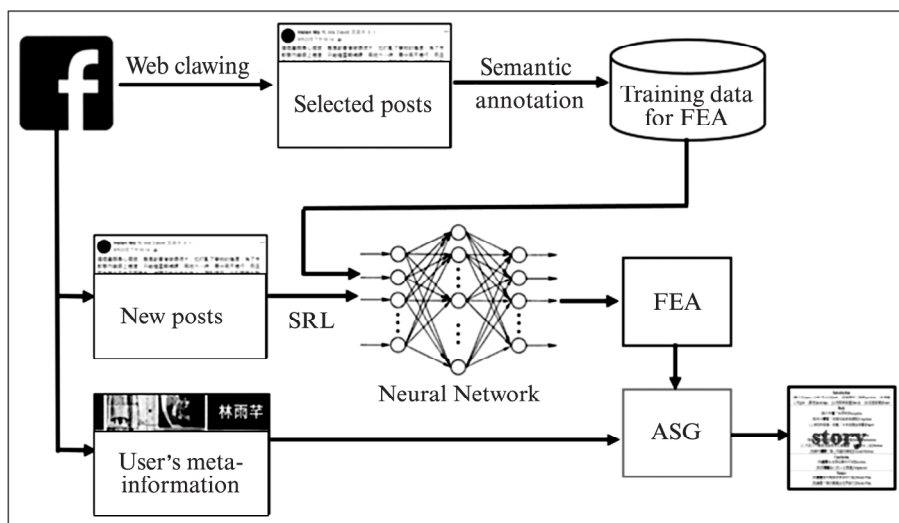
图 4 自动语义角色标记模型

该模型使用的 FNN 模型包含两个子层:非线性模型及注意力模型。在数据输入两个子层模型之前,句子被自动分词成主谓词(标记为 1)和非谓词语(标记为 0)的标记的词向量。这些带标记的特征向量被输入深度学习模型以实现语义角色的自动标记。借助于语义自动标注的实现,我们得以迅速扩充中文动词语义网的基本资源库,将“框架为本,构式为本”的语义分类及标注系统运用于广泛的中文可监督机器学习的自然语言处理任务之

中,以下文的应用为例。

### (三) 事件框架分析及自动故事生成

中文动词语义网的研究主要围绕动词展开语义标注,每个动词及其牵涉的参与角色及显性构式构成了一系列事件(event)的有效组成部件,因此,语义网中的动词语义标注及框架分类对于事件的分析、鉴别及检索提供了非常有效的语言学线索。基于此原理,我们利用中文动词语义网现有的大概10个源框架及256种语义角色标记,结合已经实现的语义角色自动标注体系(SRL: semantic role labelling),提出了一个用于社交网络(如脸书)的事件框架分析(EFA: event frame analysis)及自动故事生成(ASG: automatic story generation)系统。构建该系统的原理图如图5:



我们提出的自动故事生成系统,首先利用网上爬虫软件从社交网络自动抓取大量语料,语料包含两大信息:用户的个人信息及过往贴文。在已经抓取的语料基础之上,利用中文动词语义网现有的标注体系做事件框架建模、鉴别、分类,以及事件串联,最后生成针对网络用户的人生故事。

故事由四个主要部分组成“介绍”“主体”“结论”及“展望”。“介绍”由用户的个人信息组成,例如用户的出生日期、工作情况、家庭背景,等等。“主体”“结论”及“展望”则对应不同的事件类别。例如,“主体”部分包括了用户过往贴文中看(感官类动词)了什么,学习(认知类动词)了什么,吃(饮食类动词)了什么,说(沟通类动词)了什么,对其他人做了什么(社会互动动词),等等。表8举例说明了故事各组成部分对应的事件框架。

根据语义网的现有语义标注系统对各种动词的分类,以及事件框架分析,我们可以自动识别这些不同的事件类型,并按照以上四大部件自动串联成一个完整的人生故事。



表 8 基于中文动词语义网模型的自动故事生成系统模板

<p>介绍</p> <p>[陈**]姓名,[19**年出生]生日,[**工程师]职位,[在**工作]工作, [单身]婚否,[父母移居**]家庭背景,[好友刘**]朋友</p>
<p>主体</p> <p>[他今年看了世界杯]感官 [他年中学习了信号控制系统课程]认知 [上周他和家乐一起吃了日本放題自助餐]饮食 [他们聊天聊了一整晚]沟通 [他支持朋友年底结婚及买楼计划]社会互动 [三天前山竹台风过后香港交通瘫痪 他跑步去上班]自动 [他顺手挪开了路上被吹断的树枝]致使移动</p>
<p>结论</p> <p>[他感觉生活很枯燥而平淡]情感 [他很怀疑自己的人生意义]评判</p>
<p>展望</p> <p>[他想要趁年轻感受更多的可能]计划-愿望 [他决定下个月辞职去世界旅行]计划-愿望</p>

## 六、结 论

本文以中文动词词汇语义网的理论研究、网站构建及实践应用为例,全面性地总结和探讨了当下人工智能盛行的语言学语义资源构建的关键问题和可能的应用。本文以“框架为本 构式有用”的研究方法为基础,依循语言学家 Fillmore 提出的框架语义及 Goldberg 提出的构式语法理论,并参考以英语动词为研究对象的 FrameNet,以实际的语料为依据,具体而微、形义兼顾地对中文动词进行系统有效的分类。网站的构建经过了语料收集和筛选、人工标注、交叉验证及网站开发等步骤,构建了一个具备词汇表征、语义标注、分类标准、框架层级(包括源框架>初级框架>基本框架>微框架)等信息的语义知识库。语义网目前已经实现八个源框架的动词研究及语义标注,包括“沟通”“认知”“感知”“情绪”“社会互动”“自动”“致使移动”及“评判”类动词。同时将继续对“空间配置”“自然行为”等更多的语义框架进行系统性地分类和研究。语义网作为富含语言学智慧的中文动词词汇语义知识库,除了为中文教学及本体研究提供了包含动词语义框架及标注的基本信息,也提供了一个可有效运用于各种语义相关的中文自然语言处理的宝贵资源库。

### 附 注

- [1] <https://framenet.icsi.berkeley.edu/fndrupal/>
- [2] <https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>
- [3] <http://ucrel.lancs.ac.uk/usas/>
- [4] <https://propank.github.io/>
- [5] <http://verbs.colorado.edu/chinese/cpb/>

[6] <http://treebank.sinica.edu.tw/>

[7] <http://mega.lt.cityu.edu.hk/~yufechen/#/>

## 参考文献

1. 刘洪超 詹卫东. “A + 一 + X , B + 一 + Y” 构式的分类及释义模板. 中文信息学报 2014 28(6): 34-40.
2. Boas H C. A Lexical-constructional Account of the Locative Alternation. // Carmichael L , Huang Chia-Hui , Samiian V. ( eds. ) *Proceedings of the 2001 Western Conference in Linguistics* , Fresno , CA: California State University , 2003: 27-42.
3. Chen Keh-Jiann *et al.* The CKIP Chinese Treebank: Guidelines for Annotation. *ATALA Workshop—Treebanks* , Paris , 1999: 85-96.
4. Fillmore C J. Frame Semantics. // Linguistic Society of Korea ( ed. ) *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Company , 1982: 111-137.
5. Fillmore C J , Atkins B T. Towards a Frame-based Organization of the Lexicon: The Semantics of RISK and Its Neighbors. // Lehrer A , Kittay E. ( eds. ) *Frames , Fields , and Contrasts: New Essays in Semantics*. Hillsdale: Lawrence Erlbuan , 1992: 75-102.
6. Fillmore C J , Baker C F. FrameNet: Frame Semantics Meets the Corpus. // Robinson D , Farrell P. ( eds. ) *Proceedings of the 74th Annual Meeting of the Linguistic Society of America* , Chicago , 2000.
7. Goldberg A E. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press , 1995.
8. Goldberg A E. Relationships between Verb and Construction. // Verspoor M , Sweetser E. ( eds. ) *Lexicon and Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company , 1997: 383-398.
9. Goldberg A E. Verbs , Constructions , and Semantic Frames. // Rappaport H M , Doron E , Sichel I. ( eds. ) *Syntax , Lexical Semantics and Event Structure*. Oxford: Oxford University Press , 2010: 39-58.
10. Huang C R , Chen F Y , Chen K J , *et al.* Sinica Treebank: Design Criteria , Annotation Guidelines , and On-line Interface. // Association for Computational Linguistics. *Proceedings of the Second Workshop on Chinese Language Processing: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics* , 2000 ( Vol. 12 ) : 29-37.
11. Kingsbury P , Palmer M. From TreeBank to PropBank. // *LREC* , Las Palmas , Canary Islands Spain , 2002: 1989-1993.
12. Levin B. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press , 1993.
13. Levin B , Hovav R M. Lexical Semantics and Syntactic Structure. // Lappin S. ( ed. ) *The Handbook of Contemporary Semantic Theory* , Oxford: Blackwell: 487-507.
14. Levin B , Hovav R M. *Argument Realization*. Cambridge: Cambridge University Press.
15. Liu Mei-Chun. Motion , Direction and Spatial Configuration: A Lexical Semantic Study of HANG Verbs in Mandarin. // Shay E , Seibert U. ( eds. ) *Motion , Direction and Location in Languages: In Honor of Zygmunt Frajzyngier*. Amsterdam/Philadelphia: John Benjamins Publishing Company , 2003: 177-187.
16. Liu Mei-Chun. Lexical Information and Beyond: Meaning Coercion and Constructional Inference of the Mandarin Verb GAN. *Journal of Chinese Linguistics* , 2005 33(2) : 310-332.

17. Liu Mei-Chun. Emotion in Lexicon and Grammar: Lexical-constructional Interface of Mandarin Emotional Predicates. *Lingua Sinica*, 2016 2(4).
18. Liu Mei-Chun. A Frame-based Morpho-constructional Approach to Verbal Semantics. // Kit Chun-Yu, Liu Mei-Chun (eds.) *Empirical and Corpus Linguistic Frontiers*. Beijing: China Social Sciences Press, 2018.
19. Liu Mei-Chun, Chang Chun-Edison. From Frame to Subframe: Collocational Asymmetry in Mandarin Verbs of Conversation. *International Journal of Computational Linguistics and Chinese Language Processing*, 2005 10(4): 431-444.
20. Liu Mei-Chun, Chang Jui-Ching. Semantic Profile as a Source for Polysemy: Insight from the Spatial-configuration Verb fang in Mandarin. // Lu Qin, Gao H H. (eds.) *Chinese Lexical Semantics*. Berlin: Springer International Publishing, 2015(9332): 24-32.
21. Liu Mei-Chun, Chang Jui-Ching. Placement Verbs in Chinese and English: A Contrastive Study of Lexicalization Patterns. // Hong Jia-Fei, Su Qi, Wu Jiun-Shiung (eds.) *Chinese Lexical Semantics*, 2018 20(2).
22. Liu Mei-Chun, Chang Jui-Ching. From Caused Motion to Spatial Configuration: Placement Verbs in Mandarin. // Peck Jeeyoung (ed.) *Language and Linguistics*, 2019 20(2).
23. Liu Mei-Chun, Chiang Ting-Yi, Chou Ming-Hui. A Frame-based Approach to Polysemous Near-synonymy: The Case with Mandarin Verbs of Expression. *Journal of Chinese Language and Computing*, 2006 15(3): 137-148.
24. Liu Mei-Chun, Chiang Ting-Yi. The Construction of Mandarin VerbNet: A Frame-based Approach to the Classification of Statement Verbs. *Language and Linguistics*, 2008 9(2): 239-270.
25. Liu Mei-Chun, Hu Chia-Yin. Conceptual Schema as Semantic Link: A Frame-based Study of Mandarin Cognition Verbs. *International Journal of Computer Processing of Oriental Languages*, 2008 21(1): 55-75.
26. Liu Mei-Chun, Hu Chia-Yin. Free Alternation? A Study on Grammatical Packaging of Excessive Predication in Mandarin Chinese. *Language and Linguistics*, 2003 14(1): 47-90.
27. Liu Mei-Chun, Hu Chia-Yin, Tsai Hsin-Shan et al. The Proto-motion Event Schema: Integrating Lexical Semantics and Morphological Sequencing. *Journal of Chinese Linguistics*, 2015 43(2): 503-547.
28. Manning C D. Part-of-speech Tagging from 97% to 100%: Is It Time for Some Linguistics? // *International Conference on Intelligent Text Processing and Computational Linguistics*, 2011: 171-189.
29. Marcus G. Deep Learning: A Critical Appraisal. CoRR, abs/1801.00631 2018.
30. McArthur T G. Longman Lexicon of Contemporary English. London: Longman, 1981.
31. Palmer M, Gildea D, Kingsbury P. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 2005 31(1): 71-106.
32. Schuler K K. VerbNet: A Broad-coverage, Comprehensive Verb Lexicon. Ph. D Dissertation, University of Pennsylvania, 2005.
33. Wan Ming-Yu, Liu Mei-Chun. Supervised Word Sense Disambiguation with Frame-based Constructional Features: A Pilot Study of *fán* 'to annoy/be annoying/be annoyed'. *International Journal of Knowledge and Language Processing*, 2018 9(2): 33-46.
34. You L, Liu K. Building Chinese Framenet Database. // *Natural Language Processing and Knowledge Engineering, Proceedings of 2005 IEEE International Conference*, Wuhan, 2005: 301-306.

(香港城市大学翻译及语言学系 香港)

(责任编辑 马 沙)

## Chinese Verbs and Their Categorization: Construction and Application of the Semantic Network of Chinese Verbs

Liu Mei-Chun Wan Mingyu

**Abstract:** With the growth of AI , linguistics has become a new field in which researchers in different areas can cooperate and make breakthroughs. The construction and labeling of semantic resources have thus turned into a hot topic , as well as a difficult point to overcome. This article discusses Chinese verbs and their semantic categorization from three perspectives , namely , theoretical bases , semantic network construction , and application. The theory used for this research is framework semantics and particularly Gestalt; the construction of the semantic network is based on linguistic analysis and language data. The semantic knowledge base currently consists of eight types of verbs and has been effectively used in many kinds of Chinese natural language processing tasks , including automatic semantic disambiguation , automatic semantic role labeling , event identification , and automatic story generation.

**Keywords:** frame semantics , construction grammar , verb categorization , semantic labeling , natural language processing

## On the Origin and Growth of *wuye*

Liu Danqing Zhang Ding

**Abstract:** The word *wuye* in contemporary Chinese comes from Hong Kong , but its meanings and uses are different in China ' s mainland and Hong Kong. This article describes the different treatments of this word in various dictionaries and investigates its origin and semantic growth.

**Keywords:** *wuye* , origin , semantic evolution , language contact

## Categorization and Characteristics of the Phenomenon of “Different Tones Leading to Different Meanings” in the Wu Dialect

You Rujie

**Abstract:** Taking Wenzhou dialect as an example , this article studies the categorization and characteristics of the phenomenon of “different tones leading to different meanings” in Wu dialect. It argues that there are three major types of this phenomenon and this is different from the traditional “four tones and different word meanings”.

**Keywords:** four tones and different word meanings , different tones leading to different meanings , Wu dialect , Wenzhou dialect