

# 基于扩展依存语法模型的汉语动补结构语义标注研究\*

陈波<sup>1,2</sup>, 吕晨<sup>2</sup>, 姬东鸿<sup>2</sup>

(1.湖北文理学院 文学院, 湖北 襄阳441053; 2.武汉大学 计算机学院, 湖北 武汉430072)

**摘要:** 汉语的动补结构中主、宾、动、补四个成分之间的语义关系错综复杂, 自然语言处理很难对动补结构的语义关系进行完整的形式化描述。本文提出了一个新颖的扩展依存语法模型, 运用特征结构的方法形式化地描述动补结构中的主、宾、动、补四个成分之间的复杂语义关系。通过分析汉语动补结构的十五种语义关系, 对其进行分类概括, 并将扩展依存语法与传统依存语法的方法进行对比, 结果表明, 扩展依存语法是一个有向图结构, 允许多重节点之间的语义关联, 并允许递归, 可以比较完整地描述汉语的复杂语义关系, 提高标注精度和效率。

**关键词:** 中文信息处理, 扩展依存语法, 动补结构, 特征结构

**中图分类号:** TP391

**文献标识码:** A

## Semantic Labeling of Chinese Verb-complement Structure Based on

## Extended Dependency Grammar

Bo Chen<sup>1,2</sup>, Chen Lv<sup>2</sup>, Donghong Ji<sup>2</sup>

(1. Department of Language and Literature, Hu Bei University of Arts & Science, Xiangyang, Hubei 441053, China;

2. School of Computer, Wuhan University, Wuhan, Hubei 430072, China)

**Abstract:** Semantic relations of Chinese verb-complement structure are complicated. It's difficult to analyze semantic relations in NLP. We proposed an extended dependency grammar and applied it in representing the semantic relations among the four components which were subject, verb, object and complement. We annotated the fifteen types of semantic relations, and compared extended dependency grammar with traditional dependency grammar. The results show that the extended dependency grammar is a directed recursive graph which can describe more Chinese semantic information, and achieve higher annotating efficiency and accuracy.

**Keywords:** Chinese information processing; extended dependency grammar; verb-complement structure; feature structure

### 1 前言

语义分析是自然语言处理 (Natural Language Processing, 简称 NLP) 领域最具挑战性的课题之一, 也是当前制约语言信息技术大规模应用的主要瓶颈。在众多语义分析的问题中, 短语级的语义分析是一项最基本的任务。汉语由于具有语序灵活等特点, 它的语义分析更具挑战性。在自然语言处理中, 对汉语语句的语义标注, 特别是汉语特殊结构的语义标注, 是亟须解决的难题, 如“动补结构”、“复杂名词短语”、“把字句”、“被动句”等等。

在语言学界, 动补结构 (verb-complement structure) 是汉语中一种具有独特特点的句型, 它一般包含“主语” (subject)、“谓语动词” (verb)、“补语” (complement)、“宾语” (object) 四个成分。这四个成分之间的语义关系错综复杂, 如补语与主语之间的语义关系、补语与谓语动词之间的语义关系、补语与宾语之间的语义关系, 以及主语、谓语动词与宾语三者之间的语义关系, 都是目前研究的重点及难点。

如何寻找一种较为有效的方法, 可以对汉语动补结构进行更好的语义标注和分析, 对于语言学界和 NLP 学界, 都具有重要意义。本文提出了一种新颖的扩展依存语法模型 (Extended

\*收稿日期:

定稿日期:

**基金项目:** 国家自然科学基金资助项目(61202193); 国家哲学社会科学重大项目招标项目 (11&ZD189); 湖北省重点学科建设项目成果; 中国博士后科学基金(2013M540593)。

**作者简介:** 陈波 (1976—), 女, 副教授, 博士后, 主要研究方向为自然语言处理; 吕晨 (1989—), 男, 博士研究生, 主要研究方向为自然语言处理; 姬东鸿 (1967—), 男, 教授、博导, 主要研究方向为自然语言处理。

Dependency Grammar)——“特征结构理论”(Feature Structure Theory),选取汉语动补结构的十五种语义关系类型(吕叔湘,1980)进行了形式化的描述和语义分析,与传统依存语法的方法进行对比,得到了比较好的结果。

## 2 汉语动补结构语义研究简介

### 2.1 语言学界

动补结构是指两个或两个以上的词组合起来,发生补充与被补充的关系,即动补关系,具有这种关系组合的称动补结构,这种结构是汉语所特有的。助词“得”是这种结构的形式标志,如“累得很”。在汉语中也有大部分是直接组合的,如:“好极了”。

含动补结构的句子,一般句式是:

主+动+补+宾

形式化可以表示为:

S+V+C+O

语言学界对动补结构的研究包括形式上的分类和语义上的分类研究等,本文侧重语义关系的研究,将基于语言学家们的研究成果,对动补结构的四个成分之间的语义关系进行形式化描述和语义标注。这四个成分之间的语义关系类型可以概括为下图1:

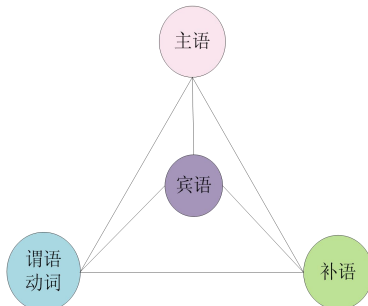


图1 动补结构四个成分语义关系概图

如图一所示,动补结构的语义关系包括两大类6小类,首先是补语与主语、补语与谓语动词、补语与补语之间的语义关系;其次是主语与谓语动词、谓语动词与宾语、主语与宾语之间的语义关系。我们可以看一组典型的动补结构语例,如例1-例3:

例1 衣服洗干净了。

例2 衣服洗完了。

例3 衣服洗晚了。

例1-例3的句法格式均可描述为:

“S(衣服)+V(洗)+C(形容词)”

这三句主语与谓语动词相同,仅补语不同,补语由不同的形容词充当,而且这三句话没有宾语,语义关系相对简单,省去了宾语与其他三个成分之间的关系。通过对比,我们可以发现,例1-例3中相同之处在于主语“衣服”与谓语动词“洗”之间的语义关系是主语是动词的受事,而三个补语在句中与其他成分产生的语义关系则各不相同。例1中,补语与主语、谓语动词都产生语义关联,主语“衣服”与补语“干净”是主谓关系,可以说成“衣服干净了”。例2中,补语与主语产生语义关联,补语“完”与主语“衣服”的数量有关,同时也补充说明谓语动词“洗”这个动作的结束。例3中,补语也与动词产生语义关联,但是补语“晚”是补充说明动词“洗”发生的时间。从例1-例3可以看出,动补结构由于句法结构相似,提高了完整而精确描述其内部语义关系的难度。

### 2.2 基于传统依存理论的动补结构分析

依存结构分析(dependency structure)是国际上语义标注采用的主流方法,运用依存结构树来标注汉语动补结构,如图2所示:

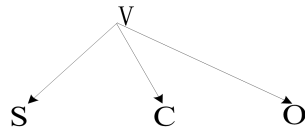


图2 动补结构传统依存分析方案

运用依存结构分析例 1-例 3，结果将都是图 3 所示：

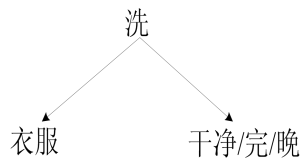


图3 例 1-例 3 依存结构分析树

依存分析法对例 1-3 的分析，漏掉了主语与补语之间的语义关系。再如例 4：

例 4 小王羞红了脸。

例 4 的依存分析树如图 4 所示：

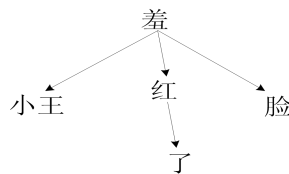


图4 例 4 依存结构分析树

根据语义关联，我们按照语义认知来分析，例 4 中具有语义关系的词语对(semantic word pairs)至少有 5 组：

[小王，羞]；[小王，脸]；[羞，红]；[脸，红]；[红，了]。

图 4 是传统依存语法对例 4 的分析结果，可以看到，运用依存分析法的分析结果，仅描述了谓语动词与主语、谓语动词与补语、谓语动词与宾语的语义关系，而主语“小王”与宾语“脸”之间的领属关系、宾语“脸”与补语“红”之间的主谓关系却没有表示出来。

因此，从例 1-4 可以看出，当前的传统依存分析法在分析汉语动补结构时丢失了很多语义信息，为进一步的自然语言处理工作带来一定的困难。

### 3 基于扩展依存语法的动补关系的语义描述策略

#### 3.1 扩展依存语法

考虑到汉语的独特特点，如灵活语序、错综复杂的语义关系，我们基于传统依存语法将其进行了改进，提出了一种扩展的依存语法描述体系 (extended dependency grammar)，这种方法用特征三元组的形式来描述汉语句子里每一对词语的语义关系，可以表示出多重节点之间的语义关系，并且允许递归 (recursive)，形式化表示为“有向的可递归的图结构” (directed recursive graph)。

冯志伟(2004)认为，找出句子中单词之间的语义关系才是机器翻译中语义分析的最终目的。这种语义关系与单词本身的语义特征不同，而且与单词在句子和文中的上下文密切相关。扩展依存语法体系的基本任务是对句子或短语内部概念关联和关联的种类的描写。

例 5 从广州飞 飞到武汉

例 5 可描述为(飞，从，广州)，(飞，到，武汉)，“飞”和“广州”、“飞”和“武汉”都存在概念关联，如果在这两个关联对中添加上关联种类的话，就是“飞-从(起点)-广州”，“飞-到(终点)-武汉”，直观上，这个三元组正好可以理解为“实体、特征、特征值”关系。例 5 可以用特征三元组表示为：

[飞，从，广州] [飞，到，武汉]

一般来说，一个短语或句子可表示成一个由实体(Entity)、特征(Feature)和特征值(Value)组成的三元组(Triple)的集合，这个集合我们称之为该短语或句子的特征结构(Feature Structure)，一个特征

三元组反映一对语义关联，实体作为中心词（the head），特征值依存于实体。例 5 可形式化表示为图 5：

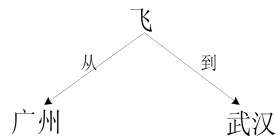


图 5 例 5 扩展依存结构分析图

例 4“小王羞红了脸。”的扩展依存结构分析图可表示为如图 6 所示：

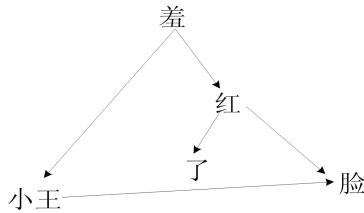


图 6 例 4 扩展依存结构分析图

如图 6 所示，除了动词中心词“羞”之外，主语“小王”、补语“红”、补语“脸”都是多重节点，同时与两个词产生语义关联。扩展依存结构分析的方法表示为有向的可递归的图，可以描述出这个动补结构中各个词语之间的语义关系。

### 3.2 扩展依存语法的形式表示

特征结构关注于描述词汇之间的语义关系，因此不用定义句法范畴，避免了语言学界关于句法问题的争端。我们通过大规模的真实文本的语义标注，特征结构的方法在形式上不仅允许嵌套 (embedded)，还可以表示词语之间语义的多重关联。一个特征三元组形式上可以表示为两个“节点”(node)和连接它们的一条“边”(edge)。实体或特征值表示为两个“节点”，特征表示为之间的“边”。这样，一个特征结构可看成是一个“图”(graph)，两个节点之间的方向是从特征值指向实体，因此表现为一个“有向图”(directed graph)。由于特征结构允许嵌套和多重关联，因此可表示为“有向的递归图”(recursive graph)，如图 7 所示：

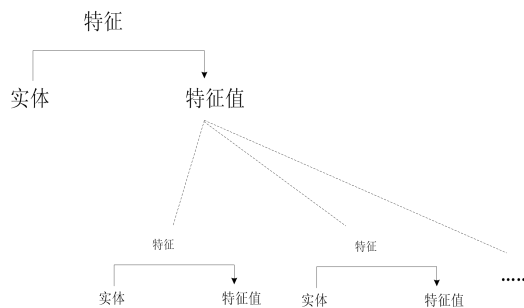


图 7 特征结构的有向递归图

### 3.3 动补结构的扩展依存结构模型分析

我们以汉语动补结构的语义关系为描述对象，选取了《含动补结构的句子的语义分析》（吕叔湘，1986）的例句和语言学分类结果，共三大类十五小类，运用扩展依存结构分析方法细致地分析了这 15 类语义关系的特征结构三元组和扩展依存结构分析图。因篇幅关系，本文仅选取其中三种语义类型的例句，分析如下：

例 6 他喝醉了酒。

例 6 动补结构的特点是补语“醉”与主语“我”产生语义关系，可以说成“他醉了”，即主语和补语构成主谓关系（吕叔湘，1986），该类动补结构分布较广。例 6 包括五个特征三元组，分别是：

[喝, , 他]; [喝, , 酒]; [喝, , 醉]; [醉, , 他]; [醉, , 了]  
它的扩展依存结构分析图如图 8 所示:

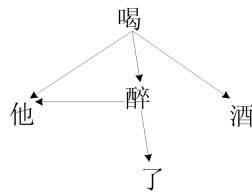


图 8 例 6 扩展依存结构分析图

例 7 (踢球, 踢球,) 一个月踢坏了三双鞋。

例 7 动补结构的特点是补语“坏”与宾语“鞋”产生语义关系,可以说成“鞋坏了”。即宾语和补语构成主谓关系(吕叔湘, 1986), 该类动补结构也较为常见。例 7 包括五个特征三元组, 分别是:  
[踢, , 一个月]; [踢, , 坏]; [坏, , 鞋]; [鞋, 双, 三]; [坏, , 了]  
它的扩展依存结构分析图如图 9 所示:

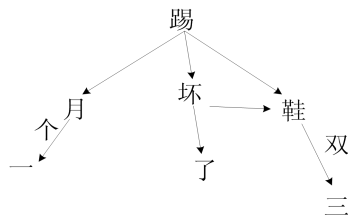


图 9 例 7 扩展依存结构分析图

例 8 你(吃浆糊)吃迷了心了。

例 8 动补结构的特点是补语“迷”与宾语“心”产生语义关系, 构成了一个述宾短语“迷了心了”。这里补语和宾语构成了述宾关系(吕叔湘, 1986), 而且作为一个整体与谓语动词“吃”发生语义关联。例 8 包括五个特征三元组, 分别是:

[吃, , 你]; [吃, , 迷了心]; [迷, , 心]; [迷, , 了]; [迷了心, , 了]  
它的扩展依存结构分析图如图 10 所示:

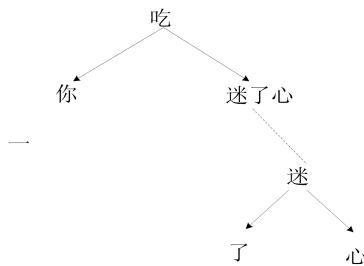


图 10 例 8 扩展依存结构分析图

我们考察了包括把字句、被字句在内的十五种动补结构的语义关系, 对典型的动补结构的扩展依存结构图进行了归纳, 每一类的扩展依存结构图可以清楚地反映出动补结构中四种成分(有时宾语并不出现)之间不同的语义关系, 详见表 1:

表 1 动补结构扩展依存结构分析图类型

典型的扩展依存结构分析图	语义关系	例句
<pre>           V          / \         S   O       </pre>	<p>该类是最常见的动补结构句型。谓语动词分别与主语、补语和宾语有语义关联。</p>	<p>他洗完了衣服。</p>

2		该类主语与补语具有主谓关系，其中分为两小类，主语和谓语之间的语义关系，一为施事和动词，二为受事和动词。	他喝醉了酒。 字写错了。 妈妈被女儿逗笑了。
3		该类补语和宾语之间具有语义关系，其中分为两小类，一种是“补语-宾语”为述宾关系；一种是“宾语-补语”之间为主谓关系。	你吃迷了心了。 他羞红了脸。 孟姜女哭倒了长城。 我把杯子打破了。

三种类型下面还包括若干小类，由于篇幅关系，这里不再赘叙。

#### 4 传统依存分析与扩展依存分析两种策略的标注分析对比

我们以例 6-7 为例，运用扩展依存结构分析与传统依存分析两种方法，将分析结果进行对比，来评测对动补结构的语义分析效果。

表 2 例 6-7 两种分析策略结果对比表

传统依存分析树	扩展依存结构分析图

如表 2 所示，传统依存分析结果没有把主语“他”和补语“醉”、宾语“鞋”和补语“坏”的语义关系表示出来，丢失了重要的语义信息。

扩展依存结构模型考虑到汉语灵活语序的特点，也考虑到一个词语与多个词语之间可能产生语义关系，因而能够描述更多的词语之间的语义信息。我们运用扩展依存结构的方法，对宾州中文树库 CTB5.1 中的动补结构语料进行了语义标注，取得了较好的效果。

#### 5 结语

本文回顾了汉语动补结构的研究成果，总结了动补结构中的主、宾、动、补四个成分之间的复杂语义关系类型，分析了面向自然语言处理时动补结构的标注难点。提出了一个新颖的扩展依存语法模型，运用特征结构的方法形式化地描述动补结构语义关系。通过分析汉语动补结构的十五种语义关系，对其进行分类概括，并将扩展依存语法与传统依存语法的方法进行对比，结果表明，传统依存语法无法表示出主语与补语、宾语与补语之间的语义关系，扩展依存语法可以比较完整地表示出各个成分之间的复杂语义关系，因而包含更加丰富的语义信息。扩展依存语法模型是一个有向图结构，允许多重节点之间的语义关联，并允许递归，可以比较完整地描述汉语的复杂语义关系，提高标注精度和效率，为面向汉语的自然语言处理提供了一种不同的语义分析方法。

## 参考文献

- [1] Chen Bo, Ji Donghong, Lv Chen. Building a Chinese Semantic Resource Based on Feature Structure [J]. International Journal of Computer Processing Of Languages, 2012, 24(01): 95-101.
- [2] Chen Bo, Wu Hongmiao, Lv Chen, et al. Semantic labeling of Chinese serial verb sentences based on feature structure[C]. Chinese Lexical Semantics, 2013: 784-790.
- [3] Ji Donghong. Semantic annotation of Chinese phrases using recursive-graph[C]. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, Hong Kong: Association for Computational Linguistics, 2000:101-108.
- [4] Zhou Q. Annotation Scheme for Chinese Treebank, Chinese Information Processing, 2004: 1-8.
- [5] Zhou M, Huang CN. Approach to the Chinese Dependency Formalism For the Tagging of Corpus. Chinese Information Processing, 1994: 35-53.
- [6] 陈波. 基于特征结构的汉语主谓谓语句标注[J].中文信息学报,2012,(3):22-25.
- [7] 陈波, 姬东鸿, 吕晨. 基于特征结构的汉语连动句语义标注研究[J]. 中文信息学报, 2013, (5): 60-66.
- [8] 冯志伟. 机器翻译研究[M].北京:中国对外翻译出版公司,2004:467-469.
- [9] 陈波. 特征结构及其汉语语义资源建设[D].武汉: 武汉大学,2011.
- [10] 高燕. 从认知角度看现代汉语动补结构的语义指向[D]. 上海: 上海外国语大学, 2005.
- [11] 吕叔湘. 含动补结构的句子的语义分析[C].第一届国际汉语教学讨论会论文集,1985.
- [12] 张嘉宾. 动补结构与其宾语之间的语义, 语法关系[J]. 求是学刊, 1984, (1): 4.