

# 基于规则的把字句语义角色标注\*

何保荣, 邱立坤, 徐德宽

(鲁东大学文学院, 山东 烟台 264025)

**摘要:** 把字句是现代汉语中一种重要的特殊句式, 本文尝试用基于知识库的规则方法对把字句进行语义角色自动标注。首先, 我们从《人民日报》语义角色标注语料库中收集把字句例句, 形成一个覆盖范围较广的把字句例句库。之后, 对例句库中把字句的句法和语义构成规律进行手工标注, 标注内容包括谓语动词的配价类型、把字句谓语结构类型、把字句句模类型等。在上述标注的基础上, 对把字句的句模构成规律进行分析, 总结出若干条语义角色标注规则。最后, 在测试数据上对前述规则进行验证, 语义角色标注的最终正确率为 98.61%, 这一结果说明本文所提出的规则在把字句语义角色标注上是有效的。

**关键词:** 把字句; 语义角色标注; 句模

中图分类号: TP391

文献标识码: A

## Rule-based Semantic Role Labeling of Ba-sentences

HE Baorong, QIU Likun, XU Dekuan

(School of Chinese Language and Literature, Ludong University, Yantai, Shandong 264025, China)

**Abstract:** Ba-sentence is one of the most important Chinese special sentence patterns. This paper proposed a rule-based method to address the task of automatic semantic role labeling, especially focusing on ba-sentences. Firstly, we collected a set of ba-sentences from our annotated semantic corpus, including texts from *People's Daily*, and thus formed a sample gallery of ba-sentences. Then, we tagged the valence type of each predicate, the syntactic structure type and semantic structure type of each ba-sentence manually. Based on this annotated corpus, we analyzed the rules of semantic formation, and summed up several rules of semantic role labeling. Finally, we evaluated these rules in a test set. The overall precision is 98.61%, which shows that the proposed method is effective in labeling the semantic roles of ba-sentences.

**Key words:** ba-sentence; semantic role labeling; semantic sentence pattern

### 1 引言

语义角色标注是一种浅层语义标注, 其主要内容是识别谓词的论元, 并为每个论元标注一个语义角色<sup>[1]</sup>。现有研究一般将语义角色标注视为分类问题或者序列标注问题, 使用最大熵模型<sup>[2]</sup>、条件随机场模型<sup>[3]</sup>等予以实现。在训练数据较为充足的情况下, 已取得较高精度。但现有自动标注方法主要使用句法信息和词汇信息, 较少考虑谓词的格框架以及语义角色与句式(句子的句法结构类型)之间的配合关系。

语言学中与语义角色标注相关的研究包括句模研究和格语法等。所谓句模, 一般定义为句子的语义结构类型<sup>[4]</sup>, 一个句模由若干个语义角色组成。比如鲁川等<sup>[5]</sup>归纳出存在句模、领属句模、特征句模等 26 大类基本句模和 122 小类特殊句模, 其中存在句模包括两个语义角色——“处所”和“当事”。将句模与特定类别的动词联系起来, 就可以形成动词的格框架。句模和格框架都将句子的语义结构作为整体来看待, 但是现有的句模研究也较少与句式结合起来, 即没有把语义与句法两个层面联系起来。

针对上述问题, 本文尝试将谓词格框架和句式结合起来, 并以现代汉语的把字句为例来

\* 收稿日期: 定稿日期:

**基金项目:** 国家社科基金重大项目(12&ZD227); 国家自然科学基金项目(61572245, 61103089); 鲁东大学人文社会科学研究项目(WY2013003)

**作者简介:** 何保荣(1990—), 女, 硕士研究生, 主要研究方向为中文信息处理; 邱立坤(1979—), 通讯作者, 男, 博士, 副教授, 主要研究方向为计算语言学; 徐德宽(1968—), 男, 博士, 教授, 主要研究方向为计算语言学。

研究特定句式中的语义角色自动标注问题。本文工作包括以下三个方面：(1) 以基于人民日报新闻的语义角色标注语料库为主，构建覆盖范围较广的把字句例句库，并对例句库中把字句的句法和语义构成进行人工标注，标注内容包括谓语动词的配价类型、谓语结构类型、句模类型等；(2) 基于人工标注的开发集，总结出若干条语义角色标注规则；(3) 使用前述规则对人工标注的测试集进行标注，以验证标注规则的有效性。

本文实验结果表明，在格框架词典的帮助下，基于少量规则可以较好地对绝大部分把字句进行语义角色自动标注。

## 2 相关研究

就语义角色标注的语料资源而言，目前较为知名的语义角色标注主要集中在英语和汉语语料上。FramNet<sup>[6]</sup>和 PropBank<sup>[7]</sup>是其中的佼佼者。近年来，汉语语义角色标注也逐渐发展起来。Chinese PropBank<sup>[8]</sup>是根据 English PropBank 的标注方法，对汉语进行的语义角色标注；Chinese FramNet<sup>[9]</sup>则是山西大学根据框架语义学理论构建起来的中文语义标注语料库。

范晓、胡裕树<sup>[10]</sup>在三个平面语法理论的基础上，进一步强调从句法、语义和语用三个层面来研究句子的结构，并首次提出了句子的基本语义结构类型——句模。句模由动核结构生成，是动核结构在生成句子时与句型结合的语义配置模式<sup>[11]</sup>。徐昌火<sup>[12]</sup>认为句模研究的起点应是各种客观存在的基本联系，形式化配价研究是句模研究的最终目的；孙道功等<sup>[13]</sup>在语料标注的基础上构建了现代汉语单句句型、句模对应关系，并总结出了句型和句模并非一一对应等几个对应特点；亢世勇等<sup>[14]</sup>在大规模标注语料库的基础上构建了现代汉语句系系统，从句型的角度考察了句型、句模的对应关系并统计出了句模的频度和分布状况；郑丽娟、邵艳秋<sup>[15]</sup>基于语义依存图库对兼语句句型和句模的对应关系进行了研究，归纳出了兼语句两个事件之间的四种语义关系。以上研究，有些对句模研究的意义、目的和方法等进行了叙述和分析，但过于概括，没有关注具体句式的句模；有些虽关注了句式、句模的对应关系，但并未考虑复杂谓语结构的句模，句模与动词、句式的对应规律也未作出实质的总结。因此，句模研究还有待进一步的深入和细化。

把字句历来是学界研究的热点。以往对于把字句的研究主要集中于其句法结构、内部深层语义联系、生成条件及原因和与其他句式的变换及变换条件的限制等方面，而对于把字句句模的研究则鲜有涉及。范晓《动词的配价与汉语的把字句》<sup>[16]</sup>分析了把字句的谓语中心词和谓语次要动词，归纳出了一价、二价和三价动词所形成的 21 种把字句句模，最后，归纳出了 10 种把字句句式并分析了句式与动词、动核结构的关系，但范文并未考虑句模与具体动词和句式的联系，难以确定一个动词与某种句模之间的具体对应关系。

## 3 研究基础

### 3.1 语义角色标注语料库

本文借鉴“融合三元空间的中文语言知识与世界知识获取和组织”973 项目提出的语义角色标注体系<sup>[17]</sup>，该体系共有 27 个语义角色标签（表 1）。其中，核心语义角色（主论元）标签共 10 个；外围语义角色（辅论元）标签共 17 个，包括工具、材料、方式、范围、事量、原因、目的、起点、终点、路径、方向、处所、起始、结束、时点、时段、比事。

基于上述语义角色标注体系，我们对 2000 年 1 月份《人民日报》语料进行了标注，总计 50000 个句子，形成了中文语义语料库（Chinese Semantic Corpus，以下简称为 CSC）。图 1 是对“这份伪声明把周末的气氛烘托得格外紧张。”这个句子进行语义角色标注的可视化结果。该句的谓语动词和得字短语各构成一个命题。谓语动词“烘托”的格框架为“当事、系事”，“紧张”为一价形容词，“这份伪声明”充当“烘托”的当事，“周末的气氛”则分别充当“烘托”的系事和“紧张”的当事。

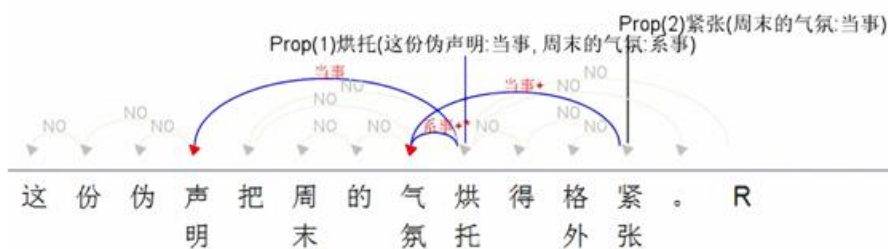


图 1 语义角色标注示例

### 3.2 格框架词典

本文语义角色标注依据 HowNet<sup>[18]</sup>中的格框架词典进行。该词典将动词分成了 785 种语义类,每个语义类均有相应的格框架,语料库中绝大多数动词按照不同义项可映射到 HowNet 中的动词语义类。HowNet 格框架词典中所包含的语义角色种类较多,达到 50 余个,而 973 标注体系中不到 30 个。因此,在使用格框架词典时,我们首先要建立从 HowNet 语义角色到 973 语义角色的映射,从而将 HowNet 的格框架词典改造为基于本文语义角色体系的格框架词典。

进一步,我们参照音系学中用于归纳音位的对立互补原则<sup>[19][20]</sup>,对核心语义角色进行归并,以形成粗粒度的语义角色体系(在第 5 节总结语义角色标注规则时必须使用粗粒度语义角色)。所谓对立互补原则指的是:(1)一个格框架中,每个粗粒度语义角色只能出现一次;(2)如果两个核心语义角色在同一个格框架中共现,则它们是对立的,不能归并;(3)如果两个核心语义角色在所有动词的格框架中均不共现,则它们是互补的,可归并为同一个粗粒度语义角色。基于上述原则,我们将 10 个核心语义角色归并为 4 个粗粒度语义角色,其对应关系如表 1 所示。

表 1 核心语义角色标签

核心角色 (主论元)	主事	施事 (agent)
		同事 (co-agent) <sup>①</sup>
		当事 (experiencer)
	客事	受事 (patient)
		系事 (relative)
		内容 (content)
		结果 (result)
	与事	与事 (dative)
		对象 (target)
	领事	接事 (indirect experiencer) <sup>②</sup>

### 3.3 例句库

本文的例句库的来源是 CSC 中的把字句,总计 2079 句。CSC 中的例句被分成两个部分,前 1000 句为开发集,用于总结句模的构成规则;后 1079 句为测试集,用于评价句模构成规则的有效性。

## 4 句法和语义信息标注

句模是对句子语义关系的描写,语义关系又受制于句法的制约。因此,研究把字句句模的生成规则,必须了解把字句谓语动词的配价、句法结构类型等句法信息。本文对 CSC 语料库中 1000 个例句(开发集)的谓词配价类型、谓语结构类型及句模构成类型进行人工标

<sup>①</sup> “共同实施”的缩略,谓词表达的事件有多人共同参与时,第一个参与者标为“实施”,其余为“同事”。

<sup>②</sup> “间接当事”的缩略,语义角色接近“领事”,与当事之间有“整体-部分”、“实体-属性”的关系。

注并对标注结果进行了归纳和总结。

#### 4.1 谓词配价类型

判断一个动词的价位，应以动词的必有论元为判定标准，一个动词有几个必有论元则该动词就为几价动词，其他非必有论元不参与动词价位的计算。

表 2 把字句谓词配价类型

谓词配价类型	例词
一价动词	一般动词：跑、玩、哭、荒、哑、站、醉、气
	形容词：苦、忙、美、高兴、累、急
	离合词：翻番、录音、过称
二价动词	一般动词：驱散、安排、搜集、还原、说、记、交、压、搞、找、做、摸、兜、背、围、炸、包、关、写、浇洒、糊、编、捏、种、上、插、扔、锯、放、停、丢、奉献、镌刻、贯穿、引、推、搡、控制、限定、洒、吃、下、抖、争取、送、打、冲、提高、嫁接、安排、叠、改造、理解、涂、商量、批、办、实现、维护、发展、结合、连接、分开、对立、组成、转向、体现、挂、聚集、改转、转、听、洗、恨、投、印、*化、减少、提高、等同、延伸、抄
	轻动词：作、加以、经过、给予、进行
	离合词：出手、过称
	代词：怎么样
三价动词	一般动词：注入、纳入、变成、作为、看成、列为、称为、叫做、比作、给予、告诉、归还、传教、归功

来自例句库中的例词显示，出现在把字句中的谓语中心语包括表示行为动作的动词、轻动词、离合词、形容词和代词。其中，二价动词居多，一价动词和三价动词较少。这些动词主要为他动词（吃、做等），但也存在少量自动词（哑、哭等）。

#### 4.2 谓语结构类型

句模是句子的语义结构表达。同一语义可用不同的句式表达，同一句式也可表达不同的语义。研究把字句句模类型及生成规则必须分析句模和句式之间的对应关系。把字句谓语结构较为复杂，根据谓语内部结构的不同，可将把字句分为不同的句法结构类型。基于前人的研究<sup>[16]</sup>，本文将把字句谓语结构类型归为 10 类，并按照句法结构由简到繁的顺序排列如下：

- (1) 光杆动词式把字句（谓语为光杆动词：V/r）晨风把江心的薄雾吹散。
- (2) 动体式把字句（谓语动词+“了/着/过”：V+了/着/过）他把书卖了。
- (3) 介动式把字句（介宾短语+谓语动词：p+n+V）别把纸往地上扔。
- (4) 动量式把字句（谓语动词+动量词：V+m）我把他批评了一顿。
- (5) 动宾式把字句（谓语动词+宾语：V+n；V+v/a+n）他把衣服交给妈妈。
- (6) 动介式把字句（谓语动词+介宾短语：V+p+n）他把车停在门前。
- (7) 动趋式把字句（谓语动词+趋向动词：V+v）他把盒子藏了起来。
- (8) 动程式把字句（谓语动词+程度补语：V+v/a）你要把课本读透。
- (9) 动得式把字句（谓语动词+“得”字组合短语：V+得）妈妈把碗洗得干干净净。
- (10) 动结式把字句（谓语动词+结果补语：V+a/v）他把小刘打死了。

需要说明的是，许多句法结构类型都带有时体助词，因此，在判断一个把字句的谓语结构类型时，可按照从第（10）类到第（1）类由后到前的顺序判断。比如，若一个把字句包含趋向补语，则判断为动趋式，即使带“了/着/过”也不会判断为动体式。

以上把字句为简单谓语结构，除此之外，还存在某些复合谓语结构的把字句，这些把字句的谓语结构由多种句法结构类型复合而成，比如：介动式+动趋式（“我们要把理论与实际结合起来”）、介动式+动程式（“我们要把理论学习同实际应用结合好”）等。

#### 4.3 句模构成类型

把字句句模的生成与谓语动词和句式有着密切的联系。谓词的配价决定了句模的主要语

义角色,句式则影响着句模是否添加边缘语义角色。动词进入把字句后,有的配价保持不变,但有些动词由于句法结构的变化或者语义表达的制约会添加边缘语义角色。

关于把字句的语法意义,邵敬敏、赵春利<sup>[21]</sup>将把字句分为“处置把字句”和“致使把字句”两类;范晓<sup>[22]</sup>认为把字句的语义有处置义和致使义两种,表致使义的把字句可以转换为“使”字句。语言事实表明,虽然大多数把字句表示处置义,但现实中仍存在很多把字句并非表示处置意义(“这件事把他急病了”),应当加以区别。下文将分别对“处置义”把字句和“致使义”把字句不同句法结构及句模的对应关系进行描写并对句模的形成进行详细的分析。

#### 4.3.1 处置义把字句

大多数把字句表示处置义。所谓“处置”是指谓语中的动词所表示的动作对“把”字引出的受事施加影响,使它产生某种结果,发生某种变化,或处于某种状态。<sup>[23]</sup>下文分别对10种把字句句式中动词的配价及句模进行描写和分析。

##### (1) 光杆动词式: 施事+把+受事/内容+动

把字句的谓语一般为复杂结构,但在某些具有一定语境的句子中,谓语可以为光杆动词,但这些光杆动词一般是二价的双音节复合动词<sup>⑤</sup>,比如“消除、记住”。疑问代词“怎么样”以及表示变性态的动词“\*化”也可作为光杆动词出现在把字句中。这些动词在进入把字句后配价保持不变。

动词重叠作把字句的谓语,也可以看作光杆动词式把字句的一类,比如“你把衣服洗洗”。离合词重叠前边的动语素后也可以进入把字句,如“把梨过过称”。动词重叠式把字句的谓语中心语的配价也不发生变化,句模与光杆动词式相同。

##### (2) 动体式: 施事+把+受事/系事/内容+动

谓语动词后带时体助词“了/着/过”,表示动作行为的进行状态。这种把字句中的谓语动词配价不发生变化。如“他们把小刘打了”,“打”始终是二价动词,语义角色为“施事”和“受事”。离合词必须以离析形式进入把字句,有些离合词中间插入了时体助词,其配价也未产生变化,如“他把酱油出了手。”

##### (3) 动量式: 施事/当事+把+受事/系事/内容/对象/终点+动

谓语动词后接表示动量的数量短语,数量短语不是谓语动词的论元,谓语动词的配价不变。比如“他把小李批评了一顿”,“批评”的主要语义角色有“施事、对象”,进入动量式把字句后配价不变。很多动词都可以进入动量式把字句,“把”的介引宾语所代表的语义角色较为复杂,但主语表示的语义角色不外乎“施事、当事”两种。

##### (4) 动趋式: 施事+把+受事/系事/内容+动

趋向动词作补语时,意义较为虚化,不构成一个命题,谓语中心语的意义就是整个谓语结构的意义。比如“你把钱交出来”,实际意思就是“你交钱”。

##### (5) 动程式: 当事+把+对象/受事/终点/内容+动

动程式把字句中的补语表示动作行为的程度,整个谓语结构构成一个命题。如“我把他恨死了”,“死”是“恨”的程度,并非“死亡”之意,“恨”进入动程式把字句后配价不变,仍是二价动词,配价为“当事、对象”。

##### (6) 动结式

A. 谓语核心为一价动词: 施事+动1; 把+施事+动2

B. 谓语核心为二价动词: 施事+把+受事/工具+动1; 把+施事/当事+动2

动结式把字句中的谓语核心动词和补语动词一般为单音节动词,都表示实在意义,可单独构成命题,两个命题之间有致使关系。该类把字句中的谓语动词一般为一价或二价动词,补语动词一般是一价的单音节动词,进入把字句后其配价不发生变化。比如“他把时间都玩

<sup>⑤</sup>单音节动词在戏曲中偶有出现,比如“夫妻双双把家还”,这里不作讨论。

没了”，“玩”的配价为“施事（他）”，进入动结式把字句后，“玩”和“没”各构成一个命题：“他玩”、“时间没了”，两个命题之间存在致使关系，由于“他玩”而导致“时间没了”。再如“他把牙齿吃坏了。”即“他用牙齿吃”；“牙齿坏了”，“他”和“牙齿”分别是“吃”的实施和工具，“牙齿”又是“坏”的当事。

### （7）动得式

A. 施事+把+受事/内容/对象+动 1；把+当事+动 2/形+（系事）

动得式把字句中，补语一般为动词性词或短语，表示事物由于动作行为而呈现出来的情态，谓语动词和补语分别构成一个命，两个命题之间是致使关系。如“他把部队搞得像一个大家庭”，“搞”和“像”分别构成两个命题：“他搞部队”、“部队像一个大家庭”，“搞”导致了“像”这一结果。

B. 施事+把+内容+动

有些动得式把字句的补语仅表示谓语动词的程度，不能单独构成一个命题，因此其句模是单动核结构，为简单句模，如“我们要把农业和农村工作做得更好”。

### （8）动宾式

动宾式把字句可分为两类：一类是“谓语动词+宾语”，另一类是“谓语动词+补语动词+宾语”。

#### A. 谓语动词+宾语

1) 施事+把+受事/系事/内容+动+结果/对象/终点/范围/工具

有些三价复合动词本身带有位移变化，如“注入、送到”，或本身表示事物“由……变化为……”，如“变成、提高”，或者是“分为、作为、叫做、告诉”等本身包含“结果、对象”这样的论元。这些动词在进入动宾式把字句后配价保持不变。比如“医生把药注入脑补”，“注入”的核心语义角色为“施事、受事、起点、终点”，宾语所代表的论元的语义角色为“终点”。

二价动词只有两个核心论元，为了引入“结果、工具、终点”等边缘论元而在谓语动词后添加宾语成分。如“敌人把墙炸了个洞”，中的“炸”的核心论元为“施事、受事”，但为了引入动作的结果而添加了宾语“洞”，表示“炸”的结果，即“敌人把墙炸出了个洞”。“他把窗户糊了纸”，即“他用纸糊了窗户”，宾语“纸”引入“工具”。

轻动词意义虚化，后面一般要有表示实在意义的动词作宾语。轻动词作谓语中心语的把字句中，轻动词不构成命题，作宾语的实义动词构成一个命题，这类把字句的句模只考虑宾语动词的配价即可。如“他把工作重心作了安排”即“他安排了工作重心”，“安排”的配价为“施事、受事”。

2) 施事+把+工具/材料/处所/终点+动+受事/系事

有些动宾式把字句的“受事”等核心语义角色由宾语充当，“工具、材料”等边缘语义角色由“把”的介引宾语引入。如“他把柳条编了框”，“编”为二价动词，核心语义角色为“施事、结果”，“把”的介引宾语引入了“材料（柳条）”这一边缘语义角色。

**B. 谓语动词+补语动词+宾语：施事+把+受事/系事/同事+动 1；把+当事/施事+动 2+系事/终点**

此类把字句的谓语由谓语中心语和补语构成，补语一般为“成、为、到、上”等单音节动词，宾语与补语有语义联系，作补语动词的“系事、终点”等。谓语核心动词和补语动词各构成一个命题，两个命题之间有致使关系。如“我们要把犯人改造成新人”，“改造”的核心语义角色为“施事（我们）”、“受事（罪犯）”，补语“成”的核心语义角色为“当事（罪犯）”、“系事（新人）”，“犯人”由于被“改造”而“成新人”。

### （9）动介式：施事+把+受事+动+处所/终点/范围/原因/结果

一些二价动词为了句子表意的完整性需要增加论元角色，因此句法上相应地要添加某些

成分，除了添加宾语或状语之外，谓语动词后加介宾短语引入动作行为的“处所、终点、范围”等也是一种重要的方式。如“他把孩子放在奶奶家”，“放”的配价为“施事、受事”，但由于“放”后面添加了介宾短语“在奶奶家”，“放”随之增加了边缘语义角色“处所”。

**(10) 介动式：**施事+把+受事/内容+终点/同事+动

谓语动词前添加介宾短语作状语，引入动作行为的“终点、方向”等，谓语动词的配价发生了变化，增加了一个语义角色。如“他把手表往桌子上一扔”，“扔”为二价动词，核心语义角色为“施事（他）、受事（手表）”，由于添加了状语成分（往桌子上），因此增加了表示“终点”的边缘语义角色。

一些“结合”类的二价动词，如“结合、分开、对立”等，前面带有状语成分，与“把”的介引宾语为并列关系，共同作谓语动词的受事。比如“我们不能把环境保护与经济发展相对立”，“环境保护”和“经济发展”共同作“对立”的受事。

#### 4.3.2 致使义把字句

使动把字句的特点是：“把”后名词表示使事，句中的“把”可以换成“使”。<sup>[4]</sup>致使义把字句变换为“使”字句之后，处置义隐退，致使义凸显，“把”后宾语兼作“使”的宾语和谓语的主语。

**(1) 动体式：**把+施事/当事+动

表示致使义的动体式把字句的谓语核心动词一般是一价动词，且大多是单音节动词，比如“跑、荒、哑”。有些离合词可以以离析形式进入把字句，比如“翻番”。“你们要小心，别把个犯人跑了”，即“你们要小心，别使犯人跑了”，“跑”仍为一价动词，配价为“施事（犯人）”。此类把字句谓语动词的配价不变。

**(2) 动量式：**把+当事+形

谓语动词后添加表示动量的数量短语构成动量式把字句，数量短语不是谓语动词的论元，不参与句模的构成。如“这件事把他忙一阵子”意思是“这件事使他忙一阵子”，“忙”的配价为“当事（他）”。

**(3) 动程式：**把+当事+动

动程式把字句的谓语动词后的补语表示动作行为的程度，不能构成一个命题。谓语动词一般是一价动词，进入把字句后配价不变。比如“这可把我美死了”可变换为“这可使我美死了”，“美”的语义角色为“当事（我）”，仍是一价形容词。

**(4) 动结式**

A. 把+当事+动 1；把+当事+动 2

动结式致使义把字句中的谓语动词和补语动词共享同一个论元，谓语动词构成的命题对补语动作行为产生致使作用。比如“这点酒就把他醉倒了”意思是“这点酒就使他醉倒了”，“醉”和“倒”共享论元“他（当事）”，“他醉”导致了“他倒”的产生。

B. 受事+把+施事+动 1；把+当事+动 2

上述动结式把字句的主语为名词性短语，谓语动词为一价动词，但有的动结式把字句的主语为动词性短语，谓语动词为二价动词。如“吃螃蟹把孩子吃吐了”，谓语动词“吃”有两个论元“孩子（施事）”和“螃蟹（受事）”，补语动词只有一个论元“孩子（当事）”。

**(5) 动得式：**把+当事+形；把+当事+动

动得式把字句谓语动词和补语各构成一个命题，两个命题之间存在致使关系。如“中药汤把他苦得只想呕吐”，“苦”的配价为“当事（他）”，补语动词“想”的配价为“当事（他）”，因为“他感到苦”，所以“他想呕吐”。

**(6) 动宾式：**把+施事+动+受事/内容/结果

与处置义把字句不同，致使义动宾式把字句中的谓语动词为二价动词，进入把字句后配价不变。比如“他把我党在太岳区的机关都挂上了八路军的牌子”，即“他使我党在太岳区

的机关挂上了八路军的牌子”，“挂”为二价动词，语义角色为“施事（机关）”、“受事（受事）”，进入把字句后配价未发生变化。

此外，“谓语动词+补语+宾语”也是动宾式把字句的一种，谓语动词和补语动词的论元有“整体-部分”的关系，谓语动词和补语动词各构成一个命题，两个命题之间有致使关系。比如“繁重的劳作把他累弯了腰”，“腰”是“他”的领属物，“他累”导致“腰弯”。

#### (7) 动介式：把+施事+动+内容/处所

致使义动介式把字句中的谓语动词一般为一价动词，添加了介宾短语后，谓语动词增加了边缘语义角色。比如“他把双脚站在现实的大地上”，“站”为一价动词，核心语义角色为“施事”，添加介宾短语“在现实的大地上”后增加了“处所”这一边缘语义角色。

#### 4.3.4 句模类型分布

在对 CSC 中的 1000 个把字句（开发集）进行语义角色人工标注和统计之后，所得把字句句模的分布情况如下：

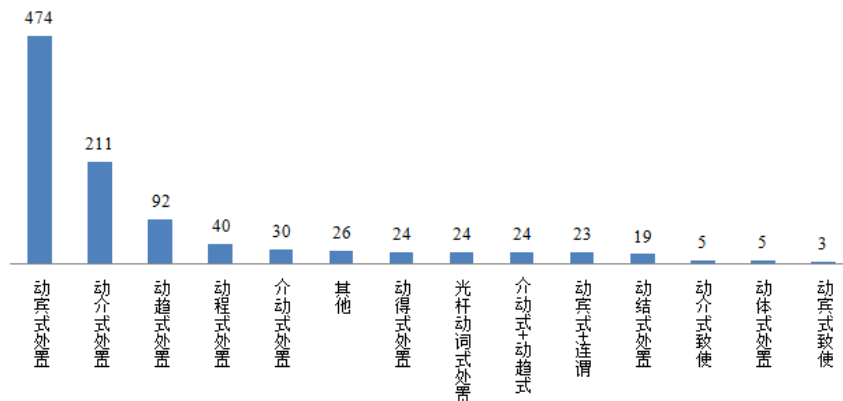


图 2 把字句句模数量分布图（开发集）

由图 2 可知，把字句句模类型复杂多样，其分布呈现出不平衡的状态：（1）处置义把字句句模在 9 种句式种都有出现，数量为 919 句，占总句模数量的 91.9%。致使义句模数量很少，只出现在动介式、动宾式之中，共 8 句；（2）处置义句模内部数量分布差别也较大。处置义句模主要以动宾式、动介式和动趋式居多，其数量分别为 474 句、211 句和 92 句，占总句模数量的 77.7%。其次是动程式和介动式句模较多，分别为 40 句、30 句，但比重仅为总数的 7%。动体式数量最少，只出现了 5 句；（3）除简单句式之外，还存在两种句式合用的把字句，如：介动式+动趋式、动宾式+连谓式等。其中，介动式+动趋式和动宾式+连谓式的数量较多，分别为 24 句、23 句。其他类型数量则很少，共 26 句。

## 5 语义角色标注规则

根据前文粗粒度语义角色的划分，本文对把字句谓语动词与句式、句模的对应规则进行了总结（表 3）。本文语义角色标注是在句法标注基础之上进行的，把字句句法结构可归纳为：

[主语]+<把后宾语>+[介词宾语]+<谓语动词>+[补语（谓词性）]+[谓语宾语]+[介词宾语]

实际句子中，这些句法成分不必全部共现，其中，把后宾语和谓语动词是必有的，其他均为可选成分（中括号内的成分）。

表 3 把字句句法成分、语义角色对应表

动词配价	句式	动核结构	主语	把宾（谓语动词）	动宾	介宾	把宾（补语动词）	例句数量
一价动词	动体式	单	主事	客事				1
	动结式	双	主事				主事	1



二价动词	光杆、动体式、动量式、动趋式、动程式、动得式、动宾（轻动词）式、介动式（组合类）	单	主事	客事				199
	动介式		主事	客事		辅论元		211
	介动式		主事	客事		辅论元		11
	动结式、动得式	双	主事	客事	客事		主事	30
	动宾（V+v/a+n）式		主事	客事	客事		主事	184
三价动词	动宾式（V+n）	单	主事	客事	辅论元/与事/结果			279

表 3 中，动结式、动宾式（V+v/a+n）和部分动得式为双动核结构，其他为单动核结构。综上，我们将把字句句法成分与语义角色之间的对应规则归纳如下：

规则 1：主语对应主事；

规则 2：若把字句谓语结构为单动核结构，把后宾语作谓语动词的客事；若为双动核结构，把后宾语兼作谓语动词的客事和补语动词的主事；

规则 3：若谓语动词为一价或二价动词，则谓语动词后的宾语对应客事；若谓语动词为三价动词，则谓语动词后的宾语对应辅论元/与事/结果；

规则 4：介词后的宾语对应辅论元；

之后，将表 1 粗粒度语义角色映射到 Hownet 中给出的谓语动词的格框架中，得出把字句的句模。

这里我们以“他把大树砍倒了”作为样例，说明上述规则的使用流程：（1）该句句法结构为：主语+把宾+P+CMP，谓语结构类型为动结式；根据规则 1，主语-主事；（2）该句谓语结构为双动核结构，根据规则 2，把宾-客事（谓语动词）；把宾-主事（补语动词）；（3）根据规则 4，Hownet 给出的谓语动词“砍”的格框架（施事、受事、工具）、“倒”的格框架（当事、处所）以及粗细粒度语义角色的映射关系，可判定该句句模为：施事+把+受事+动 1；把+当事+动 2。

## 6 实验

**实验数据。**本文所使用的方法是基于规则的方法，没有训练集，仅使用一个开发集和一个测试集。通过人工分析，从开发集中总结出若干条规则。测试集则用于评价规则的正确率。本文使用的开发集共包括 1000 条例句，测试集则包括 1079 条例句。

在训练和测试时，仅考察与把字句相关的语义角色的自动标注情况。如果一个句子中存在多个小句，我们只考察含有介词“把”的小句，对其它小句不予考虑。在评价时，正确率和召回率的值是一样的，因此下面仅报告正确率。

**实验结果及分析。**在 1079 个把字句中，除去 15 个句法结构较为复杂的句子（介动介、动量+介动介+连谓等）和 1 个动宾式致使义把字句之外，其他 1064 个把字句均能通过上述规则的测试，正确率约为 98.61%，可见上述规则在把字句语义角色标注方面是可靠的。

在对 1079 个把字句语义角色进行人工标注之后，得出把字句句模数量分布情况（图 3）。

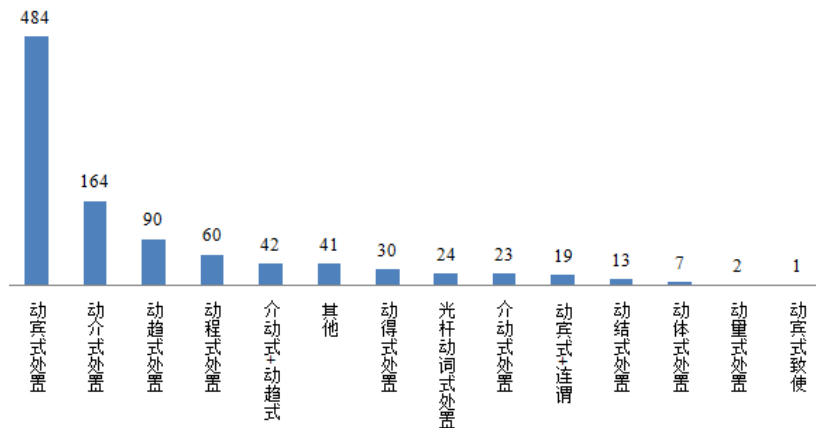


图3 把字句句模数量分布图（测试集）

对比图2可知，测试集中的把字句句模分布情况与开发集大致保持一致。1064个把字句共覆盖了10种句法结构类型，比开发集多出了动量式，但数量很少，仅有4句。动宾式处置、动介式处置和动趋式处置把字句句模数量仍占前三位，数量分别使522句、173句和93句，共占总数的74.06%。其次是动程式处置、动得式处置、介动式处置、光杆动词式处置、动结式处置和动体式处置把字句，数量分布从93到4之间，差别较大，共占总数的16.73%。复合句法结构的把字句数量最多的仍是“介动式+动趋式”和“动宾式+连谓”，占总数的6.3%，其他复合式句法结构把字句种类较多，但数量较少，总共有30句，仅占总数的2.82%。

## 7 结语

本文以基于知识库的规则方法初步解决了把字句语义角色自动标注问题。描写和分析了把字句的配价结构类型、谓语结构类型、句模构成类型及句模的分布情况，总结了把字句句模的标注规则，这些规则在把字句语义角色标注上的正确率为98.61%，这一结果表明本文所提出的方法在把字句语义角色标注上是有效的。本文所总结的标注规则可以在以下几个方面发挥作用：一、在人工标注语义角色语料库时，使用这些规则进行自动标注，规则覆盖不了的少数现象交由人工判断，从而发挥规则方法正确率高的优点，以降低人工标注的工作量；二、利用这些规则自动检查出纯人工标注过程中的错误，从而提高语义角色标注的准确率；三、为语言学对把字句的研究以及自动语义角色标注研究提供借鉴。

本文工作是一个大型语义语料库标注工程的一个组成部分，由于时间和篇幅的限制，本文仅涉及把字句的语义角色标注。在后续工作中，我们将进一步讨论现代汉语其它句式的语义构成规律，总结相应的标注规则，以提高语义角色标注工程的效率和质量，更好地为中文语义理解服务。

## 参考文献：

- [1] Gildea D, Jurafsky D. Automatic labeling of semantic roles [J]. Computational linguistics, 2002. 28(3): 245-288.
- [2] Xue N, Palmer M. Calibrating features for semantic role labeling [C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004: 88-94.
- [3] Cohn T, Blunsom P. Semantic role labelling with tree conditional random fields [C]//Proceedings of the Ninth Conference on Computational Natural Language Learning, 2005: 169-172.
- [4] 范晓. 三个平面的语法观[M]. 北京: 北京语言学院出版社, 1996: 201-209.
- [5] 鲁川, 缙瑞隆, 董丽萍. 现代汉语基本句模[J]. 世界汉语教学, 2000 (4): 11-24.
- [6] Baker C F, Fillmore C J, Lowe J B. The Berkeley Framenet project [C]//Proceedings of the 17th international conference on Computational linguistics, 1998: 86-90.

- [7] Kingsbury P, Palmer M. From treebank to propbank [C]//Proceedings of the International Conference on Language Resources & Evaluation, 2002: 1989-1993.
- [8] Xue N. Labeling Chinese predicates with semantic roles [J]. Computational Linguistics, 2008, 34 (2) : 225-256.
- [9] 刘开瑛. 汉语框架语义网构建及其应用技术研究[J]. 中文信息学报, 2011, 25 (6): 46-52.
- [10] 范晓, 胡裕树. 试论语法研究的三个层面[J]. 新疆师范大学学报, 1985 (2): 7-15.
- [11] 范晓, 朱晓亚. 论句模研究的方法[J]. 徐州师范大学学报, 1999, 25 (4): 18-23.
- [12] 徐昌火. 试论句模研究的对象、起点和基本原则[J]. 南京师范大学学报, 1999 (4): 101-108.
- [13] 孙道功, 亢世勇, 孙茂松. 面向语言处理的单句句型句模对应关系研究——基于标注语料库的定量考察[J]. 计算机工程与应用, 2006, 42 (33): 170-173.
- [14] 亢世勇, 许小星. 现代汉语句系系统的构建和研究[J]. 中文信息学报, 2010, 24 (1): 103-109.
- [15] 郑丽娟, 邵艳秋. 基于语义依存图库的兼语句句模研究[J]. 中文信息学报, 2015, 29 (6): 30-37.
- [16] 范晓. 动词的定价与汉语的把字句[J]. 中国语文, 2001 (4): 309-319.
- [17] 詹卫东, 惠志方, 常宝宝, 等. 现代汉语谓词语义角色标注语料库规范 (讨论稿) (内部资料).
- [18] Dong Z, Dong Q. Hownet and the computation of meaning [M]. World Scientific Publishing Company. 2006: 763-769.
- [19] Wells J C. Accents of English [M]. Cambridge, 1982: 86-95.
- [20] Likun Q, Yue Z, Meishan Z. Dependency Tree Representations of Predicate-Argument Structures[C]//In Proceedings of AAAI-16, 2016: 2645-2651.
- [21] 邵敬敏, 赵春利. “致使把字句”和“省隐被字句”及其语用解释[J]. 汉语学习, 2005 (4): 11-18.
- [22] 范晓. 论“致使结构”语法研究和探索 (十) [M]. 北京: 商务印书馆, 2000: 431-442.
- [23] 黄伯荣, 廖旭东. 现代汉语 (增订四版) [M]. 北京: 高等教育出版社, 2007: 87-88.

### 作者联系方式:



何保荣 (1990—) 女, 硕士研究生, 主要研究方向为中文信息处理, 地址: 山东省烟台市芝罘区红旗中路 186 号鲁东大学文学院, 264025, 17853519358, hebaorong123@sina.com



邱立坤 (1979—) 通讯作者, 男, 博士, 副教授, 主要研究方向为计算语言学, 地址: 山东省烟台市芝罘区红旗中路 186 号鲁东大学文学院, 264025, 13583566312, qiulikun@pku.edu.cn



徐德宽（1968——），男，博士，教授，主要研究方向为计算语言学，地址：山东省烟台市芝罘区红旗中路 186 号鲁东大学文学院，264025，15066387236，xudekuan@tom.com