

基于结构树库的补语位形容词语义分析及搭配库构建*

田思雨 邵田 荀恩东[†] 饶高琦[‡]
北京语言大学 北京语言大学 北京语言大学 北京语言大学
1416092064@qq.com shaotian2017@163.com edxun@126.com raogaoqi@blcu.edu.cn

摘要

在形容词充当补语的粘合式述补结构¹中，通常以两个谓词性成分连用（“形容词+形容词”、“动词+形容词”）的形式出现，由于这一结构没有形式标记，为计算机自动识别该结构带来了较大的难度，同时，形容词充当补语并不是其最基本、典型（作定语、谓语）的用法，在语言学界与计算语言学界也没有受到足够的关注。因此，该文以补语位的形容词为研究对象，从大规模的句法结构树库中抽取形容词直接作补语的述补结构，并通过编程和人工校验的方式对语料进行降噪，对补语位形容词进行穷尽式检索，得到补语位形容词词表，进一步对补语位形容词的语义进行细分类，构建相应的语义搭配库。不仅可以提升句法切分的正确率，为深层句法语义分析提供语义信息，也可以为语言学本体的相关研究提供参考。

关键词： 补语位形容词；语义分类；语义搭配库

Semantic analysis of complementary adjectives and construction of collocation database based on structural tree library

Siyu Tian Tian Shao Endong Xun Gaoqi Rao
Beijing Language Beijing Language Beijing Language Beijing Language
Culture University Culture University Culture University Culture University
1416092064@qq.com shaotian2017@163.com edxun@126.com raogaoqi@blcu.edu.cn

Abstract

The adjective-complement bonded declarative-complement structure, which usually appears in the form of two predicative components in succession (“adjective + adjective”, “verb + adjective”), poses a greater difficulty for automatic computer recognition of this structure because of its lack of formal markers, and because the adjective-complement is not its most basic and typical (as a definite article and predicate) usage, and has not received sufficient attention in linguistic and computational It has not received enough attention in the linguistic and computational communities. Therefore, this paper takes adjectives in the complement position as the object of study, extracts the descriptive-complement structure of adjectives as complements directly from a large-scale syntactic structure treebank, programmatically and manually verifies the corpus for noise reduction, performs exhaustive retrieval of adjectives in the complement position, obtains a

* 本课题为北京语言大学研究生创新基金资助项目（中央高校基本科研业务费专项资金）（23YCX135）与国家自然科学基金“中文意合图的表征与生成方法研究（62076038）”

[†] 通讯作者:荀恩东(edxun@126.com)

[‡] 通讯作者:饶高琦(raogaoqi@blcu.edu.cn)

¹指不带形式标记“得、个、不”的述补结构

lexical list of adjectives in the complement position, further subcategorizes the semantics of adjectives in the complement position, and constructs a corresponding semantic collocation database. It can not only improve the correct rate of syntactic cut and provide semantic information for deep syntactic semantic analysis, but also provide reference for the related research on linguistic ontology.

Keywords: Complement adjective, Semantic classification, Semantic collocation library

1 引言

补语位形容词是指在述补结构中充当补语的形容词, 包括粘合式述补结构中“动词/形容词(谓语)+形容词(补语)”和组合式述补结构中“动词/形容词(谓语)+得+形容词(补语)”两类。前者在语法上可能构成主谓、动宾、述补等结构, 语义上可能表示支配、修饰等关系, 句法语义关系较为复杂, 在句法语义分析时较难准确识别。由于后者具有明显的形式标记, 计算机在识别时没有较大的难度, 因此不在本文的研究范围之内。例如:

- (1) 这台机器**操作简单**。
- (2) 祖国江山**显得壮丽**。
- (3) 有些问题一定要**考虑清楚**。
- (4) 再多钱也会被**挥霍干净**。

在例(1)-例(4)中, 加粗词语的词性序列均为“动词+形容词”, 但是其句法与语义关系却不相同。例(1)中的“这台机器操作简单”是主谓谓语句, 其中“操作简单”为形容词作谓语的主谓结构是句子的谓语结构, “操作”为动词, “简单”为形容词, 词性序列与述补结构相似, 但该搭配为形容词作谓语的主谓结构。在例(2)中, “显得壮丽”与述补结构的词性序列也相同, 但该搭配为形容词作宾语的动宾结构。在例(3)中, “考虑清楚”为述补结构, “清楚”表预期的结果。而例(4)中的述补结构“挥霍干净”, 表示一种非理想结果。

因此, 本文将研究对象定位于形容词作补语的粘合式述补结构, 从句法结构树库中抽取“中心语(形容词/动词)+形容词”结构, 并通过编程、人工等方式对语料进行消歧, 获取补语位形容词词表及其搭配语料。进而以消歧后的语料为基础, 对补语位的形容词进行两种语义划分: 其一为结合形容词和情感词的分类体系, 对补语位形容词进行细致的语义分类; 其二为根据形容词补语在述补结构中的语义, 对形容词补语进行语义分类, 并进一步构建补语位形容词语义搭配库。形容词本身的语义分类能够为相关研究提供细致的近义词类聚, 比如可以为深度学习等提供语义支持, 也可以为信息检索、情感分析、关系抽取、机器翻译等提供语义聚类。形容词补语的语义分类能够更好地观察形容词在述补结构中的作用及分布状况, 从而更好地识别整个述补结构组块的语义及感情色彩。此外, 语义搭配库的构建也可以提高句法语义分析中在识别述补结构时的准确率, 并提供更丰富的语义信息。

2 研究现状

关于补语位形容词, 语言学界暂无专门研究, 但零散出现于对述补结构、形容词的众多研究中。早期语言学界并不认为形容词能够充当“补语”, 如黎锦熙(1924)等。二十世纪五十年代开始, 逐步明确了对核心谓词后的形容词成分的界定, 确认了形容词可以直接跟在动词或形容词后作补语, 如王力(1943)、吕叔湘、朱德熙(1952)、丁声树等(1961: 11)等, 朱德熙(1982)进一步根据有无“得”字, 把前一类述补结构叫作“组合式述补结构”, 后一类叫作“粘合式述补结构”。八十年代以来, 对于述补结构的专门研究较多, 且偏向于语义平面的研究, 王还(1984)认为形容词作补语多描述既成事实; 马庆株(1986)对程度补语进行三级分类; 陆俭明(1990、2001)详细探讨了“VA了”的语法意义, 并指出其表达语法意义的三种情况; 马真、陆俭明(1997)详尽地列举了作结果补语的形容词, 并说明了其语义的分布情况。进入二十一世纪以后, 对于补语位形容词的研究更加细致、全面, 比如李锦姬(2003)详细说明了结果补语、程度补语的语法意义及用法; 刘振平(2007)对单音节形容词构成动结式述补结构

的情况进行了穷尽式考察；朱文文（2008）根据形容词的语义特征对形容词进行二级分类，进而总结了形容词作补语的句位意义；蔡丽（2011）总结出程度补语五个特点，并从六种角度对程度补语进行划分；刘从文（2012）根据语义将形容词分为三类，考察了其作补语的入位情况，并对其句法、语义、语用及认知进行分析。但以上学者均针对某一类补语位的形容词或状补位对比进行深入研究，即无大规模语料库支撑的实证研究，也未对补语位形容词的全貌进行研究。

搭配这一概念自提出以来，受到了国内外学者的广泛关注。Firth（1957）首次提出搭配的概念，将其引入语言学研究的领域，认为搭配整体是表达一种意义的方式。Sinclair（1966）对搭配进行了更深入地研究，指出搭配是两个及以上的词在句子中共现。国内学者早期对搭配现象的探索主要集中于对词语搭配性质的界定，如刑公碗（1978）、林杏光（1990、1991）等。随着语料库语言学的兴起，语料库搭配研究受到了语言习得、语言分析等领域学者的极大关注。在搭配知识抽取方面，提出了基于词共现和分布值的统计方法和基于框架和规则的方法，比如，孙茂松等（1997）提出了以搭配强度、离散度、尖峰为指标的搭配判断算法；卫乃兴（2002）提出了基于数据的方法和数据驱动的方法；Kilgarriff等（2004）通过词性标签构建搭配规则建立了Word Sketch Engine系统。在搭配库建设方面，学界从语法结构、二语习得等角度构建了不同的搭配库，蔡丽（2017）基于留学生的使用情况和偏误分析，构建了程度补语与述语常用搭配库；邢丹等（2020）基于BCC语料库构建了介词结构搭配库；王雨（2022）基于《国际中文教育中文水平等级标准》构建了国际中文教育词语搭配知识库。以上搭配知识抽取方法以及搭配库构建成果，对本文的搭配知识抽取以及搭配库应用具有较大的借鉴意义。

综上，理论语言学对于形容词作补语的语言现象进行了句法、语义的深度探讨，探究了形容词作补语的句位意义，考察了形容词在述补结构中的语义表达，对本文的搭配知识构建、语义分类标准具有重要的参考价值。但同时补语位形容词，尤其是形容词作补语的粘合式述补结构也存在缺乏专门研究、缺乏数据库支持的不足。此外，计算语言学领域暂无针对无形式标记的形容词作补语的知识建设工作，但对于搭配知识抽取的方法以及搭配库建设的研究成果众多，对本文的构建工作具有重要的借鉴意义。

3 数据基础

本文的语料来源是北京语言大学句法结构树库（卢露等，2020），针对形容词充当补语的粘合式述补结构进行语料的抽取，并对语料进行消歧。

3.1 树库资源

北京语言大学句法结构树库规模大、领域多、质量高。目前已有浅层结构分析树库总字数约1300万字，小句70余万句，语料涉及报刊、文学、百科、新闻、综合等领域文本1万余。该句法树库将句子分析为块状组合序列，由构成句子基本结构的句法成分、起衔接作用的衔接成分以及表附加性语义的辅助成分构成。其特点为以组块状短语结构树为句法表示，直接根据各组块的性质及功能，标注句子骨架，突出中心词信息。

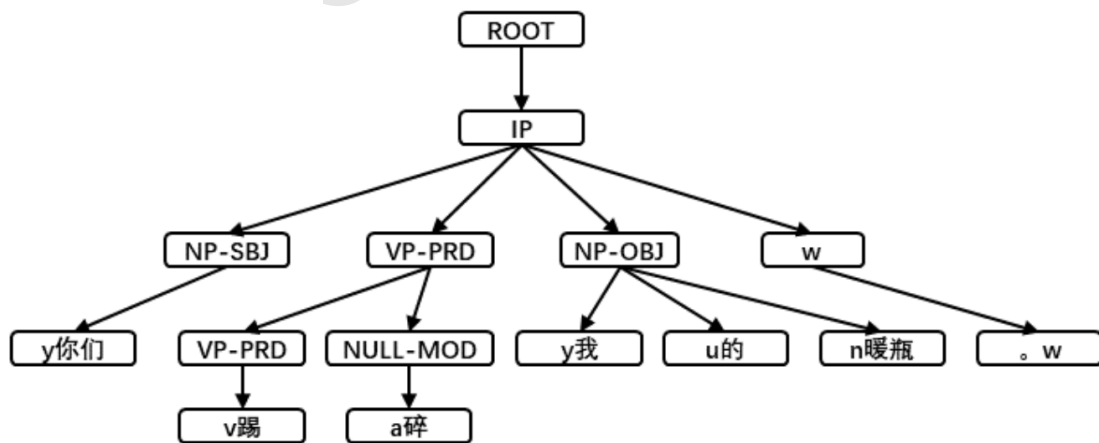


Figure 1: “你们踢碎了我的暖瓶”结构图

该树库组块句法属性标记包含组块的词性信息、功能信息以及边界信息等（见图1）。树库在对句子进行浅层结构标注时，以“谓词-论元”结构为中心的句子骨架，标注最大长度的主语块、宾语块、状语块、补语块以及主谓宾结构和状中补结构，能够准确地识别组块功能、边界及关系。

3.2 补语位形容词的抽取与消歧

本小节主要介绍从北京语言大学句法结构树库中抽取补语位形容词语料并对语料进行消歧的过程。

3.2.1 抽取过程

本文所使用的北京语言大学句法结构树库是以组块状短语结构树为句法表示，经过分词词性标注的高质量树库。因此，可以利用树库的结构信息和词性信息，结合补语位形容词构成的述补结构的句法形式构建检索式。一方面，由于直接检索“中心语（动词、形容词）+形容词”存在较多述宾结构的歧义；另一方面，在修饰性组块内部进一步限制词性信息对检索精度影响较大。因此，在进行语料的抽取时，首先将述语组块的中心语（动词、形容词）与[NULL-MOD](修饰性组块)紧连于述语块内；其次，为了考察补语位形容词的音节形式特征，再按照音节数目构建检索式：单音节动词+修饰性组块、双音节动词+修饰性组块、单音节形容词+修饰性组块、单音节形容词+修饰性组块。

同时在检索层面进行消歧。由于检索式所抽取的述补结构，未限制后部修饰性成分的词性信息，因此在检索层面可以利用词性等信息排除其他类型的补语。不是形容词充当的补语类型有趋向补语、数量补语、时处补语。参考黄伯荣、廖序东（1991）的《现代汉语》以及实际语料，整理出需要批量排除的形式特征，如表1所示：

修饰性组块	形式特征
介宾结构	于、中、入、里、内、在、至、到、在、为、给
趋向补语	来、去、上、下、进、出、回、起、起来
数量补语	0、1、2、3、4、5、6、7、8、9、一、二、三、四、五、六、七、八、九、十
时处补语	0、1、2、3、4、5、6、7、8、9、一、二、三、四、五、六、七、八、九、十、两、半
其他	过、着、有、了、开、住、成、是、而

Table 1: 后部修饰组块降噪

3.2.2 数据处理

经过检索层面降噪后的语料基本为形容词充当补语的述补结构。由于述补结构的形式特征有带形式标记（组合式述补结构）和不带形式标记（粘合式述补结构）两种。本文通过编程的方法，去除带有形式标记的述补结构之后，对抽取到的形容词充当补语的述补结构进行形式分类。将补语位形容词按照音节划分，与前部中心语相互组合。最后得到的分类结构一共有8种： $v1+a1;v1+a2;v2+a1;v2+a2;a1+a1;a1+a2;a2+a1;a2+a2$ ¹

3.2.3 语料消歧

由于分词词性标注与组块标注错误的影响，在检索层面进行消歧之后，所抽取的结果中仍包含一些不属于形容词作补语的述补结构。因此，本文利用词典与人工两种方式对检索结果进行进一步消歧。

首先，校对词性信息，进行词典消歧。在本文抽取的形容词作补语的述补结构中，中心语包含动词、形容词，补语为形容词。分别将这些词与《现代汉语词典》（第七版）中的形容词、动词进行对照，核对语料中的词性信息。具体分为两步：（1）校对中心语（动词、形容词）和补语的词性是否正确，如出现不为词或词性错误的情况，则删除该条语料。（2）校对述补结构是否正确，如所抽取粘合式补语在词典中为一个词，则删除该检索结果。

其次，核对述补结构，进行人工排查。由于字典中词语的义项较多，通过词典消歧后仍可能出现歧义的结构。因此，需要进一步通过人工判断述补结构的准确性。分为四种情况：（1）

¹注：v1为单音节动词、v2为双音节动词、a1为单音节形容词、a2为双音节形容词

述补结构中，形式标记前后的词性符合要求，但实际上并不是述补结构，比如“显得富丽”，其中“显得”为动词，该结构为动宾结构。但由于“显”和“富丽”符合动词、形容词词性，仍保留在检索结果中。(2) 所抽取的结果中频次较少的述补结构，其词性信息符合条件，但由于树库所涉及的语料涉及微博、对话等生活文本，语料中可能出现错误用法，比如“行走得性急”。(3) 一些词出现了新用法，但词典暂未收录的该义项。(4) 动词、形容词中存在许多兼类词，词典消歧不能识别出检索结果中的具体用法，造成错误检索与错误词性的漏排。以上四种情况均需要人工进行排查。本文以《现代汉语词典》(第七版)中的形容词、动词义项为标准，参考《汉语形容词用法词典》，依据上下文信息，对抽取语料进行逐个对照排查，删去歧义语料。

经过以上消歧步骤后，一共得到述补搭配14436个，补语位形容词词表包含657个形容词，其中单音节218个，双音节439个，大大超过了前人以往研究成果中的数量。

4 补语位形容词的语义分类

词在未进入语法结构之前就存在静态的、共性的意义，词汇意义是句法表现的基础，决定着句法表现的意义。本节参考现有研究成果，结合补语位形容词在述补结构中的词汇意义，对前文得到的补语位形容词词表进行语义细分，并对分类结果进行分析。

语义一级类 (数量)	语义二级类 (数量)	例句	例句
空间(112)	尺寸(10)	小	帽子给你做小了
	外观(43)	碎	你们踢碎了我的暖瓶
	光彩(30)	绿	他的脸都吓绿了
	方位(6)	高	最后一瞬间她把枪口抬高了
	距离(23)	远	两人牵着手跑远了
时间(13)	相对(4)	早	哥哥今天起床起早了
	历时(2)	久	年糕被我煮太久了
	速度(5)	慢	小徐不是因为报表做慢了被炒的
	年龄(2)	大	父母这么辛苦把我拉扯大
形态(74)	样态(67)	软	我腿都吓软了
	声音(7)	高	我抓起遥控器，把电视的音量调高
状态(146)	正面(60)	均匀	你要保证孔内涂均匀
	负面(64)	累	今天跑累了
	中性(22)	紧	用软木塞把它盖紧
情绪(20)	正面(9)	爽	今天把我吃爽了
	负面(11)	烦	家里人都听烦了
评价(210)	正面(151)	清楚	有些问题一定要考虑清楚
	负面(56)	错	我晕到连家门都走错
	中性(3)	杂	书看太杂了不好
性格(44)	正面(31)	仔细	她对事物观察仔细
	负面(13)	圆滑	要把事情做圆滑
频数(13)	频次(2)	频繁	你平时要练勤一些
	数量(11)	齐全	考试的东西备齐全
程度(32)	轻(4)	轻	学校对这事处理轻了
	重(28)	死	我真是委屈死了

Table 2: 一级类、二级类分类结果

4.1 分类标准与体系

分类是语言研究的基本方法，也是大多语义研究的基础性工作。对于形容词的语义分类，各学者的观点不尽相同。黎锦熙(1924)将形容词分为性质、形体、状态和程度形容词。刘月华等(1982)根据描述对象、感情色彩，将形容词分为描写动作者、描写动作、正向形容词、负向形容词。王惠等(2006)将形容词划分为事性值、物性值(量化属性值、模糊属性值、颜色)、人性值(年龄、品格、关系、境况)、空间值(一维值、二维值、三维值)、时间值。

5 形容词补语的语义分类

考察述补结构中形容词补语的语义，不仅可以更好地识别形容词所在述补结构的意义，也可以更好地反映出不同意义形容词在述补结构中所起的作用，为机器理解补语位形容词语义提供参考。本节参考现有研究成果，对前文得到的形容词作补语的述补结构搭配进行考察，细分补语位形容词在搭配中的语义，并对其结果进行分析。

5.1 分类标准及准则

在整个述补结构中形容词所表达的语义要从述补结构的语义分类中获得。从一般的补语分类中，粘合式补语分布在趋向补语、程度补语、结果补语、数量补语、时处补语中。由于趋向补语、数量补语、时处补语中的补语指向较为明确，且在对话料考察中皆存在形容词作程度补语和结果补语的情况。因此，本文将研究对象锁定为程度补语和结果补语中形容词的语法意义分类。马真、陆俭明（1997）将形容词作结果补语的语义概括为四种，预期结果的实现、非理想结果的出现、自然结果的出现、预期结果的偏离，并提出表不同语义的决定因素是形容词的性质、动词的性质、述语动词所表示的行为动作对作补语的形容词所表示的性质的制约作用、语境。其所提出的分类结果得到众多学者认可，本文也将参考其分类。

李锦姬（2003）研究了现代汉语全部补语，在对整体性述补结构考察中将补语分为有量补语和无量补语，并进一步在无量补语之下细分语义为自然结果、预期结果、非理想结果；并阐述了有量补语主要表达心理预期结果的偏离量。该研究和马真、陆俭明（1997）存在很多相合之处，且有更详细地分类及说明。此外，该文认为程度补语为非典型性述补结构中表程度的量，并对其进行详尽阐述。因此，本文着重参考其分类成果。

以上文得到的14436个搭配为基础对其进行人工的语义分类。参考前人对形容词作补语述补结构的语法意义分类，以真实搭配中体现的语义为基础，进行人工标注。

5.2 分类结果

经过多轮迭代后，本文将程度补语分为2类，结果补语先分有量补语和无量补语2类，再分5类。分类结果如表3所示。此外，本文还对语义三级类中形容词本义的分布情况进行了考察，由于数量较多，放入附录中展示。

语义一级类 (数量)	语义二级类 (数量)	语义三级类 (数量)	例句
结果补语(11922)	有量补语(1563)	心理预期结果的偏差-指向名词(521)	气球吹大了不好卖
		心理预期结果的偏差-指向动词(1042)	演唱会结束迟了
	无量补语(10359)	自然结果	照片存放久了不免泛黄
		预期结果(7095)	有些问题一定要考虑清楚
		非理想结果(1864)	他把这件事解释偏了
程度补语(32)	轻(4)	程度轻(4)	这个故事还是写轻了
	重(28)	程度重(28)	他真是傻透了

Table 3: 补语位形容词在述补结构中的语义分类结果

5.3 结果分析

根据以上对补语位形容词语义的三级划分，可以对其数据进行进一步统计分析。首先，通过对一类数据进行统计，可以看出形容词作结果补语的搭配远多于程度补语，形容词作补语时多为结果补语。其次，对二级类进行统计，结果补语中有量补语为1563，无量补语为10359。可以得出形容词作结果补语时多分布在无量补语，表达动作或变化引起的结果，较少表达心理预期结果的偏离量。

我们对形容词本义中能够表达五种语义的情况进行统计，但由于部分形容词本义数量较少，只选取较有代表性的种类（如图3²）。从图中我们可以看出具有褒贬意义的形容词（即正面评价、负面评价、正面状态、负面状态等）数量较多。即具有褒贬色彩的形容词更容易进入述补结构。此外，从分布上看，具有褒贬意义的形容词在述补结构中所表达的意义较为聚集，多集中分布在某一种语义上，而其他具有客观意义的形容词的分布则较为分散。因此可以推出

²图3中为了更好地观察数据的对比结果，将预期结果中正面评价类数据3338排除在外

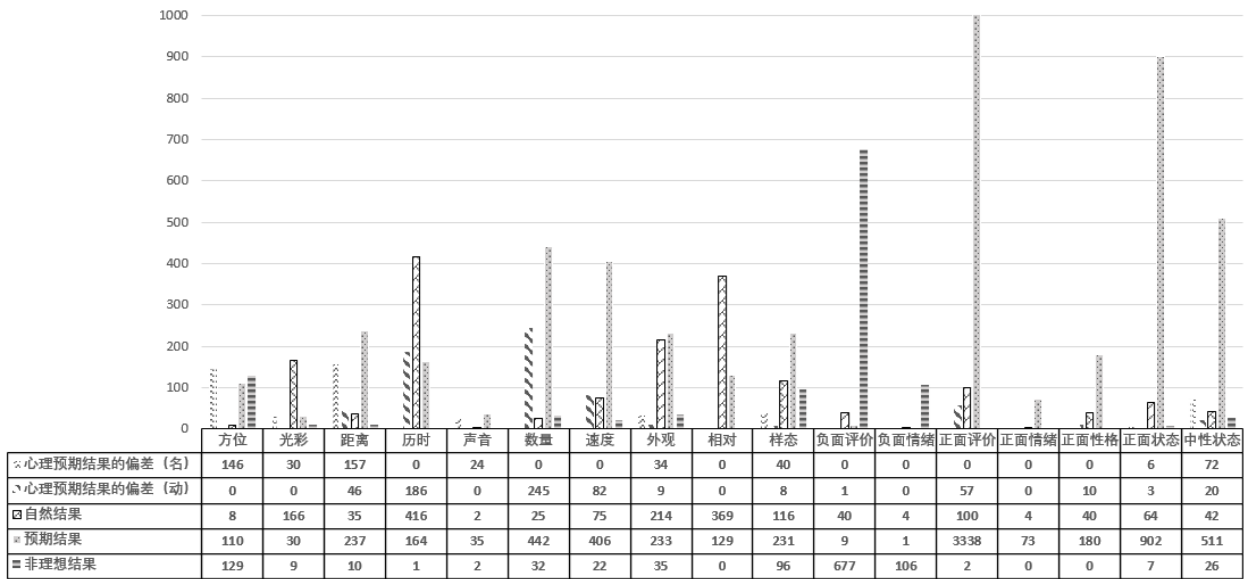


Figure 3: 形容词所在述补结构的语义的分布情况

具有褒贬义的形容词在述补结构中有其固定的表达倾向。述补结构的整体意义受褒贬义形容词的影响较大。而表客观意义的形容词在述补结构中对整体的影响较小。

综上，我们可以得出补语位形容词主要分布在结果补语中，且无量补语能够容纳更多种类、数量的形容词，而有量补语的选择则较为受限。其次，在形容词作补语的述补结构中，人们更倾向于表达预期结果，且正面评价类形容词出现率极高。此外，具有褒贬意义的形容词对述补结构的整体意义影响较大。

6 补语位形容词搭配库构建及分析

本节以补语位形容词的三级语义分类体系为基础，根据消歧后的真实语料搭配情况，构建补语位形容词的语义搭配库，并对其数据进行进一步统计分析。

6.1 构建搭配库

由于词的意义较为灵活，而搭配的意义较为固定。补语位形容词在不同的搭配情境中，其表达的语义可能不同，比如：“理解透了”和“失望透了”，“透”在前一搭配中表示“深入、透彻”，而在后一搭配中则表程度深。因此，我们希望构建语义搭配库来将语义分类更好地对接句法语义分析。本小节通过消歧后的搭配语料构建，包含述补搭配的语法信息，同时按照语义分类体系，为其添加语义信息，具体为：以中心语（动词、形容词）为中心，根据中心语和补语位形容词搭配的频次为顺序，并在补语位形容词后添加其语义类型信息。

6.2 搭配库数据分析

根据搭配库中中心语和补语位形容词的音节搭配信息，我们对其音节搭配情况进行统计。本文对粘合式述补结构进行抽取时，划分了8种音节组合情况，经过消歧后，形容词作中心语带形容词补语的语料较少，“A1+A1”、“A2+A2”的音节组合无合格语料，共得到14436个搭配。因此本文只讨论6种音节组合情况，如下表6所示。此外，我们进一步对单音节形容词和双音节形容词的组合情况进行统计，如表7所示。

从两表中可以得到，“V1+A1”格式搭配数量最多，占比35.54%。且“V1/ V2/ A2+A1”格式，占比60.03%。可以看出单音节形容词作补语的能力更强，其次“V2+A2”格式占比较高，为25.87%，体现了汉语韵律语法的音节搭配规律，双音节与双音节搭配具有普遍优势。此外，“A1+A2”、“A2+A1”格式的占比极小，说明形容词作中心语后带补语位动词的情况较少。

6.3 搭配库数据应用

本文所构建的搭配知识库对形容词作补语的粘合式述补结构进行了穷尽式检索，包含了完

音节搭配	数量	占比
V1+A1	5130	35.54%
V1+A2	2012	13.94%
V2+A1	3276	22.69%
V2+A2	3734	25.87%
A1+A2	24	0.17%
A2+A1	260	1.80%
总计	14436	100%

Table 4: 音节搭配数量及占比

音节搭配	数量	占比
V1/ V2/ A2+A1	8666	60.03%
V1/ V2/ A1+A2	5770	39.97%
总计	14436	100%

Table 5: 单双音节补语位形容词搭配对比

整的搭配数据和形容词以及述补结构的语义信息。首先，穷尽的搭配知识能够为机器自动识别无形式标记的形容词补语述补结构提供可靠的数据支撑，提升句法切分准确度。丰富的语义信息能够帮助计算机进行感情色彩的判定，进而更准确地理解自然语言。其次，对于语言学本体研究而言，详细的补语位形容词词表及搭配信息能够帮助我们验证或发现语言规律。形容词语义分类和在述补结构中的语义分类的对比，为进一步考察补语位形容词用法及规律提供了查考数据。最后，该搭配库还能够为对外汉语教学提供搭配知识支持。详细的搭配及语义知识不仅能够为智能教学设计、自动出题提供搭配知识库，也能够为对外汉语教师的教学活动提供知识支持。

7 展望

本文基于大规模语料库构建了补语位形容词词表以及补语位形容词粘合式述补结构搭配库，形容词词表包含659个补语位形容词，搭配库包含14436个搭配。通过对形容词语义以及形容词补语的语义进行细致分类，为搭配库添加了丰富的语义信息。其中形容词语义包括9个一级类，26个二级类，91个三级类，形容词补语的语义类别包括2个一级类、4个二级类、7个三级类。该搭配库数据详尽、语义信息丰富，可用于机器识别、智慧教育、自动出题以及语言研究等方面。但本文研究仍存在可进一步研究之处。首先，本文搭配资源的语义仅考察了中心语与形容词补语，未考虑副词对述补结构语义的影响。其次，由于时间原因，本文构建的搭配库缺乏对频次的考察与统计，而搭配频次也是语言研究和智慧教育的重要参考。此外，本文对于搭配库的实际应用效果没有进行详实地说明。

因此在未来的工作中，我们将从本文所抽取的述补结构中提取存在副词的搭配，进一步考察副词对该搭配语义的影响。其次，针对缺少频次信息的问题，我们将通过大规模语料库，收集各个搭配的频次信息。最后，我们将把信息丰富的搭配知识，放入智慧教学平台中，考察其在应用中的实际效果，为对外汉语教师的教学设计以及智慧教学自动出题提供搭配知识库。

参考文献

- 黎锦熙. 1924. 新著国语文法. 上海:商务印书馆, .
- 王力. 1943. 中国现代语法. 北京:中华书局
- 吕叔湘.朱德熙1952. 语法修辞讲话. 北京:中国青年出版社.
- 丁声树. 1961. 现代汉语语法讲话. 北京:商务印书馆.
- 朱德熙. 2008. 语法讲义. 北京:商务印书馆.

- 王还. 1984. 汉语的状语与“得”后的补语和英语的状语. 语言教学与研究,(04)61-66+29
- 陆俭明. 1990. 述补结构的复杂性——《现代汉语补语研究资料》序. 语言教学与研究,(01)13-20
- 陆俭明. 2001. “VA了”述补结构语义分析补议——对读者意见的回复. 汉语学习,(06)78-80.
- 马庆株. 1986. 含程度补语的述补结构. 北京: 北京大学出版社.
- 马真,陆俭明. 1997. 形容词作结果补语情况考察(一). 汉语学习, (1):3-7.
- 马真,陆俭明. 1997. 形容词作结果补语情况考察(二). 汉语学习, (04):14-18.
- 马真,陆俭明. 1997. 形容词作结果补语情况考察(三). 汉语学习, (06):7-9.
- 李锦姬. 2003. 现代汉语补语研究. 上海:复旦大学学位论文.
- 刘振平. 2007. 单音形容词作状语和补语的对比研究. 北京:北京语言大学学位论文.
- 朱文文. 2008. 现代汉语形容词状补语序选择机制研究. 北京:北京语言大学学位论文.
- 蔡丽. 2011. 现代汉语中程度补语的范围及类别. 宁夏大学学报(人文社会科学版),(04)9-14+32.
- 刘从文. 2013. 现代汉语形容词状补位对比研究. 北京:上海师范大学学位论文.
- Firth, J.R. 1957. *Papers in Linguistics 1934-1951*. London:Oxford University Press.
- Sinclair, J. 1966. *Beginning the study of lexis*. In Memory of J.R.Firth, 410-430.
- 邢公畹. 1978. 语词搭配问题是不是语法问题?. 安徽师大学报(哲学社会科学版),(04)77-84+65.
- 林杏光. 1990. 词义分类、词语搭配、语言教学. 北京语言学院出版社.
- 林杏光. 1991. 论词义分类和词语搭配. 中国人民大学学报,(05)77-82.
- 孙茂松,黄昌宁,方捷. 1997. 汉语搭配定量分析初探. 中国语文,(01)29-38.
- 卫乃兴. 2002. 现代汉语补语研究. 复旦大学博士论文.
- 李敏. 2002. 基于语料库和语料库驱动的词语搭配研究. 当代语言学,(02)101-114+157.
- Kilgarriff A, Rychlý, Pavel, Smrž, Pavel, et al. 2004. *The Sketch Engine*. London:Oxford University Press.
- 蔡丽. 2017. *AEIC Academic Exchange Information Centre(China)*. Proceedings of 2017 2nd International Conference on Humanities Science and Society Development (ICHSSD 2017) (Advances in Social Science, Education and Humanities Research VOL.155) Atlantis Press.
- 邢丹. 2020. 基于大规模语料库状中搭配库构建与应用. 北京语言大学学位论文.
- 王雨,肖叶,荀恩东等. 2022. 服务国际中文教育的词语搭配知识库建设. 语言文字应用,(02)26-37.
- 卢露,矫红岩,李梦,荀恩东. 2020. 基于篇章的汉语句法结构树库. 自动化学报,(01)2911-2921.
- 中国社会科学院语言研究所词典编辑室. 2016. 现代汉语词典(第七版). 北京:商务印书馆.
- 郑怀德,孟庆海. 2016. 汉语形容词用法词典. 北京:商务印书馆.
- 黄伯荣,廖旭东. 1991. 现代汉语. 北京:商务印书馆.
- 刘月华. 1982. 状语与补语的比较. 语言教学与研究,(01)22-37.
- 王惠,詹卫东,俞士汶. 2006. “现代汉语语义词典”的结构及应用. 语言文字应用,(1)134-141.
- 鲁川. 2010. 知识工程语言学. 北京:清华大学出版社.
- 李军. 2008. 中文评论的褒贬义分类实验研究. 北京:清华大学硕士学位论文.
- 徐琳宏,林鸿飞,赵晶. 2008. 情感语料库的构建和分析. 中文信息学报,(01)116-122.

附录:

A.1 检索式实例

```

AddTag("begs","于;中;入;来;去;上;下;进;出;回;开;起;过;着;有;了;过;至;到;成;为;给;是;住;在;里;而;0;1;2;3;4;
AddTag("ends","于;中;入;在;里;下;上;内")
Condition("len($1)=1;mid!=[modalverbs];end($1)!=[ends];beg($2)!=[begs];end($2)!=[ends]")
Handle0=GetAS("—v","","","","","","","0,1","","","")
Handle1=GetAS("$NULL-MOD","","","","","","","0,1","","","")
Handle2=JoinAS(Handle0,Handle1,"Link")
Handle3=GetAS("$VP-PRD","","","","","","","","","")
Handle4=JoinAS(Handle3,Handle2,"SameRight")
Handle=Freq(Handle4,"$Q")
Save(Handle,"v1.txt")

```

其中, AddTag 的功能为构建一个词表。Condition指的是检索式的一些限制条件。!=表示不等于。\$1表示对中心语进行限制, len表示长度,\$2表示对修饰性组块进行限制, beg表示组块开头, end表示组块结尾。因此第一行的意思就是限制作述补结构中心语的动词为单音节动词, 且不以词表中的词为开头或结尾。

Handle表示一个句柄, 以序号不同标明不同的原子查询项。GetAS表示对一个原子查询项进行限制, v表示限制该原子查询项的词性为动词, 后面的数字表示的是对该原子项的位置的描述, \$NULL-MOD限制该语块为修饰组块。Handle0与Handle1表示得到了词性为动词的原子查询项和标签为\$NULL-MOD的原子查询项。

JoinAS表示两个原子查询项在组装时的顺序和关系类型。Link表示两个原子查询项在组装时的关系是接连出现的, 即中间没有其他成分。SameRight表示两个原子查询项之间的关系为右对齐。因此Handle2表示的是动词和修饰性组块接连出现的情况; , Handle4表示的是动词和修饰性组块在一个述语块内出现, 且二者之间是右对齐的情况。

Freq表示输出检索式的频次信息。

A.2 补语位形容词分类结果

1.空间

(1)尺寸

大、小、长、短、薄、厚、矮、高大、庞大、厚实

(2)外观

烂、碎、细、粗、瘦、丰满、直、弯、圆、扁、平、斜、陡、方、立体、稀烂、烂糊、粉碎、细碎、微细、纤细、粗大、粗壮、窈窕、瘦小、丰腴、胖、臃肿、肥、平直、僵直、弯曲、鼓、尖、凹、滚圆、瘪、干瘪、平坦、平实、平整、歪斜、陡峭

(3)光彩

红、白、绿、黑、黄、青、紫、蓝、乌、灰、亮、暗、通红、白皙、苍白、惨白、黑暗、黝黑、金黄、焦黄、枯黄、铁青、明亮、明朗、亮堂、透亮、暗淡、灰暗、昏暗、浅

(4)方位

高、低、正、偏、歪、反

(5)距离

远、近、深、浅、紧、宽、松、窄、稀、稀疏、稀少、稀松、稀薄、深邃、紧凑、宽敞、宽松、宽广、宽大、宽阔、广阔、松弛、蓬松、狭窄

2.时间

(1)相对

晚、迟、早、紧

(2)历时

久、长

(3)速度

慢、缓慢、迟缓、迅速、飞快

(3)年龄

老、大

3.形态

(1) 样态

生、熟、软、硬、轻、沉、老、臭、净、湿、滑、脏、嫩、热、凉、暖、酥、咸、苦、甜、辣、涩、油、黏、鲜、胀、粗糙、干、浑、透明、浓、淡、烂熟、松软、软和、绵软、坚硬、僵硬、硬实、轻盈、轻薄、轻快、清澈、潮湿、平滑、光滑、圆滑、细嫩、白嫩、柔嫩、热乎、烫、滚热、冷、凉快、清凉、凉爽、温暖、暖和、脆、酥脆、苦涩、黏糊、新鲜、毛糙、干爽、干燥、干涩、浑浊、稠、清淡

(2) 声音

高、低、响、响亮、洪亮、大、小

4. 状态

(1) 正面

[1]牢固：牢、结实、坚实、坚固[2]均匀：均、匀、匀净[3]舒适：舒坦、舒畅、舒服、清闲、安心、安稳、幸福、香甜[4]平和：平静、从容、温和、温顺、安静[5]柔和：柔顺、柔软[6]流利：顺、顺滑、顺口、顺嘴、顺当、溜、通[7]真实：真切、确实、现实、具体[8]滋润：红润、圆润、健康[9]扎实：实、饱、满、饱满、瓷实[10]清醒：精神、清白、清新[11]齐：整齐

(2) 负面

[1]累：乏、疲劳、疲倦、疲惫、困、笨重[2]身体状态：疼、痛、晕、饿、秃、哑、沙哑、聋、僵[3]冷淡：冷漠、麻、木、麻木、沉默、消极、低迷、冷清[4]呆：楞、迷糊[5]凶猛：凶、凶狠[6]乱：混乱、花[7]破：蔫、干枯、枯萎、馊、黄、酸、枯、旧、焦、废、钝[8]疯：野、痴[9]繁忙：仓促、匆忙、忙碌[10]惨重：惨烈[11]红眼：眼红、眼馋、馋

(3) 中性

[1]密集：密、紧、密实、浓密、缜密[2]严：严肃、严格、严密、严厉[3]分散：零散、零碎[4]完：好、干净

5. 情绪

(1) 正面

[1]爽：爽快、痛快[2]乐：兴奋、愉快、高兴、快乐、快活

(2) 负面

[1]心切：急、急切[2]紧张：慌[3]暴躁：烦、倦[4]郁闷：沉闷、闷

6. 评价

(1) 正面

[1]优秀：好、突出、棒、牛、佳、绝[2]周全：周密、周到、齐备、圆[3]完满：圆满、完美、完善、成熟[4]妥当：妥、妥善、稳妥、妥帖[5]对：正确、合理、合法、标准、规整、规范、值[6]精细：细、精、细腻、详尽、详细、细致、精致[7]强：壮、强劲、强大、强壮、强健、壮大、壮实[8]稳：稳固、稳当、稳定[9]美：漂亮、好看、美丽、美艳、靓丽、俊秀、靓、酷、甜、乖[10]清楚：明显、明白、清晰、明确、明晰、明朗、白、直、亮[11]准：准确、精确[12]充足：充分、足[13]火：红、兴旺、红火、旺盛[14]生动：活、鲜活、精彩、逼真[15]顺利：通畅、通顺、顺手、顺畅、畅通[16]容易：简单、轻松、方便、通俗[17]积极：活跃[18]亲密：紧密、密切、亲密、亲切、熟[19]灵活：灵敏、敏锐、利索、巧妙[20]深刻：透彻、通透[21]平衡：平等、匀称[22]整洁：洁净、清洁、白净、干净、工整、端正[23]舒心：顺眼、顺耳、对路[24]协调：和谐、融洽、和睦[25]华丽：高级、高贵、亮丽、隆重、优雅[26]富裕：富有、富[27]熟练：娴熟、专业、高效、熟[28]犀利：尖锐、锋利

(2) 负面

[1]坏：烂、孬、糟、糟糕、贱、难听、难看、讨厌、黑[2]刁：拧、贼、抠、骄傲、可怕、残忍[3]虚：空、模糊、渺茫、玄[4]腻：腻歪、腻味、烦、滥[5]弱：脆弱、衰弱、软弱[6]错：异常、过火[7]枯燥：无聊[8]庸俗：糙、俗气、俗、平庸、平淡、丑[9]呆傻：糊涂、愚蠢、愚钝、笨、痴呆、呆板、傻、迟钝[10]艰难：复杂、麻烦、难、困难、穷、苦、危险

(3) 中性

[1]贵[2]杂[3]便宜

7. 性格

(1) 正面

[1]勤快[2]活泼：幽默[3]温柔：随和[4]勇敢：坚强[5]认真：仔细、谨慎、严谨[6]大方：坦白、慷慨[7]聪明：精明、机灵、机敏[8]爽快：干脆、果断[9]谦虚：虚心、低调[10]沉稳：稳

重、可靠[11]正经：客气、正直

(2)负面

[1]懦弱[2]猥琐[3]懒：懒惰[4]狡猾：圆滑[5]粗心：马虎[6]调皮：淘气、幼稚[7]刻薄：偏激、敏感

8.频数

(1)频次

勤、频繁

(2)数量

多、烂、广、少、全、全乎、齐全、完全、完整、丰富、广泛

9.程度

(1)轻

轻、浅、少许、轻微

(2)重

透、死、惨、深、疯、满、重、狠、甚、光、猛、绝、暴、坏、傻、严重、厉害、彻底、激烈、透顶、深厚、沉重、大发、大、远

A.3 形容词补语语义三级类中形容词本义的分布情况

