

基于结构树库的状位动词语义分类及搭配库构建*

邵田 翟世权 饶高琦 荀恩东[†]

北京语言大学 语言资源高精尖创新中心
北京市海淀区学院路15号, 100083

shaotian2017@163.com terryzhaiblcu@gmail.com raogaoqi@blcu.edu.cn edxun@126.com

摘要

一般情况下, 一个小句中只有一个动词, 但是也有两个动词同时在一个小句中出现的状况, 比如两个动词接连出现在同一小句中, 在句法上有可能构成状中、述补、动宾、连谓及并列等结构, 语义上可能表示修饰、支配、并列等关系。连续使用的两个动词构成了相对复杂的结构与语义关系, 尤其是在没有形式标记的情况下, 如何自动识别连用动词的结构及其所表达的语义关系是句法语义分析在落地过程中面对的较为困难的问题。对此, 本文将研究对象定位于直接作状语的动词, 从大规模结构树库中抽取两个动词连用的情况, 并对语料进行消歧, 提取出作状语的动词后, 进一步对其进行语义的细分类, 最后构建相应的语义搭配库。不仅为语言学本体提供了分类参考, 同时也为深层次的汉语句法语义分析提供了更多的知识。

关键词: 状位动词; 语义分类; 语义搭配库

Semantic Classification of Adverbial Verbs Based on Structure Tree Database and Construction of Collocation Database

Tian Shao Shiquan Zhai Gaoqi Rao Endong Xun[†]

Beijing Advanced Innovation Center for Language Resources, Beijing Language and Culture University
15 Xueyuan Road, Haidian District, Beijing, 100083, China

shaotian2017@163.com terryzhaiblcu@gmail.com raogaoqi@blcu.edu.cn edxun@126.com

Abstract

Under normal circumstances, there is only one verb in a clause. However, there are also situations where two verbs appear in the same clause at the same time. For example, if two verbs appear in the same clause one after another. Syntactically, it is possible to form adverbial structure, predicate-complement structure, verb-object structure, conjunction-predicate structure and coordinate structure, etc., and semantically it may express the relationship of modification, domination, and parallelism. Two consecutive verbs form a relatively complex semantic relationship, especially when there is no formal mark, how to automatically recognize the structure of the consecutive verb and the semantic relationship it expresses is a more difficult problem, especially for Chinese syntactic and semantic analysis applications. In this regard, this article locates the research object on the verbs that are directly used as adverbials, extracts the two verbs used in conjunction from a large-scale structure tree library, disambiguates the corpus, and extracts the verbs used as adverbials, and then classify it semantically, and finally

* 基金项目: 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 Corresponding Author

build the corresponding semantic collocation knowledge base. It not only provides a classification reference for the linguistic ontology, but also provides more knowledge for the deeply syntactic and semantic analysis of Chinese.

Keywords: adverbial verbs , semantic classification , collocation library

1 前言

状位动词指作状语的动词，包括动词带“着、地”等助词作状语和动词直接作状语两类。前者具有明显的标记，可以从形式上较好地加以匹配和识别，而动词直接作状语修饰中心语，由于缺乏形式标记，句法上可能构成动宾、连谓、并列、述补以及状中等结构，语义上可能表示支配、并列、修饰等关系，具有多种可能性，比如：

- eg1: 亚历山德罗维奇和利季娅·伊万诺夫娜意味深长地相视一望，于是**阅读**开始了。
- eg2: 世贸组织每一个成员都应该全面**遵守承诺**。
- eg3: 若是一条狼走近来，可以立即将它**打杀**。
- eg4: 我**回头看**了他一眼。
- eg5: 马切德让手下将他俩**带离**了战壕。
- eg6: 如果需要报销车票，还是在网络购票后**提前换取**纸质车票。
- eg7: 它将在第一时间**滚动报道**电视节和电影节的所有最重要的新闻和通告。
- eg8: 近年来先后有28名干警**在职攻读**本科。

以上例子中加粗的词语在《现代汉语词典》(第七版)⁰中都有动词词性，并且在例句中体现的也是动词词性，但是，每两个连用的动词的句法结构及其所表达的语义关系都不同。“阅读开始”句法上是主谓结构，语义上表示陈述说明的关系；“遵守承诺”是动宾结构，表示支配关系；“打杀”是并列结构，“回头看”是连谓结构，都表示并列关系；“带离”是述补结构，表示修饰关系；“提前换取”“滚动报道”与“在职攻读”是状中结构，都表示修饰关系，但是三者所表示的深层次语义不同，“提前换取”中的“提前”表示动作发生的时间，“滚动报道”中的“滚动”表示动作的方式，而“在职攻读”中的“在职”表示动作发生时的状态。对这一类无标记的动词连用结构进行结构与深层次语义的定性，在深度的句法语义分析阶段有很大的难度，同时给机器翻译等具体应用也带来了一定的困难。

©2021 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

⁰ 【阅读】{动}看(书报等)并领会其内容。

【开始】①{动}从头起；从某一点起。②{动}动手做；着手进行。③{名}开始的阶段。

【遵守】{动}依照规定行动；不违背。

【承诺】{动}对某项事务答应照办。

【打】1{动}①用手或器具撞击物体。②器皿、蛋类等因撞击而破碎。③殴打；攻打。④发生与人交涉的行为。⑤建造；修筑。⑥制造(器物、食品)。⑦搅拌。⑧捆。⑨编织。⑩涂抹；画；印。(11)揭；凿开。(12)举；提。(13)放射；发出。(14)〈方〉付给或领取(证件)。(15)除去。(16)舀取。(17)买。(18)捉(禽兽等)。(19)用割、砍等动作来收集。(20)定出；计算。(21)做；从事。(22)做某种游戏。(23)表示身体上的某些动作。(24)采取某种方式。

【杀】①{动}使人或动物失去生命；弄死。②{动}战斗。③{动}削弱；减少；消除。④同“煞”①。⑤{动}用在动词或形容词后，表示程度深。⑥〈方〉{动}药物等刺激皮肤或黏膜使感觉疼痛。

【回头】①(-//-) {动}把头转向后方。②{动}回来；返回。③{动}悔悟；改邪归正。④{副}少等一会儿；过一段时间以后。⑤{连}不然；否则(用在祈使句后的句子开头申述理由)。

【看】①{动}使视线接触人或物。②{动}观察并加以判断。③{动}访问。④{动}对待。⑤{动}诊治。⑥照料。⑦{动}用在表示动作或变化的词或词组前面，表示预见到某种变化趋势，或者提醒对方注意可能发生或将要发生的某种不好的事情或情况。⑧{助}用在动词或动词结构后面，表示试一试(前面的动词常用重叠式)。

【带】①(～儿){名}带子或像带子的长条物。②{名}轮胎。③地带；区域。④白带。⑤{动}随身拿着；携带。⑥{动}捎带着做某事。⑦{动}呈现；显出。⑧{动}含有。⑨{动}连着；附带。⑩{动}引导；领。(11){动}带动。(12){动}照看(孩子)。(13){名}姓。

【离】1①{动}分离；离开。②{动}距离①。③{动}缺少。④(Li) {名}姓。

【提前】{动}(把预定的时间)往前移。

【换取】{动}用交换的方法取得。

【滚动】{动}①一个物体(多为球形或圆柱形)在另一物体上接触面不断改变地移动，如车轮在地上的运动。②逐步积累扩展；不断地周转。③一轮接一轮地连续不断进行。

【报道】①{动}通过报纸、杂志、广播、电视或其他形式把新闻告诉群众。②{名}用书面或广播、电视形式发表的新闻稿。

【在职】{动}担任着职务。

【攻读】{动}努力读书或钻研某一门学问。

此外，对于助动词是作状语修饰中心词还是作谓宾动词学界有不同看法，但助动词数量有限，可穷尽列举，因此不论助动词是作状语还是带谓词性宾语在判断上都基本没有问题，不在本文考察范围之内。

因此，本文将研究对象定位于直接出现在动词前作状语的一般动词，以期在汉语句法语义分析时准确识别出动词作状语的状中结构并且给出该结构所表达的深层次的语义关系。

本文结合理论语言学和计算语言学，从结构树库中抽取两个动词在一个小句中连续出现的情况并进行消歧，得到作状语的动词之后，结合状位动词本身的语义及其在状中结构中所表达的语义进一步对状位动词进行语义上的细分类并构建相应的动词作状语的语义搭配库。基于语义搭配库不仅可以识别出连用的两个动词是否是状中结构，还可以进一步提供该状位动词在状中结构中的语义类型，为深层次的汉语句法语义分析提供更多的句法和语义知识，同时也为机器翻译等具体应用提供更为准确的结构知识。

2 研究现状

从语言学的角度来看，对动词作状语这一现象的认识是不断变化的。在早期的语法研究中，大多数学者认为除助动词外，其他动词不能作状语，朱德熙在《语法讲义》一书中将其视作为动词转化为副词，如“他同情地笑了笑”。80年代中后期，随着动词作状语的使用日益增多以及学者们认识的深化，对这一现象的研究也越来越多，如肖伟良（1983）、王政红（1989）、朴正实（2003）、侯雪钰（2012）、张军（2014）等，集中从形式和语义两个层面加以探讨，形式方面比如动词加“地、着”等助词作状语或者动词直接作状语，状语和中心语之间能否插入其他成分，能够进入何种句法格式，以及单双音节搭配等方面的特征。语义方面则主要考察前动或后动的语义类别和特征、动词和动词组合的语义框架以及作修饰语的动词的语义指向等等。还有的文章加入句法功能、内部结构以及语用因素等考察角度，如耿雪（2008）等。

动词直接作状语和动宾、动补、并列等结构在形式上相同，很多文章着重探讨连动式和状中式区别，如马鸣春（1984）一文中，作者列举了连动式和状中式各自的特点，然后从结构形式上加以对比，最后给出七条判断标准。还有的文章探讨某一著作或某一历史时期的动词作状语情况，如章新传（2004）与杨佳玉（2015）等。

也有的学者从不同的理论角度来看待动词作状语这一问题，从范畴化角度出发的有《非范畴化视角下的现代汉语动词作状语研究》（郭奇军 龙启艳, 2010）与《基于原型范畴理论的现代汉语连动式和状中式对比》（吴思霖, 2019）。从语法化角度出发的有《从语法化角度看动词直接作状语》（高增霞, 2004）等。

此外，还有的学者在定性分析的基础上，从相关语料库搜集语料进行定量分析，如考察能作状语的动词的数量等。比如孙德金（1997）以动词义项为单位对一份包含770个单音节动词及2266个双音节动词的词表进行封闭式考察，统计出能直接做状语的单音节动词33个，双音节动词78个，共计111个动词，并得出结论：一般修饰语是单音节的，则中心语也是单音节；修饰语是双音节的，则中心语也是双音节。但其所考察的动词未能涵盖《现代汉语词典》中的全部动词，且并未对状位动词所表达的语义进行分析，本文则从大规模的结构树库中抽取动词作状语的语料，所考察的动词范围大大超过该文，并进一步对状位动词所表达的语义进行分类。

总体来看，对于动词作状语的研究主要集中在理论语言学领域，而计算语言学领域相关的研究相对来说比较少，比如邢丹（2020）利用词性信息构建了动词作状语的搭配库，但是该文并未明确指出都有哪些动词可以作状语，且未对其所构建的动词作状语的搭配库作穷尽性的正确度考察，更未对动词作状语时所表达的语义关系进行分析，并且仅仅利用词性信息难免会受到词性标注结果的影响。本文以句法结构树库为语料来源，借助结构信息与词性信息抽取语料，在结构树的基础上，不仅可以准确地定位谓词性结构，还可以进一步减少词性标注对标注结果的影响。

3 数据基础

从本文的目标出发，需使用具有词性标注信息与结构标注信息的语料库，即北京语言大学句法结构树库（卢露等，2020），并从中进行状位动词语料的抽取和消歧。

3.1 结构树库资源

本文的语料来源是北京语言大学句法结构树库，该句法结构树库是一个大规模、多领域的

高质量树库，将句子分析为由句法成分、衔接成分、辅助成分构成的块状组合序列，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。目前已构建有质量保证的浅层结构分析树库，包含60余万小句，共近1300万个字，语料涉及百科、新闻、专利等应用领域文本1万余篇。

由于该规范包含的术语及标注符号比较多，仅介绍本研究涉及的述语及述语符号。述语是在句内相对传统主语和宾语而言可以分别陈述主语、支配宾语的句子核心部分，在该句法结构树库中，述语是包括了状语、补语的最大长度的连续谓词性短语结构，由动词（或中心语为动词的短语、动词性熟语）、形容词（或中心语为形容词的短语、形容词性熟语）、主谓短语充任。该句法结构树库认为一个述语块内部只有一个核心谓词，一个单句内可以有多个述语（复谓句），述语块用“()”整体标注，再用“()”将核心谓词标注出来。以下面的例子来说明：

eg9: 旅客(可以提前(查询))车次。

eg10: 我们(电话(联系))。

eg11: 你(怎么(说)个没完没了)。

eg12: 我((听)不懂)他们的话。

其中，“可以提前查询”“电话联系”“怎么说个没完没了”与“听不懂”都是述语块，用“()”整体标注，并且在底层用“VP-PRD”标签来标明块的性质。该述语块内部又包含修饰语和核心谓词两部分，修饰语又根据其位置的不同分为状语和补语两部分，核心谓词依旧使用“()”标注，一个述语块内只有一个核心词，即只有两层“()”，可以达到区分核心谓词与修饰语的作用。上述例子中的核心谓词依次是“查询”“联系”“说”“听”，其中位于核心谓词之前的是状语部分，比如“可以提前”“电话”“怎么”；位于核心谓词之后的是补语部分，比如“个没完没了”“不懂”。

本文根据北京语言大学句法结构树库的这一特点，限定两个连用的动词出现在带有“VP-PRD”标签的述语块内，保证连用的两个动词充当的是句子的述语成分，“提前查询”这个动词作状语修饰动词中心语的结构就可以在两个双音节的动词接连出现并且标签是“VP-PRD”的述语块中检索出来。

3.2 状位动词的抽取与消歧

对状位动词进行语义分类并构建相应的语义搭配库的前提是得到比较准确的状位动词语料，本小节主要介绍从北京语言大学句法结构树库中抽取状位动词语料并进行消歧的过程。

3.2.1 语料抽取

本文的语料抽取基于北京语言大学句法结构树库，该句法结构树库不仅是经过浅层树结构标注的树库，同时也是经过分词词性标注的树库，因此可以依据结构信息与词性信息来进行语料的抽取。考虑到后期形式分析的需要，在抽取时首先根据前后两个动词音节的不同将两个连用的动词分为以下四种情况：

- (1)单音节动词状语+单音节动词中心语：简称V1+V1
- (2)单音节动词状语+双音节动词中心语：简称V1+V2
- (3)双音节动词状语+单音节动词中心语：简称V2+V1
- (4)双音节动词状语+双音节动词中心语：简称V2+V2

同时在检索式层面对数据进行降噪。虽然助动词不在本文考察之列，但是关于助动词具体包括哪些词，各家看法不一，本文同时参考了朱德熙（1982），黄伯荣、廖序东（1991），马庆株（2005），邢福义（2011）等学者的研究成果，取其并集，最后一共得到了64个助动词，具体如表1所示：

单音节	会、可、能、好、得、应、该、愿、想、许、要、肯、敢、配、需、须、甬、别、准
双音节	准许、许可、容许、允许、宁愿、宁可、甘愿、须要、一定、显得、必须、便于、得以、能够、可以、可能、应该、应当、须得、总得、不得、必得、愿意、情愿、想要、要想、希望、企图、乐得、高兴、敢于、勇于、甘于、苦于、懒得、忍心、值得、难于、易于、善于、适于、宜于、容易、乐意

表1.助动词词表

在语料的抽取阶段，直接限制作状语的动词不在以上列出的范围之内，为下一步人工消歧降低了一定的难度。

以抽取“V1+V1”的流程图为例来介绍语料的抽取过程，详细的代码及解释请参考附录1：



图1.语料抽取流程

首先给出检索式的限制条件，即连用的前一个动词不是助动词，并且连用的两个动词都是单音节的；其次定位两个连用的动词；第三步定位一个述语块；第四步将两个连用的动词限制在一个述语块内部出现，即两个连用动词整体充当一个句子的述语部分并且该句子只有一个核心谓词；最后得到符合该检索式的样例频次及上下文信息。

综上，该流程图所表达的意思是，两个单音节的动词连用，同时作状语的动词不是助动词，并且连用的两个动词出现在一个述语块内且只有一个核心谓词的情况，最后输出符合该检索式的样例频次及上下文信息。

3.2.2 歧义消解

在经过检索式层面的降噪之后，抽取结果中仍旧包含许多有歧义的句子。根据齐夫定律，依据频次信息将结果按照二八比例分开，删去占有所有频次20%的检索结果，并对剩余的80%的高频检索结果按照以下几个步骤进行消歧，具体如下：

首先，针对词性标注错误的情况，对检索结果进行词典消歧。将连用的两个动词与《现代汉语词典》（第七版）中的动词进行对照。具体情况如下：如果其中一个不为词，则删除检索结果。对于保留的检索结果，如果“V1+V1”或者“V1+V2”或“V2+V1”整体为词典中的一个词，则删除检索结果；如果“V2+V2”的前三个字为词典中的一个词，同样将该检索结果删除。

其次，删去不为状中结构的检索结果。在词典消歧的基础上，对语料进行人工排查，并删除非状中结构的检索结果。分为两种情况，一是具有形式化特征的，比如在“V1+V2”格式中，第二个双音节的动词为趋向动词时，如“上来、上去、下来、下去、进来、进去、出来、出去、回来、回去、起来、过来、过去”，必定是做补语，可批量删除；二是没有形式化特征的，比如“笑开怀”是述补结构，“嫉妒恨”是并列结构，需要进行人工判定。

最后，针对兼类词现象进行词典义项的排查。由于许多动词是兼类词，比如动副兼类词、动形兼类词、动名兼类词等，这些词都会在检索结果中出现。动副兼类词如“拼命¹”，在例句“小明每天都拼命锻炼身体”中，“拼命”是副词而不是动词。针对这种情况，就需要人工对检索结果进行义项的排查。具体做法如下：以《现代汉语词典》（第七版）中的动词义项为标准，依据上下文信息，逐个对照检索结果中的状语位的动词是否为动词义项作状语，若不是则删去检索结果。

经过以上消除歧义的步骤之后，得到了可以作状语修饰动词中心语的动词共893个，其中单音节的有134个，双音节的有759个，这个数量大大超过了以往研究成果中的动词作状语的数量。

4 状位动词语义分类及分析

以降噪后的动词作状语的语料为基础，参考现有的研究成果，结合状位动词本身及其在状中结构中所表达的语义对状位动词进行语义细分类，并对分类结果进行统计分析。

4.1 状位动词语义分类

一般认为，动词作状语表示动作的某种方式，但是状位动词除了表示方式这一语义外，还可以表示其他多种语义，本小节针对这一问题，试图对状位动词所表达的语义信息进行详尽的

¹【拼命】pīn//mìn ①{动}把性命豁出去；以性命相拼：跟歹徒～。②{副}尽最大的力量；极度地：～地工作 | ～往山顶爬。

分类。

目前学界关于状位动词语义分类的研究成果较少，朴正实（2003）指出动词作状语的语义关系分为以下五种：方式、动作、过程、时间、情态。虽然可参考的研究成果很少，但是本文认为状语位的动词同状语位的副词与形容词的语义功能类似，在状中结构中都表达着相似的语义关系，因此参考姜汇川（1989）、邢福义（1996）、张谊生（2000）对副词的语义分类体系，对状位动词所表达的语义进行大类及次类的划分。

以上文得到的893个状位动词为基础对状位动词进行人工的语义分类。主要指导原则是：以状位动词在搭配中体现的语义为主，以状位动词本身的语义为辅。以姜汇川、邢福义与张谊生构建的副词语义分类体系为参考，经过多轮迭代之后，共将状位动词所表达的语义分为23个大类，47个小类。相较于学界目前关于状位动词的研究成果，具有较为严格的分类体系，并且语义类别的划分更加细致，为深层次的句法语义分析贡献更多更丰富的信息。

分类结果如表2所示，由于小类较多，表2中只选取大类中具有代表性的小类进行说明，全部的分类体系见附录2²：

语义大类	语义次类	例词	例句
程度	深	无比	我 无比 喜欢这棵橡树
时间	提前	提早	它能够帮你 提早 发现敌人的踪影。
范围	统括	总括	但可以 总括 说，区别都来自爱恶的不同。
频序	次序	首选	首选 使用简易器材的体育健身方法。
判断	肯定	保准	他 保准 发生过这样的事情。
语气	反问	岂可	这样的赏雪胜地 岂可 错过？
数量	超量	超标	同时违反规定，用公款 超标 接待。
速度	加速	加快	加快推进 12315行政执法体系建设。
方向	方向	反向	市场口碑也很容易 反向 操作。
结果	结果	退学	老师说下星期要交下学期的学费，否则做自动 退学 处理。
伴随	伴随	临风	然而每临黄昏总见她 临风 洒泪。
原因	原因	窒息	医生说小兵已经 窒息 死亡。
意见	同意	应许	而且你也 应许 说你一定会来。
方式	一般方式	上网	学生能 上网 了解外面的世界。
状态	状态	在职	鼓励干部 在职 学习。
情貌	比况	蜂拥	国际大公司开始 蜂拥 进入中国市场。
影响	影响	联动	本届哲学社会科学普及周将于9月20日至25日在全省 联动 举行。
目的	目的	澄清	他 澄清 说他要的是一种被重视的感觉。
限制	限制	限价	也可比照经济适用住房的价格实行 限价 销售。
距离	距离	凑近	凑近 看挺吓人的。
评价	夸奖	出彩	李莎子在场上 出彩 发挥。
态度	向上	奋发	激励学生 奋发 求学。
理应	理该	理该	灾区有困难，我们 理该 帮助。

表2.状位动词语义分类及例词例句

需要说明的是，本文将动词前的“来、去”也当作状语，认为二者在动词前出现时表示的是某种目的，比如：

eg13: 哈亚和坦皮纳**去**寻找“奥斯曼土耳其”伊斯坦布尔的忧伤形象。

²附录中，每一类仅列出一部分结果，完整的分类结果参见[https://github.com/bien-apprendre/-](https://github.com/bien-apprendre/)

eg14: 人通过接触世界来**提高**判断力。

上例中的“去寻找、来提高”在句子中表示的是某种目的，因此将“来、去”划分到了表示“目的”的语义类中。

4.2 结果分析

基于上文对状位动词的语义分类结果，本小节进一步对该分类结果进行统计分析，图1表示的是每一个语义大类所包含的状位动词的数量占比。

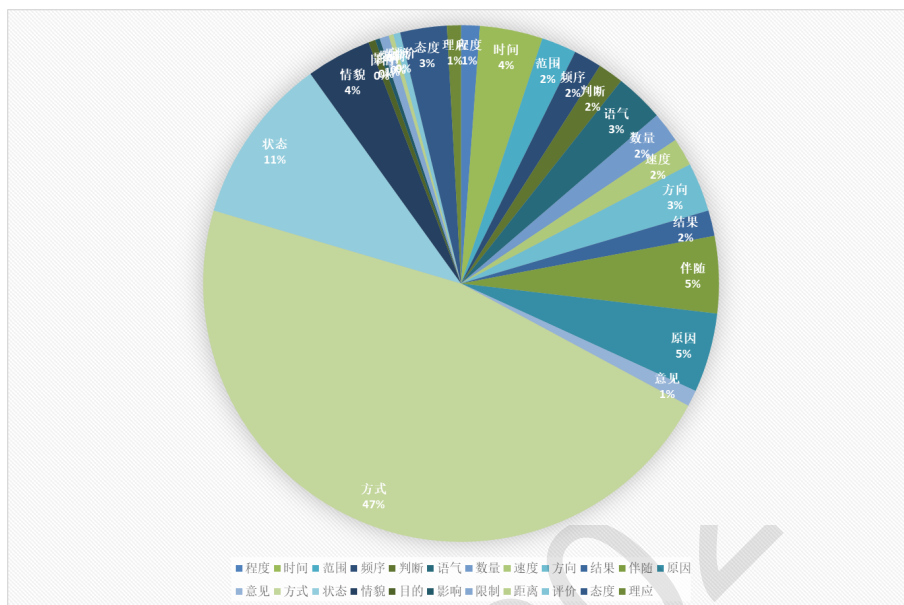


图2.状位动词大类分布情况统计

由图2可知，表示方式的状位动词占比最高，达到了47%，其次是表示状态的，有11%，其余每一类的占比都低于10%。虽然其他语义类的占比都不大，但是其总量也达到了50%多，以近一半的数量占比承载了大多数的语义。

由于方式大类的占比高达47%，方式类又按照动词本身的语义及其在状中结构中所表达的语义的不同分为：协同方式、言说方式与一般方式三小类。言说方式指说话的方式；协同方式指涉及到两个人以上的动作方式；一般方式是除去言说方式与协同方式以外的其他动作方式。三个小类的例词、例句及个数与占比如表3所示：

语义大类	语义次类	例词	例句	数量	占比
方式	协同方式	携手	LG日前结盟英特尔公司 携手 打造WiDi技术。	11	3%
	言说方式	支吾	李培斌嘴上 支吾 说工作忙走不开。	79	24%
	一般方式	自助	轻点鼠标，即可 自助 查询通话清单。	239	73%

表3.方式小类及例词例句

在表示方式的三个语义小类中，表示一般方式的最多，其次是表示言说方式的，而表示协同方式的占比最小。对于表示一般方式的类别划分不是很细致，本文也尝试对一般方式进行细分类，但是词语之间的语义共同点少，导致类别太多，类似于《同义词词林》的词簇，从应用的角度出发，将这些词簇统一归到了“一般方式”这一语义类中。

5 状位动词语义搭配库的构建与统计分析

本小节以上文动词作状语的语义分类体系为标准，以降噪后的状位动词的语料为基础，构建相应的动词作状语的语义搭配库，并进一步对该语义搭配库进行统计分析。

5.1 语义搭配库构建

得到状位动词并且对其进行语义分类之后，进一步根据降噪后的状位动词的语料及语义分类结果构建相应的状位动词的语义搭配库，标明不同的状位动词与相应的中心语之间的句法与语义关系，即本研究所构建的搭配库不仅仅包含句法信息，还包括语义信息。“搭配”一词的定义最早是由Firth (1957) 提出的，他认为自然语言中的词汇并不是孤立使用的，而是与其他词语组合使用的，并且把搭配的整体看成表达意义的方式。因此构建了状位动词的语义搭配库之后，可以将每一个搭配对看成一个整体，并且给出其整体所表达的语义关系。

语义搭配库的构建以中心语动词为中心，按照状位动词与中心语动词在语料库中连用的频次排序，并且标明每一个状位动词与中心语动词结合时所表示的语义类型，以“V2V2”格式中的核心动词“注意”为例来介绍状位动词语义搭配库的格式：

.ZZ_V2V2_注意

```
(
[
提前 时间_提前
必需 语气_坚决
足够 程度_深
理应 理应
.....
)
]
```

开头的“.”表示一个动词中心语的开始，“ZZ”表明句法类型，即状中结构的首字母缩写，“V2V2”表示修饰语与中心语的音节与词性，即双音节的动词修饰双音节的动词，最后标明该中心语是“注意”。接下来按照状位动词与中心语动词连用的频次排序，每个状位动词后面标明该状位动词与该中心语连用时所表示的语义类型，比如“提前”与“注意”连用时，表示属于“时间”大类中的“提前”小类，即所表示的语义是在时间上的提前，以此类推。但是在本文的分类体系中，有的大类没有细分小类，比如“理应”类，则直接标明大类的名称，如果大类细分小类，则在大类后面加下划线“-”后再加该小类的名称，比如“时间_提前”。

5.2 搭配库统计分析

构建状位动词的语义搭配库之后，进一步对该搭配库进行统计分析。综合“V1+V1”“V1+V2”“V2+V1”“V2+V2”四种格式，共得到10031个搭配对，并且对状位动词与中心语动词之间的音节搭配信息以及各个格式下的修饰语与中心语的个数及占比进行统计，如表4所示：

格式	搭配对数量及占比		修饰语个数及占比		中心与个数及占比	
V1+V1	1160	11.56%	134	13.1%	464	10.6%
V1+V2	2015	20.09%	2	0.2%	1345	30.7%
V2+V1	956	9.53%	354	34.6%	302	6.9%
V2+V2	5900	58.82%	533	52.1%	2276	51.8%
汇总	10031	100%	1022	100%	4378	100%

表4. 状位动词搭配库搭配数及修饰语与中心语个数统计

关于搭配对的数量，其中“V2+V2”格式的最多，占总数的58.82%，其次是“V1+V2”“V1+V1”“V2+V1”，占比分别为20.09%、11.56%、9.53%，这三个格式的占比都远远小于“V2+V2”格式的匹配数量。同时“V2+V1”格式的占比最小，而在汉语的韵律语法中，“V2+V1”这样“2+1”的音节搭配是最少的情况，说明状位动词修饰中心语的音节搭配符合汉语韵律语法中的音节搭配规律。

同时，充当修饰语的动词和充当中心语的动词都以双音节为主，充分说明了双音节词在汉语组合搭配中的优势地位，其中又以“V2+V2”最为常见，说明了双音节与双音节搭配的普遍性。而占比最低的是“V1+V2”格式，该格式的修饰语个数只有两个，通过返回原文发

现,“V1+V2”格式的修饰语只有“来”和“去”两个,在语义分析时可以通过语义搭配库直接认定“V1+V2”格式中的修饰语所表达的语义类型是“目的”,同时其他格式的也按照语义类别的不同构建状位动词与相应的动词中心语之间的语义搭配库。

为了进一步明确音节韵律组合和语义之间的交叉关系,本文对语义搭配库中的每个格式下的每一语义大类的搭配数及每一格式下每一语义类占这一语义类的比例进行了统计,如表5所示:

	V1+V1及占比		V1+V2及占比		V2+V1及占比		V2+V2及占比		汇总
程度	7	1.46%	0	0.00%	63	13.15%	409	85.39%	479
时间	30	2.49%	0	0.00%	163	13.50%	1014	84.01%	1207
范围	35	55.56%	0	0.00%	12	19.05%	16	25.40%	63
频序	9	25.71%	0	0.00%	7	20.00%	19	54.29%	35
判断	0	0.00%	0	0.00%	11	78.57%	3	21.43%	14
语气	2	0.49%	0	0.00%	79	19.46%	325	80.05%	406
数量	0	0.00%	0	0.00%	4	7.69%	48	92.31%	52
速度	0	0.00%	0	0.00%	9	5.17%	165	94.83%	174
方向	0	0.00%	0	0.00%	148	22.80%	501	77.20%	649
结果	0	0.00%	0	0.00%	0	0.00%	13	100.00%	13
伴随	36	29.03%	0	0.00%	17	13.71%	71	57.26%	124
原因	6	11.32%	0	0.00%	7	13.21%	40	75.47%	53
意见	0	0.00%	0	0.00%	11	19.64%	45	80.36%	56
方式	232	11.32%	0	0.00%	257	12.54%	1561	76.15%	2050
状态	87	33.59%	0	0.00%	20	7.72%	152	58.69%	259
情貌	3	1.20%	0	0.00%	46	18.33%	202	80.48%	251
目的	707	25.96%	2015	74.00%	1	0.04%	0	0.00%	2723
影响	0	0.00%	0	0.00%	0	0.00%	19	100.00%	19
限制	0	0.00%	0	0.00%	8	21.05%	30	78.95%	38
距离	0	0.00%	0	0.00%	3	7.69%	36	92.31%	39
评价	0	0.00%	0	0.00%	1	1.92%	51	98.08%	52
态度	3	0.56%	0	0.00%	42	7.91%	486	91.53%	531
理应	3	0.40%	0	0.00%	47	6.32%	694	93.28%	744
汇总	1160		2015		956		5900		10031

表5.状位动词搭配库语义类分布统计

由表5可知,在“V2+V2”格式中,除去“范围、判断、目的”这三个语义类之外,其他二十种语义类的搭配数占比均达到了50%以上,其中八种语义类的搭配数占比达到了90%以上。而在“V1+V1”、“V1+V2”与“V2+V1”格式中,依次是“范围”“目的”和“判断”语义类的搭配数占比达到了50%以上。因此,无论动词作状语表达的是什么样的语义关系,“V2+V2”格式总是占据着绝对的优势地位,然而是否其他结构的情况也是如此,有待进一步深入考察。

通过构建状位动词与中心语动词之间的语义搭配库,就可以正确识别两个动词在一个小句中连用时是否为状中结构并且正确判定修饰语与中心语的位置,进一步提供该状位动词在状中结构中所表达的语义信息,为深度的句法语义分析提供更多的知识。

6 总结与展望

本文以状语位的动词为研究对象,从北京语言大学句法结构树库中抽取在一个小句中接连出现两个动词的情况,经消歧后得到了893个可以直接作状语的动词,并且对这893个可以作状语的动词进行语义分类,得到了23个语义大类,47个语义小类。同时根据消歧后的语料与状位动词的语义分类结果,构建了状位动词与相应的动词中心语之间的语义搭配库,搭配库中标明了每个搭配所表示的语义关系,最后共得到了10031个搭配对。构建了这样的语义搭配库之后,可以正确识别两个连用的动词的结构,并给出状位动词在状中结构中所表达的语义,同时也为汉语深层次的句法语义分析提供了更多的知识。

但是,本文的状中结构仅详细说明了动词为中心语的情况,形容词为中心语的情况也做了相同的研究,发现VV结构中状位动词的语义分类可以完全涵盖VA结构中状位动词的语义表达,说明VV结构中状位动词的语义分类对其他结构的分类具有一定的参考价值。

同时,可以进一步深入研究的还有以下两点:

一是两个动词在一个小句中接连出现除了可能构成状中结构之外,还可能构成述补、动宾等结构,虽然在借助结构树库对述语块限制的情况下,将可能性缩小到了状中和述补两种类型,但为了使汉语句法语义分析的结果更精确,还需要进一步对述补结构做相应的系统性研究。

二是在VV结构中,由于前一动词的实际语义较为虚化,具备了充当状语的可能性,进而与核心动词(往往具有强语义)形成高低搭配关系,可以在现有数据统计分析的基础上,对两个动词的语义强度及其规律进一步研究,以揭示更深层的语言规律。

参考文献

- 高增霞. 2004. 从语法化角度看动词直接作状语. 汉语学习, (4):18-23.
- 耿雪. 1996. 现代汉语 $V1+V2$ 偏正结构研究. 南京师范大学硕士论文.
- 郭奇军, 龙启艳. 2010. 非范畴化视角下的现代汉语动词作状语研究. 集宁师专学报, 32(1):34-39+71.
- 黄伯荣, 廖序东. 2002. 现代汉语. 北京: 高等教育出版社.
- 侯雪钰. 2012. 现代汉语动词作状语问题研究. 吉林大学硕士论文.
- 姜汇川, 许皓光. 1989. 《现代汉语副词分类实用词典》. 对外贸易教育出版社.
- 卢露, 矫红岩, 李梦, 荀恩东. 2021. 基于篇章的汉语句法结构树库构建. 自动化学报, 1-12.
- 马鸣春. 1996. 连动句与动词状语句的对比. 兰州学刊, (6):94-98.
- 马庆株. 2005. 《汉语动词和动词性结构》. 北京大学出版社.
- 朴正实. 2003. 动词做状语的构成方式及语义特征. 淮阴师范学院学报(哲学社会科学版), (2):259-261.
- 孙德金. 1997. 现代汉语动词做状语考察. 语言教学与研究, (3):116-129.
- 王政红. 1989. 动词作动词的修饰语研究. 南京师大学报(社会科学版), (2):70-75.
- 吴思霖. 2019. 基于原型范畴理论的现代汉语连动式和状中式对比. 文学教育(上), (12):166-169.
- 肖伟良. 1983. 试论动词作动词的状语. 广西师范学院学报, (3):35-39.
- 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.
- 邢福义. 1996. 现代汉语. 北京: 高等教育出版社.
- 邢福义, 汪国胜. 2011. 现代汉语. 北京: 高等教育出版社.
- 杨佳玉. 2015. 晚清时期动词作状语研究. 吉林大学硕士论文.
- 张军. 2013. 现代汉语动词作状语的主要形式及动因分析. 华中师范大学学报(人文社会科学版), 53(5):100-108.

张谊生. 2000. 现代汉语副词的性质、范围与分类. 语言研究, (1):51-63.

张谊生. 2000. 现代汉语副词研究. 学林出版社.

章新传. 2004. 《史记》动词作状语考察. 江西教育学院学报(社会科学), (1):57-60.

朱德熙. 1982. 《语法讲义》. 北京: 商务印书馆.

Firth, J.R. 1957. *Modes of meaning in Papers in Linguistics*. Oxford University Press.

A 附录

A.1 检索式示例

```
Condition("$1!=[会可能好得应该愿想要肯敢配需须甬别准许];len($1)=1;len($2)=1")
```

```
Handle0=GetAS("—v","","","","","","","0,1","","","")
```

```
Handle1=GetAS("—v","","","","","","","0,1","","","")
```

```
Handle2=JoinAS(Handle0,Handle1,"Link")
```

```
Handle3=GetAS("$VP-PRD","","","","","","","","","")
```

```
Handle4=JoinAS(Handle3,Handle2,"SameRight")
```

```
Handle5=Freq(Handle4,"$Q")
```

```
Handle6=Context(Handle4,30)
```

其中, Condition指的是检索式的一些限制条件。!=表示不等于。\$1表示对作状语的动词进行限制, \$2表示对作中心语的动词进行限制, len表示长度, 因此第一行的意思就是限制作状语的第一个动词不是单音节的助动词, 并且限制两个动词的长度都为1。

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, Handle0与Handle1表示分别得到了两个词性为动词的原子查询项; “\$VP-PRD”是限制该语块为述语块, Handle3表示的是得到标签为“\$VP-PRD”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表示的是两个动词接连出现的情况; Handle4表示的是两个连用的动词在一个述语块内出现并且两个连用的动词与述语块之间是右对齐的情况。

Freq表示输出检索式的频次信息, Context表示输出检索式的上下文信息, 其中“30”表示输出检索式前后各30个词语。

A.2 状位动词分类结果

1.程度:

(1)程度深:足够 无比 致命

(2)程度浅:凑合

2.时间:

(1)提前:提前 提早 抢先 抢

(2)落后:延缓缓

(3)相继发生:接踵 跟进

(4)突然发生:突发 突现

(5)同时发生:并行 并发

(6)时间较晚:入夜 熬夜

(7)间隔:隔世 转世

3.范围:

(1)统括性:遍野 总括 兼 遍

(2)变大:扩展开 延伸 扩张 延展 扩展

(3)变小:浓缩 缩微

4.频序:

(1)次序:首选 排列 续 带头

(2)重复:复述 回放 回炉 再版

5.判断:

- (1)否定:误传
- (2)肯定:保准
- (3)总结:归结

6.语气:

- (1)坚决:决心 必需 决意 坚守
- (2)强调:突出 侧重 强化 加强
- (3)反问:岂可
- (4)揣测:估计 目测 猜
- (5)感叹:何止
- (6)担心:唯恐

7.数量:

- (1)超量:超标 超重 超值 翻番
- (2)满量:满载 期满 满仓
- (3)总量:累计 累世 累年 累进

8.速度:

- (1)加速:增速 加速 加快
- (2)减速:减速
- (3)渐进:渐进 渐变

9.方向:

- (1)趋向:起来 出来 出去 回来 进来
- (2)方向:倒回 侧移 返背 反向 相向

10.结果:和解 退学 退票 作废 报废 销毁 燃烧 停职 回收 互补 赔偿

11.伴随:拍案 临风 击掌 鼓掌 拍手 挥手 俯端 哭笑 醉颤 眯赶

12.原因:挑食 阻塞 内讧 打黑 行贿 抢劫 盗窃 偷窃 偷盗 贩毒 退瘫 晕压 跌败

12.意见:

- (1)同意:应许 应从 倾向
- (2)反对:反驳 否定 失实
- (3)居中:居中

13.方式:

(1)言说方式:发布 公布 上报 求助 提示 反映 透露 披露 投诉 造假 批复 回答 回复 回应 访问 追问 会谈 宣布 抱怨 提问 调侃 自夸 分享 讨论 叮嘱 描述 告诉 解释 攀谈 闲聊 胡诌 辩解 记述 支吾 转述 表明 规劝 讲述

(2)协同方式:联手 配合 搭档 携手 偕同 协助 搭配 联名 联结 群起 合

(3)一般方式:整合 上网 即席 抽签 鞠躬 匿名 亲临 转账 开车 自助 空降 猜拳 氧化 对接 脱水 隔行 提名 具名 匍匐 下乡 登台 起哄 拍照 开会 倒车 签字 沉淀 辨证 自筹 按压 举证 助跑 打开 致电 签名 捉拿 缴械 轮换 交换 变换 转换 整改 迂回 辗转 刷溜 镶分 铺围 吞扑 蹲嚼 抱跳 翻

15.状态:束手 在职 在场 封闭 挂牌 直立 展翅 收录 披挂 竖立 鼎立 开怀 镇定 汹涌 举债 扣押 关押 羁押 记载 涉险 飘忽

16.情貌:

- (1)情感:忍痛 蓄意 自信 违心 放心 欢喜 欢呼 感恩
- (2)比况:等同 蜂拥 貌似 宛如 好似
- (3)对比:相形 对比

17.目的:澄清 来去

18.影响:应运 联动

19.限制:

- (1)有限制:限价 节制
- (2)无限制:放手 放量

20.距离:凑近 临阵

21.评价:

- (1)褒义:出彩 创意

(2)批评:批评

22.态度:

(1)正向:尽职尽责 奋发 奋起 发奋

(2)中立:想法 警惕

(3)负向:玩笑 敷衍

23.理应:理应 理当 理该 相应

JCL 2021