

基于依存语法的偷抢类动词研究

王珊、刘晓骏

澳门大学人文学院，澳门，999078

shanwang@um.edu.mo

摘要

本文筛选了汉语“偷抢”类动词的单句，并借助依存语法的标注体系对“偷抢”类动词句法依存和语义依存进行定量分析。研究表明，当汉语“偷抢”类动词为从属词时，表现出句法功能的多样性、内部相似性和区别其他动词小类的特异性，其语义角色分布具有多样性。当汉语“偷抢”类动词为支配词时，该类动词的句法依存随其不同的句法功能而发生变化；从该类动词的语义依存来看，其客体语义密度整体低于主体语义密度，最常见的情境角色是地点和时间，在事件关系中，并列事件发生概率最高。“偷抢”类动词的句法语义特点丰富，主要的句型为主谓宾句式，而该句式中最常用的语义搭配模式是施事对受事实偷抢动作。本研究结合依存语法和框架语义学，深化了对汉语“偷抢”类动词的句法、语义和事件关系的了解，促进了对该类动词的研究。

关键词：依存语法；句法；语义；偷抢类动词

Research of Verbs of Stealing and Robbing Based on Dependency Grammar

Shan Wang, Xiaojun Liu

Faculty of Arts and Humanities,
University of Macau, Macau, 999078
shanwang@um.edu.mo

Abstract

This paper selected single sentences of verbs of stealing and robbing and conduct quantitative analysis on their dependency relations based on Dependency Grammar. Results show that when Chinese verbs of stealing and robbing are dependents, they exhibit their diversity, similarity, and specificity from other types of verbs in terms of syntactic functions; their semantic roles are also diversified. When they are heads, their syntactic dependency are different as they display different syntactic functions; from the perspective of semantic dependency, the object semantic density of this type of verb is generally lower than the subject semantic one. The most common situational roles are location and time roles. In the event relationship, eCOO is the most frequent. Chinese verbs of stealing and robbing are embedded with rich syntactic-semantics features and their main sentence pattern is subject-predicate-object; their most commonly used semantic collocation pattern is that the verbs are accompanied by Agent and Patient. This study combines dependency grammar and frame semantics, delves into the syntax, semantics and event relation of verbs of stealing and robbing, and deepens the exploration of verbs.

Keywords: dependency grammar, syntax, semantics, verbs of stealing and robbing

1 引言

移除和放置都是人类最重要的身体经验，但是目前学界大都着眼于放置类动词（如 Liu and

Chang, 2019), 对移除动词的重视还远远不够。继 Levin & Hovav (1991) 对英语移除动词进行了初步探究后, Liu (2002) 和 Lien (2006) 分别对汉语普通话和闽南语的部分移除动词进行了研究, 但是基于大规模语料的句法语义研究还有待开展。偷抢类动词是一种社会活动层面的移除行为, 对该类动词的研究有助于认识移除行为的特点, 丰富对动词的研究。本研究聚焦于移除动词的小类“偷抢”类动词, 探索其句法和语义接口问题。

本文以搜狗实验室、BCC、CCL、Chinese Gigaword、People's Daily、参考消息和腾讯新闻的语料为基础, 筛选了“偷抢”类汉语动词的单句。在依存语法的基础上, 对所选句子进行句法和语义上的依存分析, 同时从框架语义学的角度, 探讨“偷抢”类动词的句法和语义特征。我们发现“偷抢”类动词在句法功能和语义搭配上有着显著的相似性, 但也有细微的差异。我们可以“偷抢”类动词看作是一个“偷”和“抢”两个层次的连续统, 本文的贡献如下: (1) 建立了汉语“偷抢”类动词语料库;

(2) 在依存语法理论的指导下, 定量分析了“偷抢”类动词在单句使用中的句法和语义特征; (3) 从框架语义学的角度总结了汉语动词“偷抢”类的句法和语义共同特点。

2 研究概况

对动词分类的研究很多, 但学术界对动词分类的观点和依据却不尽相同, 根据动词的情感特征作为标准, 可以将动词分为活动动词、目标动词、完结动词和达成动词 (Vendler, 1967)。从动词价的角度来分析, 存在单价、二价、三价动词等 (Biber et al., 1999; 范晓, 胡裕树, 1995)。根据动词带宾语的情况来看, 动词可分为及物动词、不及物动词和连系动词 (Ralph, 1966) 或双宾语动词、及物动词以及不及物动词 (赵元任, 1979)。Levin 在 1991 年撰写的一篇短文中对动词的移除进行了初步探讨 (Levin and Rappaport, 1991), 然后根据动词的句法语义将英语动词分为不同的语义类 (Levin, 1993), 这不仅促进了英语和其他语言之间的 VerbNet 的构建, 而且推动了对不同亚类动词的研究。如运动类的动词 (Beavers and Koontz Garden, 2017)、放置类的动词 (Liu and Chang, 2019)、天气类的动词 (Huang et al., 2021)、状态变化的动词 (Beavers et al., 2017) 等等。

“偷”和“抢”作为移除动词, 也引起许多相关的讨论。Goldberg (1995) 在构式语法中研究了 steal 和 rob 这两个相近的英文单词, 并且认为这两个词的差异在于认知侧显的不同, steal 凸显的是被抢的财务, 而 rob 凸显的是被抢的人, 如图 1 所示。也有研究比较了不同语言中的“偷”类动词 (verbs of stealing), 如英语和德语 (Dux, 2018) 以及英语和西班牙语 (Fernández-Martínez and Faber, 2020)。

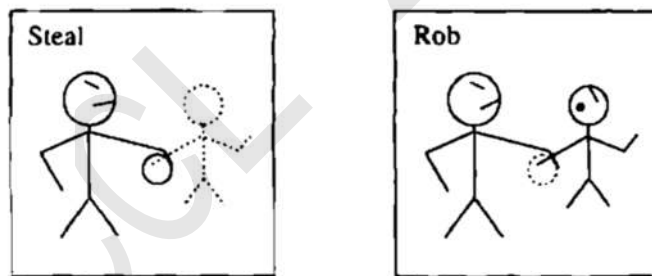


图 1. steal 和 rob 的语义侧显 (Goldberg, 1995)

汉语动词“偷”和“抢”的语义特点及其句法表现, 陆丙甫 (1998)、沈家煊 (2000)、张国宪 (2001) 和王寅 (2011) 等学者都曾有过很好的研究。沈家煊 (2000) 发表了一篇名为《说“偷”和“抢”》的论文, 他着重对比了“偷”和“抢”steal 和 rob。王寅 (2011) 对英语的 steal, rob 和汉语的“偷”、“抢”形成的构式进行了对比研究, 但是他的研究重点在于 steal 和 rob 之间, “偷”和“抢”之间的对比, 即同一种语言中两个近义词的对比。他们也都认为“偷”和“抢”都会涉及进行该动作的人和被偷/抢的物品, 但是区别在于“偷”凸显的是被偷的物, 而“抢”凸显的是被抢的受害者。

汉语中“偷”和“抢”具有不同的认知侧显, 本文对这两个单音节动词不做赘述。但是具有“偷抢”特点的汉语双音节“偷抢”类动词, 在句法和语义上是否会有一些特性呢? 为了深入研究“偷抢”类动词的句法语义特点, 本文以依存语法的理论基础, 对“偷抢”类双音节动词的单句进行大规模的句法依存标注和语义依存标注。在定量分析该类动词的句法语义特点后, 本文试图借助框架语义学的理论框架来阐释其成因。

3 研究方法

汉语在依存语法上的标注适用性高(Niu and Osborne, 2019),并且依存语法形式简洁且标注便捷(李正华, 2013)。目前刘海涛有很多基于依存语法的汉语研究(如李辉, 刘海涛, 2017; Liu, 2008; 刘海涛, 冯志伟, 2007),澳门大学依据依存语法分析了一些汉语动词,如招认类动词(王珊、汤蕾, 2021)、抱怨类动词(王珊、周洁, 2021)和许诺类动词(周洁、王珊, 2021)。故本研究将借助依存语法的标注体系对“偷抢”类动词进行大规模的语料依存分析。

第一,选取研究对象。以汉语“偷”和“抢”为语义标准,依据《现代汉语分类词典》、同义词词林(扩展版)和现代汉语常用词表,选取以下16个汉语“偷抢”类双音节动词为研究对象:扒窃、打劫、盗窃、盗用、劫持、劫夺、劫掠、抢夺、抢劫、抢掠、窃夺、窃取、偷盗、偷窃、洗劫、行劫。

第二,筛选“偷抢”类动词的单句。基于搜狗实验室(Liu et al., 2012)、BCC、CCL、人民日报和腾讯新闻的数据,根据关键词下载收集数据文段,BCC语料库只选择了报刊和文学两个子语料,下载段落的左右下载窗口宽度都是选择最大值,即,BCC语料库的情况是设置默认最大窗口值为30,CCL语料库的情况是设置的最大窗口值为1000。下载段落后,在规定好的句末标点符号分割段落,删除文段右侧窗口中不完整的句子。根据标点符号的使用,句末标点符号包括句号、感叹号以及问号。因为感叹号和问号在中英文情况下相似,在使用过程中误用率很高。所以本文所选句子末尾的标点符号包括中文句号(。)、感叹号(!!)中英文问号(??)。

鉴于复句的相互依存关系比单句的复杂很多,以及标注效率和准确性远低于单句,所以本项标注和统计分析的语言单位为单句。首先对含有关键词句子排除中文逗号(,)以及中文分号(;)通过这个手段尽可能排除复合句;然后排除含有非汉字和非汉字标点符号的单句,也就是说只保留符合中文标准的标点符号的句子;最后分词和词性标注,只保留关键词为动词的单句,构成本研究筛选的单句集。

以“行窃”为例,含有关键词的文段中的句子因为不符合筛选标准而被排除。例如:

例1.不完整的句子:行窃难免担惊

例2.有逗号的句子:我相信能够这样轻松地走遍大地,我在教堂里行窃更使我飘飘欲仙。

例3.有分号的句子:如今我已经可以思念巴黎的生活而不至于产生某种行窃的感觉;如今我与他分享着一切,哪怕那些将我们彼此分离的东西。

例4.依存关系复杂容易出错的句子:有人叛卖了葫芦,说他狂得没了边,在谢队长身上也敢行窃,把谢队长用五个青稞馒头换来的欧米茄摸走了。

第三,标注单句的句法依存和语义依存。利用哈工大语言技术平台LTP提供的API接口(刘挺 et al., 2011)开发了基于依存语法的句法-语义标自动处理脚本,主要模块及功能有:(1)分词及词性标注,(2)句法依存分析,(3)语义依存分析,(4)语义角色标注,其中(1)模块中的分词功能是使用一种CRF模型(Lafferty et al., 2001),在原本平台提供的语料基础上,添加最新的人民日报(中国)语料库作为数据训练而成的。(1)模块中的词性标注功能也是使用最新的人民日报语料,采用SVMTool3模型训练而成的。(2)(3)模块是采用高阶图模型(Che et al., 2009)在CoNLL-2009多语言共享的语法语义依赖任务(Hajič et al., 2009)基础上实现完成的。(4)模块是基于语法解析器,采用最大熵模型(Che et al., 2009)实现完成的。以上都是使用哈工大提供的LTP第四版(最新版本)(<https://github.com/HIT-SCIR/ltp>)的API接口开发完成的。使用标注脚本对语料库中的单句进行大规模依存语法标注和分类统计,总结归纳该类动词的句法语义特点。

最后,本研究的研究对象为16个“偷抢”类动词,筛选了8156条单句,具体分布参见附录。每条单句均注明了来源语料库、所属文段等元信息,以及依存句法标注、依存语义标注等依存信息。以“偷抢”类动词为关键词,分析依存标注结果,按照关键词和其他成分之间的支配和被支配关系分类进行讨论,其中关键词充当的句法成分和语义角色针对的是关键词被其他成分支

配的情况（即关键词作为从属词），而与关键词同现的句法成分及语义成分是关键词作为支配词的情况。

4 “偷抢”类动词的标注统计分析

4.1 “偷抢”类动词的句法成分分布

动词在句子中可以充当不同的句法成分，承担不同的句法功能。通过对单句进行依存句法标注，“偷抢”类动词在单句中能够充当的句法成分按照所占比例从大到小依次为定语(ATT)、并列(COO)、动宾(VOB)、谓语(HED)、主语(SBV)、介宾(POB)、状语(ADV)、前置宾语(FOB)、兼语(DBL)¹。由表1可知，“偷抢”类动词最典型的用法并不是作句子谓语，相反，该类动词最常用的句法功能是作定语和并列成分。以往研究中，都是对“偷抢”类动词作为句子核心的用法进行的分析，但是我们通过大规模语料的依存标注发现，“偷抢”类最常用的不是作为句子核心，而是作为定语，从这一点看，我们弥补了前人研究的不足。

	ADV	ATT	COO	DBL	FOB	HED	POB	SBV	VOB	总计
扒窃	1	25	12	0	0	3	7	3	13	64
	1.56%	39.06%	18.75%	0.00%	0.00%	4.69%	10.94%	4.69%	20.31%	100.00%
打劫	0	21	40	0	0	25	2	2	25	115
	0.00%	18.26%	34.78%	0.00%	0.00%	21.74%	1.74%	1.74%	21.74%	100.00%
盗窃	1	746	431	0	2	112	52	57	244	1645
	0.06%	45.35%	26.20%	0.00%	0.12%	6.81%	3.16%	3.47%	14.83%	100.00%
盗用	2	57	34	0	0	50	10	11	65	229
	0.87%	24.89%	14.85%	0.00%	0.00%	21.83%	4.37%	4.80%	28.38%	100.00%
劫持	1	801	93	0	1	281	19	26	345	1567
	0.06%	51.12%	5.93%	0.00%	0.06%	17.93%	1.21%	1.66%	22.02%	100.00%
劫夺	0	41	13	0	0	16	4	3	26	103
	0.00%	39.81%	12.62%	0.00%	0.00%	15.53%	3.88%	2.91%	25.24%	100.00%
劫掠	0	57	47	1	0	26	4	6	31	172
	0.00%	33.14%	27.33%	0.58%	0.00%	15.12%	2.33%	3.49%	18.02%	100.00%
抢夺	4	105	93	0	2	42	10	8	50	314
	1.27%	33.44%	29.62%	0.00%	0.64%	13.38%	3.18%	2.55%	15.92%	100.00%
抢劫	4	840	715	0	7	177	47	77	435	2302
	0.17%	36.49%	31.06%	0.00%	0.30%	7.69%	2.04%	3.34%	18.90%	100.00%
抢掠	1	20	50	0	0	19	0	4	18	112
	0.89%	17.86%	44.64%	0.00%	0.00%	16.96%	0.00%	3.57%	16.07%	100.00%
窃夺	0	2	1	0	0	0	0	0	1	4
	0.00%	50.00%	25.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	100.00%
窃取	1	144	65	0	2	82	26	14	167	501
	0.20%	28.74%	12.97%	0.00%	0.40%	16.37%	5.19%	2.79%	33.33%	100.00%
偷盗	0	161	92	0	0	48	16	30	52	399
	0.00%	40.35%	23.06%	0.00%	0.00%	12.03%	4.01%	7.52%	13.03%	100.00%
偷窃	2	153	95	0	2	29	24	26	102	433

¹ 本文仅在第一次提到时对使用的依存句法标签和依存语义标签同时注明中文和对应的英文标签。

	0.46%	35.33%	21.94%	0.00%	0.46%	6.70%	5.54%	6.00%	23.56%	100.00%
洗劫	0	31	25	0	0	29	3	3	82	173
	0.00%	17.92%	14.45%	0.00%	0.00%	16.76%	1.73%	1.73%	47.40%	100.00%
行劫	1	5	9	0	0	4	1	0	3	23
	4.35%	21.74%	39.13%	0.00%	0.00%	17.39%	4.35%	0.00%	13.04%	100.00%
总计	18	3209	1815	1	16	943	225	270	1659	8156
	0.22%	39.35%	22.25%	0.01%	0.20%	11.56%	2.76%	3.31%	20.34%	100.00%

表 1.“偷抢”类动词的句法功能分布（图中标红的比例是该动词各句法功能的最大占比）

从表 1 我们可以看出，“扒窃”、“盗窃”、“挟持”、“劫夺”、“劫掠”、“抢夺”、“抢劫”“窃夺”以及“偷窃”充当 ATT 句法成分的占比最多。“偷抢”类动词充当 DBL 和 FOB 的句法成分数量几乎没有。

不同的“偷抢”类动词在句法功能分布上有差异，但占比最多的都集中在定语、并和动宾三种句法功能上。为了比较“偷抢”类动词的句法功能相似性，我们计算了每个动词各自充当句法成分占自身单句数量的比例以及该类动词的句法功能占比，并以此画出“偷抢”类动词的句法功能占比热力图（如图 2 所示）。

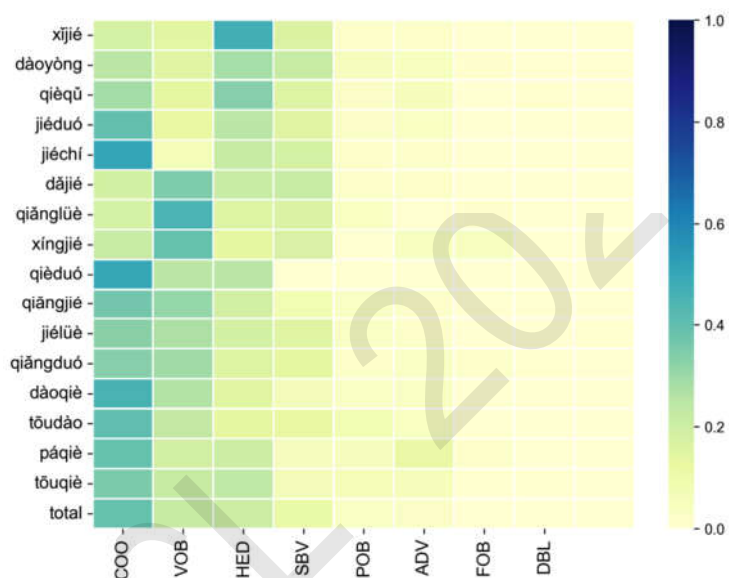


图 2.“偷抢”类动词句法功能热力图

由图 2 可知，“偷抢”类动词整体的句法功能分布十分相似，最接近该类整体句法功能分布（图 2 的 total 表示该类动词整体的句法功能分布）的动词是“偷窃”。根据其差异从大到小顺序又可细分为三小类，第一小类为“洗劫、盗用、窃取、劫夺、劫持”，第二小类为“打劫、抢掠、行劫”，第三小类为“窃夺、抢劫、劫掠、抢夺、盗窃、偷盗、扒窃、偷窃”。

为了比较不同动词小类的句法功能，我们又增加了“搜索”类动词（包括寻找、探索、窥探、探听、搜寻）（王珊 and 陈姝池，2021）和“视觉”类动词（即最常见的双音节动词“看见”）（王珊等，2021）的句法功能分布，并根据动词的句法功能分布对“偷抢”类动词、“搜索”类动词和“视觉”类动词进行聚类分析（见图 3）。

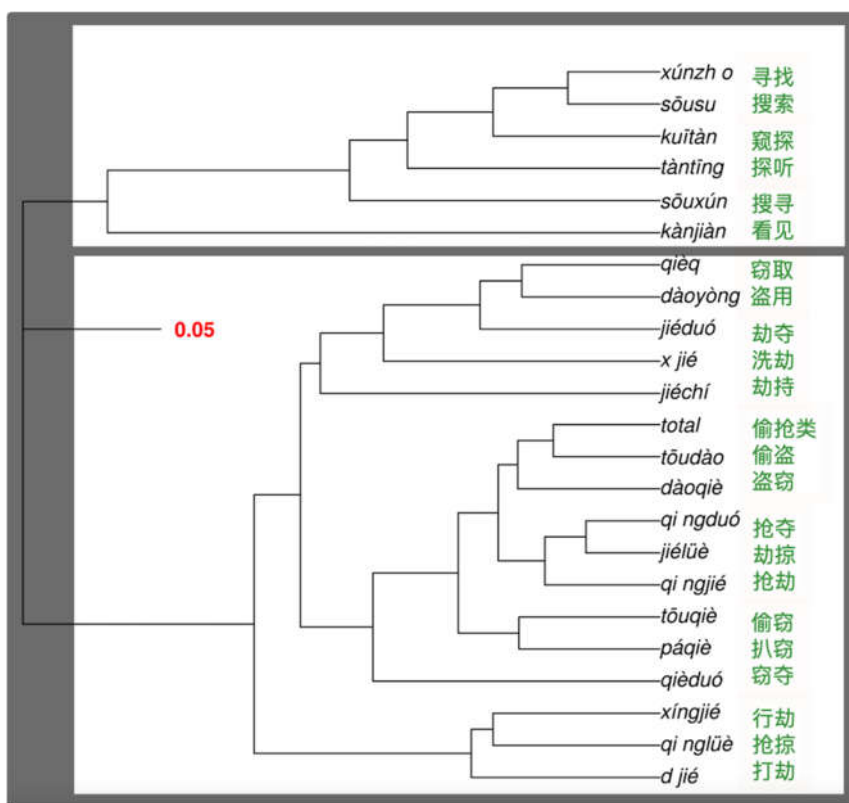


图 3.“偷抢”类、“搜寻”类与“视觉”类动词句法功能分布聚类图

图 3 使用了 R 语言的“phylogram”包进行聚类分析。从图 3 来看，聚类分析正好将“偷抢”类、“搜寻”类、“视觉”类动词区分来，其中“搜寻”类和“视觉”类动词更接近，即在这三类动词中，“搜寻”类和“视觉”类动词在句法功能分布上更相似，“偷抢”类动词与二者差异更大。

4.2 “偷抢”类动词的语义角色分布

对单句进行语义依存标注时，可以看到“偷抢”类动词在单句中也充当了不同的语义角色。一个词在一个句子中只能充当一种句法成分，但是该词担任的语义角色可能不止一个。统计表明，“偷抢”类动词充当最多的单个语义角色为并列事件（eCOO，1205 次，见例 5），其次是根（Root，1058 次，见例 6），最多的双个语义角色为修饰成分和修饰成分（FEAT，76 次，见例 7），即同时作为两个事物或动作的修饰成分，其次是修饰成分和当事（EXP）（61 次，见例 8）。

例 5：暴徒在市区滋事、捣毁门窗、打劫商家。

例 6：途中车上一名年轻男子扒窃乘客钱包。

例 7：两人还交待了近两年来在通化火车站和列车上实施的其它盗窃犯罪行为。

例 8：盗窃者是第二医院的监委王清。

4.3 “偷抢”类动词的句法依存

此处我们按照该类动词充当的不同句法成分，对“偷抢”类动词的句法同现模式进行分析。因为该类动词充当主语、介宾、状语、前置宾语、兼语等句法成分的比例均不足 5%，因此不予分析。下面仅对“偷抢”类动词在单句中充当定语、并列、动宾或谓语这四种句法成分时的句法同现模式进行分析。

当“偷抢”类动词充当定语时，和其同现的最多的四种句法成分是右附加成分（RAD）、动宾、状语、并列成分，最常用的是“偷抢”类动词单独作为定语来修饰其他成分，其次最常见的句法同现模式是“偷抢”类动词后接动词宾语以及一个右附加成分，这个整体作定语修饰其他成分，如例 9 中“木材”是“盗窃”的动词宾语，“的”是右附加成分，“盗窃木材的”作为定语修饰“人”。

例 9: 我们已经查明了盗窃木材的人。

当“偷抢”类动词充当并列成分时, 和其同现的最多的四种句法成分是顿号、动宾、左附加成分(LAD)、状语, 最常用的是“偷抢”类动词作为并列成分与其他并列成分通过标点符号相连, 如例 10 中“盗窃”与“腐化”通过顿号连接。

例 10: 中国人民银行江苏省江阴支行出纳股长孙承毅贪污、腐化、盗窃国家财产。

当“偷抢”类动词充当动宾时, 和其同现的最多的四种句法成分是动宾、状语、并列、右附加成分, 最常用的是“偷抢”类动词单独作动词的宾语, 其次最常见的句法同现模式是“偷抢”类动词后接动词宾语, 由此构成的“偷抢”类动宾结构作为另一动词的宾语成分, 如例 11 中的“普拉多博物馆的藏画”是“盗窃”的动词宾语, “盗窃普拉多博物馆的藏画”整体作该句核心动词“企图”的宾语。

例 11: 她企图盗窃普拉多博物馆的藏画。

当“偷抢”类动词充当谓语时, 和其同现的最多的四种句法成分是状语、动宾、定语、主语、前置宾语, 最常用的是“偷抢”类动词作为谓语与主语、状语、宾语同现, 如例 12 中“抢劫”是句子谓语, “他们”是主语, “肆意”是状语, “外国人”是宾语。

例 12: 他们肆意抢劫外国人。

4.4 “偷抢”类动词的语义依存

4.4.1 “偷抢”类动词的语义密度

根据依存语法, 语义依存可分为主体角色、客体角色、情境角色、语义结构关系和语义依附标记。其中语义周边论元包括主体角色、客体角色和情境角色, 语义结构关系包括反关系、嵌套事件关系和事件关系, 语义依附标记包括标点标记、关系标记、否定标记和依附标记。因为语义依附标记对句子语义的贡献相对最弱, 因此本文对语义依附标记不予分析。

由于和动词有语义依存关系的语义角色一个到多个不等, 且经常出现多个的情况, 我们对“偷抢”类动词的语义依存进行归一化处理, 并将其命名为该(类)动词的语义依存密度 D , 其计算公式为:

$$D_{\text{动词语义依存密度}} = (X_{\text{和动词直接相关的语义依存数量}} - Y_{\text{mPUNC 数量}}) \div S_{\text{动词单句数量}} \quad (1)$$

其中 X 表示和动词直接相关的语义依存数量。根据依存语法的标注规范, 和动词直接相关的语义依存也包含动词与标点符号之间的关系, 因此在计算动词的语义依存密度 D (简称 $D_{\text{动}}$) 时, 需要减去被标注为 mPUNC (标点标记) 的语义依存数量 Y 。

由此, 我们得到“偷抢”类的动词语义依存密度 (见表 2)。

	扒窃	打劫	盗窃	盗用	劫持	劫夺	劫掠	抢夺	抢劫	抢掠	抢夺	窃取	偷盗	偷窃	洗劫	行劫	偷抢类
X	79	285	2837	672	4664	303	387	751	4123	223	19	1446	736	768	417	49	17759
Y	20	57	427	63	446	28	42	120	618	37	2	148	122	110	52	4	2296
X-Y	59	228	2410	609	4218	275	345	631	3505	186	17	1298	614	658	365	45	15463
S	64	115	1645	229	1567	103	172	314	2302	112	4	501	399	433	173	23	8156
$D_{\text{动}}$	0.92	1.98	1.47	2.66	2.69	2.67	2.01	2.01	1.52	1.66	4.25	2.59	1.54	1.52	2.11	1.96	1.90

表 2.“偷抢”类动词语义依存密度 (标红的是密度最大的值, 标绿的是密度最小的值)

动词语义依存密度越高, 该动词的语义依存数量越多, 其语义搭配能力越强。由表 2 可知, “偷抢”类动词的整体语义依存密度为 1.90, 语义搭配能力最强最突出的动词是“窃夺”, 比搭配能力排在第二的动词“劫持”足足高了 1.56, 最弱的是“扒窃”。除去最高的“窃夺”和最低的“扒窃”, 其他动词的语义依存密度维持在 1.5-2.7 范围。

4.4.2 “偷抢”类动词的主体角色和客体角色

主体角色和客体角色是和动词相关的最重要的语义角色, 我们对“偷抢”类动词的主体角色和客体角色进行分析, 统计结果见表 3。

		扒窃	打劫	盗窃	盗用	劫持	劫夺	劫掠	抢夺	抢劫	抢掠	窃夺	窃取	偷盗	偷窃	洗劫	行劫	总计
主体角色	AGT	7	52	224	93	615	60	67	93	473	36	4	227	72	89	90	6	2208
	EXP	0	1	1	2	3	1	0	1	1	5	0	5	1	1	4	0	26
	总计	7	53	225	95	618	61	67	94	474	41	4	232	73	90	94	6	2234
客体角色	PAT	6	40	643	210	960	79	63	168	538	22	3	372	142	139	88	0	3473
	CONT	1	0	11	12	2	2	0	10	7	3	0	77	1	14	0	0	140
	DATV	0	2	14	3	17	2	5	21	54	0	0	39	2	5	2	0	166
	LINK	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	3
	总计	7	42	668	225	981	83	68	199	600	25	3	488	145	158	90	0	3782
总计		14	95	893	320	1599	144	135	293	1074	66	7	720	218	248	184	6	6016

表 3.“偷抢”类动词的主体角色和客体角色分布

由表 3 可知，“偷抢类”动词大多数是以客体角色为主形式存在，“偷抢”类动词的主体角色中，施事（AGT）远超过当事（EXP），这说明偷抢行为强调的是实施偷抢动作的人。“偷抢”类动词的客体角色中，受事（PAT）占绝大多数，其次是涉事（DATV）和客事（CONT），而系事（LINK）几乎没有，这说明偷抢行为产生的后果更多地会影响到受偷抢的个人或者团体，而偷抢行为的具体内容占据少数。

按照动词语义依存密度的计算方法，我们对“偷抢”类动词的主体角色和客体角色进行语义密度计算，计算公式如下：

$$D_{\text{主体语义密度}} = \frac{X_{\text{和动词直接相关的主体角色数量}}}{S_{\text{动词单句数量}}} \quad (2)$$

$$D_{\text{客体语义密度}} = \frac{X_{\text{和动词直接相关的客体角色数量}}}{S_{\text{动词单句数量}}} \quad (3)$$

“偷抢”类动词的主体语义密度和客体语义密度如表 4 所示，其中 $X_{\text{主}}$ 和 $X_{\text{客}}$ 分别表示和动词直接相关的主体角色数量和客体角色数量， $D_{\text{主}}$ 和 $D_{\text{客}}$ 分别表示动词的主体语义密度和客体语义密度。

	扒窃	打劫	盗窃	盗用	劫持	劫夺	劫掠	抢夺	抢劫	抢掠	窃夺	窃取	偷盗	偷窃	洗劫	行劫	总计
$X_{\text{主}}$	7	53	225	95	618	61	67	94	474	41	4	232	73	90	94	6	2234
$X_{\text{客}}$	7	42	668	225	981	83	68	199	600	25	3	488	145	158	90	0	3782
单句数量	64	115	1645	229	1567	103	172	314	2302	112	4	501	399	433	173	23	8156
$D_{\text{主}}$	0.11	0.46	0.14	0.41	0.39	0.59	0.39	0.30	0.21	0.37	1.00	0.46	0.18	0.21	0.54	0.26	0.27
$D_{\text{客}}$	0.11	0.37	0.41	0.98	0.63	0.81	0.40	0.63	0.26	0.22	0.75	0.97	0.36	0.36	0.52	0.00	0.46

表 4.“偷抢”类动词的主体语义密度和客体语义密度（标红的是密度最大的值，标绿的是密度最小的值）

比较“偷抢”类动词的主体语义密度和客体语义密度，我们发现整体来看，该类动词的主体语义密度低于客体语义密度；小类中，主体语义密度高于客体语义密度的动词有“行劫、窃夺、抢掠、打劫、洗劫”。主体语义密度和客体语义密度最大的分别是“窃夺”（1.00）和“盗用”（0.98），最小的分别是“扒窃”（0.11）和“行劫”（0.00）。两指数相差最大的是“盗用”（0.57），最接近的是“扒窃”（0.00）。

4.4.3 “偷抢”类动词的情境角色

情境角色也是语义周边论元。我们对“偷抢”类动词的情境角色统计如表 5。

	扒窃	打劫	盗窃	盗用	劫持	劫夺	劫掠	抢夺	抢劫	抢掠	窃夺	窃取	偷盗	偷窃	洗劫	行劫	总计
TOOL	0	1	5	2	14	3	0	5	6	0	0	10	1	2	6	2	57

MATL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MANN	4	14	96	15	46	18	16	19	153	9	0	56	10	19	10	3	488
SCO	1	1	7	0	10	1	5	1	16	0	0	1	4	2	0	2	51
REAS	0	1	3	6	3	0	0	1	6	0	0	4	0	0	1	0	25
TIME	0	8	41	5	332	5	7	8	85	6	1	30	6	10	18	4	566
LOC	7	11	70	8	377	10	11	24	155	11	0	38	25	20	8	6	781
MEAS	1	3	88	7	23	0	8	11	120	0	0	7	8	14	4	1	295
STAT	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	2
FEAT	3	1	77	9	33	1	9	6	146	11	0	2	11	16	2	2	329
总计	16	40	387	52	838	38	56	75	688	37	1	149	65	83	49	20	2594

表 5.“偷抢”类动词的情境角色分布

由表 5 可知，“盗窃”、“挟持”、“抢劫”、“窃取”最常用情景角色都是方式角色（MANN）、时间角色（TIME）、地点角色（LOC）、度量角色（MEAS）以及修饰角色（FEAT）。从总计上来看，“偷抢”类动词的最常用情境角色是地点角色（LOC），其次是时间角色（TIME）和方式角色（MANN）。这说明在偷抢行为中，地点、时间和方式是描述偷抢事件的重要情境因素。没有涉及的情境角色是材料角色（MATL），状态角色（STAT）也较少提及。

4.4.4 “偷抢”类动词的事件关系

“偷抢”类动词的语义结构关系中，反关系和嵌套事件关系都非常少。下面我们对这类动词的事件关系进行分析，其分布如表 6。

	扒窃	打劫	盗窃	盗用	劫持	劫夺	劫掠	抢夺	抢劫	抢掠	窃夺	窃取	偷盗	偷窃	洗劫	行劫	总计
eCOO	10	14	364	31	47	11	46	58	560	25	1	45	98	82	13	2	1407
ePREC	0	0	1	0	1	0	0	0	1	2	0	0	0	1	1	0	7
eSUCC	3	3	72	45	141	6	5	8	115	9	0	30	18	11	2	3	471
总计	13	17	437	76	189	17	51	66	676	36	1	75	116	94	16	5	1885

表 6.“偷抢”类动词的事件关系

并列事件（eCOO）表示偷抢和其他行为同时发生（见例 13），先行事件（ePREC）表示其他行为发生在偷抢行为之前（见例 14），后继事件（eSUCC）表示其他行为发生在偷抢行为之后（见例 15）。由表 6 可知，“偷抢”类动词最常见的事件关系是并列事件，因此我们可以判断，偷抢行为通常还会和其他如“残害”、“扣押”等行为同时发生。从偷抢和其他行为发生的先后顺序来看，偷抢行为先发生的概率是其后发生概率的 67 倍（471 句/7 句），即往往先有偷抢行为的出现，然后才出现其他行为。

例 13：这些部队劫掠和残害缅甸居民。

例 14：数万人因不满政府提高食品价格而洗劫商店和袭击警察。

例 15：有些青年还劫夺汽车来设置路障。

4.5 “偷抢”类动词的句法语义

首先我们总结了“偷抢”类动词充当不同句法成分时，各自常用的语义同现模式。和分析该类动词的句法分布一样，因为该类动词充当主语、介宾、状语、前置宾语、兼语等句法成分的比例均不足 5%，因此我们只考虑其作为定语、并列、动宾或谓语这四种句法成分时的语义同现模式。

当“偷抢”类动词充当定语时，该类动词最常见的语义角色是修饰成分，最常用的语义同现模式是该类动词+受事，如例 16 中“客车”是“抢劫”的受事，“客车抢劫”作为定语修饰“案件”。

例 16：他们已经连续两年杜绝了客车抢劫案件。

当“偷抢”类动词充当并列成分时，该类动词最常见的语义角色是并列事件。最常用的语义同现模式是该类动词+并列事件。也就是说，在该类动词充当并列成分时，最常见的语义同现模

式是三个并列事件，并且偷抢行为是第二个并列事件。如例 17 中三个并列事件分别为“破坏”、“抢劫”和“焚烧”，且“抢劫”出现在第二位。

例 17：一些公共建筑、商店和车辆遭到破坏、抢劫和焚烧。

当“偷抢”类动词充当动宾时，该类动词最常见的语义角色是降级客事，即偷抢事件整体作为另一个动词的宾语，如例 18，偷抢事件是“要抢劫城市和您的王宫”，这一偷抢事件是“发誓”的宾语。最常用的语义同现模式是该类动词+受事，这一整体共同作为另一动词的宾语，如例 19 中“金子”是“抢劫”的受事，而“抢劫金子”作为句子核心动词“打算”的宾语。

例 18：他们共同发誓要抢劫城市和您的王宫。

例 19：他们压根就没打算抢劫金子。

943 句“偷抢”类动词充当句子核心的单句中，930 句该类动词的语义依存标注为根（Root），对于该类动词的句法语义标注一致性为 98.6%（930 句/940 句）。接下来我们对“偷抢”类动词作为谓语的句法语义进行详细分析，此处仅选择句法语义标注一致的 930 条单句为研究对象。

我们根据单句是否含有主语和宾语对“偷抢”类动词的句型进行分类，并列举出每一种句型最常见的语义搭配情况（仅对频次高于 10 的句型总结最常见的语义搭配），如表 7 所示。注意，表格中“+”表示具有这一成分，“-”表示不具有这一成分。

SBV	VOB	FOB	数量	有两个及以上的事件的单句数量	最常见的语义搭配模式	例句
+	+	-	385	97	施事+动词+受事	李承晚要以武力 <u>劫夺</u> 战俘。
-	-	+	219	50	受事+动词	胜利的果实会不会被 <u>盗窃</u> ？
-	-	-	117	19	动词	不要 <u>抢夺</u> ！
-	+	-	109	16	动词+受事	不许 <u>盗窃</u> 国家财产！
+	-	-	94	34	施事+动词	吴某为何 <u>打劫</u> ？
-	+	+	5	1	-	多达一千人在这个剧院内被 <u>劫持</u> 为人质。
+	+	+	1	0	-	其义则丘 <u>窃取</u> 之矣。
+	-	+	0	0	-	-

表 7.“偷抢”类动词充当 HED 的句法语义

由表 7 可知，在 930 条“偷抢”类动词做句子谓语的单句中，从事件数量上来看，217 句为两个或两个以上的事件。从句型上来看，最常用的是主语和宾语都有，占 41.40%（385 句/930 句），最常见的语义搭配模式是动词有施事和受事，即在偷抢事件中，最常见的是实施偷抢动作的人对受事进行偷抢行为。

5 结论

本文建立了汉语“偷抢”类动词单句语料库，以依存语法对“偷抢”类动词进行句法依存和语义依存标注，对该类动词在单句使用中的句法语义分布进行了定量统计，并从框架语义学的角度归纳了汉语“偷抢”类动词的句法语义特点。本研究的创新之处在于第一次对汉语“偷抢”类动词进行大规模地依存语法的定量分析，并说明了通过依存语法自动标注，从动词的句法功能进行动词类别划分的可能性。我们还创造性地提出用“动词语义依存密度”来测量动词的语义搭配能力。具体来说，1）汉语“偷抢”类动词展现出了多种句法功能，其中最常见的是定语、并列、动宾和谓语，并且此类动词具有一定的句法功能相似性，并且可以通过其句法功能分布区分不同的动词小类；2）汉语“偷抢”类动词的语义角色分布具有多样性；3）汉语“偷抢”类动词的句法功能不同时，和句子其他成分之间的句法依存也不同；4）从“偷抢”类动词的语义依存来看，该类动词的客体语义密度整体低于主体语义密度，最常见的情境角色是地点和时间，在事件关系中，并列事件发生概率最高；5）“偷抢”类动词的句法语义特点丰富，主要的句型为主谓宾句式，而该句式中最常用的语义搭配模式是施事对受事实施偷抢

动作。本研究结合依存语法和框架语义学，深化了对汉语“偷抢”类动词的句法、语义和事件关系的了解，促进了对偷抢类动词的研究。

参考文献

- Beavers John and Koontz-Garboden Andrew. 2017. Result verbs, scalar change, and the typology of motion verbs. *Language*, 93(4), 842-876.
- Beavers John, Everdell Michael, Jerro Kyle, Kauhanen Henri, Koontz-Garboden Andrew, LeBovidge Elise, and Nichols Stephen. 2017. Two types of states: A cross-linguistic study of change-of-state verb roots. *Proceedings of the Linguistic Society of America*, 2, 38-1.
- Biber Douglas, Johanson Stig, Leech Geoffrey, Conrad Susan, and Finegan Edward. 1999. *Longman Grammar of Spoken and Written English*. Longman, London, UK.
- Che, Wanxiang, Zhenghua Li, Yongqiang Li, Yuhang Guo, Bing Qin, and Ting Liu. 2009. Multilingual dependency-based syntactic and semantic parsing. In *CoNLL 2009*, pages 49–54, Boulder, Colorado, June.
- Dux Ryan. 2018. Frames, verbs, and constructions: German constructions with verbs of stealing. *Approaching German Syntax from a Constructionist Perspective*, 367-405.
- Fernández-Martínez J. Nicolás and Faber Pamela. 2020. Who stole what from whom?: A corpus-based, cross-linguistic study of English and Spanish verbs of stealing. *Languages in Contrast*, 20(1), 107-140.
- Goldberg E. Adele. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Hajic, Jan, Massimiliano Ciaramita, Richard Johans-son, Daisuke Kawahara, Maria Antonia Mart`ı, Llu`ıs Marquez, Adam Meyers, Joakim Nivre, Sebastian Pado, Jan St`ep`anek, Pavel Stra`n`ak, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *CoNLL 2009*, pages 1–18, Boulder, Colorado, June.
- Huang Chu-Ren, Dong Sicong, Yang Yike, and Ren He. 2021. From Language to meteorology: Kinesis in weather events and weather verbs across Sinitic languages. *Humanities and Social Sciences Communications*, 8(1), 1-13.
- Levin Beth and Hovav M. Rappaport. 1991. Wiping the slate clean: A lexical semantic exploration. *Cognition*, 41(1-3), 123-151.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, University of Chicago press.
- Lien Chinfa. 2006. <<荔鏡記>>動詞分類和動相、格式. *Language and Linguistics*, 7(1): 27-61.
- Liu Haitao. 2008. The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and its Applications*, 387(12), 3048-3058.
- Liu Meichun and Chang Juiching. 2019. From caused-motion to spatial configuration: Placement verbs in Mandarin. *Language and Linguistics*, 20(2), 180-224.
- Liu Meichun. 2002. Verbs of surface contact in Mandarin: A lexical semantic study. In Su, Lily I-wen & Lien, Chinfa & Chui, Kawai (eds.), *Form and function: Linguistic studies in honor of Shuanfan Huang*, 275-303.
- Lafferty, John, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML 2001*, pages 282–289. Morgan Kaufmann, San Francisco, CA.
- Niu Ruo Chen and Osborne Timothy. 2019. Chunks are components: A dependency grammar approach to the syntactic structure of Mandarin. *Lingua*, 224, 60-83.
- Ralph B. Long. 1966. The English Verb: A Traditional View. *College Composition and Communication*, 17(2), 100-106.

- Vendler Zeno. 1967. *Linguistics in Philosophy*. Cornell University Press, New York.
- Yiqun Liu, Fei Chen, Weize Kong, Huijia Yu, Min Zhang, Shaoping Ma, and Liyun Ru. March 2012. Identifying Web Spam with the Wisdom of the Crowds. *ACM Transaction on the Web*. Volume 6, Issue 1, Article No. 2, 30 pages.
- 范晓, 胡裕树. 1995. 动词研究. 郑州:河南大学出版社.
- 亢世勇 (2015). 新编同义词词林. 上海, 上海辞书出版社.
- 课题组 (2008). 现代汉语常用词表. 北京, 商务印书馆.
- 李辉, 刘海涛. 2017. 基于句法标注语料库的汉语儿童三词句习得研究. 语言文字应用, 1.
- 李正华. 2013. 汉语依存句法分析关键技术研究 (Doctoral dissertation, 哈尔滨工业大学).
- 刘海涛, 冯志伟. 2007. 自然语言处理的概率配价模式理论. 语言科学, 32-41.
- 刘挺, 车万翔, 李正华. 2011. 语言技术平台. 中文信息学报, 25:53-63.
- 陆丙甫. 1998. 从语义、语用看语法形式的实质. 中国语文, 5.
- 沈家煊. 2000. 说“偷”和“抢”. 语言教学与研究, (01):19-24.
- 苏新春, Ed. (2013). 现代汉语分类词典. 北京, 商务印书馆.
- 王珊, 陈姝池. (2021). 搜寻类动词的句法语义研究. 第 22 届汉语词汇语义学研讨会, 南京师范大学.
- 王珊, 刘晓骏, 周洁. (2021). 基于依存语法的视觉动词“看见”研究. 第 22 届汉语词汇语义学研讨会, 南京师范大学.
- 王珊, 汤蕾. (2021). 基于依存语法的招认类动词研究. 第 22 届汉语词汇语义学研讨会, 南京师范大学.
- 王珊, 周洁. (2021). 汉语学习词典的句法语义信息: 以抱怨类动词为例. 教育语言学专业委员会第十二届年会, 上海交通大学.
- 王寅. 2011. 构式语法研究 (上下卷). 上海: 上海外语教育出版社.
- 张国宪. 2001. 制约夺事成分句位实现的语义因素. 中国语文, 6.
- 周洁, 王珊. (2021). 许诺类动词用法探究. 语言认知科学国际学术研讨会, 北京语言大学.

附录 A. “偷抢”类单句语料库分布

关键词	文段数量	筛选后的单句总数	语料库占比
扒窃	1330	64	0.78%
打劫	3464	115	1.41%
盗劫	710	0	0.00%
盗窃	30844	1645	20.17%
盗用	4425	229	2.81%
劫持	11831	1567	19.21%
劫夺	1782	103	1.26%
劫掠	2570	172	2.11%
抢夺	6548	314	3.85%
抢劫	32019	2302	28.22%
抢掠	3111	112	1.37%
窃夺	231	4	0.05%
窃取	9167	501	6.14%
偷盗	6693	399	4.89%
偷窃	6693	433	5.31%

洗劫	3775	173	2.12%
行劫	565	23	0.28%
行窃	2157	0	0.00%
总计	127915	8156	100%

JCL 2021