

施事的语义分布考察与动词的语义特征*

陈龙¹, 詹卫东^{1,2,3}

(1 北京大学中国语言文学系, 2 北京大学中国语言学研究中心, 3 北京大学计算语言学教育部重点实验室, 北京 100871)

摘要: 在自然语言处理中, 语义角色标注是一个比较重要的任务。语义角色标注任务以语义角色理论为基础。语义角色标注任务和语言学的相关理论都要求一个动词在不同语句中相应的语义角色成分保持一致。本文统计了北京大学 973 课题语义角色标注语料库及中文 PropBank 中“施事”的语义分布, 进而根据该语义分布, 加上人工筛选, 找出了在“施事”语义角色的标注上无法保持一致的动词共 85 个。本文指出, 这些动词的“施事”语义无法保持一致的原因在于它们是一类有致使语义特征, 无自主语义特征的动词。本文的研究对于处理其他语义角色(如“客体”)也存在的标注不一致问题有一定的启示。

关键词: 语义角色标注 语料库 施事 一致性 自主性

An Investigation of Inconsistency in Semantic Role Labeling: A Case Study of Agent

CHEN Long, ZHAN Weidong

(Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

Abstract: Semantic role labeling, which is based on the theories of semantic roles, is an important task in natural language processing. In semantic role labeling, it is required that the semantic role of a verb should be consistent in different sentences. We analyzed the semantic distribution of agents in the semantic role labeling corpus. Based on the distribution, we found out 85 verbs that can't meet the requirement of consistency. We point out that these verbs can't meet the requirement of consistency because they belong to a class of verbs which do not have the semantic feature of volitive verbs, but have the feature of causative verbs. The data and the method can help deal with similar problems in semantic role labeling.

Keywords: Semantic role labeling; Corpus; Agent; Consistency; Volitive

1 引言

语义角色标注是自然语言处理中比较重要的任务之一, 它有助于机器理解自然语言的语义, 对自然语言处理领域的很多其他任务, 如机器翻译、信息抽取等都有帮助^[1]。语义角色标注工作以语言学中的语义角色理论为基础。相关理论中涉及语义角色(semantic role)、论元角色(argument)、语义格(deep semantic case)、题元角色(thematic role)等名词, 它们指称的概念基本相同, 本文中我们使用语义角色的称谓。

计算机的语义角色标注任务需要一个语义角色的列表。从 Fillmore 的格语法理论^[2]起, 大量的中外学者基于语言学理论或者面向语义角色标注任务构建知识库, 提出了英语和汉语的语义角色系统。

在英语的相关理论中, Fillmore 的格理论列出了 6 个基本的格, 包括 Agentive (施事), Instrumental (工具), Locative (处所)等等^[2]; Dowty 提出原型施事受事理论, 将语义角色分为原型施事和原型受事两类^[3]。在英语的相关知识库中, VerbNet 设立了 23 个语义角色, 包括 Agent, Patient, Experiencer, Theme, Material, Instrument 等等^[4]。PropBank 用 Arg0, Arg1, ..., Arg6 来代表动词的语义角色, Arg0 相当于原型施事的角色, Arg1 相当于原型受事的角色, 其余的相当于外围的语义角色^[5]。

在汉语的相关理论中, 鲁川提出的语义角色分为 26 个小类, 包括施事、当事、领事、对象等^[6]。袁毓

*收稿日期: 定稿日期:

基金项目: 国家重点基础研究发展计划(973 项目)(2014CB340504)、教育部人文社科基地 2015 年重大项目(15JJD740002)

林设立了四级的论元层级体系，包括施事、感事、致事和主事这四个主体论元，以及受事、与事等 17 个论元角色^[7]，其中“施事”满足“不~”和“没有~”两个语法判别格式，其余主体角色只能满足其中一个格式或者两个都不满足。在汉语的语义角色标注系统和相关的知识库中，袁毓林的“北大中文网库”中包括了施事、与事、材料等 21 个语义角色标签，包括施事、感事、经事、致事、主事这 5 个主体论元^[8]；北京大学国家重点基础研究发展计划（973）项目：“融合三元空间的中文语言知识与世界知识获取和组织”，项目编号 2014CB340504 中涉及的语义角色标注的标注规范（内部资料，以下简称为 973 语义角色标注规范）包括施事、当事、受事、结果、工具等 26 个论旨角色，其中施事和当事是主体^[9]，“施事”满足“去~”的语法判别格式，而“当事”不满足。

袁毓林^[8]将语义角色分类系统按精细程度分为微观、中观、宏观三个层级。上述语义角色系统中，Dowty 的原型施事受事理论及 PropBank 的 Arg0 的标注方式属于宏观层级，Framenet 属于微观层级，其余的大多数处在中观层级。现在的语义角色标注也基本采用中观或宏观层级的标签来进行标注。

上述的语义角色的设置各有合理性和优点，但同时也存在着普遍性的问题。例如袁毓林^[8]指出，在语义角色标注工作中，衡量标注质量的最重要指标是角色标记的一致性，这个一致性的最低标准是“在不同的语句实例中，相同的谓词及谓词性短语的相应的各个论元成分的论旨角色应该一致”。例如对于动词“睡觉”，在不同的话语中，它的主体都符合施事的定义，可以一致地标为施事。其他的语义理论和知识库即使没有明确指出这一点，但实际上也都隐含或默认了这一要求。例如论元结构理论认为，一个动词的一个义项有它的论元结构，这个论元结构是它内在的词汇属性^[10]，如“学习”的论元结构是[施事，受事]。如果一个动词的一个义项的主体语义角色不是一致的或固定的，会随着动词所在的句子而变化，那就不能认为论元结构是动词本身的属性，而且不得不说该动词的某个义项有多个论元结构。袁毓林^[7]的方案中提到，四个主体角色是不同现的，没有对立的值。由此可以推断，一个动词的主体角色在不同话语中都是一致的。VerbNet 为每个动词标注了语义角色的框架，也是默认了动词的语义角色的一致性。

对于语义角色标注任务而言，语义角色的一致性可以降低标注的难度。计算机只需识别出一个成分是动词的主体，即可根据一致性，正确地标注出该成分具体是哪个角色。而在语言学上，语义角色的一致性保证了一个动词的一个义项的语义角色是由这个动词本身决定的，不会受到上下文语境的影响。那么动词的语义角色就可以在一定程度上准确地反映动词的语义，区分开语义特征不同的动词。

但是，我们发现，在动词的实际用例中，要贯彻这种语义角色标注一致性的原则，并不容易。比如，以下三个例句中，“改变”的主体是否应该标为同一个语义角色，可能是存在争议的。

- (1) 他改变了他的志向。
- (2) 这件事改变了他的志向。
- (3) 他改变了中国。

在中观层级上，这三个例句中，“这件事”和“他”是“改变”的必有论元，因此标注为原因这一外围论元并不合适。根据上文列出的袁毓林^[7]和 973 语义角色标注规范的语法判别格式，例句（1）中“他”完全符合施事的性质，而例句（2）中“这件事”更符合主事和当事的语法性质。对于例句（3），“他去改变中国”、“他不改变中国”从语感上难以判断是否合法，因此很难确定（3）中“他”是否是施事。

从语义上看，袁毓林^[7]指出，施事具有“使动性”，主事具有“变化性”，那么三个例句中“这件事”和“他”更接近施事的定义。973 语义角色标注规范^[9]则认为，典型的施事有主观意愿性，典型的当事无主观意愿性。从这一角度看，（2）中的“这件事”更接近主事或当事，（3）中“他”的语义角色仍然难以判断，既有可能是“他”自主地去“改变中国”，也有可能是“他”并没有自主的“改变中国”的意愿，但“他”的一些行为客观上“改变了中国”。由此可见，同一个谓词的语义角色在不同实例中保持一致，这一要求在中等颗粒度的语义角色系统中并不容易达到。

本文将主要考察同一个谓词的施事不一致的问题，根据语义角色标注语料库，统计动词施事的语义分布，得到存在施事角色内部不一致现象的动词，并进一步分析产生不一致现象的原因。

2 动词施事不一致的具体表现

我们观察到，按照各语义角色系统对施事和其他主体的定义，有一批动词的主体实际上很难满足语义角色的一致性。除上面提到的“改变”的例子，还有像“破坏”、“证明”和“限制”等不少例子也有类似情形。

- (4a) 他破坏了他的家乡的自然环境。
- (4b) 地震破坏了他的家乡的自然环境。
- (5a) 马克思证明了社会主义制度的优越性。
- (5b) 事实证明社会主义制度的优越性。
- (6a) 防守队员限制了他的发挥。
- (6b) 电子商务不受地域、国别限制。

例句(4a)中“他”显然有[+自主]、[+致使]、[+施动]等语义特征，通过句法判别标准也可确定“他”是“破坏”的施事。而(4b)中的“地震”并没有[+自主]、[+施动]的特征，只有[+致使]的特征，因而更接近袁毓林^[7]的主事、973 语义角色标注规范^[9]的当事等角色。将(4b)中的“地震”视作“破坏”的原因、工具角色也不合适，因为原因、工具只是外围语义角色，例句(4b)中如果“地震”是“破坏”的原因或工具，那么“破坏”就没有主体这一必有角色。而且“原因”角色语义应该是主体进行“破坏”行为的原因，“工具”是主体进行“破坏”所使用的工具，显然也和“地震”的语义不符。因此(4b)中“地震”是“破坏”的主体而非外围论元。其余两组例句也是同理。

如果认为“破坏”等动词每个都分别有两个义项，一个是有[+自主]语义特征的义项，另一个是[-自主]的，那么语义角色标注的不一致性看起来就不存在了，因为有自主性的义项的主体是施事，没有自主性的主体是当事。动词“说明”在现代汉语词典中有两个义项，一个是“解释明白”，是有自主性的，另一个是“证明”，对应的是“证明”的无自主性的情况。

但从动词的语义来看，把这些动词看做有两个义项并不合适。马庆株^[11]指出确定动词义项常用而有效的办法是看宾语，而如例句(4a)和(4b)所示，“破坏”在主体是自主和非自主的情况下都可以带相同的宾语，因而没有必要分为两个义项。而且即使分为两个义项，那么这两个义项之间的差别也仅仅是是否有自主性。有大量的、成批的动词会出现这样的语义差别，说明这样的差别是系统性的，不宜采用为每个动词增设义项的办法处理，而应该认为施事的不一致性是客观存在于这批动词上的。¹

从大规模真实语料的语义角色标注实践中也可以观察到上述不一致现象。

基于 973 语义角色标注规范^[9]，何保荣^[14]开展了中文语义角色标注语料库的建设，得到了规模为 20000 句的语义角色标注语料库，语料来源为人民日报。语料库中包含了 9168 个谓词（包括动词和形容词）及它们在句子中的语义角色。该语料库以下简称为人民日报语料库。

人民日报语料库对动词主体的判定借助了《知网》的语义角色系统来确定动词的语义角色。《知网》^[13]是董振东建设的中英文知识库，通过义原来定义词的义项。《知网》的义原分为事件、实体、属性、属性值、语义角色等 10 大类，每个的义原可能有若干下位的义原，由此构成了义原的树状结构。例如“Entity|实体”义原下位有“thing|万物”、“time|时间”、“space|空间”和“component|部分”这四个义原，“thing|万物”下位又有“physical|物质”、“mental|精神”和“fact|事情”三个义原，它们的下位有分别有一层或多层义原。例如在《知网》中动词“证明”属于义原“prove|证明”，该义原在《知网》的事件分类体系中表示为{prove|证明}{MakeOwnKnowledge|使自我感知:agent={*},content={*}}，即该义原是“MakeOwnKnowledge|使自我感知”的下位义原，该义原下的动词有 agent 和 content 两个必有语义角色，即“证明”的主体是施事。

在这种标注方式下，语料库中“证明”的主体都被统一为了施事，虽然满足了一致性的要求，但是并

¹ 也存在这样的情况：有的动词语义角色不同，对应着可以区分为不同的义项。比如“说明”在《现代汉语词典》中有两个义项，一个是“解释明白”，是有自主性的，另一个是“证明”，对应的是“证明”的无自主性的情况。“说明”之所以能区分出两个动词义项，是因为第一个义项的“说明”中有“说”的意义（可能是字面的，也可能是隐喻的），而第二个义项中“说”的意义基本消失。像这样的差别只在个别的动词上出现，没有系统性。

不符合“事实证明”中“事实”的实际语义。所以在这种标注方式下，本质上依然存在施事不一致现象。

中文 PropBank 是在宾州大学中文树库基础上构建的一个中文的语义角色标注语料库，包括 11000 个动词的超过 80000 个实例^[12]。中文 PropBank 语料库标注方式与英语的 PropBank 相同，用 Arg0, Arg1 等标签来标注。除了语料库之外，中文 PropBank 还为每个动词标注了框架集合 (Frameset)，动词的每个语义角色的具体语义都可以从这个框架集合中获取。

根据 PropBank 的标注方法，上述的不一致性现象看似被轻松解决了，因为不论是否满足施事的定义，动词的主体都会被标为 Arg0，而通过框架集合又可以得到更具体的 Arg0 的语义。但本文统计了中文 PropBank^[12]中的超过 25000 个动词的框架集合，发现施事不一致的问题依然存在。

中文 PropBank 中，有的动词不止一个义项，一个义项也不止一个框架集合，而有些动词的框架集合中并没有 Arg0，因此中文 PropBank 中总共有 26534 个包括 Arg0 的框架集合。在这 26534 个框架集合中，有 13624 个 Arg0 的具体语义为“agent”，还有 13 个 Arg0 的语义为“agents”，有 124 个 Arg0 语义为“experiencer”，有 504 个 Arg0 语义为“theme”。这意味着从中文 PropBank 的框架集合中所获取的大部分 Arg0 的语义仍然是很抽象的，和按照 973 语义角色标注规范中的标注方式中得到的语义角色的粒度相同。另一方面，有超过 1000 个框架集合中的 Arg0 的具体语义被记为了“agent/cause”、“agent, cause”、“agent or cause”、“cause, agent”及其同类。它们对应的动词有一些是 Arg0 既有施事的性质，又有原因的性质的，例如“穿孔”，但也有不少是和“证明”等动词一样有不一致现象的动词。例如“破坏”的 Arg0 被记为了“agent, cause”。那么虽然例句 (4a) 和 (4b) 被一致地标为了 Arg0，但从框架集合上看，它的实际语义仍然是不一致的，(4a) 中的 Arg0 应取“agent”的语义，(4b) 应取“cause”的语义。因此这种标记方式也只是在粗颗粒度标注的层面上避免了不一致的现象，到了更具体的语义中仍是不一致的。

综上，语义角色标注一致性的要求在“破坏”、“改变”、“证明”等一批动词上都无法得到满足。人民日报语料库和中文 PropBank 两个语义角色标注语料库也只是通过一定的标注方式回避了施事不一致现象（客观上掩盖了实际存在的施事不一致的语言事实）。以下将对语义角色标注语料库的分析，得到常见的、典型的施事的语义类别，并找出存在施事不一致现象的动词。

3 本文的考察方法

3.1 从标注语料库中提取存在施事不一致现象的动词

根据 2.2 中对人民日报语料库和中文 PropBank 的介绍，在这两个语料库中，都不能直接观察到施事不一致现象。所以为了探究和分析施事不一致现象，只能通过间接的手段来分析这两个语料库中的语料。

人民日报语料库中标注了“施事”、“当事”等语义角色，对于有施事不一致现象的动词，它们的主体都被统一标为了施事。因此本文统计的对象是那些全部或大多数主体被标为施事的动词。

袁毓林^[7]指出，论元角色难免有一定的模糊性，可以用原型理论，给出不同论元角色的典型句法、语义特征，然后通过类比确定特征不明显的语义角色成分。973 语义角色标注规范对施事和当事的定义也有类似的思想^[9]。根据原型的观点，对于动词而言，如果它的施事中大部分是“典型的施事”，那么可以认为，其余的个别不属于这些义原的主体是特征不明显的语义角色成分，可以通过类比标为施事。但如果它的施事中“典型的施事”比例不高，那么它的主体是否应该全部被标为施事值得讨论。也就是说，这个动词有可能存在施事不一致现象，它在一些实例中的主体可能只是为了满足一致性而在操作上被标为了施事。

在得到“典型施事”后，对语料库中每个动词，统计该动词的施事中“典型的施事”的比例。如果该比例低于一个特定的经验值，那么它就有可能是存在施事不一致现象的动词。

值得说明的是，这样获取的“典型的施事”可能不够精准，所以在最后得到可能存在施事不一致现象的动词之后，还需要人工筛选，根据语感判断一个动词是否确实存在施事不一致现象。此外，由于“典型的施事”比例不高只是动词施事不一致的一个充分条件，因此根据上述方法找出的动词也难免有一些遗漏。

3.2 “典型的施事”及其获取

根据上述论述，为了找出有施事不一致现象的动词，需要首先对两个语料库中统计对象的全体施事进

行语义分析，得到“典型的施事”。

“典型的施事”可以依据人的语言学知识和经验来判定，例如大量语义角色理论认为“典型的施事”是人或有生命的物体^[2,7,9]。通过这种方式得到的结论比较主观，有可能和实际的语言现象相比有一定偏差。因此本文采取了通过统计语义角色标注语料库中施事的语义类别的方式来确定哪些类别的词是“典型的施事”。

具体地，“典型施事”可以借助《知网》的义原分类来获取。对于语料库中每个动词的每个施事，首先根据它的句法树信息得到它的核心词。由于施事大多数是名词，因此对于其他词性的词，为了统计的方便和减少涉及的义原的个数，只记录了它的词性，不再对应到它在《知网》上的义原。对于名词，记录它在《知网》上的义项对应的义原。有些名词可能在《知网》上不存在。据观察，这些名词多数是人名，也有少量地名和公司、机构名称。本文暂时用“未登录词”来标记它们。

《知网》的名词主要分布于“entity|实体”和“Attribute|属性”这两大类义原下，其中“Attribute|属性”下的名词相对更为抽象，都是表示人、事、物的属性的词，如“勇气”、“年龄”等等。这两个义原下位的义原数量相当多，其中一些义原下的词数量相对较少，且根据初步统计，它们极少充当施事，因而本文对这两个义原适当做了选取，取出以下这些义原作为统计施事分布使用的义原：

entity|实体、thing|万物、physical|物质、animate|生物、AnimalHuman|动物、human|人、animal|兽、inanimate|无生物、NaturalThing|自然物、artifact|人工物、information|信息、group|群体、organization|组织、family|家庭、shape|物形、mental|精神、fact|事情、time|时间、space|空间、component|部分、Attribute|属性。

以上列表中的义原除了“Attribute|属性”之外都是“entity|实体”及其下位义原。如果一个名词的义原不在以上义原列表中，但属于“entity|实体”或“Attribute|属性”的下位义原，则列表中一定有该义原的某个上位义原。因此可以寻找距离该义原最近的上位义原来标记这个名词。如果一个名词的义原不是“entity|实体”或“Attribute|属性”的下位义原，则用“其他义原”来标记。“典型的施事”可以用这些义原来表示。

4 施事不一致的动词及其语义特征

4.1 语料中动词施事的语义分布

在语料库中有很多低频动词，它们的施事个数较少，如果计算低频动词的施事的分布可能不够有说服力，而且它们对标注错误比较敏感，一个施事的标注错误就可能致使施事属于某个义原的比例有较大变化。因此本文分析的是语料库中施事个数大于 10 的动词。人民日报语料库中符合条件的有 545 个动词、23723 个施事。由于 PropBank 将动词的主体都标为 Arg0，并不能从中直接知道主体对应到汉语的语义角色系统中是否为施事，因此在 PropBank 上本文也只统计了这 545 个动词的施事的语义分布情况，这些动词在 PropBank 中有 18425 个 Arg0。

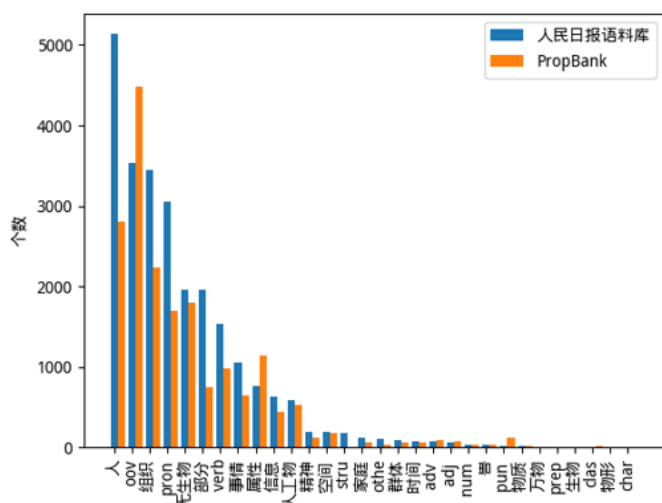


图 1 动词施事所属的义原分布

图 1 反映了两个语料库中的动词的施事在各义原/词性上的分布情况，其中蓝色的为人民日报语料库上的分布，黄色的为中文 propbank 上的分布。两个语料库中，“human|人”、“organization|组织”义原下的名词以及代词、未登录词在施事中占了多数。其余义原和词性下的施事个数都不超过 2000 个，比例均低于 10%。

各语义角色系统中一般都认为，典型的施事是人或有生命的物体。但以上数据表明，虽然“human|人”义原下的施事数量最多，但即使加上未登录词和代词这两类，也只占了所有施事的 49.4%。而其余义原下的词或其余词类的词都不能表示人类，即动词的施事中有一半以上并不是表示人类的词。

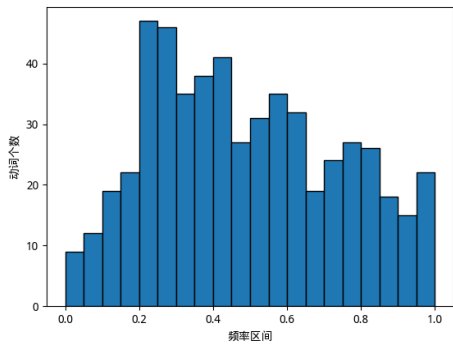


图 2 人民日报语料库的施事中人类比例与动词个数的关系

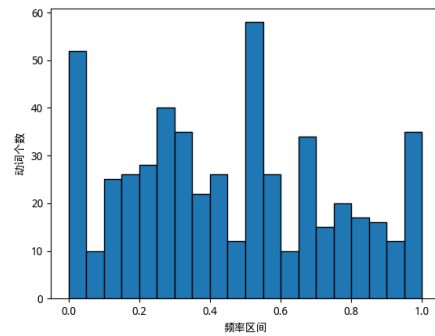


图 3 PropBank 的施事中人类比例与动词个数的关系

图 2、3 中，横坐标代表动词的施事为代词、未登录词或“human|人”义原下的词的比例，最低为 0，最高为 1，等分为了 20 段。纵坐标代表相应比例的词的个数，例如施事为人类的比例在[0,0.05]区间的动词的个数是 9 个，比例在[0.05,0.1)的动词有 12 个。如图 2 所示，施事中人类的比例超过 50%的动词还不到半数。仅有 37 个动词的施事中人类的比例超过了 90%。因此，说典型的施事是人或有生命的物体并不完全合适。图 3 表明，在 PropBank 中，同样也有类似现象，大多数的动词施事中人类的比例都不超过 55%，仅有 44 个动词的施事中人类的比例超过 90%。从动词的角度看，如果动词的典型的施事是人或有生命的物体，那么“human|人”义原下的词和代词、未登录词在大多数动词的施事中的比例应该很高。图 2、3 表明，在人民日报语料库中，并非大多数动词的大多数施事都是人类。

除了代词、未登录词和“human|人”义原下的名词之外，剩余的类别中“organization|组织”、“component|部分”、“inanimate|无生物”、“fact|事情”、“information|信息”等义原下的名词充当施事的个数较多。由于语料来源是人民日报，其中政治、经济新闻类文章中组织、机构、部门、公司、国家、报纸、会议等名词经常发生转喻现象，可以像人类一样充当施事。比如组织、机构、政府、公司类的名词都属于“organization|组织”义原，部门类的名词都属于“component|部分”义原，国家、省市、地方类的名词属于“inanimate|无生物”义原，报纸、文章类的名词属于“information|信息”义原，会议类名词属于“fact|事情”义原。

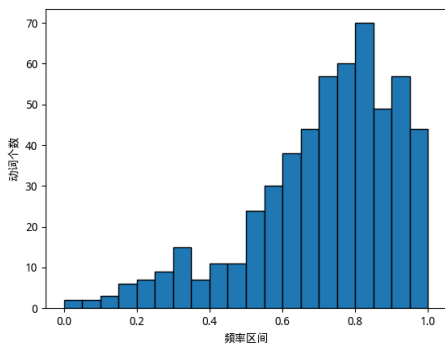


图 4 人民日报语料库的施事中典型施事

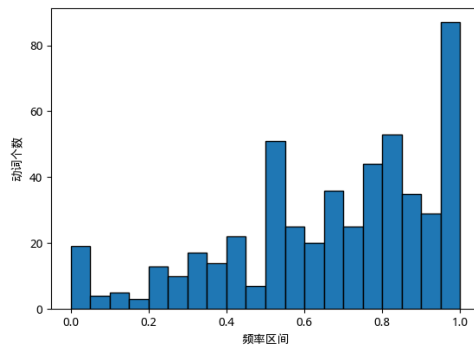


图 5 PropBank 的施事中典型施事

比例与动词个数的关系

比例与动词个数的关系

图 1-3 表明多数的施事并非仅仅是人类，还有不少可以转喻为人的词可以充当施事。在图 2、3 的统计方法的基础上，将“organization|组织”、“component|部分”、“group|群体”这几个包含较多经常转喻为人类的词的义原也考虑在内，可以得到图 4、5。根据图 4、5，显然大多数动词的大多数施事属于人类和上述几类义原，仅有 72 个动词的施事中这几类名词的比例不到 50%，这一比例低于 70% 的动词有 208 个。PropBank 中这几类名词的比例低于 50% 的也只有 91 个，低于 70% 的有 221 个。在两个语料库中施事为这几类名词的比例均低于 70% 的动词恰好有 200 个。本文将对这 200 个动词进行人工筛选，找出有施事不一致现象的动词。

由图 4、5 可见，多数动词的多数施事集中在“human|人类”、“organization|组织”、“component|部分”、“group|群体”这四个义原下。可以认为，典型的施事是有生命物体及常见的转喻为有生命物体的事物。

4.2 有施事不一致现象的动词

根据考察，4.1 中找出的 200 个“典型的施事”比例低于 70% 的动词中，包括“建交”、“建设”等动词，它们的施事多是“国家”、“会议”等不属于上述四类“典型施事”所属义原的名词，因此它们虽然施事中“典型施事”比例较低，但并没有施事不一致现象。这样的动词共有 115 个。其余 85 个动词都存在施事不一致的现象，例如在人民日报语料库中，“证明”的施事中“典型施事”的比例为 14.3%，“改变”的施事中“典型施事”的比例为 56.3%。这些动词分属于表 1 所示的《知网》上的以下若干类义原（或它们的下位义原）。为方便表示，表中义原只列出中文，不列出对应的英文名称。

表 1 存在施事不一致现象的动词

义原	动词	个数
使动	刺激、促进、促使、带动、发挥、鼓舞、激发、激励、加重、拉动、推动、引导、指导	13
变属性	巩固、缓解、计划、加快、减轻、降低、扩大、强化、缩短、缩小、拓展、限制、污染	13
创造	成立、创、奠定、开辟、建成、建立、营造	7
优化	改进、健全、满足、实现、提高、增强、振兴	7
给	赋予、给、给予、留下、提供	5
加害	干预、破坏、损害、突破	4
表示	表达、表明、说明、展示	4
认知	考验、确定、证明	3
阻动	遏制、抑制、制约	3
保证	保障、保证、确保	3
开关	打开、拉开、关	3
消除	打破、消除	2
代替	代表、替代	2
开始	开始、掀起	2
变情感	威胁、吸引	2
保持	保持、维持	2
关联	连接、团结	2
导致	带来	1
帮助	扶持	1
改变	改变	1
利用	使用	1
违背	违反	1
增加	注入	1

再来看这 85 个动词在 PropBank 中的情况。这些动词在 PropBank 的框架集中的 Arg0 主要被标为了以下几种类型：(1) “agent, cause” 及与它相似的标记，如“破坏”、“振兴”的 Arg0；(2) “agent”，如“改变”、“改进”、“带动”的 Arg0；(3) 动词的具体语义，如“违反”的 Arg0: violator 和“促使”的 Arg0: entity prodding。

事实上，存在施事不一致现象的动词数量不止 85 个。有一些动词虽然语义上会出现施事不一致现象，但在实际语料中“典型施事”的比例很高，如“帮助”等，没有被归入上述列表。还有一些动词由于在标注语料库中非自主的主体很多已经被归为了当事，或者是因为词频较低，所以也无法按上述方法直接找出，如“阻碍”、“证实”等。

4.3 动词施事不一致的原因

多数语义角色系统都将自主性作为定义施事角色的一个重要性质，有的系统中还将致使性等其他语义性质列为施事的性质。以下将从动词的自主性的角度分析，“证明”等 85 个动词施事不一致的原因。

马庆株^[11]通过语法性质，将动词分为自主动词和非自主动词两类，其中自主动词是可以进入来/去+V+O 等 8 种格式中的一种的动词，并指出，自主动词有[+自主]、[动作]的语义特征，非自主动词有[-自主]、[变化]/[属性]的语义特征，如“帮助”就是自主动词，“生病”是非自主动词。

马庆株还提到，一些自主动词的[+自主]特征极不稳定，如“咳嗽”、“眨眼”，并将这类动词非自主的情况记为“自主动词的非自主用法”。这类现象与本文讨论的现象并不相同，马庆株^[11]中并没有提及本文关注的“帮助”、“证明”等动词主体有时自主有时非自主的情况。

对于“破坏”、“证明”这些动词，根据句法格式可以判定它们为自主动词，有[+自主]特征。然而在“地震破坏”这样的短语中，“地震”的语义特征显然是[-自主]，与“证明”的[+自主]特征相冲突。而一个词的语义特征应该是稳定的，并不应该随上下文而变化。用名词的类似现象来类比，“公鸡”有[+雄性]的语义特征，“母鸡”有[-雄性]的语义特征，而“鸡”并没有[+雄性]、[-雄性]的特征，在“鸡打鸣了”和“鸡下蛋了”这样的句子中，“鸡”分别体现出了雄性和雌性的语义，但这也不意味着“鸡”有时有[+雄性]的特征，有时有[-雄性]的特征，“鸡”只是既能与有[+雄性]特征的词搭配，又能与有[-雄性]特征的词搭配，可以用[±雄性]来表示“鸡”的这一语义性质。同理，实际上“破坏”也并没有[+自主]或[-自主]的语义特征，可以说它有[±自主]的性质，不论与有[+自主]特征的主体搭配还是与有[-自主]特征的主体搭配都是合理的。所以“证明”等动词表示的事件既可以是有自主性的也可以是无自主性的，它们被归入自主动词一类或者非自主动词一类都有些不合适。

根据原型施事的理论，虽然上述 85 个动词的主体不一定有自主性，但是它们有一定的原型施事的性质，即满足自主性、感知性、致使性、移动性中的一些性质。这些动词的主体不一定是生物，而非生物一般没有感知性和移动性。因而这些动词的主体作为原型施事，应该具有致使性。所以这些动词也应该有[+致使]的特征。这些动词所属的义原也同样反映出它们有致使性。例如“使动”、“阻动”、“变属性”等义原，从义原的名称上就可以看出致使性或变化性。从更上级的义原的角度来观察，这些动词多属于“CauseToExist|使存现”、“CauseToGrow|使成长”、“MakeBetter|优化”、“MakeBad|加害”、“CauseToBeHidden|使消失”、“MakeAct|使动”等义原，也有明显的致使性。

综上，这些存在施事不一致现象的动词是一类有[+致使]语义特征，无[+自主]语义特征的动词。从语义上看，当这类动词表示的是自主性的事件时，它们的主体应为施事；当它们表示的是非自主的事件时，主体是当事/主事等角色。从语法上看，相应的判别格式也可以接受这类没有自主性的动词。这两方面的原因导致了在上述 85 个动词上主体的判定产生不一致的现象。

5 结语

语义角色标注任务中，一致性要求是一个重要的评价标准，但根据现有的语义角色体系，有一批动词很难满足一致性。这是因为有一部分动词有致使性，但并没有自主性，可以同时接受自主的主体和非自主的主体，导致在不同语句实例中，这些动词的主体有时符合施事的定义，有时不符合。这些动词的主体属

于原型施事，没有自主性的名词一般也没有位移性、感知性的性质，所以这些动词的施事不一致现象是[+致使]特征导致的。

本文对语义角色标注语料库的考察也能证明这一点。借助《知网》的义原分类，本文首先统计了两个语义角色标注语料库中施事的语义分布，指出典型的施事并非仅仅是人类或者有生命的物体，而是包括了一些常见的可以转喻为人的名词，如组织、群体、国家、会议等。进而，本文考察了 200 个典型的施事的比例低于 70% 的高频动词，通过人工筛选，得到了 85 个存在施事不一致现象的动词。这些动词在《知网》上所属的义原反映出，这些动词的确有致使性。

事实上，不仅动词的施事上存在不一致现象，在客体角色及外围角色上有同样的，甚至更为严重的问题。本文分析了施事不一致的现象及其原因，对客体的类似问题的解决可能有一定帮助。

施事的不一致现象可能还有其他类型，也可能有别的原因也会导致施事的不一致现象，例如“理发”、“洗澡”等和主体自身相关的动词也会产生施事不一致现象。这样的问题有待于进一步的考察和研究。

参考文献

- [1]Palmer, M, Gildea, D, Xue, N. Semantic role labeling[M], San Rafael: Morgan & Claypool Publishers, 2010.
- [2]C.J.菲尔墨著，胡明扬译。“格”辨[M].北京：商务印书馆，2012.
- [3]Dowty, D. Thematic proto-roles and argument selection[J]. Language, 1991:547-619.
- [4] <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> [EB/OL]
- [5]Palmer, M, Gildea, D, Kingsbury, P. The Proposition Bank: An annotated corpus of semantic roles[J]. Computational Linguistics, 2005,31(1):71-106.
- [6]鲁川，汉语的意合网络[J].语言文字应用，1998,2: 82-88.
- [7]袁毓林，论元角色的层级关系和语义特征[J].世界汉语教学，2002,3: 10-22.
- [8]袁毓林，语义角色的精细等级及其在信息处理中的应用[J].中文信息学报，2007,4: 10-20.
- [9]现代汉语谓词语义角色标注语料库规范，<http://klcl.pku.edu.cn/ShowNews.aspx?id=158> [EB/OL]
- [10]顾阳，论元结构理论介绍[J].国外语言学，1994,1: 1-11.
- [11]马庆株，自主动词与非自主动词[J].中国语言学报，1988,3: 157-180.
- [12]Xue, N, Palmer, M. Adding semantic roles to the Chinese Treebank[J]. Natural Language Engineering, 2009,15(1):143-172.
- [13]董振东，董强，知网和汉语研究[J].当代语言学，2001,1(3): 33-44.
- [14]何保荣，中文语义角色标注语料库的构建及统计分析[D].山东：鲁东大学，2017.
- 陈龙（1995—），硕士生，主要研究领域为应用语言学，汉语语言知识工程，中文信息处理。Email: 1400014101@pku.edu.cn
- 詹卫东（1972—），教授，主要研究领域为现代汉语形式语法，中文信息处理，汉语语言知识工程。本文通信作者。Email: zwd@pku.edu.cn