

事件预期属性的标注¹

邹红建 杨尔弘

北京语言大学应用语言学研究所 北京 100083

E-mail: zouhj97@163.com yerhong@blcu.edu.cn

摘要: 事件属性信息对于精细刻画事件信息是必须的。本文对事件的预期属性这一新的事件属性进行分析,证明了预期属性是独立的事件属性,有必要进行相应的标注,并将之与相关范畴进行了区分,并分析了该属性出现的事件类别,以及表达该属性的语言形式。初步的实验标注证明了标注事件的预期属性的定义的合理性以及标注的可行性。

关键词: 事件属性, 预期, 标注

Annotation of Event Attribute of Fulfillment

Zou Hongjian, Yang Erhong

Institute of Applied Linguistics, Beijing Language and Culture University, Beijing, China, 100083

E-mail: zouhj97@163.com yerhong@blcu.edu.cn

Abstract: Event attributes are vital to the precise characterization of events. We proposed to annotate one novel event attribute, the “fulfillment” attribute, and distinguished it from related semantic categories. We also discussed the necessity of its annotation, the linguistic markers indicating such attribute. Through experimental annotation, we can conclude that it is necessary and feasible to annotate such event attribute.

Keywords: event attributes, fulfillment, annotation

1. 引言

事件通常归纳为5个W和1个H,即“when”、“where”、“who”、“what”、“why”和“how”,进一步还可通过论元结构或者框架来精细刻画。除了这种框架信息,事件还包括不同的属性信息。这些属性信息使得事件的描述更加具体,若精细刻画事件信息,则属性信息不能忽略。属性的不同情况甚至会导致完全相反的意义。例如:

智利“蓝色之家”旅馆火灾烧死10名游客

火灾未致人员伤亡,起火原因正在调查中。

例子中说的都是火灾及其产生的后果,前例中是发生了伤亡,而后例则没有发生伤亡。

目前,国内尚未有事件属性信息的标注研究以及资源建设。邹红建,杨尔弘(2010)通过对真实文本标注实践的总结,归纳出一些必要的事件属性信息。其中,事件的预期属性是一个很有意思的属性,值得关注。

在一篇新闻报道中,往往涉及多个事件,这些事件中,有些事件会产生特定的结果,或者说引起了另外的事件的发生,例如地震可能导致伤亡或者引起海啸;有些事件本身蕴含了预期的结果,例如,救援这一事件本身就蕴含着一个结果信息,即救援成功了、救援失败了或者救援正

¹本文得到国家社科基金项目“面向内容计算的文本信息标注研究”(06YY047)的资助。

在进行尚未成功。

下面是恐怖袭击报道的例子：

叙利亚政府官员 12 日说，4 名不明身份武装人员当天上午企图利用自动步枪、手榴弹和至少 1 辆载满炸药的轿车对美国驻叙利亚大使馆发动袭击，但被叙利亚安全人员及时挫败，3 名袭击者在冲突中被打死。

例子中叙述的是一个恐怖袭击事件，由于该恐怖袭击事件的结果未能实现，因此，也就是说，该恐怖袭击事件“未遂”。

另一个例子是恐怖袭击达到了预期的结果。例如：

新华网莫斯科 5 月 17 日电（记者宋世益）俄罗斯车臣共和国 17 日发生军警车队遭非法武装分子袭击事件，造成 5 人死亡、6 人受伤。

例子中，恐怖袭击产生了实际的效果，也就是达到了预期的目的。

对于那种事件本身包含了一个预期结果的事件，人们不会忽略其结果如何，甚至对事件的预期结果的关注会超过了对事件本身的关注。

目前在事件信息的标注中，已有一些项目标注了事件的相关属性，但是尚未见到对事件的预期属性进行相应的研究与标注。事件的预期是否可以作为一种独立的属性，通过哪些语言范畴来表达，有什么样的形式标记，哪些类型的事件具有这一属性，在事件文本中的分布如何，有没有必要进行标注。这些是本文所要回答的问题。

2. 相关研究

2.1 ACE 对事件属性的标注

ACE (Automatic Content Extraction)² 是美国国家标准技术研究所 NIST 组织的信息提取评测项目，目的是通过大规模标注语料库的建设，促进自动内容提取技术的发展，研究能识别和描述文本所表达的意义的技术。处理的文本来源分为三种：新闻专线（普通文本）、广播新闻（ASR 得到的文本）、报纸（OCR 得到的文本）。处理的语言包括英语、汉语、阿拉伯语。在 ACE05 (ACE, 2005) 的任务中包括事件的识别和描述 (EDC, Event Detection and Characterization)；规定抽取如下 8 大类事件：Life、Movement、Transaction、Business、Conflict、Contact、Personnel、Justice。每类又分若干小类。事件的标注工作也只限于这 8 大类事件。每类事件除了标注论元成分，还标注了如下 4 个属性：1. 极性 (polarity) 说明该事件是肯定还是否定（例如：他再也不贩卖毒品了）。2. 时态 (tense) 说明该事件的发生时间相对于文本时间是过去、现在、将来还是未确定。3. 普遍性 (genericity) 说明该事件是一般性的（例如：该组织专门从事武器走私）还是具体的（例如：那只狮子在追捕羚羊）。4. 模态 (modality) 说明该事件的可能性是确定的还是其他情况（例如：歹徒威胁要杀死他）。

2.2 TimeML, EventML

TimeML (Pustejovsky etc, 2003; TimeML, 2005) 目的是对自然语言中的时序 (temporal) 信息进行标注，背后的理论是 Temporal Discourse Models (Mani and Pustejovsky, 2004)。按照 TimeML 规范标注的结果为 XML 格式。TimeML 语料库³ 目前已有多个，包括 TimeBank1.1、TimeBank1.2、

² <http://projects.ldc.upenn.edu/ace/>

³ <http://www.timeml.org/site/timebank/timebank.html>

AQUAINT TimeML 1.0。在 TimeML 中与事件相关的标记是 EVENT 和 MAKEINSTANCE，前者是事件的型 (tokens)、后者是事件的实例/实现 (instances/realisations)。在 EVENT 层面标注了事件的类别 (class)，诸如 perception, aspectual, state 等。在 MAKEINSTANCE 层面标注的属性信息包括：时态 (tense)、体态 (aspect)、极性 (polarity)、模态 (modality)。其中，时态与体态都是对表达事件的动词的语法范畴的描述。

EventML (Roser and Pustejovsky, 2008, 2009; Roser, 2008) 在 TimeML 的基础之上，对事件的事实性 (Event Factuality) 进行标注。事件的事实性即文本中表达的事件是反映了现实世界中的真实情况，或是并未发生的情况，或是并不确定。事件的事实性通过不同的语言学范畴表达出来，其中最主要的包括：1. 极性 (polarity) 即表示是肯定还是否定，取值为离散的布尔值。2. 模态 (modality) 表达了说话者对命题的真值情况的承诺的程度。其值从 uncertain/possible 到 absolutely certain/necessary。为了方便，模态也需要离散的取值。事件事实性的值表示为 <mod, pol> 这样的二元组。在 EventML 中，还注意到了不同的消息源 (sources) 以及嵌套消息源 (nested source) 对事实性的影响。根据 EventML 标注的语料库有 FactBank⁴。FactBank 建立在 TimeBank 的基础之上，增加了事件的事实性的标注信息。

3. 事件的预期属性

3.1 事件的属性信息

邹红建，杨尔弘 (2010) 参考已有的研究，并通过对文本中事件属性信息的实际分析，确定如下的事件属性信息：极性、模态、时态、体态、程度、频度、源信息、地位、预期。确定这些属性信息主要依据它们是否给事件的整体意义带来变化，以及不同的取值是否有意义上的明显对立，而暂时忽略它们是否属于一个层面。其中，“预期”这一属性在以往的事件属性标注中尚未见到，也是本文要着重讨论的对象。

3.2 事件的预期属性

3.2.1 预期属性是一种独立的范畴

与预期属性比较近似的属性主要包括极性与模态，事件之间的因果关系也与之近似。

1) 预期不等于极性

极性是表示肯定还是否定的范畴。极性说明了事件是发生还是未发生，分为“肯定、否定”。未发生的事件其极性为否定，否则为肯定。例如：

4月26日，莫斯科郊区一家工厂发生火灾，造成10人死亡。

(事件：死亡；极性：肯定)

所幸这起火灾没有造成人员伤亡。 (事件：伤亡；极性：否定)

一般人可能将预期属性与极性混淆，将达到预期目标的事件看作极性为肯定，将没有达到预期目标的事件看作极性为否定。实质上这两者有着明确的区别。

首先，没有达到预期目标的事件并不能视作极性为否定，因为极性为否定表示该事件未发生，而没有达到预期目标说明事件发生了，只是没有达到预期的目标而已。例如：

也门内政部15日说，恐怖袭击者当天企图利用汽车炸弹袭击两座石油设施，但也门官方挫败了袭击阴谋，4名袭击嫌疑人死亡，1名保安也不幸身亡。

4 <http://www ldc upenn edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009T23>

这里的恐怖袭击未达到预期的目的，但是已实际发生了。所以该事件的极性不能标注为否定。更不能认为与极性为否定是一回事情。

第二，达到了预期目标的事件，其极性为肯定，没有达到预期目标的事件，其极性仍然是肯定。预期结果或者目的的实现情况，不代表事件的极性无关。

2) 预期不等于模态

模态是对其事件的可能性的判断。目前简单分为“确定、其他”两种情况。对于确定其发生的事件，其模态为确定；否则其模态为其他。例如：

部分沿海地区居民担心地震引发海啸，携家属向市区以南的山区逃离。

(事件：海啸；模态：其他)

专家此前曾警告，震级如此高的地震有可能引发“毁灭性”的海啸，并在数小时后抵达附近海岸。 (事件：海啸；模态：其他)

因为预期是指事件可能带来的结果。一般人可能将预期属性与模态混淆。实质上，这两者的区分也是很明显的。

首先，对于模态为“确定”的事件，其预期属性可以是实现的，也可以是未实现的。前面所举的例子中的事件，其模态都是“确定”的。

其次，具有预期属性的事件，也可以具有不同的模态。例如：

俄罗斯车臣共和国17日发生军警车队遭非法武装分子袭击事件，造成5人死亡、6人受伤。

(事件：袭击；模态：确定)

美国《侨报》援引海外报道称，为庆祝自己的生日，拉登正密谋在生日当天发动一次大规模的恐怖袭击。 (事件：袭击；模态：其他)

3) 预期不等于因果关系

另外值得注意的是一种与预期非常相近的表达。例如：

墨西哥国家地震检测中心1日发布消息说，墨西哥南部格雷罗州当天凌晨发生里氏4.7级地震，但没有造成人员伤亡和财产损失。

例子中涉及到三个事件，即地震、伤亡、损失。其中，地震与后两个事件之间存在因果关系。这个例子看起来像是含有预期属性的事件。通过分析发现，事件“地震”的极性为肯定，“伤亡”和“损失”的极性为否定。同时，“地震”与“伤亡”和“损失”之间存在因果关系。而不是由“地震”一个事件蕴含的。

因果关系至少涉及到两种事件，而预期属性则涉及一种事件，即蕴含于该事件本身。

根据以上对预期属性和与之相似的属性的区分，可以清楚的看出，预期不等于也不包含于极性或者模态中，也不等于因果关系，而是一种独立的属性。预期属性信息不能由其他范畴表达，因此有必要进行标注。

3.2.2 表达事件预期属性的范畴

根据上面的分析，可以得出这样的结论，即表达事件的预期结果或者目的的属性，因为与极性、模态等相关范畴都有区别，因此，不能借用这些范畴来表示，而是需要设置为独立的属性范畴。事件的预期属性表达的是事件本身所蕴含的结果或者目的的实现情况。

3.2.3 预期属性分布的事件类别

与“极性”、“模态”这样的属性不同，“预期”不是所有类型的事件都具有的属性。例如：
5月27日，印度尼西亚爪哇岛发生6.2级地震，造成重大人员伤亡，无数房屋倒塌。

晨报讯据新华社电6月7日20时至8日7时，广西梧州市普降特大暴雨，降雨量高达250毫米，造成梧州市市区多处发生泥石流。

报道说，一辆公共汽车6日在北部林波波省首府博罗夸尼附近翻车，导致9人丧生。

据报道，这架飞机是1月24日在喀土穆机场被武装人员劫持的，并被迫更改航线，飞往乍得首都恩贾梅纳。

当地时间4月24日，埃及西奈半岛南部靠近亚喀巴湾的旅游城市宰海卜接连发生三起爆炸事件，由于目前正值旅游高峰期，爆炸至少造成23人死亡，数十人受伤。

警方25日在宰海卜附近逮捕了3名嫌犯，但并未透露更多信息。

这些例子中有自然事件，如地震、泥石流；有人为的事件，如翻车、劫机、逮捕。通过对这些例子分析，可以得出这样的结论：

1. 自然发生的事件没有预期这一属性。如例子中的“地震”、“泥石流”、“暴雨”等没有预期这一属性。需要注意的是，这些事件可能导致相关的后果，但是那属于事件的因果关系范畴，不属于事件的预期属性范畴。
2. 人为的事情有的有预期这一属性，有的没有预期这一属性。例子中，“翻车”就没有预期属性，因为这属于交通事故，一般并非有意为之（蓄意策划的除外）。而“逮捕”、“劫机”这样的事件则具有预期这一属性。
3. 人为事件具有预期这一属性，其前提条件是该事件必须是有预期的目的，事件的完成意味着预期的结果的实现或者说目的的达成，或者未实现或者说未达成。例如，“逮捕”这一事件的具体的目的就是要抓住某人，该事件必然会有一个结果，即抓住了某人或者没有抓住某人。

根据对具有预期属性的事件类型的分析，现在将事件的预期属性限定为人为的且蕴含目的和结果的事件所具有的属性。

4. 事件的预期属性的形式标记

4.1 词汇标记

事件的预期属性通过一定的语言形式表现出来。最明显的是表示预期目的的结果实现成功还是失败这样的词汇标记。诸如表示达到预期目的的“成功”，表示事件未达到预期的“失败、挫败、企图”。例如：

中国成功救出被困8昼夜的115名矿工

波罗的海通讯社报道说，救援人员赶在公寓屋顶塌陷前成功救出一些居民。

两辆装满炸药的轿车一辆炸开花 一辆引爆失败

“9·12”袭馆事件虽终未得逞，也没有造成重大伤亡。

《每日邮报》称，警方此次行动挫败了一起“重大恐怖袭击阴谋”。

埃及安全人士1日透露，疑为“基地”组织成员的5人持假护照潜入埃及，并企图在西奈半岛旅游胜地沙姆沙伊赫和宰海卜制造恐怖袭击。

4.2 时体标记

如果没有出现明显的词汇标记, 时态信息与体态信息也可以帮助确定事件的预期属性。例如, 如果具有预期属性的事件其时态为“过去”, 体态为“完成”, 其预期属性则为为实现或则达到。例如:

英国警方 31 日凌晨在英格兰中部城市伯明翰展开反恐突袭行动, 在不同地点逮捕了 8 名恐怖嫌疑犯。

苏丹一架载有 103 名旅客和机组人员的波音 737 客机 24 日在苏丹首都喀土穆上空被身份不明的持枪者劫持, 并飞往乍得首都恩贾梅纳。

屏东恒春镇的一个家具行也发生了倒塌, 8 人被困, 消防队员已救出 5 人, 其中 1 人死亡、4 人受伤, 目前仍有 3 人被困在倒塌的房屋中。

截至今日零时, 大火基本被扑灭。

5. 实验标注与结果分析

根据前面对事件的预期属性的定义, 以及对表达该属性的语言形式的分析, 对 3 类事件 60 篇文本进行了标注, 语料来源于新浪网的新闻报道。标注工作是在已经标注了事件以及事件的其他属性信息的基础上进行的。标注过程即逐个检查文本中已经标注了的事件, 确定是否为拥有预期属性的事件, 如果是, 则根据文本中的信息确定其属性值并进行标注。

初步标注的结果如下:

表 1 事件属性信息的标注的基本情况

文本类别	文本数量	句子数量	标注的句子数量	事件数量	预期		
					达到	未遂	未确定
火灾	20	309	256	442	53	0	32
地震	20	278	249	448	13	0	19
恐怖袭击	20	369	309	591	79	9	93
总和	60	956	814	1481	145	9	144

从表 1 中可以看出:

事件的预期信息在不同类别的事件文本中明显不均衡。恐怖袭击事件中预期属性信息最多, 地震事件中预期属性信息最少。这是因为预期信息仅出现在人为的且带有预期目的的事件中。在地震类文本中, 所涉及的大部分事件诸如地震、海啸、建筑物毁坏、人员伤亡等都不符合“人为的且蕴含预期目的或结果”这一要求。火灾和恐怖袭击所涉及的事件诸如恐怖袭击、救援、灭火等符合“人为的且带有预期目的”这一要求。

前面在分析拥有预期属性的事件时将预期属性总结人为的且蕴含目的和结果的事件所具有的属性。地震类事件不属于人为事件, 却出现了大量的具有预期属性的事件, 这是因为地震类事件所涉及的相关事件中, 存在诸如救援、撤离、疏散这样符合条件的事件。

另外, 在火灾与地震中未出现预期属性为“未遂”的事件, 而在恐怖袭击中出现了少量的预期属性为“未遂”的事件。分析文本发现, 这是因为恐怖袭击事件在报道时一般仅存在两种可能: 发生且达到预期, 或者被挫败, 因此存在明确表明事件“未遂”的情况。而在地震类事件和

火灾类事件文本中，以救援为例，一般存在的情况也是两种：救援成功或者正在进行，因此也就没有“未遂”的例子。这也可能与实验标注的规模小有关。

根据实验标注的情况，计划建设一个标注事件预期属性信息的 Fulfillment Bank。参考 FactBank 的建设的方法，决定在已经标注了事件词、事件的其他属性的文本上多增加一层事件预期属性信息。目前该项工作正在进行中。目前尚未报告有这种类型的语言资源，而建设 Fulfillment-Bank，对于自动发现有目的性的事件，例如救援的情况、某次袭击、政策的实施等，将会有益。

参考文献

- [1] ACE. ACE Chinese Annotation Guidelines for Events.
http://www ldc.upenn.edu/Projects/ACE/docs/Chinese-Events-Guidelines_v5.5.1.pdf. 2005.
- [2] Michele Banko. Open Information Extraction for the Web. Ph.D. Thesis, University of Washington. 2009
- [3] Inderjeet Mani and James Pustejovsky. Temporal Discourse Models for Narrative Structure. In: ACL 2004 Workshop on Discourse Annotation. 2004
- [4] James Pustejovsky, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer and Graham Katz. TimeML: Robust Specification of Event and Temporal Expressions in Text. In: IWCS-5, Fifth International Workshop on Computational Semantics. 2003
- [5] Saurí, Roser. A Factuality Profiler for Eventualities in Text. Ph.D. Thesis. Brandeis University. 2008.
- [6] Saurí, Roser and James Pustejovsky. FactBank: A Corpus Annotated with Event Factuality. In: Language Resources and Evaluation. 2009.
- [7] Saurí, Roser and James Pustejovsky. From Structure to Interpretation: A Double-layered Annotation for Event Factuality. In: Proceedings of the 2nd Linguistic Annotation Workshop. 2008.
- [8] TimeML Annotation Guidelines Version 1.2.
http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf. 2005
- [9] 邹红建, 杨尔弘. 事件属性信息的标注. 第十一届汉语词汇语义学研讨会 (CLSW2010). 2010