

# 一种汉语语义角色标注规范

An Annotation Guideline for Chinese Semantic Labeling



目前编者：刘亚慧 杨浩苹 李正华

[zhli13@suda.edu.cn](mailto:zhli13@suda.edu.cn)

2020 年 8 月 23 日版本

自 2018 年 10 月开始编制

苏州大学人类语言计算研究所

Human Language Technology Research Center, Soochow University

# 目 录

<b>1 更新记录</b>	<b>4</b>
1.1 2020.8.23 . . . . .	4
1.2 2020.6.13 . . . . .	4
1.3 2020.3.19 . . . . .	5
1.4 2020.2.19 . . . . .	5
1.5 2019.10.21 . . . . .	6
1.6 2019.8.27 . . . . .	6
1.7 2019.7.19 . . . . .	7
1.8 规范中文字颜色使用的说明 . . . . .	8
<b>2 前言</b>	<b>8</b>
2.1 语义角色标注的定义 . . . . .	8
2.2 语义角色标注和句法分析的对比 . . . . .	9
<b>3 标注技巧和说明</b>	<b>10</b>
<b>4 角色标签列表</b>	<b>10</b>
<b>5 谓词选取</b>	<b>12</b>
5.1 能愿动词通常不作为谓词 . . . . .	13
5.2 形式动词通常不作为谓词 . . . . .	14
5.3 趋向动词是否做谓语的判断 . . . . .	14
5.4 使役动词 . . . . .	16
5.5 “是/有”等不作为谓词的情况 . . . . .	16
5.6 仅起到指称作用的动词不作为谓词 . . . . .	17
5.7 常用口语不作为谓词标注 . . . . .	17
<b>6 如何从句子片段中选择一个词作为论元？</b>	<b>19</b>
<b>7 主体角色</b>	<b>23</b>
7.1 施事 (agent) . . . . .	23
7.2 共同施事 (co-agent) . . . . .	23
7.3 当事 (experiencer, expe) . . . . .	24
<b>8 客体角色</b>	<b>24</b>
8.1 受事 (patient) . . . . .	24

8.2	谓词性受事 (predicate-patient, pred-patient)	27
8.3	与事 (dative)	27
8.4	系事 (relative)	27
<b>9</b>	<b>兼主体客体 (subject-object, subj-obj)</b>	<b>29</b>
<b>10</b>	<b>主体或客体隐藏</b>	<b>31</b>
<b>11</b>	<b>附加角色</b>	<b>31</b>
11.1	工具 (tool)	31
11.2	材料 (material)	33
11.3	方式 (manner)	33
11.4	处所 (location, loc)	35
11.5	终点 (end location, end-loc)	35
11.6	起点 (begin location, beg-loc)	35
11.7	方向 (direction, dir)	37
11.8	时间 (time)	37
11.9	结束时间 (end time, end-tm)	37
11.10	开始时间 (begin time, beg-tm)	37
11.11	范围 (range)	39
11.12	原因或目的 (cause)	39
11.13	事量 (quantity)	40
<b>12</b>	<b>其它角色</b>	<b>42</b>
12.1	离合 (separated)	42
<b>13</b>	<b>难点解析</b>	<b>43</b>
13.1	深层语义 vs. 浅层语义	43
13.2	“的”字结构作论元	43
13.3	核心角色 vs. 附加角色	44
13.4	主体 vs. 客体	44
13.5	正确辨析谓词的主体角色	44
13.6	正确辨析谓词的客体角色	46
13.7	机构名、组织名等作为主体论元	47
13.8	manner vs. tool	47
13.9	range vs. loc	49
13.10	谓词为成语时的标注	49

13.11	<u>谓词为网络用语时的标注</u> . . . . .	49
13.12	主体角色中代词的选择 . . . . .	51
13.13	比较结构的论元确定 . . . . .	51
13.14	疑问词作论元 . . . . .	51
13.15	真歧义 . . . . .	52
13.16	<u>特殊谓词的标注辨析</u> . . . . .	52
13.17	不确定论元的总结 . . . . .	53
<b>14</b>	<b>分词粒度对语义角色标注的影响</b>	<b>53</b>

# 1 更新记录

如果规范有了较明显的改动，就要形成一个新的版本（以日期为版本号）。数据标注的结果也要对应保存好标注时间及修订（校正时间）。和前一个版本的差别，除了在此罗列出来，同时在正文部分用下划线突出具体内容。

## 1.1 2020.8.23

- 第5节增加名词活用的情况。
- 第5.7节增加了图20。
- 第6节增加了论元为从句和名词分辨不清时的优先级，修改了重复表达部分。
- 第7.3节中修改了图41的客体标签。
- 第8.2节强调 pred-patient 的主体可以对应施事也可以对应当事，引用了 expe 中的图41。
- 第9节增加图54。
- 第10节增加了一种不标隐藏的情况，增加了图107。
- 第11.3节增加了部分不标注 manner 的内容。
- 第11.4节增加了不标注 loc 的情况，增加了图70。
- 第11.5节修改了图71的说明。
- 第11.8节增加了部分说明，增加了图81。
- 第11.12节增加了部分 cause 的标志词。
- 第11.13节中明确了作为频率的词等部分内容，修改了图89的说明。
- 第13节去掉了“辨别共指与非共指”这一节。
- 第13.5节增加了两种不容易辨别主体的情况，增加了图104和图103
- 第13.9节修改了图111的客体标签。
- 新增第13.11节“谓词为网络用语时的标注”，增加图114。
- 新增第13.16节“特殊谓词的标注辨析”，增加图118和图119。
- 第13.17节将图50移到了第8.4节，并修改了图的客体标签。

## 1.2 2020.6.13

- 第6节增加论元片段中多个谓词时代表词的选择规则。
- 第7.3节延伸了主体为当事的谓词范围，增加了图41。
- 第8.1节增加了“等”列出多个受事论元的内容，并增加图43。
- 第9节中修改了图52的客体标签。
- 第10节增加了存在省略但不标隐藏的情况，增加了图60。
- 第11.3节中增加了图66。

- 第11.4节去掉了“抽象地点”的定义，明确“空间或地点”的定义，修改了图67的注释。
- 第11.11节强调了“范围”的标注范围。
- 第11.12节增加了 cause 的标志词“来”，增加了图85。
- 第11.13节增加图91。
- 新增第13.9节“range vs. loc”，增加图110和图111。
- 新增第13.10节“谓词为成语时的标注”，增加图112和图113。

### 1.3 2020.3.19

- 第4节中修改了论元角色数量，以及图3和表1。
- 第5.1节明确对能愿动词的判断。
- 第5.3节修改了图8的标注，标出了“穿”的隐藏主体；修改了图10的标注，标出了“来”的隐藏客体。
- 第5.4节修改图12的标注，“让”不跟双宾语。
- 第5.5节中明确“有”不作谓词的情况。
- 第6节修改了“他是一个善良的人”的论元选择规则，选择“他”为谓词的论元代词，修改了图93的注释。
- 去掉“间接当事 (indirect-experiencer)”标签。根据当事和施事的定义，“王冕死了父亲”中的“父亲”不能再标为当事，所以“王冕”标为间接当事不再适用。之前我们标注“iexpe”标签是为了能够尽可能多地标注一些信息，现在感觉标注的信息太多，跨越太大，而且定义中实体、属性的概念很难把握，因此我们将间事标签去掉。
- 第10节新增了优先标注隐藏的规定。
- 第11.8节合并“时间点 (time-point)”和“时间段 (time-span)”两个论元角色为一个论元角色“时间 (time)”，鉴于有时候时间点和时间段的界限很模糊。
- 第11.13节修改了图89的标注，标注出“解释”的隐藏客体。
- 去掉难点辨析中“time-point vs. time-span”。
- 修改第13.8节“tool vs. manner”，去掉“优先 manner 标签”的规定。
- 第13.15节新增了真歧义的例子，且标注系统中加入了“真歧义”的选项。
- 去掉第13.17节中的例子“花儿开放标志着春天到了”，增加了图50。

### 1.4 2020.2.19

- 第4节中修改了论元角色数量，以及图3和表1。
- 第5节中去掉了谓词选取中“我在家”不作谓词的描述。
- 第6节中明确由多个词构成的名词复合短语的标注方法，增加了例子23、24和30，修改了例子21中的解释。

- 第7.2节中增加了省略 co-agent 的情况，增加了图34。
- 第7.3节中修改了当事人的定义，“在”作谓词时主体论元标为当事，增加了图39。
- 第8.3节删除“罢免”和“颁奖”的例子，增加了例子“请教”，即图47。
- 第8.3节增加了双宾语中省略间接宾语（或直接宾语）的情况，增加了例图48。
- 新增第9节兼主体客体。
- 第11.13节增加了谓词“占”的客体角色标注说明。
- 第12.1节增加了离合词和一般词的区别。
- 新增第13.4节“主体 vs. 客体”。
- 新增第13.6节“正确辨析谓词的客体角色”。
- 新增第13.7节“机构名、组织名等作为主体论元”。
- 第14节修改了图123。
- 使用蓝色表示强调。

## 1.5 2019.10.21

- 第5.6节中明确强相关的论元成分为核心角色。
- 第7.1节中修改了施事的定义（扩大了施事的范围），删除了原来的例子，增加了图31和32。
- 第7.2节中修改了共同施事的定义，去掉了“具有主观行为能力（即具有主观意愿）”。
- 第7.3节中修改了当事人的定义（缩小了当事人的范围），增加了图40。
- 第8.1节中修改了受事的定义，去掉了“可能受到一定的影响”和“作为客体角色，受事通常对应的主体角色为施事，但也可以对应当事”。
- 新增第8.2节，谓词性受事，增加了图44、45和46。
- 第8.4节中修改了系事的定义。
- 新增第11.12节，原因或目的，合并原来的“原因”和“目的”角色为一个角色。
- 新增第13.15节“真歧义”。
- 删除了客体角色“内容（content）”、“对象（target）”和“结果（result）”。
- 删除了附加角色“原因（reason）”和“目的（purpose）”。
- 删除了第13节的“target vs. content”和“dative vs. target”。
- 修改了存在新增或者已删除标签的其它例子中的标签。

## 1.6 2019.8.27

- 第5节中去掉了形容词不作谓词的情况。
- 第5节中强调了动词的名词化需要标注的情况。
- 第5.1节中增加了常见的能愿动词。

- 第5.3节中增加了趋向动词作谓词的判断条件。
- 第5.6节中修改了起指称作用的动词的定义。
- 第6节中增加了形容词作补语或状语时如何标注的说明。
- 第8.3节中修改了与事的定义。
- 第11.1节中的图63：草原：relative->patient。
- “原因”角色定义中增加了判定原因论元角色的标志词及需要标注论元角色的情况。
- 第11.13节中对事量进行了补充并修改了图87下的注释。
- 第12.1节中增加了离合词判定的标准，增加了几个离合词的例子。
- 第13.3节中增加了图99。
- “dative vs. target”节中修改了例图。
- “dative vs. target”节中去掉了 dative 与 purpose 易混淆的点。
- 在“time-point vs. time-span”节中对时间词进行了补充（由于时间点和时间段的合并，已去掉该节）。

## 1.7 2019.7.19

- 修改了第7.1节“施事 (agent)”中添加了第三个例子“浪卷走了小船”（目前已删除该例子）。
- 第7.3节“当事 (experiencer, expe)”中添加了图38。
- 第11.4节“处所 (location, loc)”中添加了图69。
- 第13.3节“核心角色 vs. 附加角色”中添加了图98。
- 增添了第13.5节“正确辨析谓词的主体角色”。
- 修改了“dative vs. target”节，添加了图例。
- 增添了“time-point vs. time-span”节（由于时间点和时间段之间界限模糊，将其合并为“time”标签，已去掉该节）。
- 增添了第13.12节“主体角色中代词的选择”。
- 修改了第14节“分词粒度对语义角色标注的影响”，添加了图124和125。
- 修改了表1中开始时间和结束时间的例子：比赛七点 [开始]-> 比赛七点开始 [进行]；会议三点 [结束]-> 会议 [开到] 三点。
- 修改了图1：篮球：content->patient。
- 修改了图63：灯光：expe->agent。
- 修改了图64：墙壁：targrt->patient。
- 修改了图79：暑假：end-tm->time-point，并将该例归到时间点的例子中。
- 修改了图92：他：agent->co-agent。
- 修改了“时间点”这一节：“星期天、今天”表示时间点 -> 一般情况下“星期天、今天”表示时间点（由于时间点和时间段合并，已去掉该节）。



- 删除了第13.17节“不确定论元总结”中关于“她穿衣服好看”的图例。

## 1.8 规范中文字颜色使用的说明

蓝色表示强调。 红色表示禁止。 紫色表示优先。

## 2 前言

本规范的制定过程中，大量参考了北京大学的《现代汉语谓词语义角色标注语料库规范》<sup>1</sup>以及他们在网上公布的数据<sup>2</sup>。本规范和北大语义标注规范之间的主要区别是：

- 北大规范中，论元对应句子的片段，可以是词、词组或子句，需要标注者确定论元的边界；而我们的规范论元都具体到词，也就是说我们要求标注者从句子片段中确定唯一的代表词（见第6节）。
- 我们额外标注了省略主体和省略客体信息。

另外，我们也深入学习了宾大的 Chinese Propbank 的标注规范《Annotation Guidelines for the Chinese Proposition Bank》<sup>3</sup>。了解了 NomBank、FrameNet、和 CAMR。

### 2.1 语义角色标注的定义

给定一句话，语义分析的目标是获取句子中的所有语义信息，并以某种形式表示出来。语义分析是自然语言处理的终极目标。如果可以深入理解句子的语义，那么计算机就可以真正实现人工智能。

然而，什么是语义？语义应该如何表示？这些问题对于研究界而言仍然遥不可及。<sup>4</sup> 如果把真正意义上的语义称为深层语义的话，那么语义角色标注则尝试刻画句子的浅层语义信息。<sup>5</sup>

给定一个句子，语义角色标注首先确定句子中包含哪些谓词 (predicate)。进而，对于每一个谓词，确定与这个谓词相关的论元 (argument)，并给出论元的角色 (role)。

例如，“我喜欢打篮球”这个句子中包含两个谓词“喜欢”和“打”。关于谓词的判定，我们在第5节中给出介绍。如下图1给出了这两个谓词的论元信息。

本规范中，规定箭头的方向是从谓词到论元，弧上的标签为论元角色标签。我们会在后面详细介绍每一种角色的定义。

除了上图的形式，本规范也会用文字的方式来表示语义角色的标注结果，例如：

---

<sup>1</sup><http://klcl.pku.edu.cn/gxzy/231695.htm>

<sup>2</sup><http://klcl.pku.edu.cn/gxzy/231694.htm>

<sup>3</sup><http://verbs.colorado.edu/chinese/cpb/cpbguide.pdf>

<sup>4</sup>类似的问题：什么是概念？世界上有哪些概念？

<sup>5</sup>目前学术界提出了一些语义表示形式，尝试刻画更深层次的语义信息，例如 AMR, HSPG, UCCA 等。但是这些语义表示也仅仅是从某一些任务出发，刻画了一部分语义信息。



图 1: 句子中两个谓词“喜欢”和“打”各自的论元信息。

我 | 施事   喜欢 | 谓词   打 | 谓词性受事   篮球 。

我 | 施事   喜欢   打 | 谓词   篮球 | 受事 。

## 2.2 语义角色标注和句法分析的对比

句法分析的目的是分析句子的句法结构，通常以动词为核心，确定主谓宾定状补。当句子的语序发生变化时，句子的句法结构通常也会随之变化。与之相反，对于两个语义相同的句子，即使其句法结构不同，语义角色标注也会得到相同的语义结构。例如，图2中，三个句子的语序不同，表达方式不同，句法结构差异也很大，但是语义结构是相同。

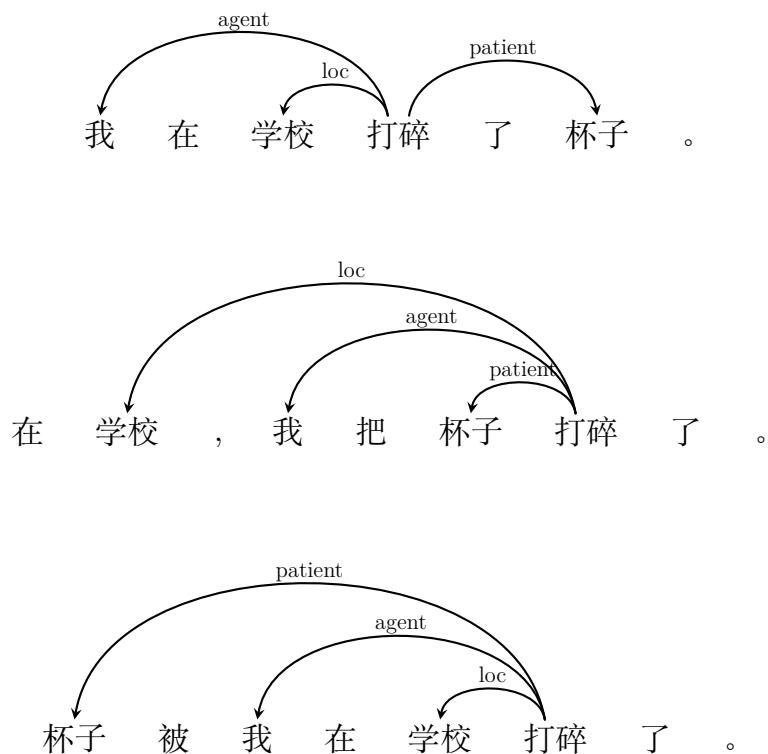


图 2: 三个句子的语序不同，句法结构不同，但是语义角色信息一致。

### 3 标注技巧和说明

(1) 规范中给出了一些典型的例子，请充分理解，举一反三。遇到困难，请尝试套用规范中的例子，并选用最合适的方式标注。遇到看不懂的句子，可以查百度 ([www.baidu.com](http://www.baidu.com))，先充分理解，再标注。请不要贸然标注，影响质量。

(2) 很多句子中作论元成分的往往是从句，需要根据规则确定一个代表词作为论元。

(3) 如果发现句子中有错别字，但不影响句子的语义理解 (例如与正确字有相同读音)，则请按照正确的字正常标注，提交之前在标注系统中的“错别字”复选框中打勾 (作为一条标注给报酬)。

(4) 如果发现句子中有和谓词或论元相关的分词错误，或者句子是古诗文、病句 (无法理解)、或具有真歧义，请直接选择标注系统中对应的按钮后直接提交，无需进一步标注。

(5) 请充分利用已有的词典资源。对于一些词或者字，不确定其含义或者词性，那么建议结合利用如下三个资源，进行搜索。

- hanyu (百度汉语): [hanyu.baidu.com](http://hanyu.baidu.com)
- baike (百度百科): [baike.baidu.com](http://baike.baidu.com)
- 5156 (无忧无虑中学语文网): <http://xh.5156edu.com/>

### 4 角色标签列表

本规范目前采用了 24 个论元角色，可以分为三大类，如图3所示。

核心角色是谓词所表达概念中的核心参与者，通常不可缺少。如果缺少了核心角色，那么谓词所表达的概念就不完整。值得注意的是，这并不是说核心角色不能省略。事实上，汉语表达非常灵活，可以随意省略各种语言成分，包括核心角色。由于有上下文语境，这种省略一般不会影响理解和沟通。考虑到核心角色的重要性，本规范要求标注者将省略的核心角色标注出来 (我们在第10节介绍具体规则)。

附加角色又称之为非核心角色或外围角色<sup>6</sup>，可以让谓词所表达概念更加详细和完整。

其他角色目前只有一个离合词角色，因为真实语料中存在很多离合词，例如“洗了澡、上了班”等，“洗澡、上班”本来是作为一个谓词，分开之后的字通常是语素词，无法通过每个字的含义得到谓词的意思。所以我们将离合词归为一类，规定为离合词角色。

角色标签的个数比较多，之后我们将对论元角色标签进行详细解释。

上述的语义角色对应的标签 (及英文和缩写) 如表1所示， (之后我们会加入和其他规范中使用标签的大致对应关系)。

---

<sup>6</sup> 北京大学的《现代汉语谓词语义角色标注语料库规范》称之为外围角色

论元角色标签	论元角色	例句	标注论元
agent	施事	我 [打] 他	我
co-agent	共同施事	我和他 [讨论]	我、他
expe (experiencer)	当事	天气真 [好]	天气
hidden-subject	隐藏主体	[吃] 饭了吗	隐藏主体
patient	受事	他被 [打] 了	他
pred-patient(predicate-patient)	谓词性受事	他 [喜欢] 打篮球	打
dative	与事	[给] 他书	他
relative	系事	这 [是] 他的书	书
hidden-object	隐藏客体	你 [吃] 了吗	隐藏客体
subj-obj(subject-object)	兼主体客体	温度计 [伸入] 水中	温度计
tool (instrument)	工具	用毛笔 [写] 字	毛笔
material	材料	用颜料 [刷] 墙	颜料
manner	方式	按计划 [执行]	计划
loc (location)	处所	在学校 [学习]	学校
beg-loc (begin location)	起点	从学校 [出发]	学校
end-loc (end location)	终点	[流] 入大海	大海
dir (direction)	方向	向西 [流]	西
time	时间	星期天去 [打篮球]	星期天
beg-tm (begin time)	开始时间	比赛七点开始 [进行]	七点
end-tm (end time)	结束时间	会议 [开] 到三点	三点
range	范围	在数学上 [有] 天赋	数学
cause	原因或目的	我因为爱你才 [撒谎]	爱
quantity	事量	我 [跑] 了一圈	一圈
separated	离合	我们 [见] 过面	面

表 1: 语义角色标签（及英文和缩写）汇总表

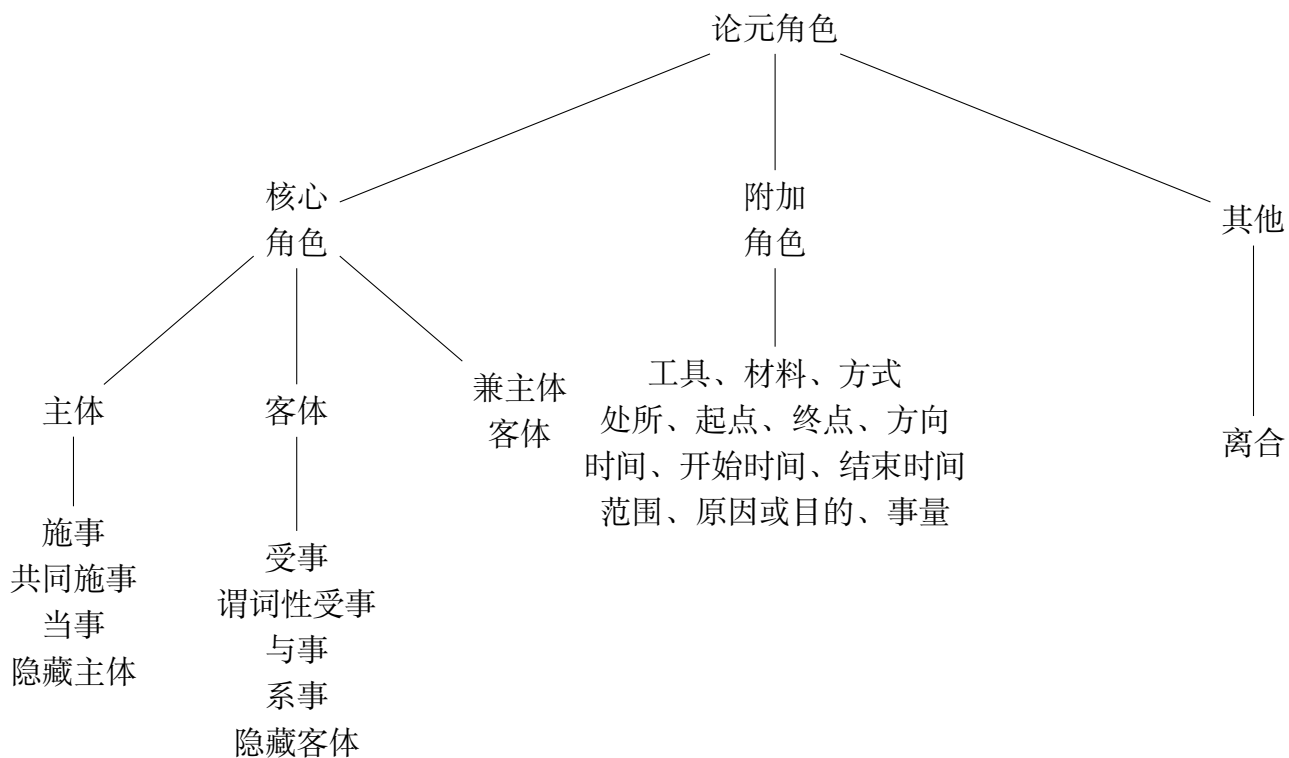


图 3: 所有论元角色列表和层次关系。

## 5 谓词选取

在标注页面中，我们会给出一个分好词的句子，同时用红色方框将唯一的一个候选谓词框出来。首先，标注者需要根据本小节中的规则，判断候选谓词是否属于我们规定的谓词范围。如果不是，直接选择系统页面中“不是谓词”的选项，然后提交即可。如果候选谓词是谓词，则继续进行论元标注，最后提交。

由于我们的标注系统每次只处理一个谓词，当句中包含多个谓词时，我们将逐一对每个谓词进行标注，从而获得完整的语义信息，如图4所示。如果句子中没有谓词，如“今天周五”，我们不会标注。

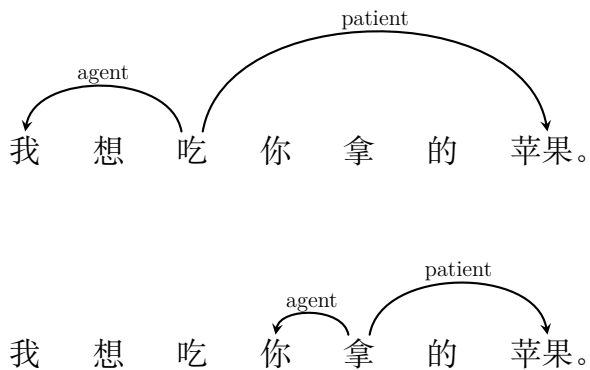


图 4: 句中有多个谓词的标注示例。

结合不同语言学理论和语义标注规范，我们认为动词和形容词是句子语义中最核心的单元，因此规定除了下面列出的一些情况，绝大部分动词和形容词都作为谓词。大的原则是，如果动词表达的语义比较虚，在句中主要起到满足韵律或语法结构的需求，则不作为谓词标注。

需要注意的是，名词化的动词也是需要进行标注的。例如“我们的决定是去爬山”中的“决定”也是需要当作谓词进行标注的，因为有强相关论元成分（核心论元成分）“我们”、“去爬山”。

另外，活用为动词或形容词的名词也需要作为谓词进行标注。例如“这件衣服很淑女”中的“淑女”，词典上只有名词的解释，但这里活用为形容词。又如“我粉你”中的“粉”名词活用为动词，我们规定类似的词作为谓词进行标注。但是是否将其判断为活用，我们需要根据词语及句子的实际意思进行判定，有些句子只是缺少谓词，比如“这个东西好评”中缺少了“给”这个谓词，“好评”并不是活用，我们不将“好评”作为谓词进行标注。

## 5.1 能愿动词通常不作为谓词

能愿动词在句中和其他谓词一起出现，表示客观的可能性、必要性和人的主观意愿等。我们将能愿动词分为如下几种情况：

- 表示可能：能、能够、会、可能、可以、可、得、得以。
- 表示必要：要、需要、应、应该、应当、该、该当、必须、须得、必得、得、一定。
- 表示意愿：愿、愿意、乐意、乐得、情愿、肯、要、愿、想、想要、要想、敢、敢于、乐于、希望、企图、懒得、甘于、试图。
- 其他：配、高兴、甘于、犯得着、犯不着、来得及、免不了、值得、好意思、准、许、允许、许可、容许、容易、难于、勇于、有助于、苦于、便于、宜于、善于、好、适于、忍心、不见得、甘愿、宁愿、宁可、总得、允准、免不得、须要、足以、甘心、怪不得、恨不得。

需要注意的是，能愿动词也可以作为一般动词单独使用，如图5，此时需要作为谓词进行标注。

此外，判断一个动词是否是能愿动词，需要满足下面的条件：

- 能愿动词后面会直接跟动词短语或形容词短语，例如“我想开开心心地工作”中“想”是能愿动词，后面“开开心心地工作”是动词短语；“我想天天都开心”中“想”是能愿动词，后面的“天天都开心”是形容词短语。
- 能愿动词后面不能跟一个完整的主谓宾从句，如“我希望他能早点毕业”中的“希望”后面是一个完整的宾语从句，不作能愿动词。
- 能愿动词后面不能跟“着、了、过”。
- 能愿动词在句子中常作状语，但与作状语的副词不同，它有“V 不 V”或“不 V 不”式。例如“他能不能来（V 不 V）”、“他不会不来（不 V 不）”等。

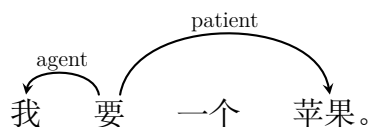


图 5: 此处“要”没有和其他动词一起出现，是“讨要”的意思，作为一般动词。类似的例子如“我希望我能早点毕业”。

## 5.2 形式动词通常不作为谓词

常见的形式动词有：进行、加以、予以、给以、给予等。形式动词后面通常跟另一个动词性谓词。我们规定不把形式动词作为谓词，而把后面的动词作为谓词，如图6。值得注意的是，形式动词也有普通动词的用法，如图7中的这种情况，就应该作为谓词处理。

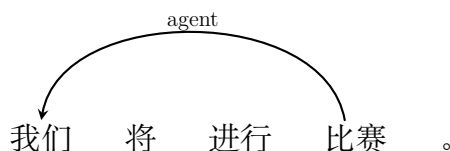


图 6: 形式动词“进行”不作为谓词，仅将“比赛”作为谓词。

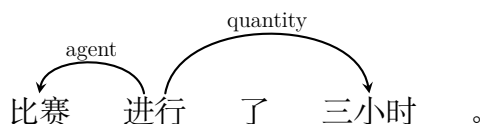


图 7: 此处的“比赛”是抽象名词，起到指称作用，没有明显的动词含义，不作为谓词，而“进行”作一般动词，因此作为谓词标注。

## 5.3 趋向动词是否做谓语的判断

常见趋向动词有“来、去、起、起来、下来”等。如果趋向动词和其他动词一起出现，在句中主要满足语法需求或是对其他动词的含义补充，通常不作为谓词标注，例如图8和9。

趋向动词可能会出现在动词前，也可能出现在动词后，是否作为谓词进行标注，我们分别对其做出了详细的判断规定。

(1) 趋向动词用在其它动词后。若趋向动词 V2 是动词 V1 的补语，则不作谓词，具体判断规则如下：

- 动词 V1+ 趋向动词 V2+ 宾语 N，N 是动词 V2 的宾语，即趋向动词 V2 和宾语 N 形成动宾结构，则趋向动词作谓词标注。例如：“走过邮局”中，邮局是过的宾



语，“过”作谓词。类似的还有“走出家门”、“走到学校”，其中“出”和“到”都作谓词标注。（“走”可以标注为 manner）

- 动词 V1+ 趋向动词 V2+ 宾语 N，N 只是动词 V1 的宾语，而不能作 V2 的宾语，则 V2 不作谓词标注。例如：“呈现出丰收景象”、“服下药片”中“出”、“下”不作为谓词进行标注，其中宾语 N 只能是动词 V1 的宾语。
- 动词 V1+ 宾语 N+ 趋向动词 V2，趋向动词 V2 没有实际意义，不作谓词。例如：“他回学校去”、“飞进苍蝇来”中的“去”和“来”。
- 动词 V1+ 趋向动词 V2，后面没有宾语，趋向动词不作谓词。例如：“苹果掉下来”、“跑起来”中的“下来”和“起来”不作谓词。

(2) 趋向动词用在其它动词前。不区分下面两种情况，均不作谓词标注：

- 充任连谓结构前一个直接成分表示目的，例如“我去买菜”，“我来看看”。
- 只起连接的作用，没有什么实在的意思，例如“我们打算用这个方法去帮助他”，“我们打算用这个方法来帮助他”。

值得注意的是，如果趋向动词没有和其他动词一起出现，如图10，则作为谓词标注。

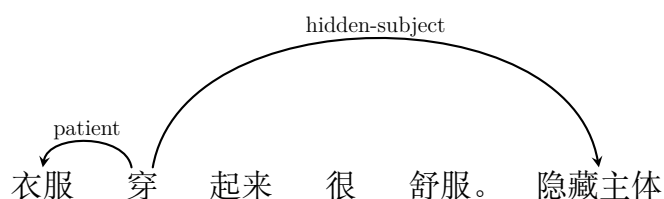


图 8: 此处的趋向动词“起来”无实义，满足语法需求。



图 9: 此处“下来”虽然有一定的含义，但和动词“掉”一起出现，补充动词的含义，因此不作为谓词。

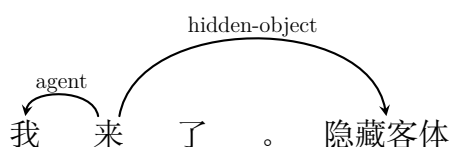


图 10: 此处的“来”没有和其他动词一起出现，作谓词。



## 5.4 使役动词

使役动词，如“使、让、令”等，通常标记“致使语法结构”在句子中的范畴，描述前后内容的因果关系，其动词义被虚化了，因此不作为谓词标注，如图11。注意，有时使役动词也是可以作一般动词使用，需作为谓词标注，例如图12。

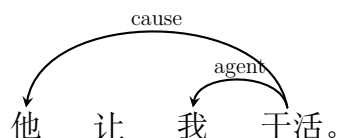


图 11: “让”是使役动词，“他”是“我干活”的原因。

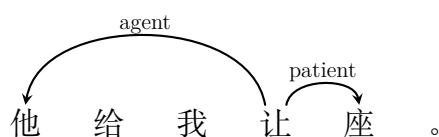


图 12: “让”在这里作普通动词进行标注。

## 5.5 “是/有”等不作为谓词的情况

在现实生活中“是”字结构的句子很多，但是我们只把意义的“是”作为谓词标注，如图13。相反，如果句子中有其他谓词和“是”一起出现，则不作为谓词标注，如图14。

“有”字和“是”字类似，我们只将有意义的“有”字作为谓词进行标注。“有”（“没有”、“无”类似）不作为谓词进行标注的情况：

- 形容词“有”作定语：
  - 用在“人”、“时候”、“地方”前面，表示一部分或者不定指，跟“某”作用相近，如“没有人喜欢你”、“有传闻说你公司倒闭了”。
  - 表示过去有一段时间，如“有一年”。
- 副词“有”：“我有爱过你”、“水有一米深”（“有”修饰量词“一米”）。
- 表示比较的情况：如“她没有你好看”。

其余的情况可作为谓词，如“桌子上有一朵花”、“他有机会上台演唱”和“家里有客人吃饭”等。

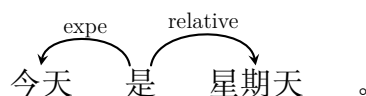


图 13: 这里的“是”为系动词。类似的例子还有“这本书是爸爸的”

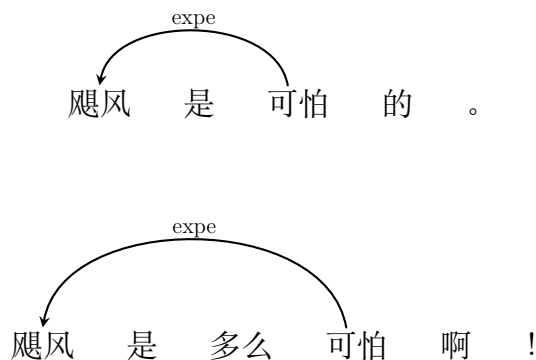


图 14: 这里的“是”可以理解为助词，通常是韵律、语法的需要，或者表达强调（例如“我是不喜欢你，怎么了”“苹果是红色的”“这个电脑是旧的”）等。

## 5.6 仅起到指称作用的动词不作为谓词

很多谓词在某句话中仅仅起到指称作用，即主要作为名词使用，且句中没有与该词强相关的论元成分（核心角色），则不作为谓词标注，例如图15、16。判断一个动词是否属于这种情况的一个重要线索是看句子中是否存在和这个词有强相关的论元成分（核心角色）。如果有，那么通常作为谓词标注，如图17。

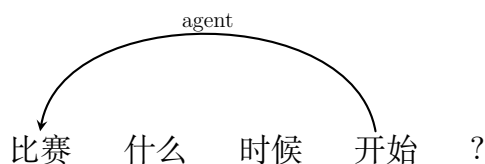


图 15: 这句话中单个动词“比赛”作为主语，起到指称作用。

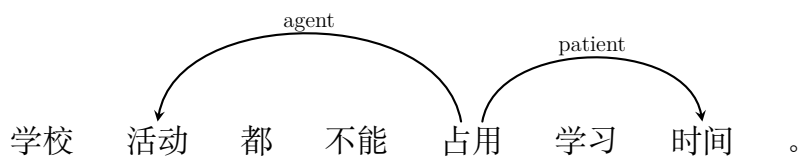


图 16: “学习”修饰“时间”，作名词，且句中也没有强相关的论元成分（核心论元成分），因此不作为谓词。类似，“活动”也不是谓词。相反，如果是“学生的学校活动...”中，“活动”作为谓词，“学生”作为“活动”的施事。

## 5.7 常用口语不作为谓词标注

生活中也有很多常用口语，比如问候语，是人们生活中最常用的重要交际语言，没有谓词性作用，不进行标注。例如图18、19和图20。

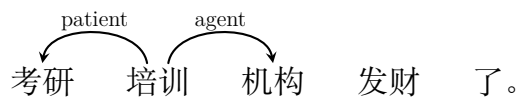


图 17: “考研”这个词在复合名词短语中作名词使用，且没有强相关的论元成分，故不作为谓词；而“培训”则有强相关的论元成分（核心论元成分），故作为谓词。



图 18: “你好”是问候语，故“好”不作为谓词标注。

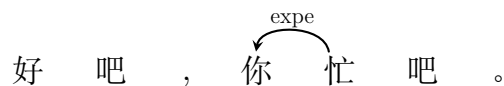


图 19: ”好吧”是常用口语，表示感叹，“好”不作为谓词标注。

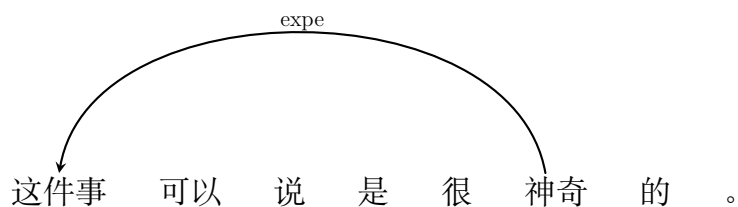


图 20: “可以说”算是插入语，“说”在句中没有实意，不作为谓词标注。

## 6 如何从句子片段中选择一个词作为论元？

在本规范中，我们规定谓词的一个论元对应一个词，即要求标注者从句子中确定唯一的词代表一个论元。通常情况下，论元对应一个句子片段，下面我们给出了相关的规则，从片段中选择唯一的代表词作为论元（并列的情况时通常为多个代表词）。基本的原则是，选择片段中语义上最重要的词作为代表词；但如果语义上重要程度无法区分时，则考虑句法上的修饰关系。

具体而言，句子片段中语义上重要程度的顺序为：

- 谓词 > 其它词。谓词（通常是动词）在语义上最重要，例如21。
  - 句子片段中动词和形容词一起出现时，优先选择动词为代表词作论元。比如句子片段“工作很努力”，选择“工作”为论元代表词，“这批基金操作稳健”中选择动词“操作”为代表词而不是形容词“稳健”。
  - 句子片段中有多个并列的动词时，每个谓词都作为论元标注。例如，句子“我喜欢看书、跑步”中，“看书”和“跑步”都是“喜欢”的 pred-patient 论元；又例如，句子“他说，如果你不走，我就走”中，两个“走”都是“说”的 pred-patient 论元。
  - 句子片段中有多个不并列动词时，我们标注最外围的一个动词，即在该片段中不再作为其他谓词的论元的谓词。例如句子片段“小明认为经济发展迅速”，选最外围的谓词“认为”为代表词，其中的“发展”是“认为”的客体论元，而“认为”不作其它谓词的论元。
  - 句子片段中有多个不并列动词，且前一条规则“最外围”不适用时，优先标注先出现的谓词。例如句子片段“出口到一指定目的地”中，“出口”和“到”是不并列的两个谓词，且两者都是“最外围”谓词，优先选择先出现的谓词“出口”为代表词。
  - 句子片段中的名词和动词（或形容词）都可以是谓词的客体论元时，我们优先选择从句中的动词为代表词。例如“屋内有光闪动”，“有”的客体标为“闪动”。
- 实词 > 虚词。例如在介宾短语中，因为介词是虚词，我们选取宾语部分作为论元成分。此外，在图22中，因为“们”是助词，即虚词，我们将“老人”选作论元成分。
- 由多个词构成的名词复合短语，根据以下几个原则（前面的原则优先级更高）来处理。
  - 如果有一个词可以基本涵盖整个短语的含义，则以这个词为代表词；
  - 如果某一个词很明显比其他所有词都更重要，则选这个词为代表词；
  - 其他情况，将最后一个词作为代表词。

下面给出一些例子：

- “白求恩大夫救死扶伤”，“救死扶伤”的施事标在“白求恩”上而不是“大夫”。如果标注“大夫”，句子意思是“大夫救死扶伤”，和原义的“白求恩救死扶伤”不符。类似的例子还有，“江苏/省”中，“江苏”更重要，“他自己

很马虎”中，“他”更重要。此外，还有图26，这样的标注使得句子意思更加清晰明确。

- “阿里巴巴 网络 技术 有限 公司”、“北京 京东 世纪 贸易 有限 公司”等以名称标识的实体标在明确的词上，标注如图23所示。
- “京津 城际 铁路”、“上海 纺织 有限 公司”等出现较长的名词复合短语，我们只标在最后一个词上，标注如图24所示。
- 方位词短语中，规定“方位词 > 地点或时间词”，例如图25。又例如“从开学以来”中“以来”为时间（time）论元。
- 数量词短语中，规定“量词 > 数词”，例如句子“我走了一趟”中“一趟”为谓词“走”的 quantity，当“一趟”分开时，我们标在量词“趟”上。
- 表示范围的短语中，如“在学习上”或“在学习方面”，“学习”为论元代表词，而“上”或“方面”不作为代表词。
- 并列名词短语中，将并列的多个代表词同时标注为论元。例如，句子“我和爸爸一起跑步”中，“我”和“爸爸”都是 agent 论元。
- 句子片段中，如果存在重复表达同一实体的情况，我们只标注谓词直接修饰的主客体。具体的做法按照第13.12节的方法进行判断。类似的例子还有图26和27。还有一种情况，例如“他是一个善良的人”，当“善良”作谓词时，只标注“他”为当事，即“A 是 B 的 C”结构，如果“B”作为谓词，则优先标注“A”作为“B”的论元成分，因为 A 往往比 C 更具体。类似的还有“这是一本有趣的书”、“这是美丽的风景”，都标注“这”作为谓词的主体论元。
- 很多情况下，形容词和谓词一起出现，并作为状语或补语修饰谓词。在状语或补语位置的形容词我们也是统一作谓词进行标注，当事通常标注到动词上。例如“他很努力地工作，他工作很努力”中，“努力”标为谓词，“工作”标为 expe，如图28。类似的还有“我们很艰难的生活”、“我们生活很艰难”、“食物的获取很困难”、“获取食物很困难”，我们取“生活”作为“艰难”的论元成分词，获取“作为“困难”的论元成分词，都标为 expe。
- 其它情况下，名词短语中通常最后一个词作为论元代表词。例如图29和图30。



图 21: “不许”的客体论元是“船员离开船”，选取子句中最重要动词“离开”作为论元。

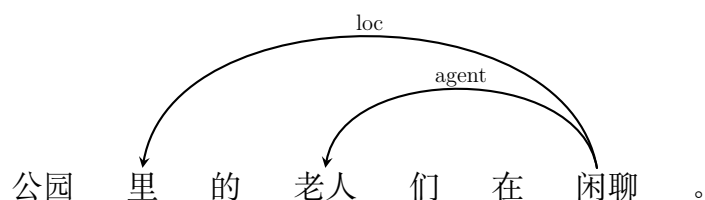


图 22: 类似的还有甲方, 甲队, 标在“甲”上。



图 23: 真实场景中,“阿里巴巴网络技术有限公司”一般简称为“阿里巴巴”。在标注时,我们将“阿里巴巴”选为论元代表词。

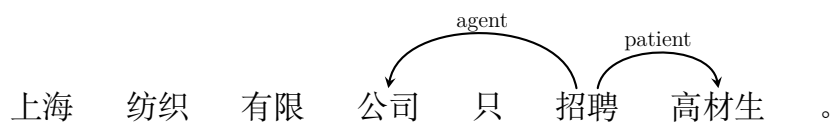


图 24: 将名词复合短语“上海 纺织 有限 公司”的最后一个词“公司”作为论元代表词。

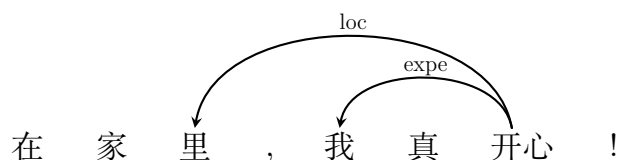


图 25: 标注在“里”上,因为“里”是方位词,还有类似的例子,如“房子前、桌子上”等, location 标注在“前”、“上”。

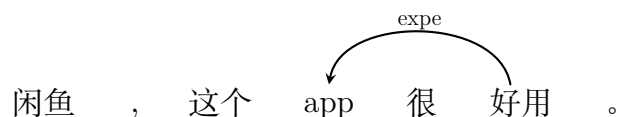


图 26: “闲鱼”和“这个 app”属于重复表达选择直接修饰的“app”作为“好用”的主体。

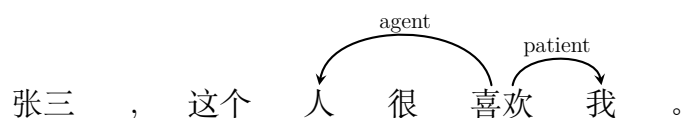


图 27: “张三”和“这个人”重复表达,标注直接修饰的词“这个人”作为主体。

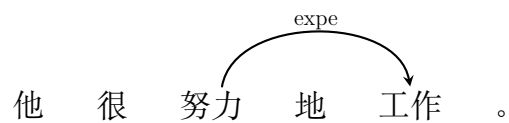


图 28: “努力”标为谓词时, 当事标在谓词“工作”上。

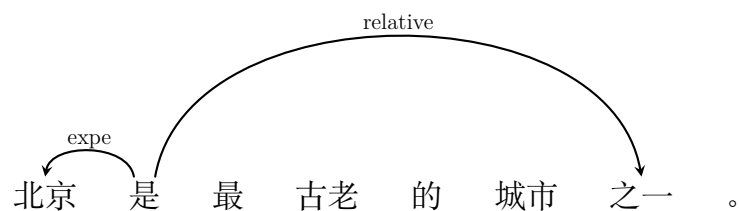


图 29: 按照句子表述含义, 更想说明的是“之一”。

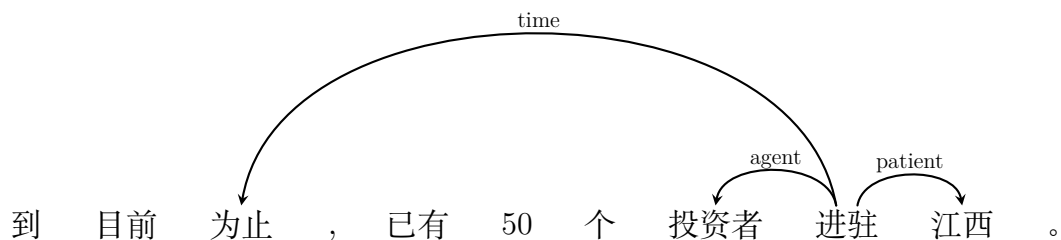


图 30: 从整体理解句子的意思, 即不将“目前”和“为止”分开理解, 把表示时间的从句“到目前为止”看成一个整体, 选择最后一个词作为论元词。

## 7 主体角色

### 7.1 施事 (agent)

施事指谓词所表达的动作或行为的发出者、事件的引起者（引起/导致/出现）或经历者。例如，“我喜欢你”这个句子中，“我”是“喜欢”这个行为的发出者。例如图31和32。

值得注意的是，我们规定施事是范围最广的一种主体角色，如果其他主体角色都不适用，均标注为施事。

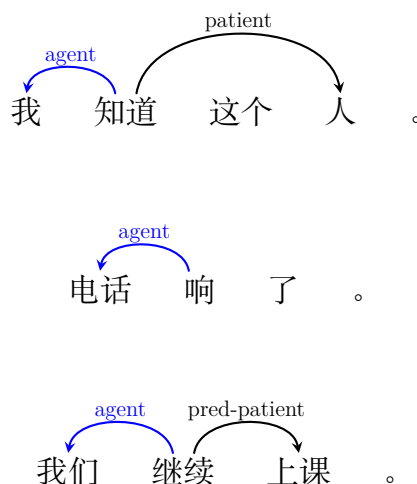


图 31: “我”是“知道”的经历者；“电话”是“响”的事件的经历者；“我们”是“继续”的发出者。



图 32: 类似的还有“他的失误导致事故”、“她出了车祸”、“他发生意外”等。

### 7.2 共同施事 (co-agent)

共同施事是指谓词所表达的动作或行为的其他发出者。有些谓词所表达的动作或行为需要多个人共同参与才能完成，为了和其他情况区分，<sup>7</sup>我们规定句子中首先出现的施事标注为 agent，其他后续的施事标注为 co-agent。例如图33。值得注意的是，当谓词所表达的动作或行为需要多个人共同参与才能完成，而句中只出现一个施事时，将其标为 agent，并标注隐藏主体角色以表示省略的其它施事参与，例如图34。

<sup>7</sup>其他情况主要指一个谓词有多个并列的施事，但是这个谓词不一定需要多个施事参与。例如，“讨论”、“开会”通常需要多个人参与，而“邀请”、“说话”则不一定需要多个人参与。我们参考北大规范，进行这样的区分，希望未来会对语义理解产生帮助。



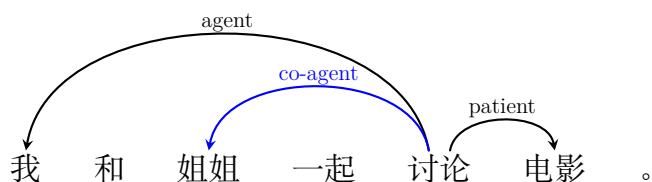


图 33: “讨论”需要就某一问题交换意见，一个人无法完成。

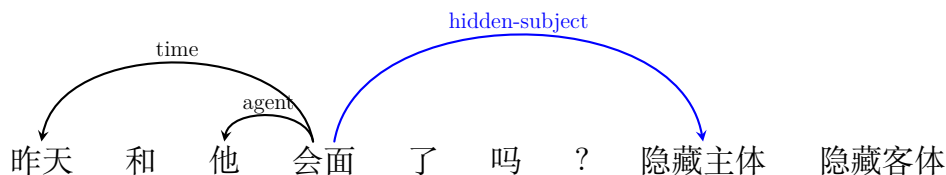


图 34: “会面”需要多个人共同参与才能完成，句中只出现了一个主体论元“他”，将首先出现的“他”标为 agent，同时标出隐藏主体，表示 co-agent。

### 7.3 当事 (experiencer, expe)

标注当事的情况：1) 谓词是形容词（包括作状语或补语的情况）；2) 谓词“是、当、作为”、“有”或“在”等。例如图35、36、37、38、39和40。值得注意的是，谓词“拥有、具有、占有、持有”和“存在”等是谓词“有”和“在”的延伸含义，其客体论元都标注为当事，如图41。

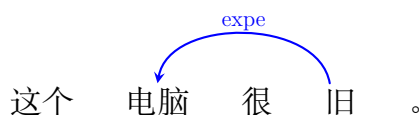


图 35: 类似的例子还有“我喜欢穿旧的衣服”中的“旧”。

## 8 客体角色

客体角色和主体角色对应。

### 8.1 受事 (patient)

受事是指因谓词所表达的动作或行为的承受者，通常可以由介词“被”引出。在实际句子中，作为客体角色，受事对应的主体角色通常为施事，例如图42。当存在多个列举的论元，并用“等”结尾时，“等”前后的部分都需要标注，例如图43。

值得注意的是，我们规定受事是范围最广的一种客体角色，如果其他客体角色都不适用，均标注为受事。

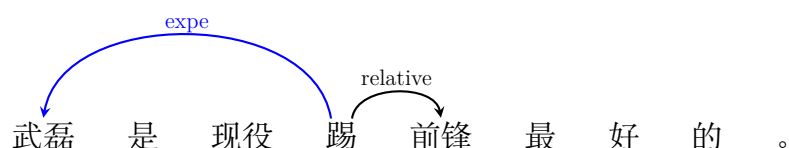


图 36: “踢”在此不是表示“脚踢”的动作,而是代表比赛中担任的角色,有“是”的意思。

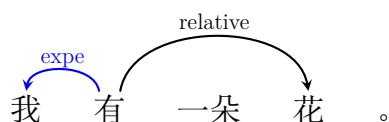


图 37: 类似的例子还有“我要当班长”中的“当”,“我是班长”的“是”。

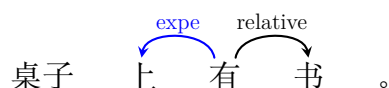


图 38: 类似的例子还有类似的:“花园里有一个人”的“花园里”,都统一认为是核心角色当事。

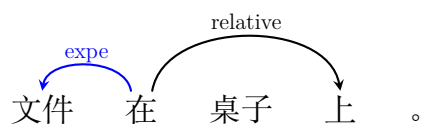


图 39: 此句中“在”是动词,意为“存在”,作谓词标注。需要注意的是,“你在干什么”中“在”是副词,“事情发生在去年”中“在”是介词,其中“在”均不作谓词标注。

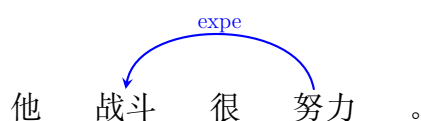


图 40: “努力”是谓词“战斗”的补语,“努力”作谓词进行标注时,标注其修饰的“战斗”作为当事。

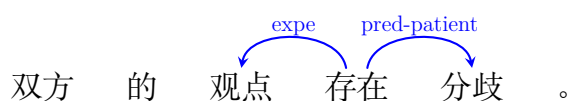


图 41: “存在”表示事物持续地占据着时间和空间时,有“有”的延伸含义。句中存在“分歧”的主体论元“双方”,因此主体客体分别标为 expe, pred-patient。

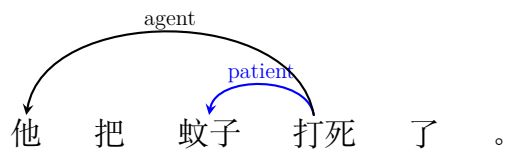
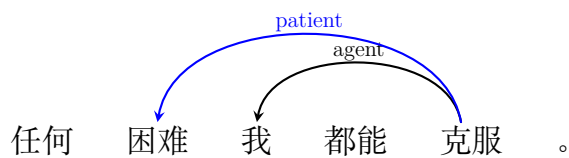
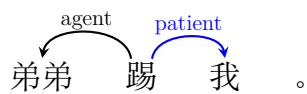


图 42: 受事的典型例子：谓词的承受者。

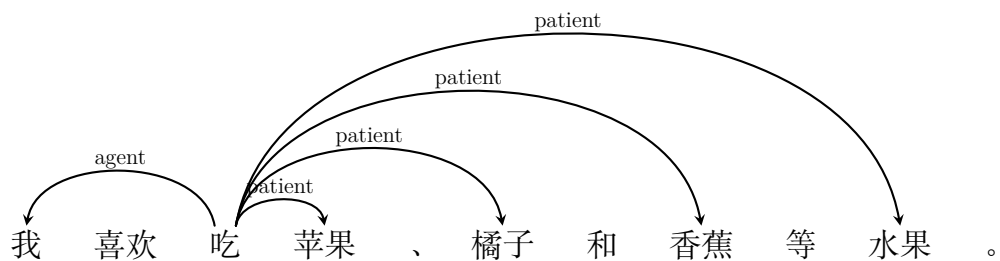


图 43: “苹果”、“橘子”、“香蕉”、“水果”都需要作为谓词“吃”的客体论元进行标注。

## 8.2 谓词性受事 (predicate-patient, pred-patient)

谓词性受事是指客体论元为形容词、动词、动词短语或从句的情况，主体可以对应施事也可以对应当事。例如图44、图45和图41。

需要注意的是，当客体论元为名词性动词或只起到指称作用的动词时，参考第5.6节，若句中有与客体论元强相关的论元成分时（客体论元作谓词时句中存在核心论元），标注为谓词性受事，否则标注为受事，例如图46。

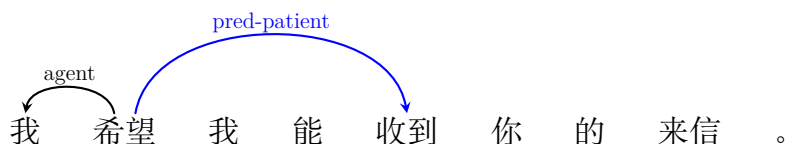


图 44: 将从句标注为谓词性受事，类似的还有“我知道你喜欢我”中的“喜欢”。

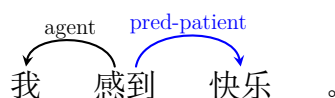


图 45: 将形容词标注为谓词性受事。

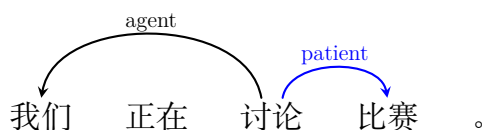


图 46: “比赛”作谓词时，其主体论元不一定是“我们”，句中没有与其强相关的论元成分（核心角色），则不标注为谓词性受事。

## 8.3 与事 (dative)

与事指与谓语所表示的动作或行为没有明显施受关系的客体论元，本规范中限定与事为双宾语结构中的间接宾语。例如“送给她一朵花”中的“她”。其他例子如图47。此外，当双宾语结构中的只出现了间接宾语（或直接宾语），省略直接宾语（或间接宾语），标为隐藏客体，例如图48和图59，其中，图59是省略了间接宾语。

## 8.4 系事 (relative)

系事是与当事对应客体角色，当谓词是“是、当、作为”或“有”等时（当事人的第二种情况），客体论元标注为系事。例如图49和图50。

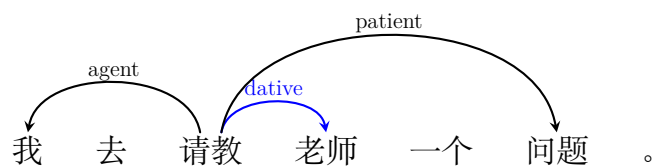


图 47: 该句子是典型的双宾语结构，其中”老师”是间接宾语。

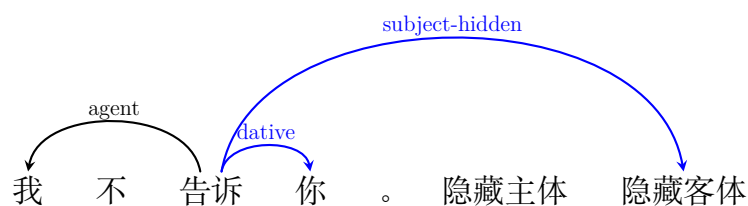


图 48: 这句话是双宾语结构，“你”是“告诉”的间接宾语，省略了直接宾语，即省略了受事，标为隐藏客体。

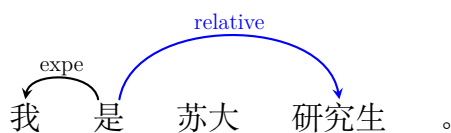


图 49: 类似的例子还有“我以前当过班长”、“我有一朵花”。

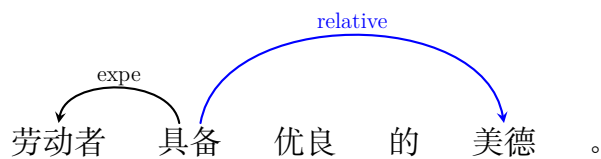


图 50: “具备”类似于“具有”，有“有”的意思。

## 9 兼主体客体 (subject-object, subj-obj)

有一些复合动词 (V)，如“伸入”、“投入”、“列入”、“译成”、“载入”、“写成”等，事实上由两个动词 (V=V1V2) 复合而成。这样的动词往往需要三个核心角色 r1、r2、r3，其中 r1 为 V1 的主体，r3 为 V2 的客体，r2 为 V1 的客体兼 V2 的主体。此时，我们将 r2 称为兼主体客体<sup>8</sup> (subj-obj)。这样的动词通常以把字句或被字句出现，如图51图52所示。

其中，r1 可能会被省略，如图53所示。注意，如果 r2 被省略，我们规定不标注隐藏主体或隐藏客体，如图54。

注意，图55中的例子不属于兼主体客体的情况。

我们将例子按照谓词结尾的字分为以下几类（以后将不断补充）：

- 以“入”字结尾：伸入，插入，投入，列入等；
- 以“成”字结尾：写成，译成，做成，变成等；
- 以“作”字结尾：写作，当作，认作等。

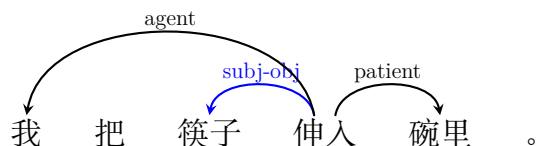


图 51: “伸入”是由谓词“伸”和“入”复合而成的动词，这句话中可以理解为“我 伸 筷子 入 碗里”，存在“我”、“筷子”和“碗里”三个核心论元成分，分别对应我们定义中的 r1、r2 和 r3。谓词“伸”的主体论元是“我”，客体论元是“筷子”，谓词“入”的主体论元是“筷子”，客体论元是“碗里”，我们将“筷子”标注为兼主体客体。

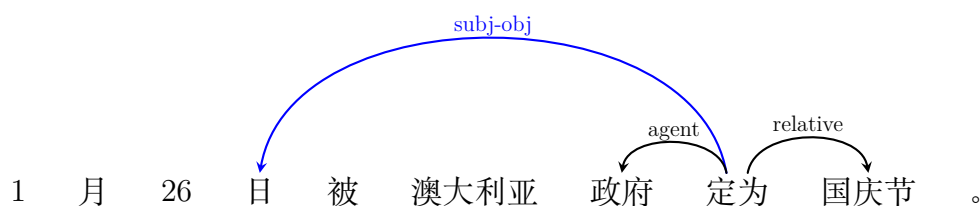


图 52: 这句话可以理解为“澳大利亚政府 定 1月26日 为 国庆节”。谓词“定为”中的“定”和“为”有不同的主体论元，论元成分“澳大利亚政府”、“1月26日”和“国庆节”分别对应上述定义中的 r1、r2 和 r3。“定”的主体是“澳大利亚政府”，客体是“1月26日”；而“为”的主体是日期“1月26日”，客体是“国庆节”，我们将“1月26日”标注为兼主体客体。

<sup>8</sup> “我把你当作亲人”这句话中，中文 frame 将标注“你”为 arg1，“我”标为 arg0，“亲人”标注为 arg2；北大规范中将“把你”标为当事，“我”不进行标注，“亲人”标注为系事。

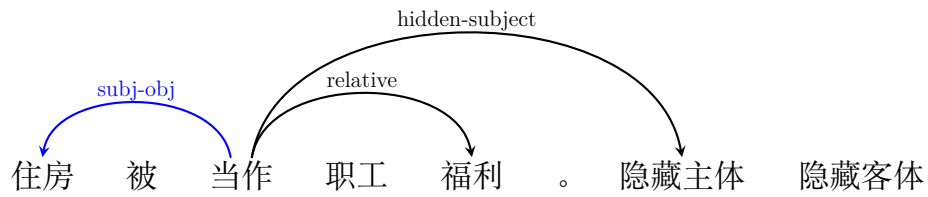


图 53: “当”的主体论元被省略，客体论元是“住房”，“作”的主体论元是“住房”，客体是“福利”，我们将在谓词“当作”中既做主体论元又作客体论元的“住房”标注为兼主体客体。

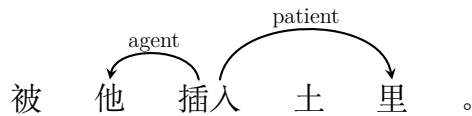


图 54: 句中缺少了插入土里的东西 r2，因为它既是“插”的客体，又是“入”的主体，我们暂时不标注 r2。

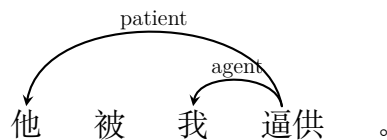


图 55: 虽然“逼供”也是复合而成的词，“逼”的主体论元是审问者，“供”的主体论元是被捕者，同时又是“逼”的客体论元，但是“逼供”这个词是不及物动词，“供”没有客体论元，因此“逼供”不需要三个核心论元，我们直接将“他”标为客体角色受事。

## 10 主体或客体隐藏

在真实语言交流过程中，为了避免重复并突出新的信息，人们通常会省略一些语言成分，包括核心论元角色。交流者可以根据上下文语境，自动将缺失的成分补全。

为了更好的理解句子的语义结构，我们要求将句子中省略的核心论元标注出来。为了降低标注难度，<sup>9</sup>我们只规定了“隐藏主体”和“隐藏客体”这两个标签，不要求标注者进一步区分更细的角色。例如图56和57。需要注意的是，如果句子中明显省略了一个论元成分，我们标省略主体或省略客体，例如图58。如果句子中省略了多个主体论元或多个客体论元，我们也只标一个隐藏主体或隐藏客体，例如图59。

另外，当句中谓词有不及物动词词性也有及物动词词性，我们需要根据句子的意思来判定是否进行标注隐藏客体，例如图107。不能明确判断是否缺少客体时，优先标注隐藏客体。如“他在认真学习”中的“学习”，虽然“学习”后面没有客体论元，但优先判断该词有隐藏客体。

需要注意的是，列举多个论元时，虽然有省略，没有列出所有可能的论元，但已标注了必不可少的核心论元，不再标注隐藏主体或客体。例如“我不单单喜欢香蕉和苹果”，根据句子的意思可以看出喜欢的不仅只有“苹果”、“香蕉”，但是也不再标注隐藏客体，其它例子如图60。

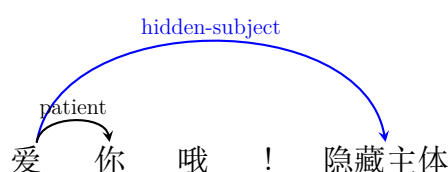


图 56: “某个人”爱你，省略了“爱”的施事，标注为隐藏主体。

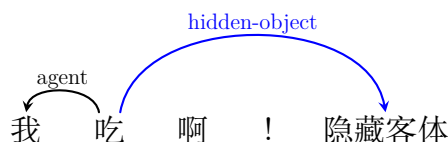


图 57: “吃”后面省略了吃的受事，标注为隐藏客体。

## 11 附加角色

### 11.1 工具 (tool)

工具是指完成谓词动作或行为所使用的器具，一般可以用“用、拿”等介词引出，例图61。不用介词引出的情况如图62。

<sup>9</sup>事实上，有时候也无法细分缺省成分的角色。如“我在看”中，仅根据当前句子，无法确定省略的是“受事”还是“谓词性受事”。



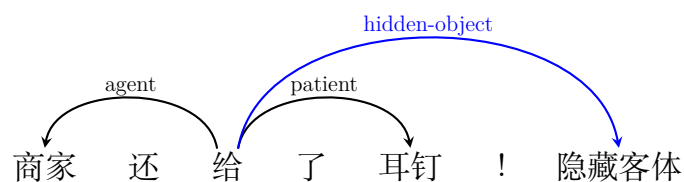


图 58: 句中省略了“给”后面的间接宾语，即省略了与事，标为隐藏客体。

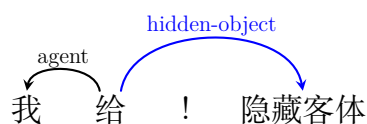


图 59: 这句话可能是“你还给我钱吗”的回答，“给”后面隐藏了与事“你”和受事“钱”两个客体论元，但只标注一个隐藏客体标签。

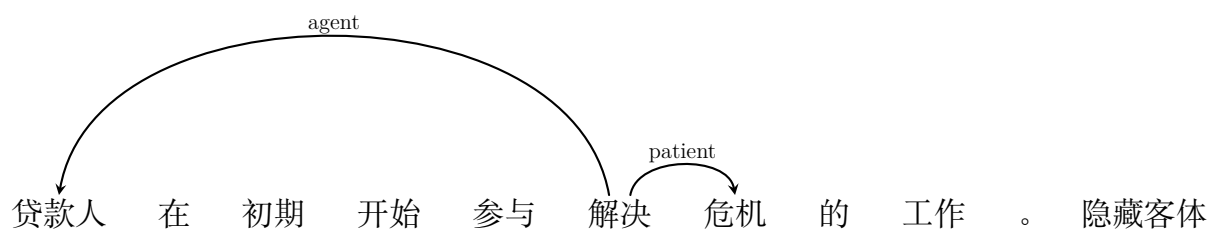


图 60: “解决”的主体论元可能不止贷款人，但句中已经存在必不可少的主体论元，所以不再考虑标隐藏主体。

需要注意的是，在句子中缺少主体论元时，部分可作工具的词也可标作主体论元成分。例如图63，标注时无需过分联想句子的深层语义，而只根据句子本身体现出来的语义进行标注。

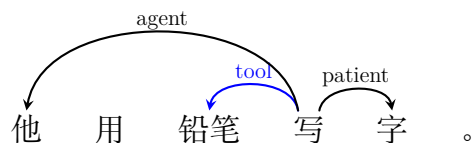


图 61: 写字的工具是“铅笔”，用介词“用”引出。

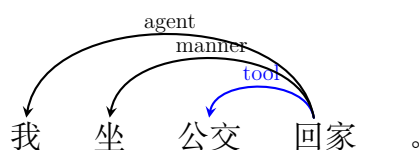


图 62: “公交”是“回家”的工具，“坐”是“回家”的方式，需要注意的是，当“坐”为谓词时，“公交”是“坐”的客体角色，标为受事。

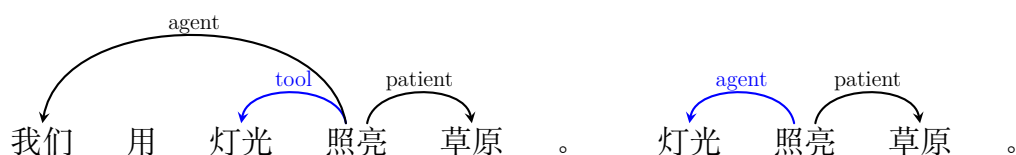


图 63: 下图中没有提到“我们”，理解为“灯光”自身的光照亮了草原。类似于“太阳照亮大地”中，“太阳”是“照亮”的施事。

## 11.2 材料 (material)

材料指谓词动作或行为在发生过程中所使用并且消耗的物品。材料和工具很容易混淆，其区别在于：工具在谓词作用过程中不发生显著的性状（性质或状态）改变。例如“用铅笔写字”中，“铅笔”不会有显著的性状改变。与之相反，材料在谓词作用过程中会发生显著的性状变化，甚至转化成新的物质或形态，或被完全消耗掉，如图64。

## 11.3 方式 (manner)

方式指谓词动作或行为发生过程中所采取的方式、方法、姿态、形式，通常可以由“用、以、按”等介词引出，可以进一步扩展为“用/以……方式/方法”来修饰谓词，例如图65和图66。

需要注意的是，像“轻轻地”“慢慢地”等词只是表示谓词的程度或状态，不标注为 manner。

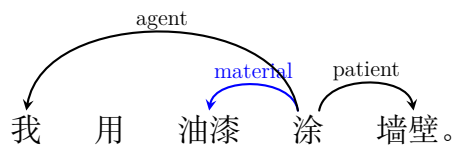


图 64: “油漆” 在涂墙壁的过程中被消耗掉。

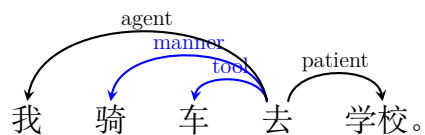


图 65: 去学校的方式是“骑车”，类似的例子有“西瓜按斤卖”中“斤”为方式论元。

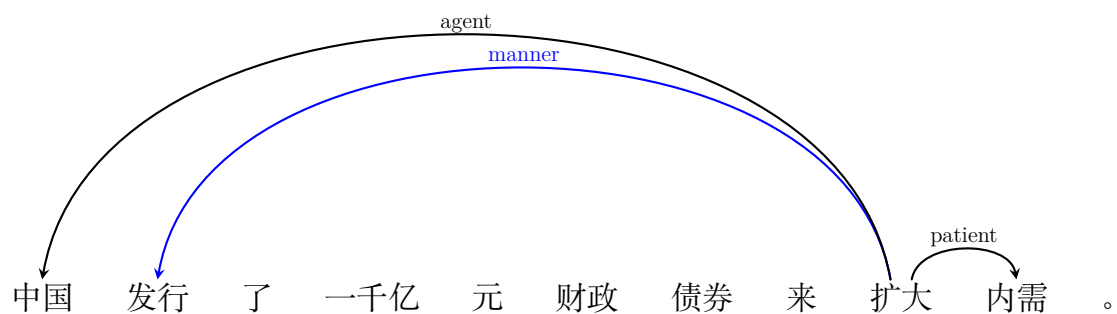


图 66: “扩大内需”的方式是“发行一千元财政债券”，需要注意的是，当“发行”为谓词时，“扩大内需”是“发行”的 cause。

## 11.4 处所 (location, loc)

处所是指谓词动作或行为发生的空间或地点，是空间中的某一个位置。例如图67、68和69。

需要注意的是，如果地点只是用来修饰主客体，那么我们不标注地点为谓词的地点。例如图70。

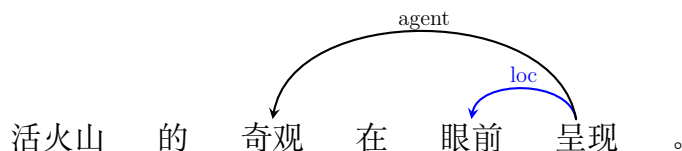


图 67: 奇观出现在眼前，“眼前”顾名思义是眼睛前面的意思，是一个空间位置。

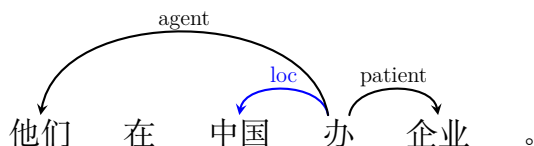


图 68: 中国是一个地点名词。

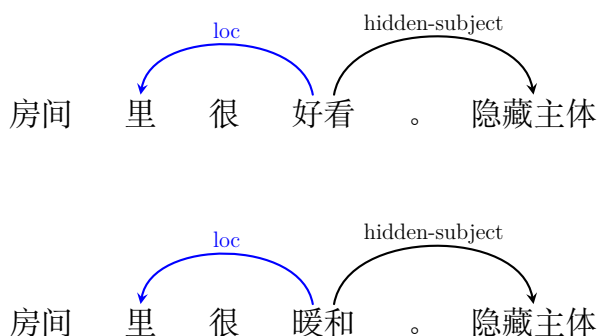


图 69: 选择方位词“里”为地点论元。另外“房间里好看”指房间里的布置好看，而“房间里暖和”指在房间里人感到暖和，在这个例子中有推理的成分，但是要适当推理。

## 11.5 终点 (end location, end-loc)

终点指谓词动作或行为发生的结束地点。谓词通常是直接描述物体位移的动词（如“回家”），或造成物体空间位置改变的动词（如“搬”）。例如图71和72。句中同时出现起点和终点的情况，如图73。

## 11.6 起点 (begin location, beg-loc)

起点是指谓词动作或行为发生的初始地点，和终点对应。例如图74。



图 70: 这里的“地上”主要是用来修饰“尸体”的地点,“冰凉”是个形容词,不需要标注地点。

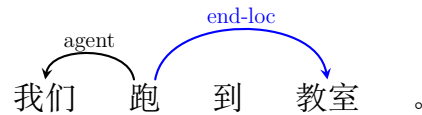


图 71: “跑”是位移动词,“教室”是“我们”的最终目的地。

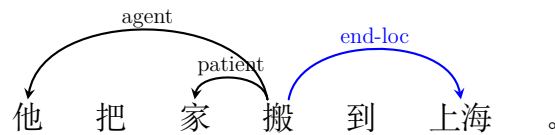


图 72: “家”的位置空间发生了改变。

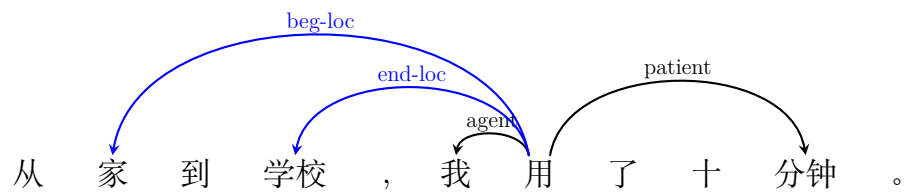


图 73: 同时标注起点和终点的情况。

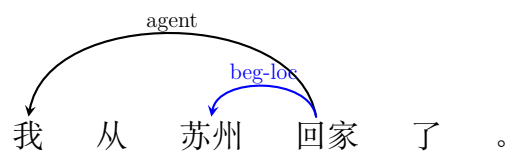


图 74: “回家”是一个位移动词,直接描述“我”的位移,标在实词“苏州”上。

## 11.7 方向 (direction, dir)

方向是谓词动作或行为发生的空间上或抽象意义上的方向。方向一般可由“向、朝”等介词引出。例如图75和76。

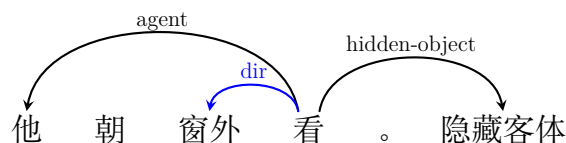


图 75: “窗外”是“看”的方向。

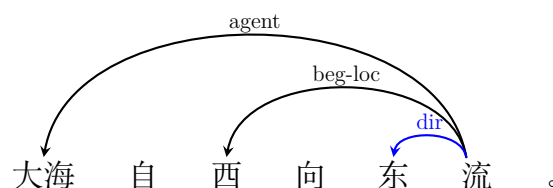


图 76: “自”后面引出的是起点，“向”后面引出的是方向。

## 11.8 时间 (time)

时间指谓词动作或行为发生的时间，如“星期天、今天、近日”等。例如图77、78、79和80。有的时间词是局限于句中某个谓词的时间，而有的时间词范围比较大，需要根据语义来判断，例如图81。

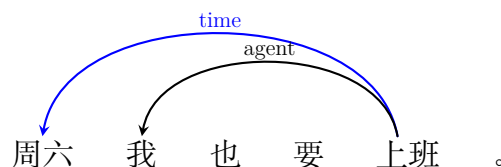


图 77: 周六是“上班”的时间。

## 11.9 结束时间 (end time, end-tm)

结束时间是谓词动作或行为发生的结束时间点，一般是由“到”等介词引出。如图82。

## 11.10 开始时间 (begin time, beg-tm)

开始时间指谓词动作或行为发生的起始时间点，一般是由“从或自”引出，主要是有“从……起”等。例如图82。

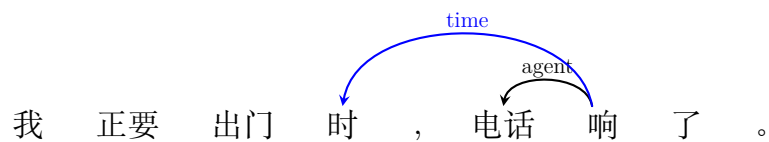


图 78: 句子片段“正要出门时”中，“正要出门”是修饰“时”的定语，我们规定选择“时”作为论元。

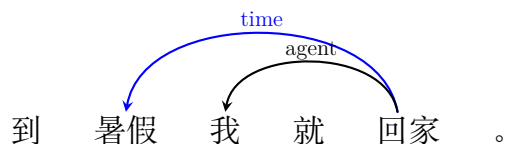


图 79: “暑假”是“我回家”发生的时间。

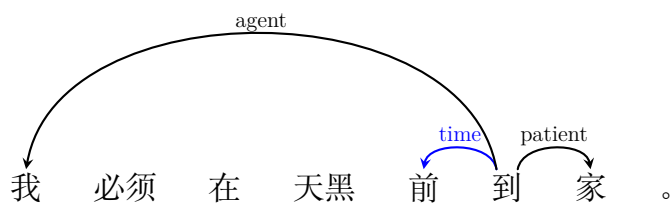


图 80: 类似的有“自开学以后”，将“以后”作为论元代表词。

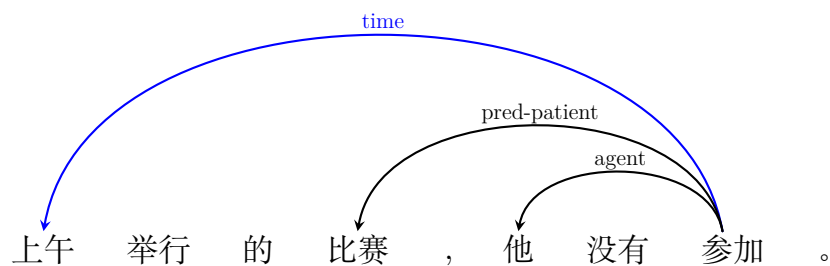


图 81: 根据句子的意思，可以很明显的看出“没有参加比赛”和“举行比赛”的时间都是“上午”，“上午”需要标为“参加”的时间论元。

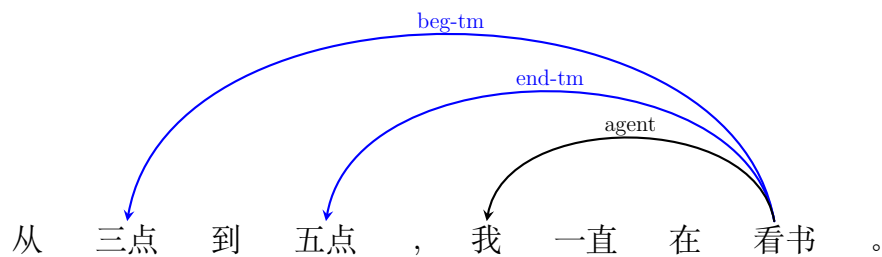


图 82: 同时标注开始时间和结束时间的情况。

### 11.11 范围 (range)

范围指谓词动作或行为所涉及的方面或范围。一般可以用“在……方面”来对范围论元进行提问。例如图83。范围不局限于“在……方面”结构，例如“在半决赛中，我拿到两分”句中，“半决赛”可以是限制谓词“拿到”的范围。类似的还有“在某电影中担任主演”，“电影”也是“担任主演”的范围。

需要注意的是，**修饰主体或客体角色的成分，即主体或客体本身所属的范围，通常是用于比较，不标注为范围**。例如“在所有同学中，他战斗最努力”中，“所有同学”是主体角色“他”所属的范围，不标注为谓词“努力”的范围。“他接受了两项任务中困难较小的一项任务”，“两项任务”是客体角色“任务”所属的范围，不标注为谓词“接受”的范围。

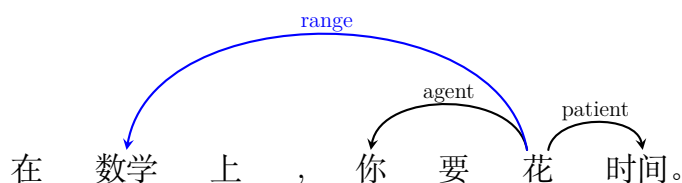


图 83: 谓词“花”的受事是“时间”，范围是在“数学”。

### 11.12 原因或目的 (cause)

原因或目的指谓词所表示动作或行为发生的原因或目的。原因或目的论元一般可以通过标志词进行判定，如通过介词或连词“因为、以、由于、因、为、随着、以、要不是、鉴于、根据、为了、好、以便、从而、来、要、造成、导致、引起”等引出，例如图84和85。

此外还有以下情况需要标注出原因角色：

- 由“使、让”等动词形成的致使结构中，“使、让”前的主语成分往往也充当“cause”角色，例如图86。
- 疑问词“为什么、怎么、干嘛”等，例如句子“宝宝为什么 [哭] 了”、“干嘛把我 [拉] 进去”中，标注疑问词为“cause”角色。
- “在... 下”的中间成分，例如句子“在老师的帮助下，他 [取得] 了很大进步”中，标注“帮助”为“cause”角色。

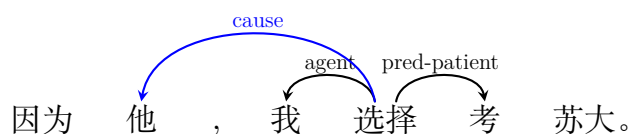


图 84: 注意，“他”也是“考”的原因论元。类似的，“因为我喜欢他，所以选择考苏大”中，“喜欢”也是“选择”和“考”的 cause 论元。





图 85: 句中有标志词“来”，从句意看“来”引出的是“捐献物资”的目的。

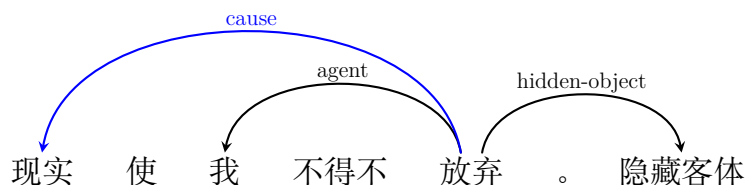


图 86: 致使结构中“致使者”充当原因成分。注意，“使”不作为谓词标注。

### 11.13 事量 (quantity)

事量是谓词动作或行为所涉及的数量、频率、时长、幅度等，通常是数量短语，例如图87、88和89。当句子中数词和量词分开为两个词时，量词作为论元代表词，例如图90。标注多个事量的例子，如图91。

- 数量：表示动作发生次数词，标为事量。
  - 修饰主体或客体的数量不标为事量。“画了两幅画”中，“两幅”不作为“画”的事量，因为“两幅”在语义上直接修饰“画”。类似的，“他吃了很多”中，“吃”的客体省略，“很多”不作为“吃”的事量，而在语义上直接修饰省略的客体。
  - 另外，“睡一觉”中，“一”是事量论元，“觉”是分离论元，标注为分离词，不标注为事量。
  - “她微微一笑”中的“一”仅仅表示动作是短暂的，不作为事量进行标注。
- 频率：“从没、常常、快、每天、天天、每个周末、从来不、从不、老是、常常、永远、一直”等都是频率，标为事量。
- 时长：表示动作持续时间的词，标为事量。例如“呆一天”中“一天”不作为时间，而作为事量，表示“呆”持续的时长。另外，仅仅表示动作发生突然的短暂时长，是为了突出动作的突然性，不是时长。比如“他出现了，一瞬间又消失了”中的“一瞬间”不是“消失”的时长，仅仅表示“消失”这个动作很突然，但是它是“出现”的时长，标注“出现”持续了“一瞬间”。类似的例子还有“他只睡了一瞬间，仿佛过了很久”中“一瞬间”是“睡”的时长，但“他一瞬间就睡着了”中“一瞬间”不是“睡”的时长。
- 幅度：指谓词变化的大小，用来强调变化。例如“小麦产量增长较大”：“大”需要作为“增长”的事量进行标注。

值得注意的是，“占”和比例相关，将所占的比例或数额标为客体角色受事，如“工资 [占] 收入的百分之五十，将“百分之五十”标为受事，而不标为事量。

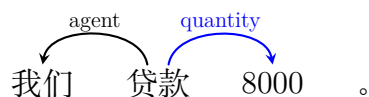


图 87: 句子“我想贷 8000”中，“8000”也标注为事量，“贷”省略了客体角色。

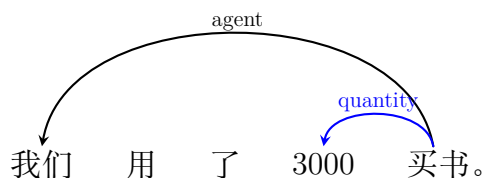


图 88: “3000”同时是谓词“用”的受事。类似的例子如，“用了三年学习弹琴”中，“三年”是谓词“学习”的时长，因此是事量，同时也是“用”的受事。

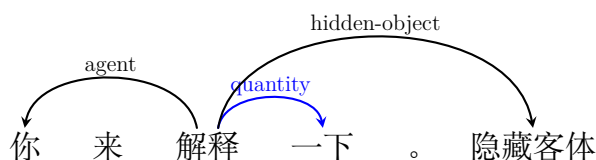


图 89: “一下”可以表示短暂的时间，如“一下热，一下冷”，由于短暂的时间，往往都是为了说明情况比较突然，我们不标注为 quantity。但有时也表示数量。该例子中的“一下”表示数量。

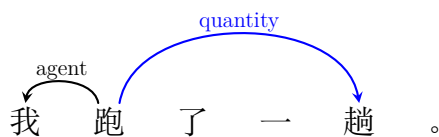


图 90: 类似的例子如“跑了 8000 米”、“跑得很多”、“跑了很多次”、“跑得很频繁”、“跑了两圈”、“跑得快”、“跑得久”、“跑了三小时”。

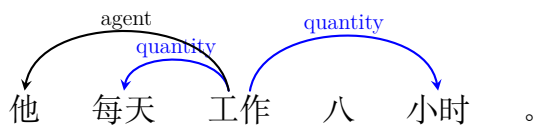


图 91: 例子中的“每天”是频率，“八小时”是时长。

## 12 其它角色

### 12.1 离合 (separated)

汉语中有一类词，具有两个特点：1) 词中的字搭配非常紧密，通常一起出现，但也可以根据一定的形式插入其他语言成分；2) 无法通过每个字的基本含义组合得到词语的含义。这种词在语言学上称为离合词。<sup>10</sup>

根据相关资料<sup>11</sup>，以下是对离合词判定的详细规定：

- 组合成分中是否含有粘着语素来鉴定离合词。粘着语素不能自由活动，只有和前面或后面的语素合成一个单位后才能自由活动。
- 搭配严格受限，即一个动词性成分只能与一个名词性成分组合，或一个名词性成分只能与一个动词性成分组合的动名组合是离合词。
- 非动宾式而用如动宾式的动名组合是离合词。如“鞠躬、洗澡、游泳、睡觉、游行、考试、登记”等。
- 可以扩展又兼属名词或形容词的是离合词。如“签证、刹车、移民、命题、导游、存款、贷款”等。

凡符合上述四个鉴定标准中任何一个标准的动名组合，都应划入离合词。除上述的离合词外还有“上班、上夜班、上学、上大学、见面、操心、放心、道歉、帮忙”等。当离合词中间插入其他成分后，如“洗了一个澡”等，由于“洗”和“澡”之间的语义关系非常强烈，但是无法确定合适的论元角色，我们规定为离合角色。但像“打电话、挨揍”等可以通过每个字的基本含义组合得到词语的含义、“从事、企图、奔跑”等中间不能加入任何成分，则不标注为离合词。另外，离合词分开后，我们固定只有前一个字作为谓词标注。例如图92。

离合动词与一般动词的区别：<sup>12</sup>

- 是否可带宾语：一般动词可以，如“表演节目”、“打扫房间”等；离合词不可以，如“跟他结婚”、“向他道歉”等；
- 加“着、了、过、起来等”：一般动词必须加后面，如“讨论过”、“交谈起来”等；离合词可以加在词语后面也可以加在词语中间，如“结过婚”、“唱着歌”等；
- 重叠形式：一般动词是 ABAB 形式的重叠，如“商量商量”、“介绍介绍”等；离合词是 AAB 形式的重叠，如“聊聊天”、“散散步”等。

由于离合词和非离合词的界限也不是很清晰，我们也无法给出一个完整的离合词列表，因此需要标注者根据如上我们总结的离合词的判定方法或特点分辨。模棱两可的时候，倾向于非离合词。

<sup>10</sup><http://www.cnki.com.cn/Article/CJFDotal-YYJX601.002.htm>

<sup>11</sup>赵淑华, 张宝林. 离合词的确定与离合词的性质 [J]. 语言教学与研究, 1996(1):40-51

<sup>12</sup>吕明. 对外汉语教学离合词研究——以《桥梁》为例 [D]. 辽宁师范大学, 2011

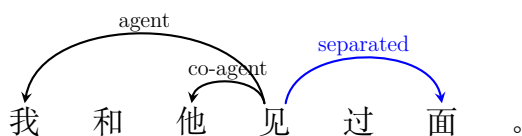


图 92: 类似的例子如“洗了一下午澡”、“他帮了我的大忙”等。

## 13 难点解析

在实际标注过程中，由于文本的来源非常广泛，包括网络文本，标注时难免会存在歧义，无法明确判断如何标注。某些情况下，多个角色标签同时适用于一个论元，似乎都满足标注规范，界限很模。为了提高标注的一致性和质量，我们将针对不同的情况，逐一给出明确的优先级规定及例子，便于标注者的抉择。

需要注意的是，如果同时存在两个角色标签 (A 和 B) 都满足规范，但是标注标签 A 明显比标注标签 B 更加合理，也就是说标注标签 A 可以更加准确地表达语义信息。那么不属于模棱两可的情况，一定要选择标注标签 A。

### 13.1 深层语义 vs. 浅层语义

本规范的主要目标是标注句子的浅层语义结构，因此标注时不需要进行深层次的语义推理和联想。如图93所示。



图 93: 归为“A 是 B 的 C”结构，如果标“B”为谓词，则优先标注“A”作为“B”的论元成分，因为 A 往往比 C 更具体。类似的例子如：“秋季是收获的季节”中将“秋季”标为谓词“收获”的时间，“我的爸爸是一个合格的爸爸”中将“我的爸爸”中的“爸爸”标为“合格”的当事。

### 13.2 “的”字结构作论元

“的”字结构在汉语中是非常常见的一种表达省略的方式。如果省略的成分为主体或客体，我们可以选择标注为省略主体或省略客体，但是所表达的语义会有一定的损失。因此，我们规定，这种情况下，直接将“的”作为论元，标注为具体的主体或客体角色，如 agent 或 patient 等。如图94所示。

注意，不是所有的“的”字结构都标注为论元，如图95所示。

另外，需要注意的是，其他省略情况要正常标注为省略主体或省略客体，如图96所示。

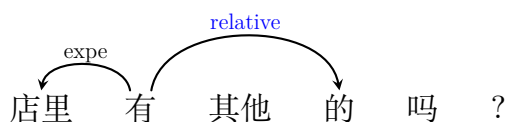


图 94: 此处不是标注为隐藏客体, 而直接将“的”作为系事标注。类似的例子还有“这本书是爸爸的”、“我喜欢红色的”等。



图 95: 这里的“是”不作为谓词标注, 因此这里的“的”字结构并不表达省略。

### 13.3 核心角色 vs. 附加角色

在辨析角色标签时, 首先要区分核心角色和附加角色。通常情况下, 谓词表达信息时不可或缺的成分为核心角色。比如及物动词的宾语均为核心角色, 如图97。

此外, 我们给出了及物动词和非及物动词的对比示例, 如图98和图99。

### 13.4 主体 vs. 客体

在句子中可能存在既可作主体论元也可作客体论元的成分, 例如“经济发展”, 可以理解为“经济在发展”, 将“经济”标为主体角色 agent, 也可以理解为“国家发展经济”, 即“经济被发展”, 将“经济”标为客体角色 patient。

我们规定, 如果句中同时有主体和客体论元, 我们将主客体角色都标注出来; 如果句中只存在一个核心论元, 且它同时可以理解为主体论元和客体论元, 就将其标为主体角色, 例如图100。

### 13.5 正确辨析谓词的主体角色

在标注过程中, 我们需要正确辨别谓词的主体角色是什么, 有的时候虽然主语是物, 但是主体角色可能并不是物, 如图101、102所示。类似的还有句子“婴儿 [看上去] 很快活、健康”中, “婴儿”标注为受事, 主体角色为隐藏主体; 句子“婴儿看上去很 [快活]、健康”中, “婴儿”标注为当事。“他打仗 [像] 个疯子一样”中, “他”标注为主

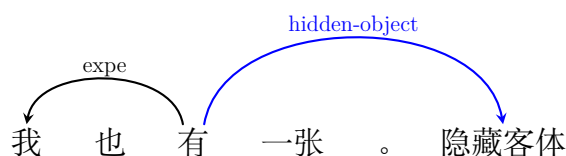


图 96: 不能像“的”字结构一样, 将“一张”标注为“系事”。

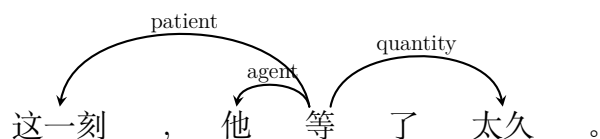


图 97: “这一刻”是“等”的宾语，因此为核心角色，因此一定不能标为非核心角色“时间”。类似的例子还有“我到学校了”中的“学校”为“到”的受事，“我马上要过马路了”中的“马路”是“过”的受事而不是抽象地点。

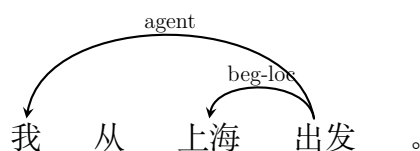
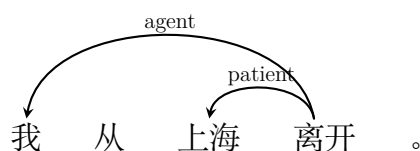


图 98: “离开”是及物动词，后面缺少核心角色，所以将“上海”标为客体角色 patient，而“出发”是个不及物动词，后面不缺少核心角色，所以把“上海”标为附加角色 beg-loc。

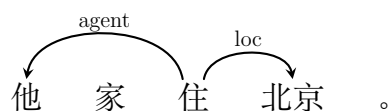
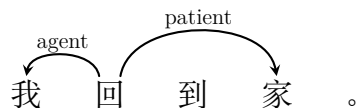


图 99: “我回到家”和“我回家”中“回”标注的论元成分一致。因为“回”是及物的，需要客体角色，标注“家”为 patient，同时“家”也是“到”的客体角色。“他住北京”和“他住在北京”中“住”标注的论元成分一致。因为“住北京”是口语表达，实际上“住”是不及物的，没有客体角色；另外“他家住北京”中“家”是口语表达，因此标“他”为主体角色。



图 100: 可以理解为“事故发生”，虽然也可以理解为“汽车发生事故”，但标注“事故”为主体角色 agent。

体角色。还有一些词，明显修饰的是人，例如图103。

另外，有一些谓词存在主体不明确的情况，例如“面料质量好”，“好”作为谓词时，可以表达“面料好”的意思，也可以表达“质量好”的意思。这种情况，我们规定标注直接修饰的“质量”作为“好”的主体论元。又如图104。

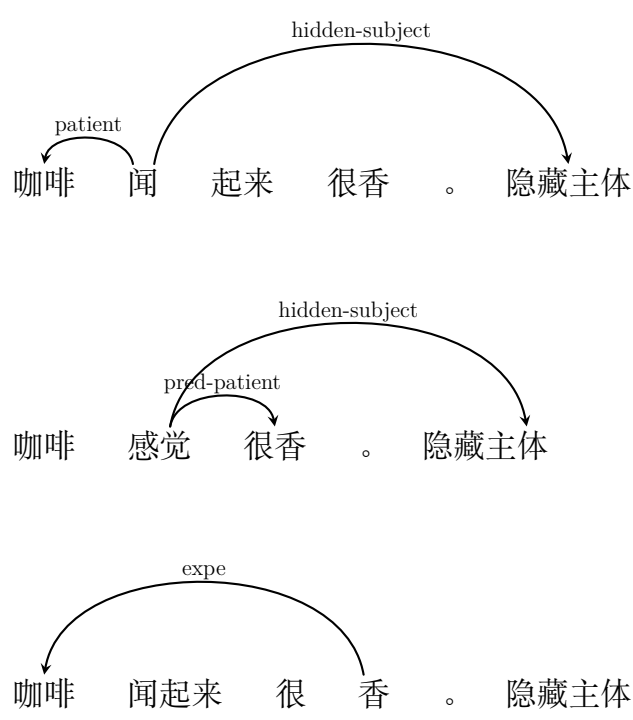


图 101: “闻”和“感觉”是“人”发出的动作或行为，所以应该标为隐藏施事，而“香”是用来形容“咖啡”的，所以将“咖啡”标为“香”的主体角色当事。

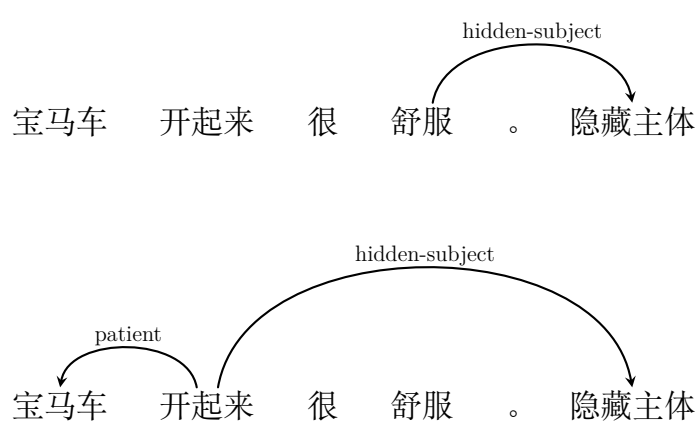


图 102: “舒服”形容“人”的感觉，此处应标注为隐藏主体。

13.6 正确辨析谓词的客体角色

很多情况下，客体角色得选择比较复杂，在难以抉择得情况下，我们规定从句子片段中选择谓词为客体论元，例如图105。值得注意的是，“我推荐你去参加面试”这个

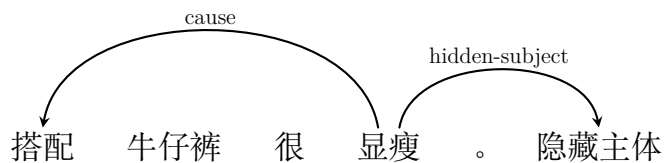


图 103: “人”显瘦，标注主体为人，虽然没有标志词，但是“显瘦”的原因是“搭配牛仔裤”。

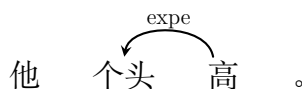


图 104: 虽然这句话可以理解为“他高”，但这句话主要是来强调“他的个头高”，我们标注“个头”作为谓词“高”的主体。

句子中，谓词“推荐”的客体角色是“你”而不是“你去参加面试”，标注“你”为受事，类似的例子如图106，类似的动词还有“请求、劝、教、选”等。

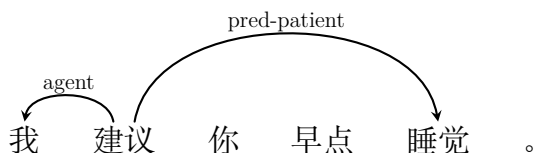


图 105: 建议的客体论元可以是“你”，也可以是宾语从句“你早点睡觉”，优先选择宾语从句为论元成分，标注宾语从句中的动词“睡觉”为客体论元 pred-patient。

### 13.7 机构名、组织名等作为主体论元

有些情况下，当句子中没有明确给出具体的主体论元，而用机构名（或组织名等）等作为一种模糊表达时，则以机构名作为主体论元。如图108中的例子，“种植”的施事应该是“植物园的员工”，“植物园”可以理解为一种抽象的表达。类似的例子还有“中国发现了一种新的病毒”。

### 13.8 manner vs. tool

在标注过程中，manner 和 tool 常常会混淆，例如图109中的“英语”作为谓词“写”的 manner 或 tool 都可以解释得通，在这种模棱两可的情况下，我们规定若论元成分是名词性的词，标为 tool，将动词性的词标为 manner。



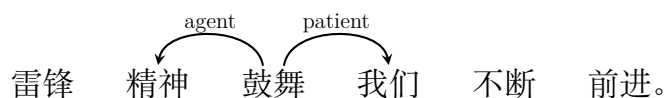


图 106: 标注和谓词“鼓舞”联系更紧密的论元“我们”为 patient。类似的例子还有“命令他刷碗”、“我有个弟弟叫小江”。



图 107: “笑”在作嘲笑的意思时才有客体，根据句子的意思，这个“笑”不是嘲笑的意思，我们不标注隐藏客体。

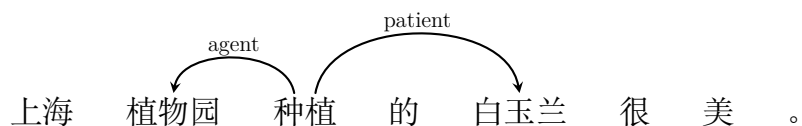


图 108: 严格意义上来说，“种植”的主体论元应该是人，但是在我们日常生活中，“植物园”可以抽象为“种植”的主体论元“植物园的管理人”，类似的还有“我们的国家击退了敌人”中的“国家”。

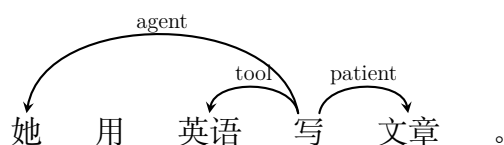


图 109: “英语”是名词，标为 tool。

### 13.9 range vs. loc

在标注中，会存在一些 range 和 loc 容易混淆的情况。当两者都适用时，优先标注为 loc。例如图110和111。

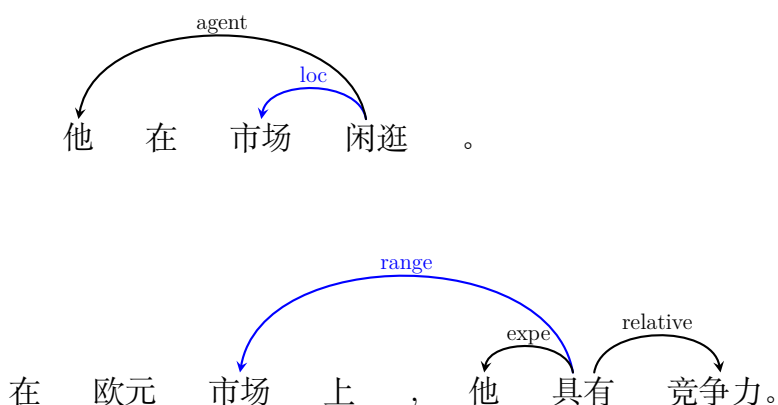


图 110: 前者的市场指“买卖的场所”，是具体的地点，标为 loc；后者的市场指“流行的范围”，标为 range。

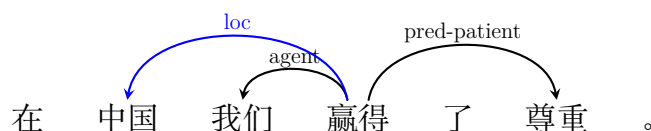


图 111: 这里的“中国”也可以是“赢得”的范围，也可以说是地点，我们可以优先标注为地点 loc。

### 13.10 谓词为成语时的标注

在标注中，我们也会遇到谓词为成语的情况。有的成语比较好理解，是形容型的成语，例如“安居乐业”、“出人意料”等。这样的成语作谓词时，只需要标注主体角色，通常主体论元标签为当事，例如图112。

还有一些成语已经包含了主体和客体论元，例如“众所周知”，其含义是“大家普遍知道”，成语中已包含了主体论元“大家”，在句子“众所周知，苏州大学的校园很美”中，我们可能会将成语理解为动词，将后面的成分标为客体论元。但是这句话也可以理解为“苏州大学的校园很美这件事非常著名”，成语是形容词。当遇到这种情况时，我们规定优先将论元标注为主体论元，这里将“事”标为当事，例如图113。

### 13.11 谓词为网络用语时的标注

在一些网络文本中，我们会经常遇到新兴的词。对于这样的词，我们按照网络上的理解正常标注即可。如图114。

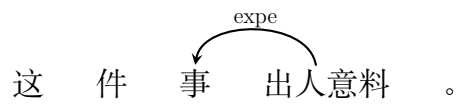


图 112: “出人意料”指事物的好坏、情况的变化、数量的大小等出乎人们的意料。是形容词形的成语

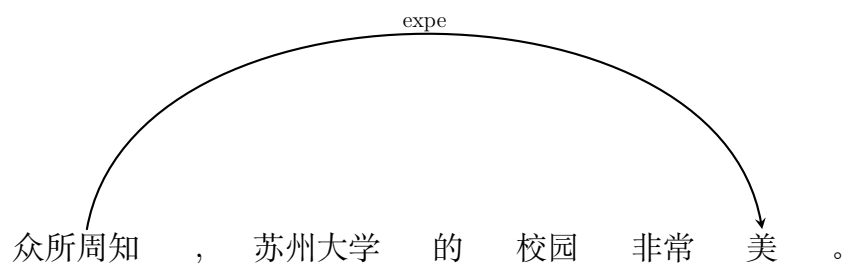


图 113: “众所周知”既可以作动词理解又可以作形容词理解，我们将它的主体论元标注为 expe。

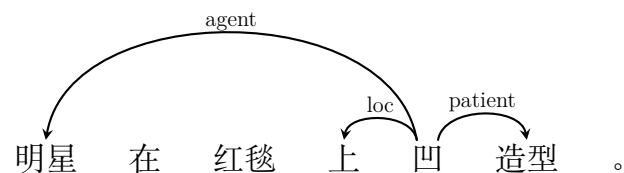


图 114: “凹造型”是一个网络用语，我们按照正常理解，把“凹”作为“摆”标注即可。

### 13.12 主体角色中代词的选择

出现多个 agent (expe 也是类似的), 两种情况: 1) 多个同等级别的参与者, 并列, 均标为 agent/expe (和 co-agent 不一样); 2) 一个参与者, 共指。

如果不确定是否共指, 那么只标注明确和谓词有关系的, 作为标注对象(agent/expe)。对确定的共指的多个代词的情况, 按照如下优先级, 确定唯一的主体角色:

- 如果当前子句中包含了谓词的主体角色 A, 那么就直接标注。
- 如果句子中包含一个实体词 B 作为主体角色, 则将 B 标注为主体角色 (如果有多个, 找最近的, 优先左侧)
- 如果句子中包含一个代词 C 作为主体角色, 则将 C 标注为主体角色 (如果有多个, 找最近的, 优先左侧)
- 缺省主体

例如对于任务“那朵花是红色的, 它很 [好看]”, 在子句“它很好看”中, “它”是谓词“好看”的直接主体角色, 直接标注, 而不标注“那朵花”。如果是“那朵花是红色的, 很 [好看]”, 子句中不存在谓词, 则标注实体词“花”作为主体角色。而对于任务“它是红色的, 很 [好看]”, 句中不存在直接主体角色也不存在实体词, 标注共指的其它代词“它”作为主体角色。

### 13.13 比较结构的论元确定

在一些日常场景中, 我们经常会用到“比较结构”, 其中谓词通常为形容词, 我们只把与谓词语义最相关的成分作为论元标注, 例如图115和116。

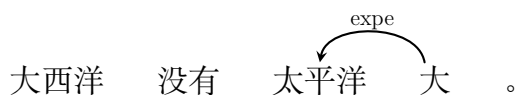


图 115: 主要说明“太平洋更大”。

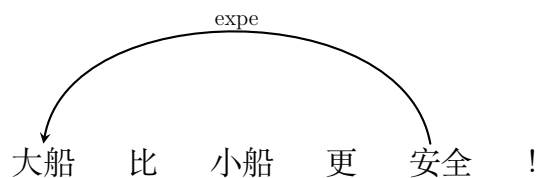


图 116: 主要是说明“大船安全”。

### 13.14 疑问词作论元

通常情况下, 句中的疑问词也作为标注论元, 如“谁、哪里、怎么”等词。当无法判断疑问词应该为标注哪个角色时, 我们规定: 1) 核心角色中, 优先标注为施事或受

事；2) 时间角色中，论元优先标为时间；3) 空间类角色中，论元优先标注为为处所。例如图117所示。

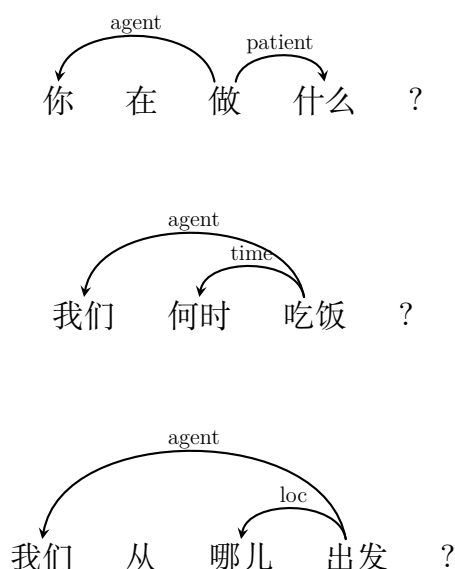


图 117: 疑问词按照优先级顺序进行标注的示例。

### 13.15 真歧义

实际标注任务中，对具有真歧义的句子，没有上下文时无法判断其含义。因为不同的理解会导致不一致的标注结果，直接选择真歧义提交即可。

例如句子“我为他 [写] 信”，存在含义 1) “我替他写信”，“他”不作为论元成分标注；2) “我给他写信”，“他”不作为论元成分标注；3) “我为了他写信”，表示“我因为他的事写信”，“他”标注为 cause 角色。这样的情况下，直接选择真歧义提交即可。类似的例子还有“她正在进行治疗”，只根据这个句子无法判断是“她被治疗”还是“她在治疗别人”，选择真歧义。

需要注意的是，一些存在真歧义的句子，多数人的理解相同，则统一按照多数人的理解标注。例如句子“这件衣服她穿起来 [好看]”存在真歧义，“好看”的主体取决于语境、语调，我们统一规定，对这样句子，优先标注“衣服”为主体角色。

### 13.16 特殊谓词的标注辨析

在商品数据中，经常遇到“搭配”这个词，标注过程中会出现主体客体分不清的情况。我们作了以下规定：如果是“某人搭配了一套衣服”，那么“某人”是主体，“衣服”是客体；如果是“衬衫搭配短裙”这样的 a 搭配 b，ab 两者应该是等同的，ab 都是主体，也就是分别标为 agent 和 co-agent。因此“衣服”可以作为主体也可以作为客体。上面的句子里只出现了一个核心论元，且它可以理解为主体也可以理解为客体，优先标注为主体论元，若省略了应有的 co-agent，那么我们需要标出隐藏主体。如果是“衬衫和裙子很搭配”，那么“搭配”是形容词，我们标“衬衫”和“裙子”为 expe。如图118所示。类似于“搭配”的词作同样的处理。

另外，还有一些谓词看似涉及三个及以上的核心角色，但仔细分析会发现，还是只有两个核心角色，例如图119。

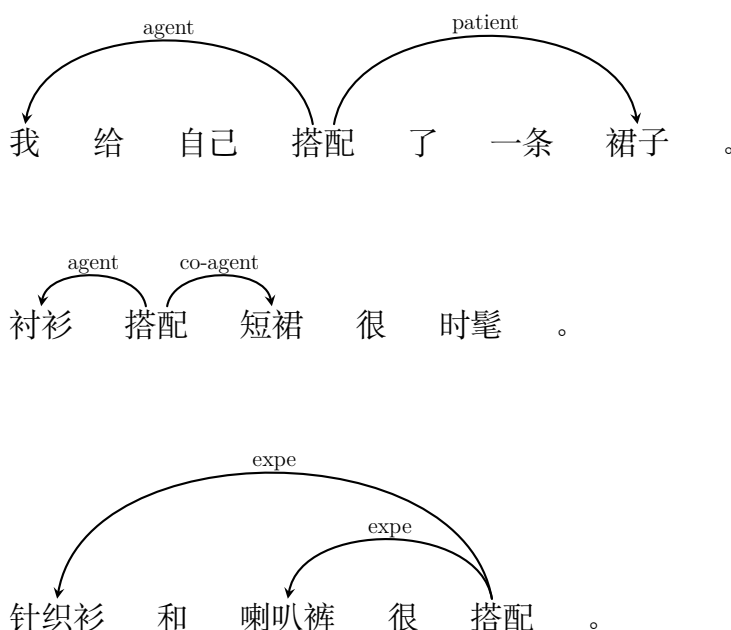


图 118: “搭配”作为谓词进行不同标注的示例。

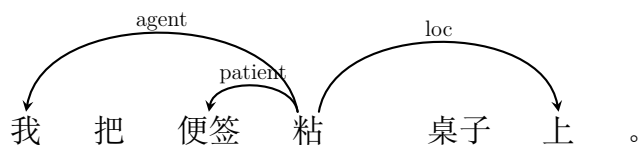


图 119: “粘”是指用粘的东西使物件连接起来，涉及到粘的东西，粘东西的人，连接的东西。这里的“桌子上”其实是“粘”的地点，类似的谓词还有“加”、“贴”等。比如“他将水加桶里”中“桶里”是地点，“他给车贴膜”中的“膜”是受事，“车”是地点。

### 13.17 不确定论元的总结

一些论元我们还不太确定是否需要进行标注，如图120，我们对于类似的情况暂不进行标注。

## 14 分词粒度对语义角色标注的影响

我们在标注系统中会给出分好词的句子。如果标注者发现和谓词、论元相关的内容有明显的分词错误，不合乎语义信息，标注者可以直接点击“分词错误”按钮，然后提交，而不需要尝试进一步标注。

在这一节，我们主要讨论不同分词粒度下语义标注的问题。

不同的分词粒度通常会导致不同的标注结果，如图121、122和123。

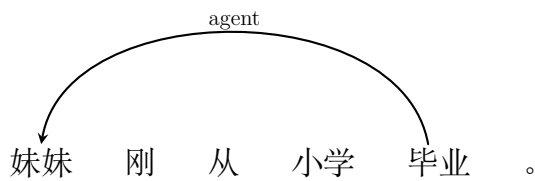
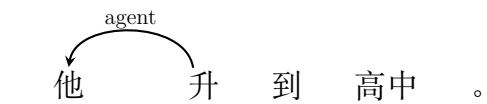


图 120: 句中的“高中”和“小学”不确定标注为什么论元成分，暂时不进行标注。

不同的分词粒度也能有相同的标注结果，如图124和125。

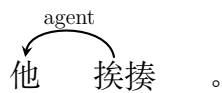
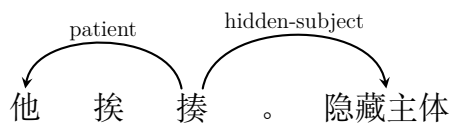
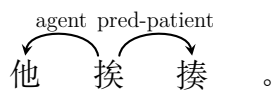


图 121: “挨”和“揍”分开时，作为两个谓词标注。

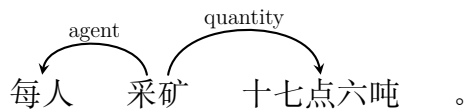
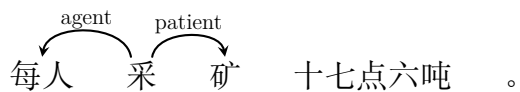


图 122: “采”和“矿”对应两个词时,“十七点六吨”直接修饰名词“矿”,和“采”没有直接语义关系,因此不作为“采”的论元。当“采矿”为一个词时,则作为事量论元。

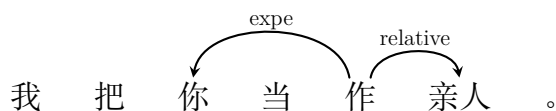
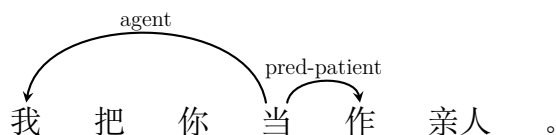
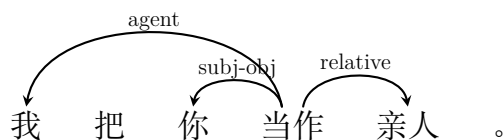


图 123: “当作”作为一个词时,三个相关的核心论元同时标注。“当”和“作”分开时,作为两个谓词分别标注,其中“当”作谓词时,“当”的含义是认为,理解为“我当你作亲人”,标注宾语从句“你作亲人”中动词“作”为 pred-patient。“作”作谓词时,类似“是”标注。



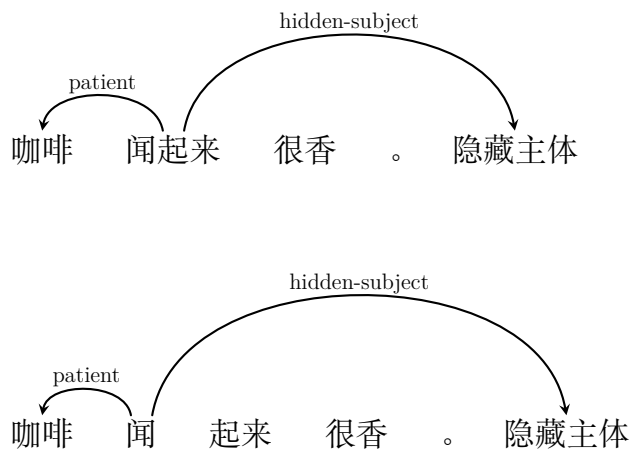


图 124: 不区分“谓词 + 起来/着”和“谓词”的标注, 类似的还有“宝马车 [开起来] 很舒服”和“宝马车 [开] 起来很舒服”, 其中“宝马车”都标为受事, 同时都标注隐藏主体角色。

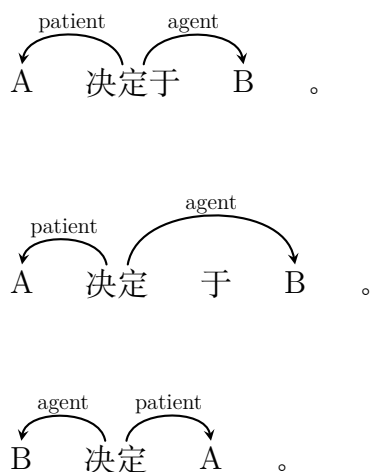


图 125: 不区分谓词是否带“于”的情况的标注, 类似的还有“局限于”。