

## 19. 语义角色标签

在公元前 7 至 4 世纪之间的某个时候,印度语法学家 Pānini<sup>1</sup>写了一篇著名的梵语(Sanskrit)语法专著《Aṣṭādhyāyī》(共 8 本书),该专著被称为“人类智力最伟大的纪念碑之一”(Bloomfield, 1933, 11)。该作品以 3959 部经典(sutras)的形式描述了梵语的语言学,每一种都非常有效地(因为必须记住!)表达了形式规则系统的一部分,该系统很好地预言了形式语言理论的现代机制(Penn 和 Kiparsky, 2012 年)。与本章中的讨论相关的一组规则描述了 kārakas, 动词和名词论元之间的语义关系,诸如施事者、工具或目标之类的角色。Pānini 的作品是我们所知道的最早的作品,旨在了解事件及其参与者的语言实现。理解参与者与事件之间的关系这一任务(能够回答“谁对谁做了什么”(也许还有“何时何地”))是自然语言理解的核心问题。



让我们向前迈进 2.5 千年,并考虑一下了解 XYZ 公司购买股票的文字这一非常平凡的目标。可以通过各种表面形式来描述此购买事件及其参与者。事件可以用动词(sold, bought)或名词(purchase)来描述,XYZ Corp 可以是(of bought)句法主语, (of sold)间接宾语,或所有格或名词复合关系(带有名词 purchase)尽管在所有这些对象中在概念上都具有相同的作用:

- XYZ corporation bought the stock.
- They sold the stock to XYZ corporation.
- The stock was bought by XYZ corporation.
- The purchase of the stock by XYZ corporation...
- The stock purchase by XYZ corporation...

在这一章中,我们将介绍一个层次的表达,来抓住这些句子之间的共性:有一个购买事件,参与者是 XYZ 公司和一些股票,XYZ 公司是买家。这些浅层语义表示,语义角色,表达了一个谓词的论元在事件中所扮演的角色,在 PropBank 和 FrameNet 等数据库中进行了编码。我们将介绍语义角色标签,在句子中分配角色的任务,选择限制,谓词表达其论元的偏好,例如 eat 的主题通常是一些可食用的东西。

### 19.1. 语义角色

想想在第 15 章中,我们是如何为以下句子表达论元的含义的:

(19.1) Sasha broke the window.

(19.2) Pat opened the door.

新戴维森(neo-Davidsonian)事件对这两句话的表示是:

$$\exists e, x, y \text{ Breaking}(e) \wedge \text{Breaker}(e, \text{Sasha}) \wedge \text{BrokenThing}(e, y) \wedge \text{Window}(y)$$

$$\exists e, x, y \text{ Opening}(e) \wedge \text{Opener}(e, \text{Pat}) \wedge \text{OpenedThing}(e, y) \wedge \text{Door}(y)$$

在这个表示法中, break 和 open 的主语角色分别是 Breaker 和 Opener。这些深层角色(deep roles)是特定于每个事件的;Breaking 事件有 Breakers 主语角色,Opening 事件有 Openers 主语角色,如此等等。

如果我们要回答问题,进行推理,或者对这些事件做任何进一步的自然语言理解,我们需要知道更多关于这些论元的语义。Breakers 和 Openers 有一些共同点。他们都是自愿行为者,通常是有生命的,并且对其事件负有直接的因果责任。

**主题角色(thematic roles)**是捕捉 Breakers 和 Openers 之间语义共性的一种方式。我们说这两个动词的主语都是**施事者(agents)**。因此,AGENT 是代表抽象概念的主题角色,如自愿因果关系。同样,这两个动词的直接宾语,BrokenThing 和 OpenedThing,都是典型的无生命的宾语,它们在某种程度上受到动

<sup>1</sup> 图中显示的是来自克什米尔(Kashmir)的桦树皮手稿鲁帕瓦塔(Rupavatra),一种基于 Pānini 梵语(Sanskrit)语法的语法教科书。图片来自威康收藏(Wellcome Collection)。

作的影响。这些参与者的语义角色是**主题(theme)**。

Thematic Role	Definition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The beneficiary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

图 19-1: 一些常用的主题角色及其定义

尽管主题角色是最古老的语言模型之一,但正如我们在上文所见,它们的现代表述是由 Fillmore(1968)和 Gruber(1965)提出的。虽然没有普遍认可的角色集,图 19.1 和 19.2 列出了各种计算论文中使用的一些主题角色,以及粗略的定义和例子。大多数主题角色集都有十几种角色,但我们会看到数量更少的角色集,但意义更抽象,以及数量非常多的角色集,这些角色都是特定于情境的。我们将对所有的角色集合使用通用术语**语义角色(semantic roles)**,无论大小。

Thematic Role	Example
AGENT	<i>The waiter</i> spilled the soup.
EXPERIENCER	<i>John</i> has a headache.
FORCE	<i>The wind</i> blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke <i>the ice</i> ...
RESULT	The city built a <i>regulation-size baseball diamond</i> ...
CONTENT	Mona asked " <i>You met Mary Ann at a supermarket?</i> "
INSTRUMENT	He poached catfish, stunning them <i>with a shocking device</i> ...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations <i>for her boss</i> ...
SOURCE	I flew in <i>from Boston</i> .
GOAL	I drove <i>to Portland</i> .

图 19-2: 各种主题角色的一些原型例子

## 19.2. 动词替换

计算系统使用语义角色的主要原因是作为一种浅层的意义表示,它可以让我们的做出简单的推断,而这些推断是不可能从纯粹的表面单词字符串,甚至从解析树中得出的。为了扩展前面的例子,如果文档说 *Company A acquired Company B*,那么我们想知道这回答了 *Was Company B acquired* 的查询吗?尽管这两个句子的表面句法非常不同。类似地,这种浅层的语义可以作为机器翻译中有用的中间语言。

语义角色有助于概括谓词论元不同表面实现。例如,虽然 AGENT 通常被认为是句子的主语,但在其他情况下,THEME 也可以是主语。考虑一下动词 *break* 的主题论元的可能实现:

(19.3) *John broke the window.*

AGENT      THEME

(19.4) *John broke the window with a rock.*

AGENT      THEME      INSTRUMENT

(19.5) *The rock broke the window.*

INSTRUMENT      THEME

(19.6) *The window broke.*

THEME

(19.7) *The window was broken by John.*

THEME                      AGENT

这些例子表明 *break*(至少)有可能的论元 AGENT、THEME、和 INSTRUMENT。由动词构成的主题

角色论元集合通常被称为**主题格(thematic grid)**、**θ-格**或**格框架(case frame)**。我们可以看到，实现 break 论元有以下几种可能：

AGENT/Subject, THEME/Object  
 AGENT/Subject, THEME/Object, INSTRUMENT/PP<sub>with</sub>  
 INSTRUMENT/Subject, THEME/Object  
 THEME/Subject

事实证明，许多动词可以在不同的句法位置上实现它们的主题作用。例如，像 give 这样的动词可以用两种不同的方式来实现 THEME 和 GOAL 论元：

- (19.8) a. *Doris gave the book to Cary.*  
           AGENT      THEME      GOAL  
 b. *Doris gave Cary the book.*  
           AGENT      GOAL  THEME

这种多重论元结构的实现(break 可以以 AGENT, INSTRUMENT, 或者 THEME 为主语, give 可以以任意顺序实现其 THEME 和 GOAL)称为**动词替换(verb alternations)**或**素质替换(diathesis alternations)**。上面我们展示的 give 的替换,即**承受格替换(dative alternation)**,似乎出现在特定语义类的动词中,包括“将来时态动词”(advance, allocate, offer, owe),“发送动词”(forward, hand, mail),“投掷动词”(kick, pass, throw),等等。Levin(1993)为 3100 个英语动词列出了它们所属的语义类(47 个高级类,分为 193 个更具体的类)和它们参与的各种变化。这些动词类列表已经被合并到在线资源 VerbNet 中(Kipper 等人, 2000),它将每个动词连接到 WordNet 和 FrameNet 词条。

### 19.3. 语义角色：主题角色的问题

在主题角色层面表示意义似乎对处理诸如素质替换等复杂现象(complications)很有帮助。然而,事实证明,提出标准的角色集是相当困难的,同样也很难对角色(如 AGENT, THEME, 或者 INSTRUMENT)给出正式的定义。

例如,试图定义角色集的研究人员常常发现,他们需要将角色(如 AGENT 或 THEME)分割成许多特定的角色。Levin 和 Rappaport Hovav(2005)总结了许多这样的案例,例如,似乎至少有两种 INSTRUMENTS, 中介工具可以作为主语出现,而使能工具则不能:

- (19.9) a. The cook opened the jar with the new gadget.  
        b. The new gadget opened the jar.  
 (19.10) a. Shelly ate the sliced banana with a fork.  
        b. \*The fork ate the sliced banana.

除了分裂问题之外,在某些情况下,我们希望对语义角色进行推理和归纳,但有限的离散角色列表不允许我们这样做。

最后,很难对主题角色进行正式界定。考虑 AGENT 角色;大多数情况下,AGENTS 是有生命的,有意志的,有知觉的,有因果关系的,但任何一个单独的名词短语可能都不具备所有这些属性。

这些问题导致了使用更少或更多角色的替换**语义角色(semantic role)**模型。

第一个选项是定义抽象于特定主题角色的**广义(generalized)语义角色**。例如,原型施事者(PROTO-AGENT)和原型受动者(PROTO-PATIENT)都是表达大致类似施事者和大致类似受动者含义的广义角色。这些角色不是由必要条件和充分条件来定义的,而是由一组启发式特征来定义的,这些特征具有更像施事者或更像受动者的含义。因此,一个论元显示出类似施事者的属性(自愿参与事件,在另一个参与者中引起事件或状态改变,被感知或有意参与,移动)越多,则该论元可以被标记为 PROTO-AGENT 的可能性就越大。属性越像受动者(经历状态的改变,受到另一个参与者的因果影响,相对于其他参与者静止不动,等等),则可以将该论元标记为 PROTO-PATIENT 的可能性就越大。

第二个方向是定义一些语义角色,它们要么特定于一个独有的动词,要么特定于一组独有的语义相关的动词或名词。

在接下来的两节中，我们描述了两个常用的词汇资源，它们利用了语义角色的这些替代版本。PropBank 同时使用原型角色和特定于动词的语义角色。FrameNet 使用特定于一般语义概念(称为框架)的语义角色。

## 19.4. 命题库(PropBank)

**命题库(Proposition Bank)**，一般称为 **PropBank**，是具有语义角色标签的句子资源。英语 PropBank 对宾州树库中的所有句子进行了标签；在宾州汉语树库中，汉语 PropBank 标签句子。由于定义一套普遍的主题角色的困难，PropBank 中的语义角色是根据单个的动词意义来定义的。因此，每个动词的每种意义都有一组特定的角色，这些角色只给出数字而不是名称:Arg0、Arg1、Arg2，等等。通常，Arg0 表示 PROTO-AGENT，Arg1 表示 PROTO-PATIENT。其他角色的语义不太一致，通常是专门为每个动词定义的。尽管如此，还是有一些泛化;Arg2 通常是有益、工具、属性或结束状态，Arg3 通常是起始点、有益、工具或属性，Arg4 通常是结束点。

在 PropBank 中，对于每个动词 agree 和 fall 的一个意义，都包含一些稍微简化的词条目。这样的 PropBank 词条称为**框架文件(frame files)**。请注意，框架文件中每个角色的定义(“其他实体同意”，“范围，数量下降”)是旨在供人类阅读的非正式注释，而不是正式定义。

### (19.11) agree.01

Arg0: Agreeer

Arg1: Proposition

Arg2: Other entity agreeing

Ex1: [Arg0 The group] *agreed* [Arg1 it wouldn't make an offer].

Ex2: [ArgM-TMP Usually] [Arg0 John] *agrees* [Arg2 with Mary] [Arg1 on everything].

### (19.12) fall.01

Arg1: Logical subject, patient, thing falling

Arg2: Extent, amount fallen

Arg3: start point

Arg4: end point, end state of arg1

Ex1: [Arg1 Sales] *fell* [Arg4 to \$25 million] [Arg3 from \$27 million].

Ex2: [Arg1 The average junk bond] *fell* [Arg2 by 4.2%].

注意，fall 没有 Arg0 角色，因为 fall 的正常主语是一个 PROTO-PATIENT。

PropBank 语义角色在恢复关于动词的(verbal)论元的浅层语义信息方面是有用的。考虑动词 increase:

### (19.13) increase.01 “go up incrementally”

Arg0: causer of increase

Arg1: thing increasing

Arg2: amount increased by, EXT, or MNR

Arg3: start point

Arg4: end point

PropBank 语义角色标记可以让我们推断出以下三个例子的事件结构中的共性，即在每个例子中，Big Fruit Co.是 AGENT，而 the price of bananas 是 THEME，尽管表面形式不同。

(19.14) [Arg0 Big Fruit Co. ] increased [Arg1 the price of bananas].

(19.15) [Arg1 The price of bananas] was increased again [Arg0 by Big Fruit Co. ]

(19.16) [Arg1 The price of bananas] increased [Arg2 5%].

PropBank 还有许多非编号论元，称为 ArgM(ArgM-TMP, ArgM-LOC, 等)，它们表示修改或附加含

义。它们在谓词之间相对稳定，所以不会在每个框架文件中列出。使用这些修饰符标签的数据在训练系统中有助于检测谓词之间的时间、位置或方向修改。ArgM 的一些项目包括：

<b>TMP</b>	when?	yesterday evening, now
<b>LOC</b>	where?	at the museum, in San Francisco
<b>DIR</b>	where to/from?	down, to Bangkok
<b>MNR</b>	how?	clearly, with much enthusiasm
<b>PRP/CAU</b>	why?	because ... , in response to the ruling
<b>REC</b>		themselves, each other
<b>ADV</b>	miscellaneous	
<b>PRD</b>	secondary predication	...ate the meat raw

PropBank 侧重于动词，而相关项目 **NomBank**(Meyers 等人, 2004)则在名词谓词上添加注释。例如，在 Apple's agreement with IBM 中的名词 agreement，和 Apple 一起被贴上标签 Arg0，和 IBM 一起被贴上标签 Arg2。这允许语义角色的贴标签者给动词类和名词性谓词的论元分配标签。

## 19.5. 框架网(FrameNet)

虽然用 **increase** 推断不同句子的语义共性是有用的，但如果我们能在更多的情况下，在不同的动词，以及动词和名词之间进行这样的推断，那就更有用了。例如，我们想提取这三个句子之间的相似之处：

(19.17) [Arg1 The price of bananas] increased [Arg2 5%].

(19.18) [Arg1 The price of bananas] rose [Arg2 5%].

(19.19) There has been a [Arg2 5%] rise [Arg1 in the price of bananas].

请注意，第二个例子使用了不同的动词 **rise**，第三个例子使用了名词而不是动词 **rise**。我们希望一个系统认识到 **the price of bananas** 是上升,这 5%是上升,无论 5%作为动词 **increased** 的宾语出现或作为名词 **rise** 的名词性修饰语出现。

**FrameNet** 项目是另一个语义角色标记项目，它试图解决的正是这类问题(Baker 等人 1998, Fillmore 等人 2003, Fillmore 和 Baker 2009, Ruppenhofer 等人 2016)。PropBank 项目中的角色是特定于单个动词的，而 **FrameNet** 项目中的角色是特定于一个框架的。

什么是框架?考虑下面的一组单词：

*reservation, flight, travel, buy, price, cost, fare, rates, meal, plane*

在这个列表中的许多单词之间有许多下位关系、同义关系等单独的词汇关系。然而，由此产生的一组关系并不能完整地说明这些词是如何关联的。它们显然都是根据一组连贯的有关航空旅行的常识背景信息来定义的。

我们把把这些词结合在一起的整体背景知识称为**框架(frame)**(Fillmore,1985)。在人工智能和认知科学中，根据一些背景信息来定义单词组的想法很普遍，除了框架之外，我们看到相关的工作就像一个**模型(model)**(Johnson-Laird,1983)，甚至**脚本(script)**(Schank 和 Abelson,1977)。

FrameNet 中的框架是一个背景知识结构，它定义了一组特定于框架的语义角色，称为**框架元素(frame elements)**，并包括一组使用这些角色的谓词。每个词都能唤起一个框架，并勾勒出框架及其元素的某些方面。FrameNet 数据集包括一组框架和框架元素，与每个框架相关联的词汇单元，以及一组被标签的例句。例如，change\_position\_on\_a\_scale 框架定义如下：

This frame consists of words that indicate the change of an Item's position on a scale (the Attribute) from a starting point (Initial\_value) to an end point (Final\_value).

图 19.3 定义了框架中的一些语义角色(框架元素)。注意，这些角色被分为核心角色(core roles)和非核心角色(non-core roles)，核心角色是特定于框架的，非核心角色更像 PropBank 中的 Arg-M 论元，表示时间、位置等更一般的属性。

下面是一些例句：

(19.20) [ITEM Oil] rose [ATTRIBUTE in price] [DIFFERENCE by 2%].

- (19.21) [ITEM It] has increased [FINAL\_STATE to having them 1 day a month].
- (19.22) [ITEM Microsoft shares] fell [FINAL\_VALUE to 7 5/8].
- (19.23) [ITEM Colon cancer incidence] fell [DIFFERENCE by 50%] [GROUP among men].
- (19.24) a steady increase [INITIAL\_VALUE from 9.5] [FINAL\_VALUE to 14.3] [ITEM in dividends]
- (19.25) a [DIFFERENCE 5%] [ITEM dividend] increase...

Core Roles	
ATTRIBUTE	The ATTRIBUTE is a scalar property that the ITEM possesses.
DIFFERENCE	The distance by which an ITEM changes its position on the scale.
FINAL_STATE	A description that presents the ITEM's state after the change in the ATTRIBUTE's value as an independent predication.
FINAL_VALUE	The position on the scale where the ITEM ends up.
INITIAL_STATE	A description that presents the ITEM's state before the change in the ATTRIBUTE's value as an independent predication.
INITIAL_VALUE	The initial position on the scale from which the ITEM moves away.
ITEM	The entity that has a position on the scale.
VALUE_RANGE	A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.
Some Non-Core Roles	
DURATION	The length of time over which the change takes place.
SPEED	The rate of change of the VALUE.
GROUP	The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.

图 19-3: 框架元素

图注: change\_position\_on\_a\_scale 框架中的框架元素, 来自 FrameNet 标签指南 (Ruppenhofer 等人, 2016)。

从这些例句中可以看出, 框架包含了目标词, 如 rise, fall 和 increase。事实上, 完整的框架由以下几个字组成:

<b>VERBS:</b>	dwindle	move	soar	escalation	shift
advance	edge	mushroom	swell	explosion	tumble
climb	explode	plummet	swing	fall	
decline	fall	reach	triple	fluctuation	<b>ADVERBS:</b>
decrease	fluctuate	rise	tumble	gain	increasingly
diminish	gain	rocket		growth	
dip	grow	shift	<b>NOUNS:</b>	hike	
double	increase	skyrocket	decline	increase	
drop	jump	slide	decrease	rise	

FrameNet 还对框架之间的关系进行编码, 允许框架之间相互继承, 或者表示框架之间的关系, 比如因果关系 (不同框架中框架元素之间的泛化也可以通过继承来表示)。因此, 存在一个 Cause\_change\_of\_position\_on\_a\_scale 框架, 该框架通过 cause(因果) 关系连接到 Change\_of\_position\_on\_a\_scale 框架, 但该框架添加了 AGENT 角色, 并用于引起这种情况的如下示例:

(19.26) [AGENT They] raised [ITEM the price of their soda] [DIFFERENCE by 2%].

这两个框架在一起, 将允许一个理解系统去提取所有动词类和名词性因果关系用法和非因果关系用法的共同事件语义。

FrameNets 还针对许多其他语言进行了开发, 如西班牙语、德语、日语、葡萄牙语、意大利语和汉语。

## 19.6. 语义角色标签(SRL)

**语义角色标签(Semantic role labeling, SRL)**是指自动找出句子中每个谓词的每个论元的语义角色。当前的语义角色标签方法是基于监督机器学习的, 通常使用 FrameNet 和 PropBank 资源来指定什么可以作为谓词, 定义任务中使用的角色集, 并提供训练和测试集。

回想一下, 这两种语义角色模型之间的区别是: FrameNet(19.27)使用许多特定于框架的框架元素作为角色, 而 PropBank(19.28)使用较少的编号论元标签, 它们可以被解释为特定于动词的标签, 以及更通用的 ARGUMENT 标签。一些例子:

(19.27) [You] can't [blame] [the program] [for being unable to identify it]

COGNIZER TARGET EVALUEE REASON

(19.28) [The San Francisco Examiner] issued [a special edition] [yesterday]

ARG0 TARGET ARG1 ARGM-TMP

### 19.6.1. 基于特征的语义角色标签算法

一种简化的基于特征的语义角色标签算法如图 19.4 所示。基于特征的算法——来自最早的系统，如 (Simmons, 1973)——从解析开始，使用广泛的解析器将解析分配给输入字符串。图 19.5 显示了上面(19.28)的解析。然后遍历解析以查找所有谓词的单词。

对于这些谓词中的每一个，该算法都会检查语法分析树中的每个节点，并使用监督分类来确定其在该谓词中所扮演的语义角色(如果有)。给定带标签的训练集(例如 PropBank 或 FrameNet)，使用下一部分描述的特征模板为每个节点提取特征向量。然后训练 1-of-N 分类器，以预测给定这些特征的每个成分的语义角色，其中 N 是潜在语义角色的数量加上非角色成分的额外 NONE 角色。可以使用任何标准分类算法。最后，对于被标签的每个测试语句，对每个相关成分运行分类器。

```
function SEMANTICROLELABEL(words) returns labeled tree
  parse ← PARSE(words)
  for each predicate in parse do
    for each node in parse do
      featurevector ← EXTRACTFEATURES(node, predicate, parse)
      CLASSIFYNODE(node, featurevector, parse)
```

图 19-4：一个通用的语义-角色-标签算法

图注：CLASSIFYNODE 是一个 1-of-N 的分类器，它指定一个语义角色(对于非角色成分，则为 NONE)，训练的对象是有标签的数据，如 FrameNet 或 PropBank。

与图 19.5 所示的单阶段分类器训练不同，节点级分类任务可以分解为多个步骤：

- 1.修剪:由于一个句子中只有少量的成分是什么给定谓词的论元，许多系统使用简单的启发式来修剪不太可能的成分。
- 2.识别:每个节点的二元分类，作为被标签的论元或 NONE。
- 3.分类:前一阶段被标签为论元的所有成分的 1-of-N 分类。

区分识别和分类可以更好地使用特征(不同的特征可能对两个任务有用)或提高计算效率。

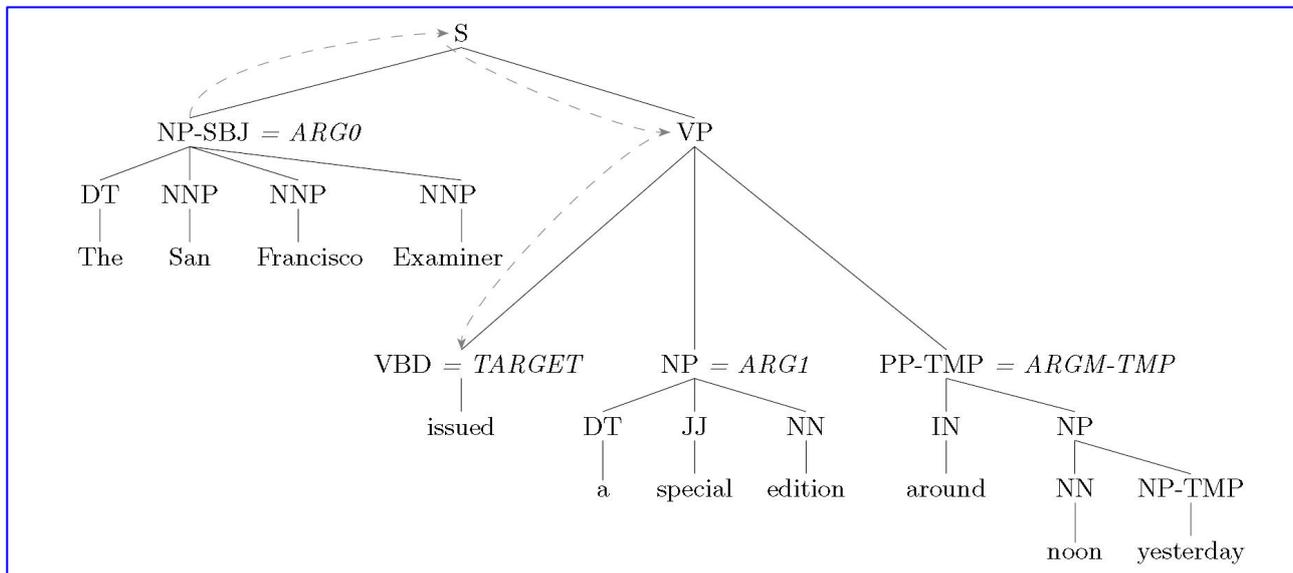


图 19-5：PropBank 句子的解析树

图注：显示了 PropBank 的论元标签，虚线显示 ARG0 的路径特征 NP ↑ S ↓ VP ↓ VBD，NP-SBJ 成分 The San Francisco Examiner。

### 全局优化

图 19.5 的分类算法对每个论元分别进行分类(“局部”), 简化假设一个谓词的每个论元都可以独立地被标签。这种假设是错误的;论元之间存在着交互, 需要对成分进行更“全局”的标签分配。例如, FrameNet 和 PropBank 中的成分必须是不重叠的。更重要的是, 成分的语义角色并不是独立的。例如, PropBank 不允许多个相同的论元;同一个动词的两个成分不能同时被标签为 ARG0。

因此, 角色标签系统通常会增加第四个步骤来处理句子中所有标签的全局一致性。例如, 本地分类器可以返回与每个成分的概率相关联的可能标签列表, 并且可以使用第二次 Viterbi 解码或重新排序方法来选择最佳一致标签。整数线性规划(ILP)是另一种选择最符合多个约束的解决方案的常用方法。

### 语义角色标签的特征

大多数系统使用 Gildea 和 Jurafsky(2000)介绍的核心特征集的一些概括。常见的基本特征模板(在 NP-SBJ 成分的 The San Francisco Examiner 演示, 见图 19.5)包括:

- 支配谓词, 在这种情况下是动词 issued。谓词是一个至关重要的特性, 因为标签只针对特定的谓词定义。

- 成分的短语类型, 在这里是 NP(或 NP-SBJ)。一些语义角色以 NP 的形式出现, 另一些以 S 或 PP 的形式出现, 等等。

- 成分的中心词, Examiner。成分的中心词可以用标准的中心词规则来计算, 例如第 12 章图 12.12 中给出的规则。某些词(如代词)对它们可能扮演的语义角色施加了强烈的限制。

- 成分的中心词词类, NNP。

- 解析树中从成分到谓词的路径。此路径以图 19.5 中的虚线标示。根据 Gildea 和 Jurafsky(2000), 我们可以使用一个简单路径的线性表示, NP↑S↓VP↓VBD。↑、↓分别表示树的向上、向下运动。作为成分和谓词之间的多种语法功能关系的紧凑表示, 路径非常有用。

- 在这种情况下, 从句的语态是主动语态(与被动语态相对)。与主动句相比, 被动句的语义角色与表面形式之间的联系有很大的不同。

- 成分相对于谓词的二元线性位置, 在前面或后面。

- 谓词的子范畴化, 动词短语中出现的一组预期的论元。我们可以使用短语结构规则来提取这个信息, 该规则扩展了谓词的直接父元素; 图 19.5 中的谓词为 VP→VBD NP PP。

- 成分的命名实体类型。

- 成分的第一个词和最后一个词。

因此, 下面的特征向量代表了我们例子中的第一个 NP(回想一下, 大多数观察值都是 NONE, 而不是 ARG0, 因为解析树中的大多数成分都不承担语义角色):

ARG0: [issued, NP, Examiner, NNP, NP↑S↓VP↓VBD, active, before, VP →NP PP, ORG, The, Examiner]

此外, 还经常使用其他特性, 比如成分内部的 n-gram 集合, 或者路径特性的更复杂版本(向上或向下的一半, 或者特定节点是否出现在路径中)。还可以使用依存解析而不是成分解析作为特征的基础, 例如使用依存解析路径而不是成分解析路径。

## 19.6.2. 语义角色标签的神经算法

SRL 的一种简单的神经方法是使用 BIO 方法将其视为序列标签任务, 就像命名实体识别一样。让我们假设我们得到了谓词, 任务只是检测和标签跨度。回想一下, 对于 BIO 标记, 我们为每个可能的角色都有一个开始和结束标记(B-ARG0, I-ARG0;B-ARG1, I-ARG1, 等等), 加上一个外部标记 O。与所有标记一样, 给定单词  $w$  的输入序列, 目标是计算标记序列  $\hat{y}$  的最高概率:

$$\hat{y} = \operatorname{argmax}_{y \in T} P(y|w)$$

图 19.6 为 Shi 和 Lin(2019)以及 He 等(2017)的标准算法草图。在这里, 每个输入单词被映射到预先训练的嵌入, 并且还和一个标志(0/1)变量的嵌入相关联, 该变量指示该输入单词是否为谓词。这些连接的

嵌入通过一个双向 LSTM。每个符记的输出嵌入与谓词的输出嵌入拼接，然后传递给带有 softmax 的 MLP，该 MLP 在每个 SRL 标签上输出分布。

对于解码，可以使用 CRF 层来代替位于双向 LSTM 输出之上的 MLP 层(或者我们甚至可以不使用双向 LSTM，直接将拼接的 BERT 输出传递给 CRF)。Viterbi 解码算法(第 8 章)可以用来从 CRF 解码，这可能有助于利用标记之间的全局约束，例如标记 I-ARG0 必须遵循另一个 I-ARG0 或 B-ARG0。

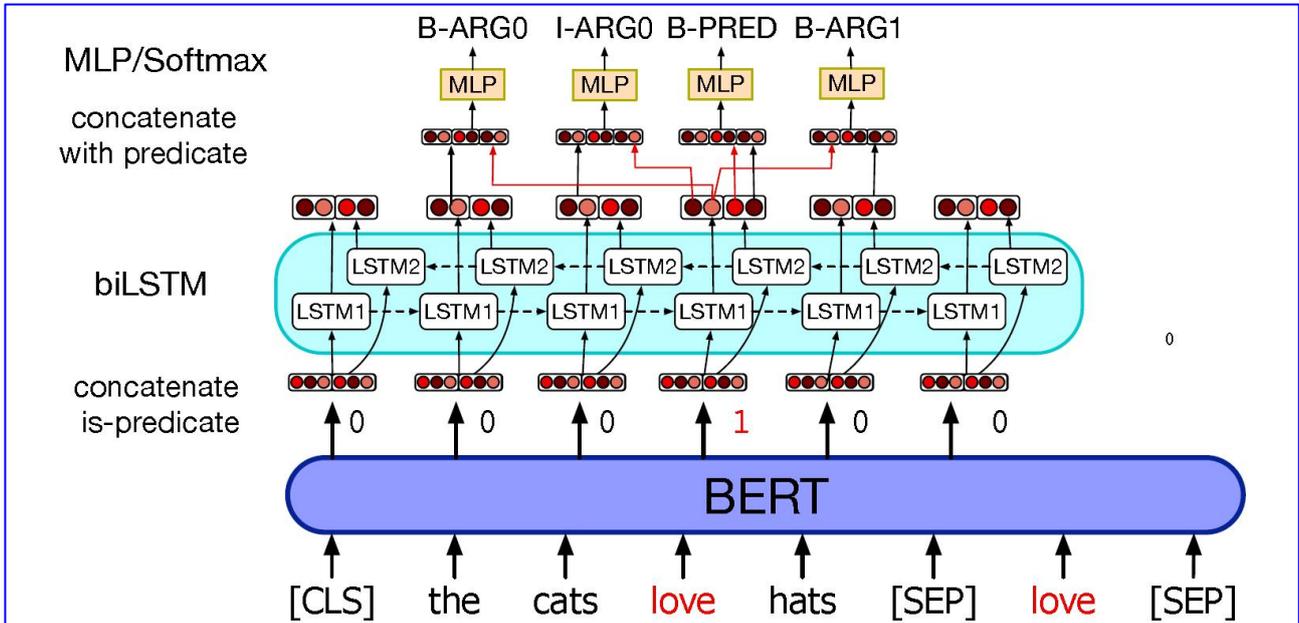


图 19-6: BERT + 双向 LSTM 语义角色标签方法

图注：输入句子后面跟着 [SEP] 和谓词的额外输入，在本例中是 love。BERT 输出拼接到一个指示符变量(谓词为 1，其他所有单词为 0)，通过一个双向 LSTM 传递，然后嵌入到每个符记位置的输出与谓词的嵌入拼接，并通过单层 MLP 传递。

### 19.6.3. 语义角色标签的评价

语义角色标签的标准评估是要求每个论元标签必须被分配到完全正确的单词序列或解析成分，然后计算精度、召回率和 F-测度。识别和分类也可以分别评价。两种常用的评估数据集是 CoNLL-2005 (Carreras 和 M'arquez, 2005) 和 CoNLL-2012 (Pradhan 等人, 2013)。

## 19.7. 选择限制

在本节中，我们将转向另一种方式来表示谓词和论元之间关系。**选择限制(selectional restriction)**是动词对允许填充其论元角色的概念类型施加的语义类型约束。考虑与以下例子相关的两种含义：

(19.29) I want to eat someplace nearby

这个句子有两种可能的解析和语义解释。在合理的解释中，eat 是不及物动词，短语 somewhere nearby 是一个附加词，给出了吃饭事件的地点。在无意义的[speaker-as-Godzilla]的解释中，eat 是及物动词，短语 somewhere nearby 是吃的直接宾语和主语，如下面句子中的 NP Malaysian food:

(19.30) I want to eat Malaysian food.

我们怎么知道 someplace nearby 不是这句话的直接对象？一个有用的提示是语义上的事实，即 THEME of EATING 事件往往是 edible(可食用的)。动词 eat 在其 THEME 论元的填充词上施加的限制是一种选择限制。

选择限制与意义相关，而不是与整个词素相关。我们可以在以下词素 serve 的示例中看到这一点：

(19.31) The restaurant serves green-lipped mussels.

(19.32) Which airlines serve Denver?

例(19.31)说明了 serve 的提供食物的意义，通常将其 THEME 限制为某种食物。例(19.32)说明了为 serve 的提供商业服务意义，从而将其 THEME 限制为某种合适的位置。

选择限制的特殊性差别很大。例如，动词 *imagine* 对其 **AGENT** 角色提出了严格的要求(将其限制为人类和其他有生命的实体)，但对其 **THEME** 角色很少提出语义要求。另一方面，像对角化(*diagonalize*)这样的动词对其 **THEME** 角色的填充物施加了非常具体的限制:它必须是一个矩阵，而形容词 *odorless* 的论元则仅限于可能具有气味的概念:

(19.33) In rehearsal, I often ask the musicians to imagine a tennis game.

(19.34) Radon is an odorless gas that can't be detected by human senses.

(19.35) To diagonalize a matrix is to find its eigenvalues.

这些例子说明了我们用来表示选择限制的一系列概念(矩阵，能够拥有气味等)是非常开放的。这将选择限制与其他词汇知识的表示特征区别开来，比如数量相当有限的词类。

### 19.7.1. 选择限制的表示

捕获选择限制语义的一种方法是使用和扩展第 15 章的事件表示。回想一下，*neo-Davidsonian* 对事件的表示由代表事件的单一变量、表示事件类型的谓词、以及事件角色的变量和关系组成。忽略 $\lambda$ 结构的问题，使用主题角色而不是深层事件角色，像 *eat* 这样的动词的语义贡献可能如下:

$$\exists e, x, y \text{ Eating}(e) \wedge \text{Agent}(e, x) \wedge \text{Theme}(e, y)$$

有了这个表示，我们对 *y*(**THEME** 角色的填充物)的全部了解就是它通过 **Theme** 关系与一个 **Eating** 事件相关联。为了规定选择限制 *y* 必须是 *edible*，我们简单地增加了一个新术语:

$$\exists e, x, y \text{ Eating}(e) \wedge \text{Agent}(e, x) \wedge \text{Theme}(e, y) \wedge \text{EdibleThing}(y)$$

当遇到像 *ate a hamburger* 这样的短语时，语义分析器可以形成以下几种表现形式:

$$\exists e, x, y \text{ Eating}(e) \wedge \text{Eater}(e, x) \wedge \text{Theme}(e, y) \wedge \text{EdibleThing}(y) \wedge \text{Hamburger}(y)$$

这种表示是完全合理的，因为假设知识库中有一组合理的事实，*y* 在 **Hamburger** 类别中的成员资格与其在 **EdibleThing** 类别中的成员资格是一致的。相应地，由于诸如 *ate a takeoff* 这样的短语的表述形式会不正确，因为类似事件的类别诸如 **Takeoff** 中的成员资格将与 **EdibleThing** 类别中的成员资格不一致。

虽然这种方法充分地捕获了选择限制的语义，但直接使用它存在两个问题。首先，使用 **FOL** 执行强制选择限制的简单任务是过分的。其他简单得多的形式可以用更少的计算成本完成这项工作。第二个问题是，这种方法预先假定了一个关于构成选择限制的概念的事实的庞大的、逻辑的知识库。不幸的是，虽然这些常识性的知识库正在开发中，但目前没有一种具有这项任务所必需的覆盖范围。

一种更实用的方法是使用 **WordNet** 语法集来声明选择限制，而不是使用逻辑概念。每个谓词只是指定一个 **WordNet** 同义词库作为对其每个论元的选择限制。如果角色填充词是该同义词库的下位词(从属)，则含义表示形式良好。

例如，对于 *ate a hamburger* 的例子，可以将动词 *eat* 的 **THEME** 角色的选择限制设置为同义词集{食物, 营养}，解释为 *any substance that can be metabolized by an animal to give energy and build tissue*(任何可以被动物代谢以提供能量和构建组织的物质)。幸运的是，图 19.7 所示的汉堡包的上义词链显示汉堡包确实是食物。同样，角色的填充不需要完全匹配约束同义词集;它只需要将同义词集作为其上级之一即可。

```
Sense 1
hamburger, beefburger --
(a fried cake of minced beef served on a bun)
=> sandwich
  => snack food
    => dish
      => nutriment, nourishment, nutrition...
        => food, nutrient
          => substance
            => matter
              => physical entity
                => entity
```

图 19-7: WordNet 的证据表明汉堡包是可以食用的

我们可以将这种方法应用于前面讨论过的动词 *imagine*、*lift* 和 *diagonalize* 的 THEME 角色。让我们把 *imagine* 的 THEME 限制为同义词集{entity}，*lift* 的 THEME 为{physical entity}，*diagonalize* 的 THEME 为{matrix}。这种安排正确地允许 *imagine a hamburger* 和 *lift a hamburger*，同时也正确地排除了 *diagonalize a hamburger*。

## 19.7.2. 选择偏好

在最早的实现中，选择限制被认为是对谓词可以接受的论元类型的严格限制(Katz 和 Fodor 1963, Hirst 1987)。例如，动词 *eat* 可能要求其 THEME 论元为[+FOOD]。早期的词义消歧系统使用这个想法来排除那些违反其主导谓词的选择限制的词义。然而，很快，这些选择限制被更好地表示为偏好，而不是严格的限制(Wilks 1975c, Wilks 1975b)。例如，违背选择限制(如 *eat* 的 *inedible* 论元)经常出现在结构良好的句子中，例如因为它们被否定了(19.36)，或者因为选择限制被夸大了(19.37)：

(19.36) But it fell apart in 1931, perhaps because people realized you can't eat gold for lunch  
if you're hungry.

(19.37) In his two championship trials, Mr. Kulkarni ate glass on an empty stomach,  
accompanied only by water and tea.

因此，用于选择偏好的现代系统使用某种软约束来指定谓词及其可能论元之间的关系。

### 选择关联

最有影响的一种是 Resnik(1993)的选择关联(selectional association)模型。Resnik 将选择偏好强度(selectional preference strength)的概念定义为谓词告诉我们有关其论元的语义类别的一般信息量。例如，动词 *eat* 告诉我们很多有关其直接宾语的语义类的信息，因为它们往往是可食用的。相比之下，动词 *be* 告诉我们的直接宾语更少。选择偏好强度可以通过两个分布之间的信息差异来定义：预期语义类别  $P(c)$  的分布(直接宾语落入类别  $c$  的可能性有多大)；特定动词  $P(c|v)$  的预期语义类别的分布(特定动词  $v$  的直接宾语落入语义类别  $c$  的可能性有多大)。这些分布之间的差异越大，该动词为我们提供的有关可能宾语的信息就越多。这两个分布之间的差异可以通过相对熵(relative entropy)或 Kullback-Leibler 散度(divergence)来量化(Kullback 和 Leibler, 1951 年)。KL 散度  $D(P||Q)$  表示两个概率分布  $P$  和  $Q$  之间的差异：

$$D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (19.38)$$

选择偏好  $SR(v)$  使用 KL 散度来表示动词  $v$  表达了多少关于其论元可能的语义类的信息(以比特为单位)。

$$\begin{aligned} SR(v) &= D(P(c|v)||P(c)) \\ &= \sum_c P(c|v) \log \frac{P(c|v)}{P(c)} \end{aligned} \quad (19.39)$$

然后，Resnik 将特定类别与动词的**选择关联(selectional association)**定义为该类别对动词的一般选择偏好的相对贡献：

$$AR(v,c) = \frac{1}{SR(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \quad (19.40)$$

因此，选择关联是谓词和支配谓词论元的类之间关联强度的概率度量。Resnik 通过分析语料库来估算这些关联的概率，计算每个谓词与每个论元词一起出现的所有次数，并假设每个词都是对包含该词的所有 WordNet 概念的部分观察。下表来自 Resnik(1996)，展示了动词高、低选择性关联的一些样本，以及它们直接对象的一些 WordNet 语义类。

Verb	Direct Object Semantic Class	Assoc	Direct Object Semantic Class	Assoc
read	WRITING	6.80	ACTIVITY	-.20
write	WRITING	7.26	COMMERCE	0
see	ENTITY	5.79	METHOD	-0.01

### 通过条件概率的选择偏好

在动词和它的论元的 WordNet 类之间使用选择性关联的另一种方法是使用给定谓词动词的论元词的条件概率，直接建模一个动词(谓词)与一个名词(论元)的关联强度。

条件概率模型可以通过解析一个非常大的语料库(数十亿个单词)和统计共现计数来计算：给定动词与给定名词在给定关系中出现的频率。给定一个特定关系  $P(n|v, r)$  的动词，一个论元名词的条件概率可以用作该词对  $(n, v)$  的选择偏好度量(Brockmann 和 Lapata 2003, Keller 和 Lapata 2003):

$$P(n|v, r) = \begin{cases} \frac{C(n,v,r)}{C(v,r)} & \text{if } C(n, v, r) > 0 \\ 0 & \text{otherwise} \end{cases}$$

反向概率  $P(v|n,r)$  被发现在某些情况下有更好的性能(Brockmann 和 Lapata, 2003):

$$P(v|n, r) = \begin{cases} \frac{C(n,v,r)}{C(n,r)} & \text{if } C(n, v, r) > 0 \\ 0 & \text{otherwise} \end{cases}$$

一个更简单的方法是使用带有论元  $\log \text{count}(v,n,r)$  的谓词的简单对数共现频率，而不是条件概率：这似乎能更好地提取语法主语的偏好，而不是宾语(Brockmann 和 Lapata, 2003)。

### 评估选择偏好

评估选择偏好模型的一种方法是使用**伪词(pseudowords)**(Gale 等人 1992c, Schutze 1992a)。伪词是将一个测试词(比如香蕉)和一个混淆词(比如门)连接在一起而产生的一个人造词(比如香蕉门)。该系统的任务是识别两个单词中哪个是原单词。为了评估选择偏好模型(例如动词和直接宾语之间的关系)，我们使用一个测试语料库并选择所有动词符记。对于每一个动词符记(如 **drive**)，我们选择直接宾语(如 **car**)，然后连接一个与其最近的混淆词，即频率与原始词(如 **house**)最近的名词，构成 **car/house**。然后，我们使用选择偏好模型来选择 **car** 和 **house** 中哪个是更受 **drive** 偏好的宾语，并计算模型选择正确原始宾语(如 **car**)的频率(Chambers 和 Jurafsky, 2010)。

另一个评估指标是获得人类对一组动词-论元对的测试集的偏好，并让他们对其可信程度进行评级。这通常是通过使用幅度(magnitude)估计来完成的，这是一种来自心理物理学的技术，在这一技术中，受试者对一个与模数成比例的论元的可信性进行评级。选择偏好模型可以通过其与人类偏好的相关性来评估(Keller 和 Lapata, 2003)。

## 19.8. 谓词的原始分解

我们在本章中讨论的语义角色的一种思考方式是，基于有限的主题角色列表(agent、patient、instrument、原型 agent、原型 patient 等)，它们帮助我们以分解的方式定义论元所扮演的角色。这种将含义分解为一组原始语义元素或特征的想法，称为**原始分解(primitive decomposition)**或**成分分析(componential analysis)**，已经得到了进一步的发展，并且特别关注谓词。

让我们来看看 **kill** 这个动词的例子：

(19.41) Jim killed his philodendron.

(19.42) Jim did something to cause his philodendron to become not alive.

从真值条件(命题语义学)的角度来看，这两个句子具有相同的含义。假设这是等价的，我们可以将 **kill** 的含义表示为：

(19.43)  $KILL(x,y) \Leftrightarrow CAUSE(x, BECOME(NOT(ALIVE(y))))$

因此，使用像 do、cause、become not 和 alive 这样的语义原语。

事实上，一组这样的潜在语义原语已经被用来解释 19.2 节(Lakoff 1965, Dowty 1979)中讨论的一些言语变化。考虑下面的例子。

(19.44) John opened the door.  $\Rightarrow CAUSE(John, BECOME(OPEN(door)))$

(19.45) The door opened.  $\Rightarrow BECOME(OPEN(door))$

(19.46) The door is open.  $\Rightarrow OPEN(door)$

分解方法断言，与 open 相关联的单个状态类谓词是所有这些示例的基础。这些例子含义上的差异，是由于这个单一谓词与原语的“CAUSE”和“BECOME”相结合而产生的。

虽然这种原始分解方法可以解释状态和动作之间的相似性，或者因果谓词和非因果谓词之间的相似性，但它仍然依赖于拥有大量像 open 这样的谓词。更激进的方法也选择分解这些谓词。在早期自然语言理解系统中发挥作用的动词谓词分解方法之一是**概念依存(conceptual dependency, CD)**，它由 10 个原始谓词组成，如图 19.8 所示。

Primitive	Definition
ATRANS	The abstract transfer of possession or control from one entity to another
PTRANS	The physical transfer of an object from one location to another
MTRANS	The transfer of mental concepts between entities or within an entity
MBUILD	The creation of new information within an entity
PROPEL	The application of physical force to move an object
MOVE	The integral movement of a body part by an animal
INGEST	The taking in of a substance by an animal
EXPEL	The expulsion of something from an animal
SPEAK	The action of producing a sound
ATTEND	The action of focusing a sense organ

图 19-8：一组概念依存(CD)原语

下面是一个示例句子及其 CD 表示。动词 brought 被翻译成两个原语 ATRANS 和 PTRANS，表示 waiter 既亲自把支票传递给 Mary，又把控制权传递给她。请注意，CD 还将一组固定的主题角色与每个原语关联起来，以表示动作中的各种参与者。

(19.47) The waiter brought Mary the check.

$$\begin{aligned} \exists x, y \text{ Atrans}(x) \wedge \text{Actor}(x, \text{Waiter}) \wedge \text{Object}(x, \text{Check}) \wedge \text{To}(x, \text{Mary}) \\ \wedge \text{Ptrans}(y) \wedge \text{Actor}(y, \text{Waiter}) \wedge \text{Object}(y, \text{Check}) \wedge \text{To}(y, \text{Mary}) \end{aligned}$$

## 19.9. 总结

- **语义角色**是一个论元在谓词描述的事件中所扮演的角色的抽象模型。

- **主题角色**是基于单一有限角色列表的语义角色模型。其他语义角色模型包括每个动词(per-verb)语义角色列表和**原始施事/原始受事(proto-agent/proto-patient)**，它们都在 PropBank 中实现，以及在 FrameNet 中实现的每个框架角色列表。

- **语义角色标签**是为句子的成分指定语义角色标签的任务。该任务通常被视为有监督的机器学习任务，在 PropBank 或 FrameNet 上训练模型。算法通常从解析一个句子开始，然后用语义角色自动标记每个解析树节点。神经模型直接从单词端到端映射。

- **语义选择限制**允许单词(特别是谓词)对其论元单词的语义属性进行限制。**选择偏好**模型(如**选择关联**或简单条件概率)允许为谓词和论元词或类之间的关联分配权重或概率。

## 19.10. 文献和历史说明

虽然语义角色的概念可以追溯到 Panini, 但 Gruber(1965), Fillmore (1966)和 Fillmore (1968)将语义角色重新引入现代语言学。Fillmore 通过研究 Lucien Tesnière 开创性的《句法结构》(Éléments de Syntax Structurale, 1959)对论元结构产生了兴趣, 在这本书中引入了“依存关系”一词, 并为依存关系语法奠定了基础。根据 Tesnière 的术语, Fillmore 首先将论元角色称为 actants(Fillmore, 1966), 但很快就转到格(case)这个词(参见 Fillmore(2003)), 并提出了一个通用的语义角色或格列表(Agent, Patient, Instrument 等), 这些语义角色或格可以由谓词的论元来承担。动词将与它们的格框架一起在词汇中列出, 即强制(或可选)的格论元列表。

语义角色可以提供中间级别的语义表示的想法, 可以帮助从句法分析结构映射到更深的, 更全面的含义表示, 这在自然语言处理中很快被采用, 并且创建了提取格框架的系统以进行机器翻译(Wilks, 1973), 问题回答(Hendrix 等人, 1973), 口头语言理解(Nash-Webber, 1975)和对话系统(Bobrow 等人, 1977)。开发了通用语义角色标签器。最早的(Simmons, 1973)首先通过 ATN(Augmented Transition Network, 增强过渡网络)解析器解析了一个句子。然后, 每个动词都有一组规则, 这些规则指定解析应如何映射到语义角色。这些规则主要参考语法功能(主语, 宾语, 特定介词的补语), 但还检查了内部的成分特征, 例如中心词名词的生命力(animacy)。后来的系统又通过使用带有动词特定格框架的字典来从预建的解析树中分配角色(Levin 1977, Marcus 1980)。

到 1977 年, 格表示在人工智能和自然语言处理课程中被广泛使用和教授, 并在 Winston(1977)的教科书《人工智能》第一版中被描述为自然语言理解的标准。Fillmore 在 20 世纪 80 年代提出了他的框架语义模型, 后来对直觉的描述如下:

框架语义学背后的理念是, 说话者意识到可能相当复杂的情境类型, 联系期望的包, 它们有不同的名称——框架、图式、场景、脚本、文化叙事、模因(mememes)——我们语言中的词汇被理解为以框架作为预设背景。(Fillmore, 2012, p. 712)

框架这个词似乎浮现于一系列相关概念提出的同一时间, Minsky (1974), Hymes (1974), 和 Goffman (1974), 以及与其他相关概念如脚本(Schank 和 Abelson, 1975)和图式(Bobrow 和 Norman, 1975)(见 Tannen(1979)比较)。Fillmore 也受到语义场理论学家和耶鲁人工智能实验室的影响, 在那里, 他注意到早期信息提取系统(如 DeJong(1982), Schank 和 Abelson(1977))使用的槽和填充物的列表。20 世纪 90 年代, Fillmore 利用这些见解, 开始了 FrameNet 语料库评注项目。

与此同时, Beth Levin 利用她早期的格框架字典(Levin, 1977)来发展她的书, 其中总结了由共享论元实现所定义的动词类集(Levin, 1993)。VerbNet 项目建立在这项工作的基础上(Kipper 等人, 2000), 随后很快导致了由 Martha Palmer 及其同事创建的 PropBank 语义角色标签语料库(Palmer 等人, 2005)。

FrameNet 和 PropBank 实例化的丰富的语言标注和基于语料库的方法的结合, 导致了语义角色标签的自动方法的复兴, 首先是在 FrameNet (Gildea 和 Jurafsky, 2000), 然后是在 PropBank 数据(Gildea 和 Palmer, 2002, 除其它之外)。这个问题最早是在 20 世纪 70 年代由手写规则解决的, 因此现在通常被重新定义为由大型且一致的数据库支持的有监督的机器学习。Gildea 和 Jurafsky(2002)、Surdeanu 等人(2003)、Xue 和 Palmer(2004)、Pradhan 等人(2005)、Che 等人(2009)和 Zhao 等人(2009)定义了许多常用的角色标签特征。在 CoNLL-2008 共享任务(Surdeanu 等人, 2008)中引入了依存而不是成分解析的使用。调查见 Palmer 等人(2010)和 Marquez 等人(2008)。

Collobert 等人(2011)率先使用神经方法对语义角色贴标签, 他们在卷积网络上应用了 CRF。Foland, Jr. 和 Martin(2015)的早期工作主要集中在使用依存特征上。后来的作品完全避开了语法特征; Zhou 和 Xu (2015b)介绍了堆叠(6-8 层)biLSTM 架构的使用, 并且(He 等人, 2017)展示了如何利用高速网络来增强 biLSTM 架构, 并将 CRF 替换为 A\*解码, 从而可以应用各种各样的全局约束于 SRL 解码。

大多数语义角色标签方案只在一个句子中起作用, 集中在动词谓词(或名词类谓词, 在 NomBank 的情况下)的宾语上。然而, 在许多情况下, 动词或名词类谓词可能有一个**隐含论元(implicit argument)**: 它只出现在上下文句子中, 或者可能根本就没有, 必须进行推断。在 This house has a new owner 和 The sale

was finalized 10 days ago 这两个句子中，第二句中的 sale 没有 ARG1，但理性的读者会推断 ARG1 应该是前面句子中提到的 house。为了发现这些论元，Gerber 和 Chai(2010)以及 Ruppenhofer 等人(2010)引入了**隐式论元检测(iSRL)**。参见 Do 等人(2017)，了解最新的神经模型。

为了避免需要大量的标签训练集，无监督的语义角色标签方法试图通过对论元的聚类来诱导语义角色集。该任务由 Riloff 和 Schmelzenbach(1998)以及 Swier 和 Stevenson(2004)率先提出;参见 Grenager 和 Manning (2006), Titov 和 Klementiev (2012), Lang 和 Lapata (2014), Woodsend 和 Lapata (2015), Titov 和 Khoddam(2014)。

近年来框架标签的创新包括内涵框架，它标记了关于谓词的论元的更丰富的信息。内涵框架标志着作者或读者对论元的情感(例如，在 he survivors a bombing 中使用动词 survive)表达作者对主语 he 的同情和对 bombing 的负面情绪。更多细节见第 20 章。

除了 Resnik(1993)和 Resnik(1996)的选择关联模型外，选择偏好也得到了广泛的研究。方法包括聚类(Rooth 等人, 1999)、判别学习(Bergsma 等人, 2008a)和主题模型(Seaghdha 2010, Ritter 等人 2010), 约束可以在单词或类的水平上表达(Agirre 和 Martinez, 2001)。选择偏好也被成功地整合到语义角色标签中(Erk 2007, Zapirain 等人 2013, Do 等人 2017)。

## 19.11. 练习

(无)