

载《解放军外国语学院学报》2006年，第3期

从格语法到框架网络*

冯志伟

(教育部语言文字应用研究所, 北京, 100010)

摘要: 20世纪末菲尔摩在框架语义学的基础上提出了框架网络(FrameNet)。框架网络的目的在于研究英语中语法功能和概念结构(也就是语义结构)之间的关系,建立用于自然语言处理的词汇知识库。框架网络是格语法的最新发展,它可以更加准确地描述语言中的题元角色关系,对于语言理论研究和自然语言处理都有十分重要的意义。

关键词: 格语法; 题元角色关系; 框架语义学; 框架网络; 框架元素; 定价模式。

From Case Grammar to FrameNet

FENG Zhi-wei

(Institute of Applied Linguistics, Ministry of Education, 100010, Beijing)

Abstract: In the end of 20th century, based on the frame semantics, Charles Fillmore proposed FrameNet. The goal of FrameNet is to study the relation between grammar function and conception structure in English language, then to build the lexical knowledge database. The FrameNet is the new development of case grammar; it can more concisely describe the thematic role relation in the language. The FrameNet is very important for theoretical linguistics and natural language processing.

Key words: case grammar; thematic role relation; frame semantics; Framenet; frame elements; valance model.

1. 格语法与题元角色关系

在理论语言学和自然语言处理中,句子中单词与单词之间的语义关系,有许多不同的表示方法:例如,我们可以用“格”(case)来表示语义关系,采用AGENT(施事者),PATIENT(受事者),BENEFICIENT(受益者)等深层格作为标记;我们也可以用定价语法来表示语义关系,采用行动元(actant)和状态元(circonstant)等作为标记;我们也可以用谓词论元关系来表示语义关系,采用Arg0, Arg1, Arg2, Arg3等作为标记。

这些表示方法虽然各有不同,但是都可以归结为“题元角色关系”(thematic role relation)。

题元角色的标记基本上来自菲尔摩(Ch. Fillmore)20世纪60年代在《“格”辨》(The case for case, 1968)中提出的格语法(case grammar)的格,菲尔摩提出的格有施事格(A=Agentive)、工具格(I=Instrumental)、客体格(O=Objective)、处所格(L=Locative)、承受格(D=Dative)、使成格(F=Factitive)。菲尔摩本人从来没有说过他提出的格一共有多少个。经过我们归纳,在1966年到1977年间,菲尔摩一共提出了13个格。除了原来的施事格、工具格、客体格、处所格、承受格之外,还增加了感受格(E=Experiencer)、源点格(S=Source)、终点格(G=Goal)、时间格(T=Time)、行径格(P=Path)、受益格(B=Benefactive)、伴随格(C=Comitative)、永存格/转变格(essive/translative)。原来的使成格并入了终点格。

— 施事格(Agentive):表示由动词确定的动作能察觉到的典型的动作发生者,一般为有生命的人或物。

— 工具格(Instrumental):表示对于动词所确定的动作或状态而言,作为某种因素而牵涉到的、无生命的力量或客体。

— 承受格(Dative):表示由动词确定的动作或状态所影响的有生物。“承受格”常常被翻

* 本文得到国家社会科学基金资助,项目编号03BYY019。

译为“给予格”，后者的字面含义容易引起误解。

— 使成格(Factitive):表示由动词确定的动作或状态所形成的客体或有生物,或者是理解为动词意义的一部分的客体或有生物。

— 处所格(Locative):表示由动词确定的动作或状态的处所或空间方向。

— 客体格(Objective):表示由动词确定的事物或状态所影响的事物,它是由名词所表示的事物,其作用要由动词本身的词义来确定。客体格后来改称“受事格”(Patientive)。

— 受益格(Benefactive):表示由动词所确定的动作为之服务的有生命的对象。

— 源点格(Source):表示由动词所确定的动作所作用到的事物的来源或发生位置变化过程中的起始位置。

— 终点格(Goal):表示由动词所确定的动作所作用到的事物的终点或发生位置变化过程中的终端位置。

— 伴随格(Comitative):表示由动词确定的、与施事共同完成动作的伴随者。

格是格语法解释语义和句法关系的基本工具,可是确定一个格的清单却十分困难。菲尔摩本人从来就没有列出一个完整而明确的格清单,在不同的文章中,格的数目各不相同,连名称也经常改变。我们上面举出的是菲尔摩经常使用的13个格。

格语法在自然语言处理中广为使用,在机器翻译、人工智能等领域发挥了作用,是语言信息处理重要的基础理论。

20世纪70年代中期以后,格语法的发展进入了第二阶段。第二阶段的格语法主要作了如下修改:菲尔摩把第一阶段表示格角色的结构叫做底层结构,底层结构由格角色构成,在第一阶段的格语法中,底层结构经过转换就得到表层结构;而在第二阶段,由格角色构成的底层结构,在转换之前还必须经过深层主语和深层宾语等语法关系的分配,从而得到深层结构,深层结构进入转换部分,经过转换得到表层结构。这样一来,每一个句子就有格角色和语法关系两个分析平面,这两个平面把句子和句子所描述的事件联系起来,解释句子的语义和句法现象。

菲尔摩提出,句子描述的是场景(scene),场景中各参与者承担格角色,构成句子的底层结构。底层结构经过“透视域”(perspective)的选择,一部分参与者进入透视域,成为句子的核心成分(nucleus),每一个核心成分根据突出的等级体系(saliency hierarchy)确定其语法关系,其他的参与者不一定能进入句子,即使它们出现在句子中,也只能成为外围成分(periphery)。

场景是语言之外的真实世界,如物体、事件、状态、行为、变化,以及人们对于真实世界的记忆、感觉、知觉等。语言中的每一个词、短语、句子都是对场景的描述。当人们说出一个词、一个短语、一个句子、或者一段话语,都是确定一个场景,并且突出或强调那个场景中的某一部分。例如,动词“写”描写的是这样一种场景:一个人在某个物体的表面握着一个顶部尖锐的工具使其进行运动,在物体表面留下痕迹。在这个场景中有4个实体(即4个参与者):发出这个行为的人、实施这个行为所凭借的工具、承受这个行为的物体表面、这个行为在物体表面留下的痕迹。这是在没有上下文的时候,单独一个动词“写”所描述的全部场景,也就是当我们没有遇到任何其他的上下文条件时,一个单独的动词“写”所产生的全部想象,这也就是“写”这个词给我们引发的全部想象。句子的功能在于突出被描述的主体。假如我对你说,“小王正在写”,那么,这个句子所引发的场景就不同了。根据这个句子,你可以知道这是真实世界中一个事件的场景,当你听到这个句子时,你会建立起这样一个场景:小王正在握着一个笔,在某一物体表面移动,并且在物体表面留下痕迹。这个场景仍然有4个实体:书写人(小王)、书写工具(笔)、书写物体的表面(纸)、在表面留下的痕迹(字),但是,在这个场景中突出了书写人小王这一个实体。如果我说“小王正在写信”,那么,这个句子引出的场景仍然只有4个实体,但是突出了书写人(小王)和在表面留下的痕迹(信)2个实体。如果我说“小王用粉笔在黑板上写”,这个句子引发的仍然是4个实体,但是突出了书写人(小王)、书写工具(粉笔)和物体表面(黑板)3个实体。如果我说“小王用粉笔在黑板上写了一个数学公式”,这个句子引发的实体仍然是4个,不过,与前面3个句子不同的是,这4个实体都突出了:书写人(小王)、在表面留下的痕迹(数学公式)、书写工具(粉笔)、物体表面(黑板)。

语义联系着场景，但是场景并不等于语义，场景必须通过语言使用者的透视才能进入语言，才能与语义发生联系。我们说出每一个句子或者每一段话语，都有一个特定的透视域。在一段话语的任何一个地方，我们都是从一个特殊的透视域去考虑一个场景，当整个场景都在考虑之中的时候，我们一般只是注意场景的某一部分。例如，商务事件有4个参与者：买主、卖主、款项和货物，款项有时还可以再进一步分析为现金和赊帐两种情况。一个原型商务事件应该包括上述的内容，但是，当我们谈论这个事件时，所使用的单个句子要求我们对于事件选择一个特殊的透视域。例如，想把卖主和货物置于透视域，就用动词“卖”；想把买主和款项置于透视域，就用动词“购买”，如此等等。这样，任何人听见并理解他所听到的某一句话时，心目中就有一个包括商务事件的全部必要方面的场景，然而，只有事件的某些方面被确定下来，并且被置于透视域中。

进入透视域的成分成为句子的核心成分。每一个核心成分在深层结构都常有一种语法关系，担任句子的主语或直接宾语。没有进入透视域的成分不一定出现在句子中，即使出现的话，也只是作为句子的外围成分。外围成分通常由介词、状语或者小句引入。

核心成分的突出情况是不同的，菲尔摩提出如下原则来确定核心成分的突出等级：

1. 主动成分级别高于非主动成分；
2. 原因成分级别高于非原因成分；
3. 作为人的（或有生命的）感受者的级别高于其他成分；
4. 蒙受改变的成分的级别高于未蒙受改变的成分；
5. 完全的或个性化的成分的级别高于一个成分的某一部分或无个性化的成分；
6. 实际形体的级别高于背景成分；
7. 肯定成分的级别高于不定成分。

这里的等级是按照突出程度递减的顺序来排列的，因此，主动成分的级别高于其他任何成分，原因成分的级别高于除了主动成分之外的任何一种成分，作为人的感受者的成分的级别高于除了主动成分和原因成分之外的任何一种成分，依此类推。

因此，在确定核心成分的语法关系时，应该按照突出程度的顺序来考虑。

当核心成分确定为一个时，场景中最高的成分就是主语。当确定核心成分有两个时，应该按照它们在等级中的相对位置来分配主语和直接宾语，级别高的成分为主语，级别较低的成分为直接宾语。当一个动词的主语已经确定，可以在其他两个事物中选择一个作为直接宾语时，在突出等级中级别高的事物占有优先地位。如果两个成分的突出程度相同，那么，它们中的任何一个都可以进入透视域。不过，这种突出等级的划分还处于假设阶段。正如菲尔摩所说的：“在现阶段，这一切还纯属推测。”

格语法中的深层格具有普遍性，适用于描写各种自然语言的语句。一旦用格语法对句子结构进行了格的描写，就能对句子的表层关系和性质做出各种推断，例如，推断主语是什么，能否形成一个主谓结构，如何安排句子中的词序，等等。

菲尔摩在 1977 年指出，能够描述同一商业事件的不同动词可以选择不同的方式来表达事件的参与者。例如，在 John 和 Tom 之间涉及 3 美元和 1 个三明治的交易可以用下面的任何一种方式来描述：

- a. John **bought** the sandwich from Tom for three dollars.
(John 花三美元从 Tom 处买了那块三明治。)
- b. Tom **sold** John the sandwich for three dollars.
(Tom 以三美元卖给 John 那块三明治。)
- c. John **paid** Tom three dollars for the Sandwich.
(John 付给 Tom 三美元来买那块三明治。)

在这些句子里，动词 buy、sell 和 pay 从不同的视角来表达商业事件，并选择潜在参与者与题元角色的不同的映射来实现这种视角。我们可以看出，这三个动词具有完全不同的映射。这个事实告诉我们：动词的语义角色必须在动词的词典条目中列出，从潜在的概念结构是不能预测的。

根据这些事实，许多研究者认为，在自然语言处理系统的词典中，需要分别列出每个动词的句法和语义组合的可能性，不能完全依靠句法功能和语义关系之间的对应，简单地进行逻辑推理来解决语义分析问题，而动词的句法和语义组合的可能性应该通过“框架”(frame)来描述。

2. 框架网络

由于语言中句法功能和语义结构之间的对应关系因单词的不同而不同，因此，菲尔摩深切地认识到需要针对具体的单词来描述句法功能和语义结构之间的对应关系，建立描述句法和语义结构的框架。基于这样的认识，在 20 世纪末年，菲尔摩提出了“框架语义学”(frame semantics)，从格语法进一步走到了框架网络。

框架网络(FrameNet)是菲尔摩主持的一个课题，课题主要成员是 Srinu Narayanan, Dan Jurafsky, Mark Gawron，项目经理是 Collin Baker，词典编纂顾问是 Sue Atkins。这个课题的目的在于研究英语中语法功能和概念结构(也就是语义结构)之间的关系，建立用于自然语言处理的词汇知识库。这个课题得到美国国家科学基金(U.S. National Scientific Foundation, NSF)的多年持续资助。课题名称是 NSF ITR/HCI # 0086132: “框架网络++: 一个在线的词汇语义资源及其在语音、语言科技方面的应用”，2000 年 9 月-2003 年 8 月。由于这个课题影响很大，2003 年 8 月之后仍然在继续进行，不断取得新的成果。

这个框架网络根据框架语义学的理论，依靠语料库的支持，正在建立一个在线(online)的英语词汇资源。截至 2005 年 10 月，整个框架网络的规模至少包含 7600 个词元(lexical unit)，包括动词，名词，形容词，覆盖很广的语义领域，对于每一个词位(lexeme)的每一个涵义(sense)都要详尽地描述它的语义和句法的各种结合可能性，也就是它的配价(valences)。这些配价是通过手工标注例句以及自动地对标注结果加以组织和整理而得到的。

框架语义学的中心思想是词的意义描述必须与语义框架相联系。框架是信仰、实践、制度、想象等概念结构和模式的图解表征，它为一定言语社团中意义的互动提供了基础。

框架网络为自己确立的任务是：

- (1) 描述给定词元所隶属的概念结构或者框架；
- (2) 从语料库中抽取包含某个词的句子，并从中挑选能够例示我们所要分析的具有某种给定意义的词元的例子；
- (3) 通过把与框架相关的标记(也就是“框架元素”)指派到包含词元的句子中的短语上，使挑选出来的句子得到标注；
- (4) 准备最终的标注总结报告，简明显示每个词元在组合上的可能性；这些被称作“配价描述”(valence descriptions)。

框架网络数据库的格式是独立于开发平台的，因而可以通过网络和其他交互手段进行显示。

下面，我们通过分析一个简单的例子，使大家对语义框架的做法有一个较好的理解。这里请看一组与称之为“复仇(Revenge)”框架相关的词。唤起“复仇(Revenge)”意义的词元包括：avenge(复仇)，avenger(复仇者)，get back(at)(实行报复)，get even(with)(和……算帐)，retaliate(报仇)，retribution(报应)，revenge(报仇，名词)，revenge(报仇，动词)，以及vengeance(报仇)。“复仇(Revenge)”必须与为了回应某个不应该的遭受而施加的某种惩罚相关。一个“复仇者(AVENGER)”对一个“冒犯者(OFFENDER)”施加某种“惩罚(PUNISHMENT)”，以回应冒犯者早期所做的坏事，即某种“伤害(INJURY)”。“复仇者(AVENGER)”也许就是“被伤害方(INJURED PARTY)”，即遭受伤害的人，也许不是。对“冒犯者(OFFENDER)”所造成的“伤害(INJURY)”的裁断与法律无关；这就要求要把复仇概念与法律上许可的“惩罚”区分开来。复仇情景实例中的事件和参与者，如“复仇者(AVENGER)”和“惩罚(PUNISHMENT)”，被称作“框架元素”(Frame Elements, 简称 FEs)。

请看下列包含“Revenge(复仇)”框架词元的做了标注的例句：

1. [Ethel ^{AVENGER}] eventually **got even** [with Mildred ^{OFFENDER}] [for the insult to Ethel 's family ^{INJURY}]
(Ethel 最终向侮辱她家的 Mildred 报了仇。)

2. Why hadn't [he **AVENGER**] sought to **avenge** [his child **INJURED PARTY**] ?

(他为什么还没有试图为他的孩子报仇?)

3. Yesterday [the Cowboys **AVENGER**] **avenged** [their only defeat of the season **INJURY**] [by beating Philadelphia Eagles 20-10 **PUNISHMENT**].

(昨天, 牛仔们以 20 比 10 战胜费城老鹰队, 从而为他们赛季的唯一失利报了仇。)

4. The Old Bailey was told [he **AVENGER**] was desperately in love and wanted to **get back** [at the woman **OFFENDER**] ["for ending their relationship" **INJURY**]

(据说, 那个老 Bailey 在恋爱中绝望, 并且想向那个结束他们恋爱关系的女人复仇。)

5. [The USA **AVENGER**] **retaliated** [against the harassment of its diplomats **INJURY**] [by expelling 36 staff from the Iraqi embassy in Washington on Aug. 27 **PUNISHMENT**]

(通过驱逐驻华盛顿的伊拉克大使馆的 36 位工作人员, 美国为其外交官所受的折磨报了仇。)

从上述例子可明显看出, 我们拥有所需的用以标注主要参与者的各种框架元素。现在我们可以考虑不同的框架元素在语言上是怎样实现的, 即: 框架元素怎样与句法成分相关。有时不同的词元会有不同的可能性。

以上述框架中的动词为例, 在主动语态的句子中, “AVENGER (复仇者)” 是主语。“OFFENDER (冒犯者)” 典型地出现在介词短语当中。介词词汇形式的不同, 取决于词元: 与 **get even** 搭配的是 **with**, 如例 (1) 所示, 与 **get back** 搭配的是 **at**, 如例 (4) 所示。“INJURY (伤害)” 大多数出现在 **for** 前置词短语中, 但也可以是动词 **revenge** 和 **avenge** 的直接宾语。“INJURY (伤害)” 的表达可以从原始事件 (如: **my brother's murder**, 我哥哥的谋杀) 的角度理解, 也可以从对被伤害方的影响上理解 (如: **my brothers' death**, 我哥哥的死)。“PUNISHMENT (惩罚)” 典型地表现为一个包含动名词补足语的 **by** 短语。最后, “INJURED PARTY (被伤害方)” 有时表现为一个独立成分, 特别是像例 (2) 那样充当 **avenge** 的直接宾语。

在带有动词核心的句法结构中, 相比之下, 某些成分与动词框架之间具有更为特定的语义联系。因此, 框架网络区分了**中心框架元素 (core FEs)**和**非中心框架元素 (non-core FEs)**。尽管与句法学家传统所做的论元 (**argument**) 与修饰语 (**adjunct**) 的区分有相当一部分的重合, 这二者并不相同。传统的区分主要是基于诸如提取 (**extraction**) 这种有关句法配置和句法现象所作的假设。框架网络的概念主要是语义的, 关注某个概念对于框架的意义理解是否必要。在框架网络中, 与动词描写密切相关的配价模式只建立在中心元素的基础上。非中心元素包括各种类型的外围修饰语, 它们或多或少地与各种类型的事件或者状态相协调。非中心元素的例子如上述例 (3) 中的时间副词 **yesterday**。尽管任何“复仇”行为很明显地都有空间和时间的属性, 但是时间修饰语 **yesterday** 与动词 **avenge** 没有特定的意义联系。尽管框架网络的二级目标是对所考察的句子至少提供部分的语义分析, 标注者经常给这些成分标上适当的框架元素标记 (时间、地点等), 但是, 在框架网络中, 对于相关动词的基本的配价的描述只包括那些中心框架元素。

由此可见, 语义框架 (**semantic frame**) 是一个类似于“脚本” (**script**) 那样的结构, 结构中的各个成分由词汇单元的意义联系起来。

每一个框架是框架元素的集合。框架元素包括框架的参与者 (**participant**) 和框架的道具 (**props**), 它们是题元角色。词汇单元的框架语义要描述在所给定的涵义下, 框架元素的结合方式和框架元素在框架中的分布情况。

每一个涵义都要描述它的配价, 配价不仅要表示出框架元素组合方式的集合信息, 而且还要表示出在有关语料库中检验过的语法功能信息和词组类型信息。

标注好的句子是数据库的一个组成部分。它们是用 **XML** 语言置标的, 这些句子是词汇条目的基础。这样的格式可以支持采用框架、框架元素以及它们的组合来进行搜索。

框架网络数据库既可以作为词典 (**dictionary**) 来使用, 也可以作为叙词表 (**thesaurus**) 来使用。

作为词典来使用时, 词典中单词条目的信息包括:

- 该单词的定义: 大部分的定义来自简明牛津词典 (*Concise Oxford Dictionary*, 第 10 版, 简称 *COD*)。

• 标注好的例句：这些例句来自语料库，它们应该是语言学家精选过的，在词典的“标注报告”中加以说明。

• 框架元素表：这个表中要说明框架元素在标注报告中的出现情况以及它们表示的句法关系。

• 定价模式：要说明该单词可以具有的定价模式，并说明每一个定价模式中的框架元素相应的词组类型和句法功能。

• 索引：按照字母顺序排列。

作为叙词表来使用时，每一个单词都与它们所参与的语义框架相链接，而框架反过来又与词表和其他相关的框架相链接。

框架网络所使用的语料库是包含 1 亿词的英国国家语料库 (British National Corpus, 简称 BNC), 取得了牛津大学出版社 (Oxford University Press, 简称 OUP) 的使用许可。语义标注是使用 MITRE 公司的 Alembic 工作台 (Alembic Workbench) 进行的, 句法标注是使用他们自己的标注程序进行的, 这个程序可以给每一个短语标注上语法功能信息和短语类型信息。框架网络中的每一个条目都可以与其他的词汇资源相链接, 这些词汇资源包括词网的 SYNSET 和 COMLEX 的次范畴化框架。

框架网络中的每一个条目要列出该条目的所有论元, 包括题元角色以及它们的词组类型和语法功能。

框架网络包括若干个领域 (domains), 每一个领域又包括若干个框架 (frames), 每一个框架由若干个题元角色来定义。

例如, 在前期的框架网络中, COGNITION (认知) 这个领域包括如下的 3 个框架:

• STATIC COGNITION (静态认知) 框架: 如 believe (相信), think (考虑), understand (理解) 等;

• COGITATION (沉思) 框架: 如 brood (细想), ruminare (反复推敲)

• JUDGMENT (判断) 框架: 如 respect (尊重), accuse (控告), admire (赞美), rebuke (指责)。

在领域 COGNITION 的各个框架中都有题元角色 COGNIZER (认知者), 这个题元角色在不同的框架中可以使用不同的名字来引用。例如, 在 JUDGMENT 框架中, 引用 COGNIZER 的名字叫做 JUDGE (判断者), 此外, 在 JUDGMENT 框架中的题元角色还有 EVALUEE (被评价者), REASON (原因) 和 ROLE (作用)。这些题元角色的意思从下面的关于动词 respect 的例句中可以看出 (表示题元角色的单词用方括号标出):

JUDGE: [John] respects Kim for being so brave.

EVALUEE: John respects [Kim] for being so brave.

REASON: John respects Kim [for being so brave].

ROLE: John respects Kim [as a scholar].

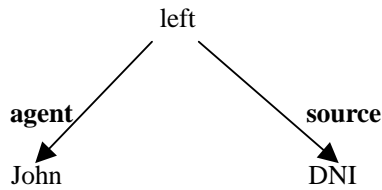
这些题元角色也就是相应框架的框架元素。

在框架网络中, 每一个条目还要标注词组类型 (如 NP, PP) 和句法功能 (如 Subj, Obj)。

例如, 表示判断的动词 appreciate 有动态认知的涵义和静态认知的涵义, 它的框架如下。

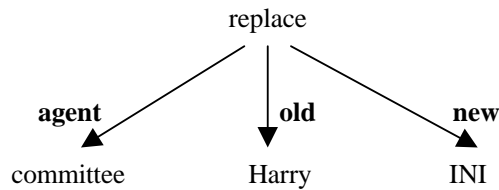
——动态认知的涵义, 表示 “to be thankful or grateful for”:

a. JUDGE	REASON	EVALUEE
NP/Subj	NP/Obj	PP(in)/Comp
I still appreciate	good manners	in men.
b. JUDGE	EVALUEE	REASON
NP/Subj	NP/Obj	PP(for)/Comp
I could appreciate	it	for the music alone.
c. JUDGE	REASON	
NP/Subj	NP/Obj	
I appreciate	your kindness.	
d. JUDGE	EVALUEE	ROLE
NP/Subj	NP/Obj	PP(as)/Comp



c. 无定零形式框架元素 (INI)：缺省的元素的自然类型或语义类型都能够被理解，没有必要找回或者建立一个特定的篇章所指。

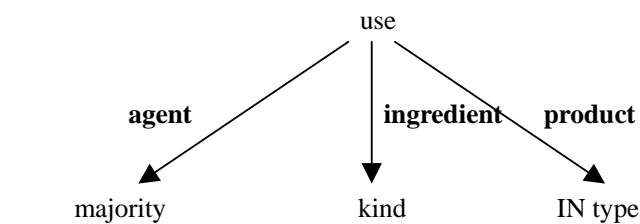
例如，“The committee replaced Harry with Susan”省略 with Susan 之后，变为“The committee replaced Harry”，核心框架图如下：



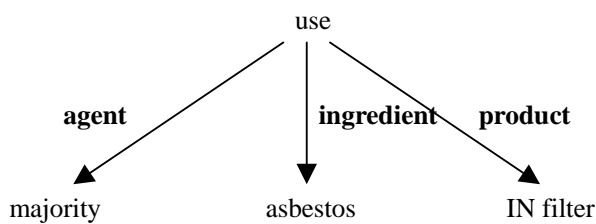
3. 透明名词 (transparent noun)：一个名词短语的句法核心成分代表了数量成分、类型或者容器，它的补足语则包含了这个名词短语的语义核心。

例如，several **pints** of water 中的 pints (品脱)，a **kind** of asbestos (一种石棉) 中的 kind，this **type** of filter 中的 type 都是透明名词。在核心框架图中，我们应当注意挑选与透明名词在语义上相关的名词作为核心。

句子 “The majority of tobacco producer use a kind of asbestos in this of filter” 的核心框架图如果画为如下形式，其语义就很模糊：

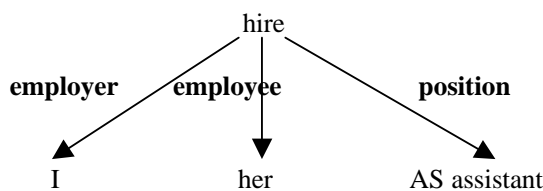


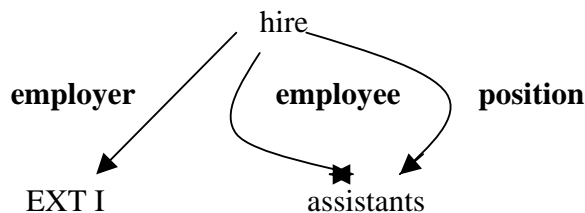
但是，如果画为如下形式，提供的信息就多得多：



4. 框架元素融合 (frame element fusion)：与两个框架元素相关的信息由一个成分来表达。在有些框架中，成对的框架元素非常紧密地联系在一起，因此，语法上可以容许省略其中的一个，因为被省略的那一个可以从另一个实现了的框架元素中推出来。

例如，“I hired [her **EMPLOYEE**] [as my assistant **POSITION**]”与 “I expect to hire two new assistants **EMPLOYEE + POSITION**”。在第二个句子中框架元素 **EMPLOYEE** 与 **POSITION** 融合了。它们的核心框架图如下：





在框架网络中，需要对于已经标注好的句子以及这些句子的配价模式进行深入的研究。框架网络课题组为此开发了相应的软件工具，这样的软件工具可以从标注语料库中自动地生成两个报告：一个报告叫做“词元标注报告”（Annotation by LexUnit Report），一个报告叫做“词条报告”（Lexical Entry Report）。这两个自动生成的报告可以帮助研究人员进行进一步的深入研究。

在自动生成的“词元标注报告”中，首先列出该词元的框架元素表（Frame Element Table），然后，展示出用这些框架元素标注的包含该词元的例句，这些例句是从语料库中自动抽取出来的。

例如，“复仇（Revenge）”框架中词元 *avenge* 的标注报告如下：

框价元素表为：

- AVENGER**: 复仇者
- INJURED PARTY**: 被害方
- INJURY**: 伤害
- OFFENDER**: 冒犯者
- PUNISHMENT**: 惩罚

包含词元 *avenge* 的标注例句为：

1. [Swegen **AVENGER**] is also to have invaded England later to AVENGE [his brother **INJURED PARTY**] . [DNI **PUNISHMENT**] [DNI **OFFENDER**]

（后来 Swegen 也进犯英格兰为他的弟弟复仇。）

2. With this, [ElCid **AVENGER**] at once AVENGED [the death of his son **INJURY**] and once again showed that any attempt to reconquer Valencia was fruitless while he still lived. [DNI **PUNISHMENT**] [DNI **OFFENDER**]

（采用这样的方式，ElCid 马上为他的儿子报了仇，并且再次说明，只要他还活着，任何试图征服 Valencia 的尝试都不会有结果。）

3. His secret ambition was for the Argentine ban to be lifted so [he **AVENGER**] could get to England and AVENGE [Pedro's death **INJURY**] [by taking out the England and especially one pocker-faced Guards Officer **PUNISHMENT**]. [DNI **OFFENDER**]

（他暗暗下了决心来消除 Argentine 的禁令，这样，他就有可能进入英格兰并且通过破坏英格兰特别是杀死一个麻脸侍卫官的方式为 Pedro 的死复仇。）

4. In article 3 of the agreement, [each **AVENGER**] had promised to AVENGE [the violent death of the other **INJURY**] [with the blood of the murderer **PUNISHMENT**]. [DNI **OFFENDER**]

（在协议的第三条中，每个人都承诺要用杀人者的鲜血来为其他人暴烈的死复仇。）

5. Suddenly he walked back to me and said [I **AVENGER**] ought to AVENGE [my father's death **INJURY**] and that he could help me. [DNI **PUNISHMENT**] [DNI **OFFENDER**]

（突然他转过身来对着我，并且说，我应该为我父亲的死报仇，他可以帮我的忙。）

6. [The Trojans **AVENGER**] wish to AVENGE [the death of Hector **INJURY**]; their misplaced values mean that patience in adversity is impossible. [DNI **PUNISHMENT**] [DNI **OFFENDER**]

（Trojans 希望为 Hector 的死复仇；他们错误的估计意味着在逆境中忍耐是不可能的。）

7. "We know the conditions here and [we **AVENGER**] want to AVENGE [that World Cup defeat **INJURY**]," he said, referring to South Africa's 64-run with in New-Zealand. [DNI **PUNISHMENT**] [DNI **OFFENDER**]

（他引用在新西兰进行的南非的第 64 场比赛，并且说，“我们知道这里的条件，我们想为世界杯比赛中的失败复仇”。）

我们可以看到，在这些自动抽取出来的例句中，都进行了框架元素的标注，其中，DNI 是

有定零形式框架元素，尽管在例句中没有出现，但仍然应当标出。

通过上面带标注的 7 个例句，我们可以归纳出词元 *avenge* 的两个配价模式：

1. [AVENGER]-[INJURED PARTY]-[PUNISHMENT]- [OFFENDER]
2. [AVENGER]-[INJURY]-[PUNISHMENT]- [OFFENDER]

如果我们从语料库中自动地抽取更多的标注例句，还可以归纳出第三个配价模式：

3. [AVENGER]-[INJURED PARTY]-[INJURY]-[PUNISHMENT]- [OFFENDER]

这三个配价模式反映了 *avenge* 这个词元的句法语义特性。显而易见，这样的配价模式对于自然语言处理是非常有价值的。

框架网络的软件工具还可以自动地生成“词条报告”。自动生成的词条报告包括“框架元素句法实现表”和“词元的配价模式表”，这两个表格分别总结了框架元素的句法实现情况以及词元的配价模式。

“框架元素句法实现表”可以展示出某一词元的全部核心框架元素、被标注的例子的数目以及它们的句法实现情况。

例如，词元 *avenge* 的框架元素句法实现表如下：

框架元素	在例句中的标注实例数	句法实现情况
AVENGER	33 exx	NP. Ext 25 exx 7 exx Poss. Ext 1 exx
INJURED PARTY	14 exx	NP. Ext 4 exx NP. Obj 11exx
INJURY	21 exx	NP. Ext 4 exx PP. Comp 2 exx NP. Obj 13 exx 2 exx
OFFENDER	33 exx	PP. Comp 3 exx 30 exx
PUNISHMENT	33 exx	PPing. Comp 5 exx PP. Comp 3 exx 25 exx

其中，exx 表示在语料中出现的实例。例如，第一行中的 33 ext 表示在语料中 AVENGER 这个框架元素出现了 33 个实例。

“词元的配价模式表”可以分别说明模式中框架元素的词组类型（如 NP, VP 等）和句法功能（如 Ext, Obj, Comp 等）。

下面是词元 *avenge* 的配价模式表，它说明了在 *avenge* 的三个配价模式中的词组类型和句法功能的分布情况。

标注实例数	配价模式				
2 exx TOTAL	[Avenger] - [Injured party] - [Injury] - [Punishment] - [Offender]				
2 exx	NP	NP	PP	PPing	--
	Ext	Obj	Comp	Comp	--
12 exx TOTAL	[Avenger] - [Injured Party] - [Punishment] - [Offender]				
2 exx	--	NP	--	--	--
	--	Ext	--	--	--
1 exx	--	NP	--	--	PP
	--	Ext	--	--	Comp
6 exx	NP	NP	--	--	--
	Ext	Obj	--	--	--
1 exx	NP	NP	PP	--	--
	Ext	Obj	Comp	--	--
1 exx	NP	NP	PPing	--	--
	Ext	Obj	Comp	--	--
1 exx	NP	NP	PPing	PP	--

		Ext	Obj	Comp	Comp
19 exx	TOTAL	[Avenger] -	[Injury] - [Punishment] - [Offender]		
3 exx		--	NP	--	--
		--	Ext	--	--
1 exx		--	NP	PP	--
		--	Ext	Comp	--
1 exx		NP	--	--	--
		Ext	--	--	--
11 exx		NP	NP	--	--
		Ext	Obj	--	--
1 exx		NP	NP	PP	--
		Ext	Obj	Comp	--
1 exx		NP	NP	PPing	--
		Ext	Obj	Comp	--

这些通过软件自动地生成的“词元标注报告”和“词条报告”，以直观的形式为我们提供了充分的语言信息，有助于我们对相关词元的句法和语义功能进行深入的分析和研究。

除了这些能够自动地生成以标注语料库为基础的报告的软件工具之外，框架网络课题组还开发了一个强大的以网络为基础的数据库查询工具，叫 **FrameSQL**，这个工具是由日本 **Senshu University** 的 **Hiroaki Sato** 教授协助开发的，可通过链接框架网络的网页得到。**FrameSQL** 能够帮助使用者实现多个搜索参数的数据库查询，如框架名称、框架元素名称、语法功能等。例如，可以查询被称作“惩罚 (PUNISHMENT)”的框架元素以介词短语的形式出现的任意框架的所有句子。

对框架网络有兴趣的读者可以访问下面的网址：[http:// www.icsi.berkeley.edu/~framenet](http://www.icsi.berkeley.edu/~framenet)

上面我们描述了从格语法到框架网络的发展过程，由此我们可以看出，菲尔摩对于题元角色关系的研究工作有了长足的进步，我们认为，这些进步主要体现在如下三方面：

- 第一，框架网络中使用的框架元素比格语法中使用的 13 个格更加丰富，更加具体，因而也更加便于用来描述单词的句法语义功能，使我们对于题元角色关系获得更加深刻的认识。
- 第二，格语法研究所依赖的语言事实主要是根据语言学家本人的语言知识以及语言学家对于语言的直观感受，难免带有主观性和片面性，而框架网络的研究则是在大规模标注语料库的基础上进行的，能够客观地反映语言现象的真实面貌，有助于避免主观性和片面性。
- 第三，格语法的研究方法主要是靠语言学家的内省和对于语言现象的洞察力，而框架网络的研究则使用计算机提供各种软件工具，如“词元标注报告”和“词条报告”的自动生成工具，以网络为基础的数据库查询工具等等，这些软件工具成为了研究人员的有力助手，提高了研究工作的效率。

参考文献

- [1] 冯志伟. 机器翻译研究[M], 北京: 中国对外翻译出版公司, 2004 年.
- [2] Fillmore, Charles. The case for case [A]. In Emmon Bach and Robert Harms. Universals in Linguistic Theory [C]. New York: Holt-Rinehart-Winston. 1968. 1-88.
- [3] Fillmore, Charles J. An alternative to checklist theories of meaning [J]. BLS (Berkeley: Berkeley Linguistics Society), 1975, 1: 123-131.
- [4] Fillmore, Charles J. Scenes-and-frames semantics [A]. In Antonio Zampolli, ed. Linguistic Structures Processing: Fundamental Studies in Computer Science, No. 59 [C]. North Holland Publishing. 1977.
- [5] Fillmore, Charles J. Frame semantics [A]. In Linguistics in the Morning Calm [C]. Seoul, Korea: Hanshin Publishing Company, 1982. 111-137.
- [6] Fillmore, Charles J. 1985. Frames and the semantics of understanding [A]. Quaderni di Semantica [C]. 1985. 6.2: 222-253.

- [7] Fillmore, Charles J. and B. T. S. Atkins. Towards a frame-based lexicon: the semantics of RISK and its neighbors [A]. Proceedings from the 1991 Nobel Symposium on Corpus Linguistics [C], Stockholm: Mouton de Gruyter. 1992. 35-66.
- [8] Gildea, Daniel and Daniel Jurafsky. Automatic labeling of semantic roles [J]. Computational Linguistics, 2002. 28(3): 245-288.
- [9] Mohit, Behrang and Srini Narayanan. Semantic extraction with wide-coverage lexical resources [A]. Paper delivered at the 3rd Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL) [C]. Edmonton, Canada: May 2003.
- [10] Somers, Harold. Valency and Case in Computational Linguistics (Edinburgh Information Technology Series 3) [M]. Edinburgh: Edinburgh University Press, 1987.

作者简介：冯志伟（1939-），男，云南昆明人，教育部语言文字应用研究所研究员，博士生导师，研究方向为计算语言学、应用语言学、理论语言学。

通信地址：100010 北京 朝内南小街 51 号 语言文字应用研究所，email 地址：zwfengde@public.bta.net.cn

