

同源角色框架的表达和计算¹

董振东 董强 郝长伶

中国科学院计算机语言信息工程研究中心 北京 100083

E-mail: support@keenage.com

Representation and Computing of Cognate RoleFrame

Zhendong Dong Qiang Dong Changling Hao

Research Center of Computer & Language Information Engineering, CAS, Beijing, 100083

E-mail: support@keenage.com

Abstract: Borrowing from the term of “cognate object”, we use “Cognate RoleFrame” to reveal a kind of semantic relations between nouns and verbs like “hatred” and “hate”. By Cognate RoleFrame, we mean that the noun has the same role frame as its correspondent verb. In HowNet we use “CoEvent” as the identifier to describe all the nouns of Cognate RoleFrame. We demonstrate two HowNet-based tools to evaluate our treatment of Cognate RoleFrame.

Keyword: HowNet, valency grammar, event role, event role frame, cognate RoleFrame

1. 前言

我们认为词汇语义表达的灵魂是关系，词汇语义处理（特别是面向计算机）的关键是意义的计算。这一主张贯穿于我们十多年来研究词汇语义的全过程。这一主张是《知网》建构的基本理念，它全面地体现在《知网》的各个方面，进而成为了《知网》的精髓。我们认为这样的主张和实践不仅对于面向计算机的研究是正确的，而且对于面向人的研究也会是有益的。众所周知，《知网》描述了多种类型的词汇语义关系，涉及了词汇语义的各个方面。大家也知道，与英文的 WordNet 不同，第一，《知网》是基于由中英文两种词语所代表的概念为基础的；第二，《知网》着重描述了不同词性的词语所代表的概念之间的语义关系，其中特别重视名词所代表的概念与动词所代表的概念之间的语义关系，也即我们通常称作实体与事件之间的语义关系。这样的语义关系最普遍的是所谓的“涉入”关系，在《知网》中称为角色关系，例如作为实体的“医生”和作为事件的“医治”，两者有着“事件”与“施事”的关系。本文所谈及的是另外一种类型的实体与事件之间的关系，例如“意见”和“认为”、“意愿”和“愿意”等的关系。有的语法学家把这类的现象归结为“二价名词” [2]。《知网》把这样的关系称为“同源角色框架”。本文将讨论和回答：什么是“同源角色框架”，它有多少种类型，这样的名词占名词总量的比例约多少，《知网》是如何表达和处理它的，并且通过《知网》的二次资源，即《概念相关性计算》和《概念相似度计算》两个软件的演示来显示如此表达和处理的正确性和优越性。

2. 角色框架 (Role frame)

在讨论同源角色框架之前，自然应该先介绍一下角色框架。我们所说的角色框架属于动词的语义范畴，它与句法无关，它是不依赖特定语言的。现今动词框架在一些面向计算机的语义资源中很流行，如 WordNet、VerbNet、FrameNet 等都有自己的专门的动词框架，但它们是很不相同的，例如 WordNet 的动词框架应该说只是句法的。更多的不同主要在于对于角色的定义和它们在框架中的规定。《知网》

¹基金项目：国家语言文字应用“十五”项目 YB105-50；国家自然科学基金项目 60372064

的角色框架是：它包含着当某一个事件发生时，除时间和空间以外的全部的、必不可少的参与角色。

例如：

“买”	agent, possession, source, cost, beneficiary
“打仗”	agent, instrument, partner, cause
“污染”	agent, patient, instrument, PatientValue={dirty 齷}
“打”	agent, patient, instrument, PartOfTouch

这里有两点要特别强调：所谓“必不可少”，不是说在实际的语言文字中一定要出现，而是在逻辑上必然会有，并且如果缺少了，这个事件就不成立了。例如当“买”这个事件发生时，一定会有“施事”（买主）、“领属物”（买的东西）、“来源”（如商店等）、“代价”（付出的货款）、“受益人”（如家人和朋友等，或者自己）。在实际语言文字中很少有把上述各个角色都说出来的，但这绝不等于没有这些角色参与。例如，“我又买了一本词典”，这句话里

虽然只有“施事”和“领属物”，但这绝不等于实际上不包含“来源”、“代价”、“受益人”。

《知网》的事件类别有 815 个，每一个类都有一个类似上述的角色框架。同时《知网》还列出了 815 个框架中的各个角色的典型演员，例如：

“打仗”：

{fight|争斗} {HaveContest|较量:agent={human|人}{group|群体->},instrument={weapon|武器},
partner={human|人}{group|群体->},cause={*}}

所谓角色的典型演员(typical actor)很类似于传统的“语义约束(selectional restriction)”，我们摒弃“语义约束”，只是要避免这种理想主义的，实际却是不灵验的“约束”。

3. 同源角色框架 (Cognate role-frame)

同源角色框架是不同的词类的词语共享完全相同的角色框架的现象。这样的情形在名词和动词之间表现得尤为充分。同源角色框架是不依赖于特定语言的。在英文中有如下语言现象：“live a happy life”，“sleep a sound sleep”，语言学把这里的“life”、“sleep”就称作“同源宾语(cognate object)” [1]，因为它们跟支配它们的动词在词源上和语义上有联系。汉语中诸如“睡觉”的“觉”，“洗澡”的“澡”，“洗浴”的“浴”，“打仗”的“仗”，等等，也完全可以认为是“同源宾语”。下面我们以名词“愿望”和动词“盼望”为例说明。

动词“盼望”的角色框架是：“experiencer, degree, content”，这就是说当“盼望”所代表的事件发生时，必然会有如下角色参与即：谁“盼望”，“盼望”什么，“盼望”的程度。例如“她非常盼望儿子能平安归来”。而名词“愿望”也包含着同样的角色框架。例如，我们可以说，“儿子能平安归来是她最大的愿望”。从这个例子中我们看到了同样的角色框架是：“experiencer, degree, content”。上面两个句子的句法结构完全不同，但它们却有着完全相同的深层语义框架（或图解）。

4. 同源角色框架概念和同源角色框架词语

同源角色框架概念(Cognate role-frame concept)，是指共享完全相同的角色框架的不同词类的词语所代表的概念。这里我们要特别强调的是“概念”，这是因为：第一，同源角色框架是不依赖特定语言的；第二，同源角色框架是建立在概念的基础之上的，严格地说“战争”所对应的同源角色框架，并不是“打仗”这个词语，而应该是“打仗”这个词语所代表的概念。同源角色框架词语(Cognate role-frame word)，是指共享完全相同的角色框架的不同词类的词语。它们主要是两类：同源角色框

架名词如“婚”和同源角色框架动词如“结婚”。

我们必须对它们加以科学的研究分析，给出明确的解释，并在语义资源中有充分的反映。

如前指出的，我们的任务不仅仅是要说明一种语言现象，更主要的是要揭示内部存在的关系，同时还要使之便于进行意义的计算。从建设《知网》的研究中，我们发现：同源角色框架的现象更多地存在于我们称之为一种“动画-定格”的转换中。我们在回答一个经常被提及的问题时我们将采用“动画-定格”这一观念。什么是“动画-定格”？举例来说，“汽车在路上飞驶，把一个路人撞倒”，如果把这样一个连续的过程拍下来，这就是我们说的“动画”，但当我们把“汽车撞倒路人”的镜头加以凝固，这就是“定格”。再如我们常听到的例子，“出版这本书”和“这本书的出版”的区别，以及这里的两个“出版”有何不同。我们说前者是“动画”，是“电影”，有过程，我们称它为“事例”；而后者是一张静止的画，没有过程，我们称它为“事实”。无论是“事例”，还是“事实”，归根结底它们都是“事件”。它们将会有相同的角色参与。在这里我们顺便说明一点，我们不会把上面的两个“出版”认为是两个不同的词性。再举一个例子做进一步说明，我们认为“学习英语”短语中的“学习”与“英语学习”短语中的“学习”，在词性上是没有区别的，都是动词。但是“学习英语”这个短语，在语义上是“事例”，整体上是动词性短语；而“英语学习”这个短语，在语义上是“事实”，整体上是名词性短语。再强调一下，《知网》认为：这里的“学习”的词性仍是动词，而不是什么“名物化”或“名动”等。

5. 《知网》中现有的同源角色框架的类型和数量

在《知网》现有的4万2千多个汉语名词中，包含同源角色框架名词的名词约2380多个，约占5.6%。但这里要特别指出：同源角色框架名词的数量是动态的、开放的，它并不是一个封闭的集合。请看下面的例子，左侧的词是动词，而右侧的词语是同源角色框架名词语。它们表示同源角色框架名词是一个开放的集合。

“播”	“春播”	“接触”	“身体接触”
“防”	“边防”	“管理”	“行政管理”
“救”	“急救”	“援助”	“军事援助”
“治”	“法治”	“污染”	“水污染”

《知网》发现，同源角色框架名词主要有如下语义类型：

(1) 表示“事务”、“业务”，例如：

信贷 生意 交易 信贷 丑业 渔猎 通讯 炊事 防务
兵役 工作 买卖 体育 针线 财税 车险 森工 消防
建筑业 饮食业 预备役 色情业 针线活 博彩业 专业课 典当业 棉纺织
土木工程 医疗卫生 脑力劳动 人身保险 科学研究 幼儿教育
餐饮连锁业 航空意外险 汽车制造业 乳制品工业

(2) 表示“事情”、“行动”，与“事务”或“业务”不同，它是一个单一的事例，例如：

婚折亲约刑吻举丧变
婚嫁 亲事 暴行 酷刑 响指 团拜 乘务 急弯 官司
动静 丧葬 举措 手脚 壮举 言行 追尾 起居 酬酢
对台戏 爆炸案 申诉案 重婚罪 性贿赂 贸易战 处女航 夜生活 大手笔
性骚扰 慢动作 头三脚 翻身仗 发布会 回马枪 攻坚战 耳刮子 开幕式
衣食住行 洞房花烛 家庭作业 军备竞赛 搜救行动 信息检索
空头人情 电子游戏 短期行为 临终关怀 术后护理 惊人之举

恐怖袭击 贸易往来 法律诉讼 自我牺牲 生老病死 生死存亡
不正当竞争 最惠国待遇 吃喝拉撒睡 常规尿分析 记者招待会 珍珠港事件

(3) 表示“意愿”，例如：

意 欲 愿 志 色
嗜好 烟瘾 洁癖 抱负 呼声 希望 理想 盼头 人心
心声 想头 意思 欲求 志向 众望 夙愿 胃口 情欲
上进心 领袖欲 好奇心 性冲动 美国梦 求知欲 倾向性 进取心
狼子野心 一相情愿 非分之想 民心所向 共同愿望 从众心理

(4) 表示“情感”

忿 愤 火 怒 怨 嫌 信 孝 疑 仇 恨 冤 汽 讪 醋 爱 情 歉
愤慨 肝气 民愤 敌意 把握 迷信 孝心 恶感 反感
外心 存疑 炽情 激情 暖流 丹心 恐怖 余悸 谢忱
意思 恩仇 仇怨 私仇 宿怨 醋劲 爱心 春晖 青睐
认同感 负罪感 信任感 自信心 优越感 事业心 恐惧感 人情债 妒忌心
宗教信仰 敬业精神 恐惧心理 恩恩怨怨 恋母情结 深仇大恨

(5) 表示“感受”，即某种经验、体验，例如：

感觉 听觉 无知 童蒙 快感 乐趣 余兴 眼波 动觉
美感 冷眼 冷遇 倦意 觉悟 疙瘩 苦楚 痛楚 痛苦
新鲜感 安全感 冷板凳 耳目之娱 酸甜苦辣 悲欢离合 燕尔之乐

(6) 表示“现象”，包括“疾病”、“症状”，“天象”，以及“结果”，例如：

锈 囧 呃 梦 热 年 稔 荒 幸 汛 脉 屈 冤 妆
水文 假死 矛盾 潮汛 春潮 故障 毛病 胎动 停电
梦幻 低烧 年成 余缺 枯水 房荒 失业 失重 全食
浓妆 贡献 败局 败仗 进球
千年虫 弛张热 闭门羹 肠绞痛 关节痛 乌龙球 磁感应 光电导 日全食
通货紧缩 内部矛盾 逆反心理 宾客如云 人口老化 信息爆炸
成矿作用 一夫多妻 能源危机 水土流失 双喜临门 政企不分
就业不足 骨牌效应 条件反射 尖端放电 接触不良 供过于求

(7) 表示“思想”，“念头”，“规划”，“制度”，例如：

意 论 见
归心 乡情 乡思 美梦 异议 鄙见 看法 论断 想法
愚见 主张 设想 假想 心裁 念头 意见 人治 法治
思乡病 主心骨 拜火教 拜物教 革新论 分红制 兵役制
无政府主义 投降主义 冒险主义 南柯一梦 一得之愚 折衷方案
集会自由 种族主义 老年健康保险

这里有一点是要特别强调的：上面列出的 7 大语义类型，是指这些名词自身的，而不是它们所对应的动词的。名词所对应的动词是个别化的。详细内容参见下节。

6. 《知网》对于同源角色框架的表达和计算

6.1 一般原理

《知网》对于同源角色框架的表达，体现在如下几个方面：（1）确立这样的现象是一种普遍存在，必须在《知网》中予以反映；（2）务必以概念为基础，而决非以词语为基础；（3）明确一个原则，即务必一个词语一个词语地逐一建立起语义同源的关系，例如，“战争”、“意见”、“爱情”、“乡思”、“生意”等等。也就是说找出与它们各自相对应的事件概念，并以某种形式表达出来。只说它是一种什么现象是不够的。这从一个方面也体现了面向计算机的语言研究与面向人的语言研究是不同的。

6.2 《知网》的同源角色框架关系符 -- CoEvent

《知网》采用一个同源角色框架关系符 – CoEvent 来建立同源角色框架概念（通常是名词与同源角色框架动词）之间的关系。“CoEvent”的意思是：**Cognate_Roleframe_Event**，即同源角色框架事件。下面是《知网》中的两个记录：

W_C=爱心	W_C=爱护
G_C=N	G_C=V
E_C=	E_C=
W_E=love	W_E=cherish
G_E=N	G_E=V
E_E=	E_E=
DEF={emotion 情感:CoEvent={like 爱惜}}	DEF={like 爱惜}

关于“爱心”这一词语的定义的意思是：它是“情感”，它的同源角色框架事件概念是“爱惜”。再看下面简化的例子，包含三部分，分别是：（1）同源角色框架名词；（2）《知网》中对该名词的定义的实际的标注；（3）相对应的事件的角色框架；（4）事件角色与典型演员。

（1）“仗” / “战争”

（2）DEF={fact|事情:CoEvent={fight|争斗},domain={military|军}}

（3）{fight|争斗} {HaveContest|较量:agent={*},instrument={*},partner={*},cause={*}};
{HaveContest|较量:coagent={*},instrument={*},cause={*}}

（4）{fight|争斗} {HaveContest|较量:agent={human|人}{group|群体->},
instrument={weapon|武器},
partner={human|人}{group|群体->},cause={*}};

采用“CoEvent”这个同源角色框架关系符，我们实现了逐一地把同源角色框架名词与同源角色框架动词建立了关系。试比较动词“打仗”，它的标注是：DEF{fight|争斗:domain={military|军}}。这样很显然，就可以实现同源角色框架关系的意义计算了。

另外，如前所述，《知网》发现和归纳了同源角色框架名词的语义类型，同时也指出同一语义类型的名词完全可能对应不同的动词，尤其是属于“事务”和“事情”类的名词。

试看下面《知网》中实际的例子：

W_C=环卫
G_C=N

