

文章编号:

基于新闻语料库的越南语框架语义标注研究*

林丽

(解放军外国语学院, 河南 洛阳, 471003)

摘要: 越南是中国的重要邻国, 相应的越南语海量信息处理正日益凸显其重要性。参考国内外有关框架语义标注的研究和实践, 我们在构建越南语新闻语料库, 对越南语文本进行分词和词性标注、命名实体标注等研究的基础上, 尝试构建越南语框架语义知识库并初步探索了框架语义标注在越南语新闻事件抽取中的应用。

关键词: 框架语义; 标注; 越南语; 新闻

中图分类号: TP391

文献标识码: A

A Study on Vietnamese Frame Semantic Annotation

Based on the News Corpus

LIN Li

(PLA University of Foreign Languages, Henan Luoyang, 471003)

Abstract: Vietnam is an important neighboring country of China, the corresponding massive information processing has become increasingly necessary and important. By referring to the relevant studies and practices on the Frame Semantic annotation at home and abroad, we built a Vietnamese News corpus. On the basis of text segmentation and part of speech tagging, named entity tagging, we tried to build Vietnamese FrameNet and initially explored the application of the Frame semantic annotation in the Vietnamese news event extraction.

Key words: Frame semantic; annotation; Vietnamese; News

1 引言

越南是中国的重要邻国, 在当前信息呈海量爆炸型增长的背景下, 针对越南语的海量信息处理也日益凸显其重要性。其中, 军事领域的需求尤为突出。如何有效利用现代语言学和信息技术发展成果深入研究越南语军事领域新闻文本的知识表示、知识获取对于及时、准确地追踪和发现对象国信息具有重要意义, 是一个值得关注和投入的课题。

目前制约自然语言信息处理发展的重要“瓶颈”之一就是计算机对自然语言的语义理解。自然语言的语义分析和内容信息的理解, 离不开语义知识库的支持, 它是帮助计算机了解人类语言的一个媒介和手段, 也是让计算机逐渐智能起来的物质前提。可以说, 词汇语义知识库是自然语言信息处理领域中的核心工程之一。不论是使用基于规则的方法还是基于统计的方法, 信息提取和检索、词义排歧、机器翻译、自动文摘、自动问答系统的研究终究都离不开词汇语义知识资源基础上的语义分析。

在目前的研究条件下, 越南语-汉语机器翻译难度非常大。这一方面由于机器翻译理论发源于欧美, 而作为源语的越南语属于意合型语言, 缺乏形式化标记, 难以实现高精度的句法语义分析, 另一方面, 越南也缺乏相关理论和实践上的积累, 可以借鉴的资源很少。相比之下, 信息抽取可以说是自然语言理解技术和实际应用相折中的产物, 其目标是对文本的有限理解, 只关心特定领域, 特定类型的信息。事实上, 对海量文本进行信息挖掘离不开高质量的事件内容分析技术, 而这些技术的开发又需要高质量的事件语义标注资源支持。“大规模”、“覆盖面广”的语义知识库若能够描述一种(或多种)语言的全部词语, 覆盖普遍的语

* 收稿日期:

定稿日期:

基金项目: 中国-东盟研究中心(广西科学实验中心)课题“越南语通用语料库建设”(20120146)

作者简介: 林丽(1979-), 女, 四川成都人, 解放军外国语学院讲师, 在读博士。主要研究方向为语言信息处理和越南语句法语义分析。

义领域固然十分理想。然而,适用于普遍领域的语义知识库构建还存在着大量的基础问题需要解决,试图将这样的通用词典直接应用于实用系统,似乎还欠成熟。相比之下,针对某些特定领域展开细致深入的探索研究,用以解决社会需要解决的实际问题,在一定程度上更具现实意义。

本文的研究目标是结合越南语语言特点进行框架语义研究,构建一个具有一定规模的越南语南海新闻领域框架语义知识库,为信息抽取,尤其是事件抽取提供事件语义资源,同时也为今后完整地建构越南语军事领域框架语义知识库和进行越南语-汉语机器翻译研发奠定基础。

2 越南语相关研究现状

在越南语研究界,尽管尚未见到有专门基于框架语义学的研究,但已有一些针对动词或词汇语义方面的研究成果,如: *Động từ trong tiếng Việt* /《越南语动词》(Nguyễn Kim Thán, 1976), *Từ vựng-ngữ nghĩa tiếng Việt* /《越南语词汇语义》(Đỗ Hữu Châu, 1999), *Vị từ hành động tiếng Việt và các tham tố của nó-so sánh với tiếng Nga và tiếng Anh* /《越南语行为动词及其论元-与俄语和英语对比》(Nguyễn Thị Quy, 1995), *Các mô hình kết trị của động từ tiếng Việt* /“越南语动词的配价模型”(Nguyễn Văn Lộc, 2008)等等。它们虽然不是直接面向计算机的,但其中有不少研究一定程度上与语义知识库建设工作相关,只不过这些研究通常都带有比较强的描述色彩,或者所述现象都比较微观,而不是显性地像知识库那样整体形式化呈现出来。以上列举的这些研究成果在越南语框架语义知识库的构建过程中具有一定的借鉴价值和指导作用。

越南语 FrameNet 构建研究相关成果目前可见的仅有一篇论文“*Xây dựng hệ thống FrameNet tiếng Việt*”^①(构建越南语 FrameNet 系统)。该论文由胡志明市越南国家大学信息技术学院计算机科学系的硕士 Nguyễn Văn Toàn(阮文全),在校本科学生 Phạm Minh Quân(范明君), Lê Anh Sơn(黎英山)完成,为2011年该校青年科学会议参会论文(未见公开发表),主要内容为介绍 FrameNet 的结构和内容,并计划将 FrameNet 本地化。他们提出的技术路线为原样照搬 FrameNet 的结构,将词元(Lexical Unit)部分通过自动或半自动方法译成越南语。采用的英越翻译资源为: LLOCE (Longman Lexicon Of Contemporary English), 英越双语词典。对于词元一一对应的情况,基于英越词典直接翻译;对于词元意义一对多的情况,参照中国构建 CFN(汉语框架语义网)的经验,提出三种策略。就文章提出的目标而言,就是将 FrameNet 完全本地化,甚至连例句也一同翻译。这种做法没有考虑到越南语的特点,同时规模也将非常庞大,作者未提到还有更大的团队,目前也尚未见到后续的成果。

3 研究目标与工作流程

本研究以框架语义学为理论基础,以框架语义学理论指导下的 FrameNet 和我国山西大学、上海师范大学等机构合力共建的汉语框架网络(Chinese FrameNet, CFN)为重要参考,以框架语义分析方法为研究手段,以越南语军事新闻语料库为研究范围,目标为探索越南语框架语义分析方法、构建越南语南海新闻框架语义知识库、服务于越南语-汉语信息处理,尤其是针对越南语南海新闻文本的信息抽取。

具体研究目标是系统研究越南语南海新闻动词词元,建立“词元库”;借鉴先贤研究成果,对词元进行语义分类,构拟出领域框架体系;细致描写各框架通用核心框架元素和通用非核心框架元素,建立领域“框架元素库”;在系统描写的基础上,开发例句辅助标注工具,对真实文本语料进行框架语义标注,构建例句库;研究各类语义框架的句法实现规律和规则;并在此基础上进行具体应用探索。

技术路线图如下:

^① <http://uits.vn/khoahoc/author/chientx>

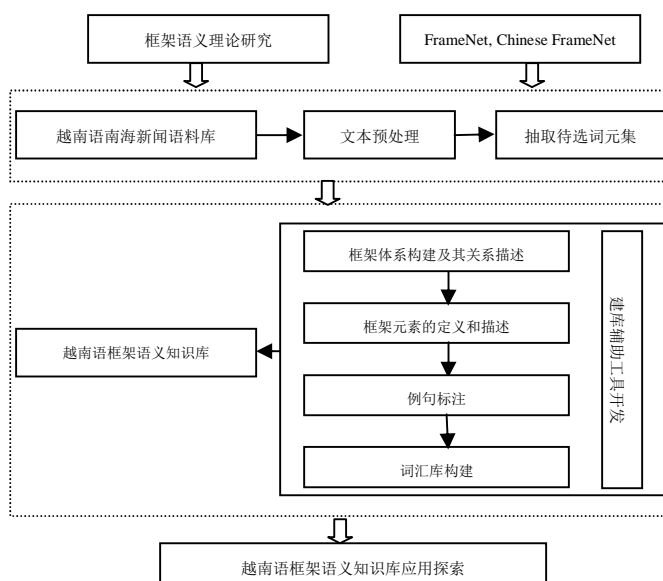


图 1. 越南语框架语义知识库技术路线图

越南语框架语义知识库构建部分是工作的重点，过程具体如下：

(1) 采集目标词元。对领域语料进行词频统计，将语料中出现的高频动词抽取出来作为“待选词元集”；

(2) 通过领域专家的审阅，对待选词元进行归类，拟构出大致的框架体系，然后扩充各个框架的词元集合；

(3) 我们根据特定语义框架的场景并结合实际语料，参照 FrameNet 中相应的框架式所设立的框架元素，在分析越南语句法语义结构的基础上，为每一个框架确定核心框架元素和非核心框架元素并进行详细地描写；

(4) 从语料库中抽取包含该词元的句子，并按照其义项选择句子加以示例；对所选的句子进行框架元素标注；

(5) 汇总框架元素标注结果，显示每个词元在组合上的可能性，即“配价描述”。

现阶段，我们并不致力于描述越南语南海新闻语料中所有出现的词语，而是以满足当前应用需要为准，按照词频和领域专家知识优先的原则，先期进行实验。之后当扩大语料范围时，再相应地增加新的词语描述。

4 研究进展与应用探索

4.1 语料库构建与文本预处理

由于我们开展的研究面向新的语种、新的领域，考虑到时间和人力的局限，将在已建成的“越南语军事新闻语料库”（规模：99M，约 13500 篇）中抽出数量相对有限、重要性突出、时效性强的“南海新闻语料库”（目前规模：4.28M，899 篇）作为研究的对象和语料来源。主要来源为越南国防网^①、越南人民军队网^②、越土报网^③、BBC 越南新闻^④等。

基本标注工具方面，我们主要使用越南河内国家大学 Lê Hồng Phương 博士开发的越南语文本处理工具包 vnToolkit 3.0。该工具包包括分词软件 vnTokenizer、词性标注软件 vnTagger 以及分句软件 vnSentDetector。据 Lê Hồng Phương 个人主页^⑤介绍，vnTokenizer 的准确率和召回率介于 96%—98% 之间。

4.2 越南语南海新闻动词词元库构建

^① <http://quocphong.vn/>

^② <http://www.qdnd.vn/qdndsite/vi-VN/43/Default.aspx>

^③ <http://www.baodatviet.vn/>

^④ <http://www.bbc.co.uk/vietnamese/>

^⑤ <http://www.loria.fr/~lehong/tools/>

我们对 899 篇越南语南海新闻分词后进行词频统计, 得到 17870 个词项, 从中选取高频 (10 次以上) 动词 855 个作为“待选词元集”。经过领域专家的审阅后, 参照汉语南海新闻语料库对待选词元进行了释义。统计结果如表 1 所示。

表 1. 越南语南海新闻语料库高频动词词表 (节选)

| 序号 | 词条 | 词性 | 汉语释义 | 词频 | 频率统计 |
|----|------------|-----|--------|-----|--------|
| 1 | tranh chấp | V/N | 争议 | 999 | 0.1559 |
| 2 | đưa | V | 传送; 列入 | 830 | 0.1295 |
| 3 | tuyên bố | V | 声明 | 682 | 0.1064 |
| 4 | thấy | V | 看 | 678 | 0.1058 |
| 5 | muốn | V | 想 (要) | 643 | 0.1003 |
| 6 | thăm | V | 访问 | 619 | 0.0966 |
| 7 | biểu tình | V | 抗议, 示威 | 576 | 0.0899 |
| 8 | gây | V | 导致 | 552 | 0.0861 |
| 9 | giao | V | 交 (给) | 503 | 0.0785 |
| 10 | hoạt động | N/V | 活动 | 476 | 0.0743 |
| 11 | giải quyết | V | 解决 | 470 | 0.0733 |
| 12 | biết | V | 知道 | 468 | 0.073 |
| 13 | thu | V | 收, 集 | 464 | 0.0724 |
| 14 | tranh | V | 竞争 | 409 | 0.0638 |
| 15 | liên quan | V | 相关联 | 406 | 0.0634 |
| 16 | phát triển | V | 发展 | 403 | 0.0629 |
| 17 | tham gia | V | 加入 | 395 | 0.0616 |
| 18 | tăng | V | 增加, 提高 | 393 | 0.0613 |
| 19 | chống | V | 反对 | 387 | 0.0604 |
| 20 | bảo vệ | V | 保护 | 383 | 0.0598 |

4. 3 领域框架体系构建

语义框架是对场景类型的图式呈现, FrameNet 中的框架选取主要依据语料库内容。新的框架通常是从已有框架中的多义词的其他意义衍生出的。就某一领域而言, 概念较为零散、缺乏系统全面性。

我们参照 FrameNet 以及越南语动词和语义研究成果, 初步将研究语料分为三个子领域, 每个子领域又分别包含有若干子类: 军事域 (军演、军购、冲突...)、经济域 (勘探、开采、招标、投标...)、政治&外交 (访问、抗议、申诉、反对...)。在每个子类中再做动词语义分类, 以突出领域特征。目前, 主要针对“军购”构建相应的框架, 包括 THƯƠNG MẠI_BÁN (商业_卖), THƯƠNG MẠI_MUA (商业_买), THƯƠNG MẠI_GIAO DỊCH (商业_交易) 等。

可以说, 语义分类和框架体系的构建都不是一蹴而就的, 而是一个“之”字形的探索过程。随着我们对待选词元的分类和分析不断深入, 框架体系也会面临一系列的合并、压缩、调整, 最终才可能形成一个较为完整和合理的领域框架体系。

4. 4 领域框架元素库构建

本研究中的框架元素即待选动词词元的语义论元。我们根据特定语义框架的场景并结合实际语料, 参照 FrameNet 中相应的框架式所设立的框架元素, 在分析越南语句法语义结构的基础上, 最终构拟出越南语框架的框架元素。以 THƯƠNG MẠI_BÁN (商业_卖) 框架为例, 其核心框架元素为: 出售者 (người bán)、商品 (hàng hóa)、购买者 (người mua); 非核心框架元素为: 方式 (cách làm)、结算类型 (loại hình thanh toán)、交易金额 (tiền)、持

续时段 (giai đoạn)、地点 (nơi)、目的 (mục đích)、商品使用目的 (mục đích sử dụng)、定价 (sự định giá)、转卖 (bán lại)、结果 (kết quả)、倒回 (ngược lại)、时间 (thời gian)、商品计量单位 (đơn vị)。

| | | | |
|----|--|-----------|-----------------------|
| 名称 | THƯƠNG MẠI BÁN/商业 卖 | | |
| 定义 | Đây là những từ mô tả các giao dịch thương mại cơ bản liên quan đến một người mua và người bán trao đổi tiền, hàng, lấy quan điểm của người bán. Các từ khác nhau có phần trong các mô hình khung thực hiện yếu tố họ cho phép. Ví dụ, các mô hình điển hình cho BÁN: BÁN HÀNG bán cho Bên mua cho tiền. | | |
| 缩写 | Commerce_sell | | |
| | 核心元素 | 非核心元素 | 词元 |
| | 名称 | 缩写 | 说明 |
| | 方式 | Man | cách làm |
| | 短语类型 | Man | Loại hình thành phần |
| | 继续时段 | Alter | giải đoạn |
| | 地点 | Place | nơi |
| | 目的 | Purp | mục đích |
| | 商品使用目的 | POG | Mục đích sử dụng Hàng |
| | 定价 | Rate | sự định giá |
| | 转卖 | Relay | bán lại |
| | 结果 | Result | kết quả |
| | 倒回 | Reversive | ngược lại |
| | 时间 | Time | thời gian |
| | 商品计量单位 | Unit | đơn vị |

图 2. 越南语框架语义知识库框架构建界面

框架元素描写分为“名称”、“缩写”和“说明”三个部分。“名称”用汉语描述、“缩写”尽量使用和 FrameNet 一致的英文表达,“说明”为越南语释义。这种描写方式既保证了研究的通用性,也拓展了其实用性。

4. 5 框架语义标注与例句库构建

例句标注以框架库为基础,给定一个词元,自动从语料库中抽取相应例句,切换到该词元所属框架,分框架元素 (Frame Elements, FE)、短语类型 (Phrase Type, PT) 和句法功能 (Grammatical function, GF) 三个层次进行标注,同时关注未登录命名实体的识别和标注。例句标注界面如图 3 所示。

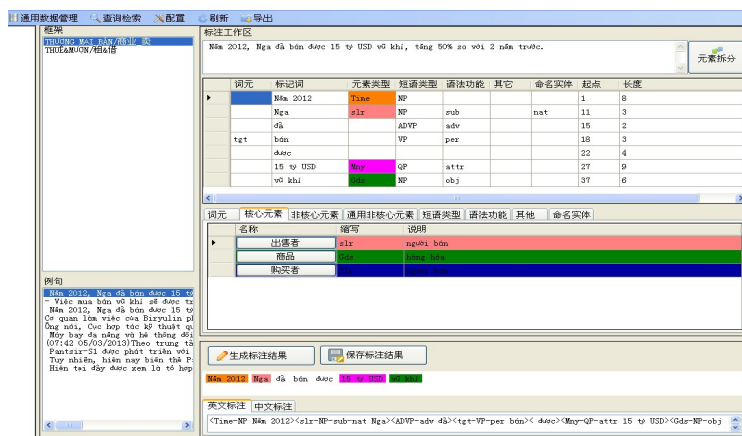


图 3. 越南语框架语义标注界面

以例句 (1) 为例, 标注结果如下:

(1) Nga sắp giao siêu tàu ngầm tàng hình đầu tiên cho Việt Nam.

俄罗斯即将将首艘隐形潜艇交付给越南。

英文标签标注结果:

< del-NP-sub-nat Nga >> ADVP-adv sắp >> tgt-VP-per giao >> Theme-NP-obj siêu tàu ngầm tàng hình đầu tiên >> goa-PP-Dep cho Việt Nam >

其中: del 表示 deliver; tgt 表示 target; goa 表示 goal。

中文标签标注结果:

< 交付者-名词短语-主语-国家名 Nga >> 副词短语-状语 sắp >> 目标词元-动词短语-谓语

giao >> 交付物-名词短语-宾语 *siêu tàu ngầm tàng hình đầu tiên* >> 对象-介词短语-依存项^① cho Việt Nam >

从标注结果来看，其信息模式抽象程度还是比较高的。我们计划在手工标注的例句达到一定的规模以后，把基于规则的方法和机器学习的方法结合起来尝试对真实文本进行语义角色的自动标注。

4. 6 词元配价模式与核心依存图

框架网络资源包含了大量的词汇搭配信息，其中标注的例句可以显示句法语义联系方面的信息。配价模式统计需要在完成一定规模的例句之后才能实现。具体的配价模式统计表以例句（1）为例展示如下：

表 2. 例句（1）配价模式统计表

| | | | | |
|---------|------|-----|-------|-----|
| 标注数量/个 | 配价模式 | | | |
| total 1 | del | tgt | Theme | goa |
| 1 | NP | VP | NP | PP |
| | sub | per | obj | Dep |

但事实上，框架网络项目的中心工作不是关注语义合成原则本身，而是关注为了信息抽取而在框架网络语料运用的相关项目中发展出来的原则（Mohit and Narayanan, 2003）。Fillmore（2004）指出，核心依存图（Kernel Dependency Graphs, KDGs）就是从框架语义资源中抽取出来的新的资源，是 FrameNet 服务于信息抽取的有力工具。通过抽取句子中最凸现的核心依存图，能够发现其所在篇章段落的语义线索。在特定文件中得到确认的 KDGs 能够作为一种标志以显示该文件特定段落的主题事件及其基本主张。

KDGs 由多个词项的结构化的串构成，每个这样的串都包含一个“控制项”（governor）（如动词短语的核心动词）以及它的所有“依存项”（dependents）的词汇核心，每个依存项的词汇核心在其相对于控制项的语义角色方面都做了标注。用“槽-填充项”（slot-filler）的术语讲就是，一个 KDG 的核心唤起一个以分支标签命名的“槽”的结构，而这些“槽”的“填充项”就是依存项的词汇核心。

以例句（2）The puppy drank the milk（小狗喝奶）为例，所属框架 Ingestion（摄取）的核心依存图如图 4 所示^②。

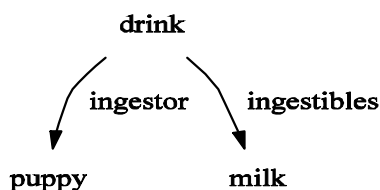


图 4. Ingestion（摄取）框架核心依存图示例

4. 7 事件抽取应用探索

事件抽取（Event extraction, EE）是信息抽取领域一个重要的研究方向。下面分别以简单句和复杂句为例讨论框架语义标注在越南语南海新闻事件抽取中的应用。

简单句即一个句子只包含一个目标词。如例（3）。

(3) Năm 2012, Nga đã bán được 15 tỷ USD vũ khí.

2012年，俄罗斯卖出了150亿美元的武器。

英文标签标注结果：

< Time-NP Năm 2012 >> slr-NP-sub-nat Nga >> ADVP-adv đã >> tgt-VP-per bán >>

^① 根据 FrameNetII: Extended Theory and Practice 第 66 页，双及物句中间接宾语直接处理为依存项（Dependent），不再二次指派语法功能。

^② 该例引自俞士汶，黄居仁（2005）46 页。

được >< Mny-QP-attr 15 tỷ USD >< Gds-NP-obj vũ khí >

其中: slr 表示 seller; tgt 表示 target; Mny 表示 money; Gds 表示 Goods。

中文标签标注结果:

< 时间-名词短语 Năm 2012 >< 出售者-名词短语-主语-国家名 Nga >< 副词短语-状语 đã >< 目标词元-动词短语-谓语 bán >< được >< 交易金额-数量短语-定语 15 tỷ USD >< 商品-名词短语-宾语 vũ khí >

抽取流程如下:

(1) 选定触发词为: < tgt= bán >;

(2) 构建信息模式: < slr >< tgt= bán >< Mny >< Gds >;

(3) 硬性约束条件: 信息模式中 < slr > 和 < Gds > 框架元素为必有, < Mny > 框架元素为可有;

(4) 根据框架层级和关系以及框架元素, 人工建立抽取规则如下:

{ Type=出售者: < slr >; Type=交易金额: < Mny >; Type=货物: < Gds > } < slr >< tgt= bán >< Mny >< Gds >。

复杂句中的目标词超过 1 个。如例 (4) 所示。

(4) Để phục vụ mục tiêu này, Chính phủ Mỹ cần sớm bỏ hoàn toàn lệnh cấm bán vũ khí cho Việt Nam và đóng góp tích cực hơn nữa trong việc giải quyết hậu quả chiến tranh. (为此目标, 美国政府应尽早彻底取消对越南出售武器的禁令, 为解决战争遗留问题作出积极贡献。)

根据目前的例句标注设置, 一个例句中只能按照一个目标词的框架元素、短语类型及语法功能进行标注, 对于多个目标词同时存在的情况, 受空间维度的限制, 只能分多个句子分别进行标注。由于框架网络可以生成核心依存图, 这使得复杂句中的语义关系变得“有据可循”。标注流程如下: 首先识别出句中的目标词, 并确定其所属的语义框架。处理结果如下:

ĐỂ PHỤC VỤ **Function** mục tiêu này, Chính phủ Mỹ cần sớm **BỎ Activity_stop** hoàn toàn **LỆNH CẤM Deny_permission** **BÁN Commerce_sell** vũ khí cho Việt Nam và **ĐÓNG GÓP Giving** tích cực hơn nữa trong việc **GIẢI QUYẾT Resolve_problem** hậu quả chiến tranh.

上句中以大写的形式显示了句中的目标词, 共六个, PHỤC VỤ (服务), BỎ (放弃), LỆNH CẤM (禁令), BÁN (出售), ĐÓNG GÓP (贡献), GIẢI QUYẾT (解决)。黑体字部分是目标词所属的框架, 本文中为说明方便, 先使用对应的英文 FrameNet 中的框架名。其中, LỆNH CẤM (禁令) 为名词, 但它能够激活语义框架式 **Prohibiting**, 起到引发事件的作用, 因此也纳入目标词之列。句中的动词 phục vụ^① (服务) 与情态动词 cần (应当、需要) 暂时不在我们的标注范围内。

例 (4) 生成的核心依存图如图 5 所示。在该图中, 目标词被表示为黄色背景的背景的节点及其依存项。节点由 <Frame name>.<LU name>.<POS> 三个部分组成。指向依存项的箭头标注了框架元素名称。紫红色高亮显示部分为命名实体。虚线框中的目标词出现在定语或状语中。虚线箭头表示有语义关系, 但不能直接抽取, 在抽取规则中以圆括号表示。

抽取规则如下:

(1) { Type=目的: < Purpose >; Type=施事: < Agent >; Type=当事: < Theme > } < Purpose >< Agent >< tgt= bỏ >< Theme >;

(2) { Type=目的: < Purpose >; Type=施事: < Agent >; Type=当事: < Theme > } < Purpose >< Agent >< tgt= đóng góp >< Theme >;

(3) { (Type=权威方: < Authority >); Type=行为: < Action >; (Type=当事方: < Protagonist >)} < Authority >< tgt= lệnh cấm >< Action >< Protagonist >;

^① phục vụ 虽然被识别为目标词, 但位于目的状语“Để phục vụ mục tiêu này”中。目前的标注范围是主句中的目标词。

(4) { (Type=卖方: < Seller >); Type=商品: < Goods >; Type=买方: < Buyer > }(< Seller >)
 < tgt= bán >< Goods > cho < Buyer >;

(5) { (Type=施事: < Agent >); Type=问题: < Problem > }(< Agent >)< tgt= giải quyết >< Problem >。

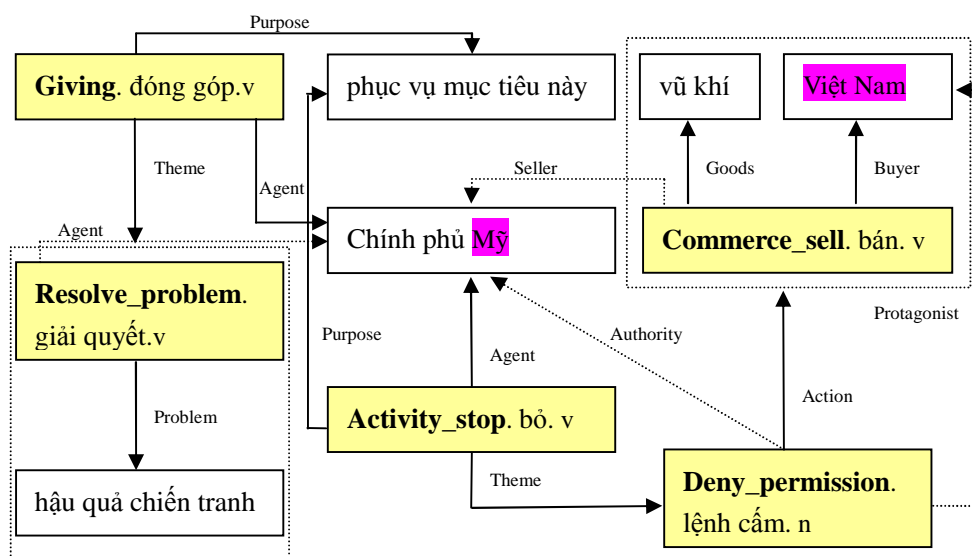


图 5. 越南语复杂句的核心依存示意图

5 总结与展望

针对越南语军事新闻的框架语义研究有着客观的迫切需求，但碍于相关研究的滞后，目前尚无较有规模和体系的成果。本文在总结越南相关研究现状的基础上尝试构建越南语框架语义知识库，运用框架语义标注方法抽取特定事件信息，初步探索了框架语义分析方法在越南语文本中的应用。

目前，越南语框架语义知识库尚处于起步阶段，已构建框架 16 个，手工标注例句约 500 句。本研究的理论基础相对扎实，但实践方面各项资源准备尚不够充分，尤其是越南语框架语义标注的规模还远远不够。因此，我们所希冀的结果是我们在理论方法上的探索能有所裨益，实践方面将继续加大力度进行拓展，以期为今后的自动标注和机器学习打下较为厚实的基础。

参考文献

- [1] Nguyễn Kim Thân. *Động từ trong tiếng Việt* [M]. Hà Nội: Nhà xuất bản khoa học xã hội, 1976.
- [2] Đỗ Hữu Châu. *Từ vựng-ngữ nghĩa tiếng Việt* [M]. Hà Nội: Nhà xuất bản khoa học xã hội, 1999.
- [3] Nguyễn Thị Quy. *Vị từ hành động tiếng Việt và các tham tố của nó-so sánh với tiếng Nga và tiếng Anh* [M]. Hà Nội: Nhà xuất bản khoa học xã hội, 1995.
- [4] Nguyễn Văn Lộc. Các mô hình kết trị của động từ tiếng Việt [J]. *Ngôn ngữ*, 2008, 2.
- [5] 周岭顺. 汉语移动域框架语义分析[M]. 北京: 社会文献出版社, 2012.
- [6] 郭丹丹, 刘伟. 汉语框架网络数据库例句辅助标注系统的设计与实现[J]. *科技情报开发与经济*, 2010, 032: 98-102.
- [7] 俞士汶, 黄居仁. 计算语言学前瞻[M].北京: 商务印书馆, 2005: 9.
- [8] 刘开瑛. 汉语框架语义网构建及其应用技术研究[J]. *中文信息学报*, 2011, 25(6): 46-52.
- [9] Mohit, Behrang, and Srini Narayanan. Semantic extraction with wide-coverage lexical resources[A]. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers-Volume 2. *Association for Computational Linguistics*[C], 2003: 64-66.
- [10] Fillmore, Charles J., Josef Ruppenhofer, and Collin F. Baker. Framenet and representing the link between

- semantic and syntactic relations[A]. *Frontiers in linguistics 1* [C], 2004:19-59.
- [11] Fillmore, Charles J., Christopher R. Johnson, and Miriam RL Petruck. Background to framenet[J]. *International journal of lexicography*, 2003, 16.3: 235-250.
- [12] Ruppenhofer, Josef, et al. FrameNet II: Extended theory and practice[DB/OL]. 2006. https://framenet.icsi.berkeley.edu/fndrupal/the_book