

doi:10.16104/j.issn.1673-1883.2021.04.003

# 信息处理用彝语语义网络知识表征的实践研究

田宣宣<sup>a,b</sup>

(西南民族大学 a. 民族语言文字信息处理重点实验室; b. 中国语言文学学院, 四川 成都 610041)

**摘要:**借鉴语义网络关于语义知识表征的理论和实践,提出彝语语义网络表征的思路。首先描述彝语语义网络表征的语义单位和方法;然后分析彝语语义网络表征的分类方式,阐明彝语的静态聚合网络、动态组合网络及知识网络的激活过程;最后探讨彝语语义网络在机器翻译和知识推理中的可能应用。彝语语义网络表示方式的探索,为构建富含彝语信息资源的语义知识库奠定了基础,以期推动彝文信息处理智能化应用和自然语言处理等工作的创新发展。

**关键词:**彝语;语义网络;知识表示;形式化;自然语言处理

**中图分类号:**H217;TP182 **文献标志码:**A **文章编号:**1673-1883(2021)04-0012-06

## Practical Study on Semantic Network Knowledge Representation of Yi Language for Information Processing

TIAN Xuanxuan

(a. Key Laboratory of Ethnic Language and Character Information Processing;

b. College of Chinese Language and Literature, Southwest Minzu University, Chengdu 610041, China)

**Abstract:** Based on the theory and practice of semantic knowledge representation in semantic network, this paper puts forward the idea of the representation of semantic network of Yi language. Firstly, the semantic units and methods of semantic network representation in Yi language are described. Secondly, it analyzes the classification of semantic network representation in Yi language, and clarifies static aggregation network, dynamic combination network and the activation process of knowledge network. Finally, the possible applications of Yi language semantic network in machine translation and knowledge reasoning are discussed. The exploration of semantic network representation of Yi language lays a foundation for the construction of semantic knowledge base with rich Yi language information resources, so as to promote the innovative development of intelligent application of Yi language information processing and natural language processing.

**Keywords:** Yi language; semantic network; knowledge representation; formalization; natural language processing

### 一、引言

语义网络作为计算机能够理解的表示知识资源的最重要的手段之一,是指用可视化方式呈现知识,把各种线性的、离散的、非结构化的知识,用图这种数据结构形式组织起来,从而描述关于世界万物的实体(entities)、概念(concepts)、事件(events)及其之间的关系<sup>[1]</sup>。研究面向机器处理的语义网络知识表示十分重要,表现在两个方面:其一,语义网络表示将客观世界抽象成一个概念层次体系,建立语义表示框架,通过语义属性准确地描述事物特

征,能够对整个世界或者某一领域的知识进行详细描述<sup>[2]</sup>;其二,以真实文本为基础构建的语义网络模拟人类语义或概念知识的组织、人的语义处理机制、语义检索过程<sup>[3]</sup>,为计算机的自然语言处理提供更接近人类语言行为的特征模型,是实现信息化、智能化的必然趋势。目前,语义网络表达知识的理论和方法日渐受到关注,在自然语言处理的机器翻译、信息检索、情感分析等各领域都具有潜在优势,为深入开展此类工作提供不可或缺的基础性关键资源。

近年来,在我国大力推进民族语言文字信息化

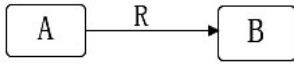
收稿日期:2021-06-21

基金项目:教育部社科规划项目(17YJA740051);四川省社科规划项目(SC20B130);西南民族大学中央高校基本科研业务费专项资金优秀学生培养工程项目(2020YYX506)。

作者简介:田宣宣(1996—)女,河南鹿邑人,硕士研究生,主要研究方向:中国少数民族语言、民族语言资源库建设。



指明它所联结的结点间各种语义联系<sup>[7]</sup>。具体可表示为:



其中,A、B 分别表示两个结点,通常用概念表示,R 表示 A 与 B 之间的某种语义关系。语义网络中概念之间的联系多种多样,常用的一般关系如表 2 所示。

表 2 几种常用的概念之间的关系

关系名	关系标记	关系描述	实例
类属关系	ISA、A-kind-of	类的实例与类之间的关系	小李(人) → ISA → 人
属性关系	Have、Can	事物和其属性之间的关系	鸟 → Have → 翅膀
聚类关系	Part-of	类之间部分与整体的关系	大脑 → Part-of → 人体
时间关系	Before、After	事件发生的时间先后次序关系	火把节 → Before → 彝族年
位置关系	Located-on、Located-at	事物在位置方面的关系	书 → Located-on → 桌子
因果关系	If-then	事件的发生导致另一事件的发生	晴天 → If-then → 玩耍
相近关系	Similar-to、Next-to	事物相近或相似	猫 → Similar-to → 老虎

对于错综复杂的关系,可以通过多个语义基元的组合来实现,构造出表示语义关系的有层次的网络。图 1 所示的语义网络给出了概念之间的某些语义关系,显然,一个概念可以和多个概念同时发生联系,这实际上是“语义组合系统”的形式化图解,也是智能电脑的“知识表示”<sup>[8]</sup>。该语义网络表示“Sᳵ(动物)”能运动、会吃;“ᳵ(鸟)”是一种“Sᳵ(动物)”,它有翅膀,会飞;“ᳵ(鱼)”是一种“Sᳵ(动物)”,它生活在水中,会游泳。

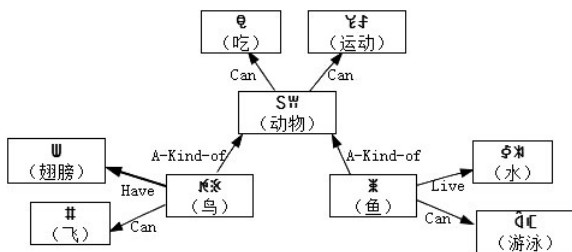


图 1 复杂关系的语义网络

(三) 彝语语义网络表示的分类

实现计算机理解和处理自然语言,必须模拟人脑对客观世界的认知过程以及对语言知识的理解和描述,用计算机能够理解的形式化方式表示内部语义知识,进而使其获取和利用与人相近或相同的语言知识和世界知识,整个过程如图 2 所示。语义网络具有很好的知识表示逻辑性和技术实现可操作性<sup>[9]</sup>。它通过模拟人脑联结机制中的“组合机制”和“聚合机制”,形成“动态的语义组合网络”和

“静态的语义聚合网络”<sup>[10]138</sup>,是实现自然语言深层理解的重要基础。下文将依据这两种语义网络,探讨彝语语义网络表示的具体内容。

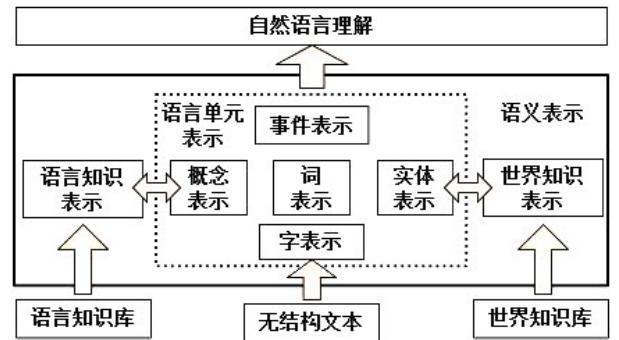


图 2 计算机自然语言理解的全过程

1. 动态的语义组合网络

语义组合网络是能够清晰地表示概念之间和事件之间全部语义组合关系和语义指向的有层次的动态语义主网络<sup>[11]30</sup>。

1) 用语义网络表示事实

(1) 一般表示,如图 3 所示。

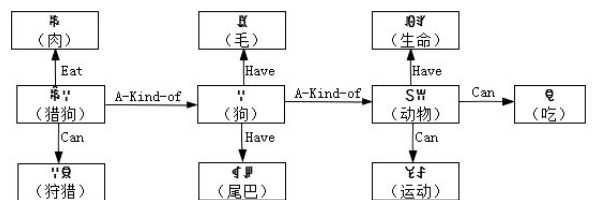


图 3 语义网络表示一般事实





以上述事件“ $\text{H}\text{C}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}$ ”为例,图7“动态的语义组合主网络”明确标注了原文和译文的“主体”“客体”“动作”“数量”等对应角色信息,按照彝语“宾语在谓语之前”“数量词组作定语在中心语之后”和汉语“宾语在谓语之后”“数量词组作定语在中心语之前”的不同的句法规则,生成的译文序列为“张三给李四一只鸟”。

## (二) 彝语语义网络在知识推理中的应用

从服务于彝文信息处理的需求出发,语义网络可以直观高效地描述事物的语义结构和相关的百科知识,该种语义知识表示形式使得计算机充分理解人类自然语言的意义、掌握常识并据此进行知识推理<sup>[15]</sup>。语义网络的推理机制一般基于网络的匹配,根据提出的问题构造一个局部网络,其中有些结点或弧的标识为空,代表待求的问题。比如,如果问: $\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\text{O}$ ? (谁给了李四一只鸟?)该事件的语义网络与图7“动态的语义组合主网络”类似,不同的地方是标有“ $\text{HC}$ (张三)”的结点变成了标有“?”的结点。则解决该推理问题首先需在知识库中寻找与之可匹配的语义网络,当局部网络中待求的问题与知识库中图7的网络匹配时,匹配的部分就是问题的解,即张三。

进一步地,调用上述事件“ $\text{HC}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}$ (张三给李四一只鸟)”的“静态的语义聚合子网络”,根据语义网络的属性继承性和关系可传递性,计算机通过推理可获得下列知识:

①  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  是  $\{\text{S}\text{W}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$ ,  $\{\text{S}\text{W}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  是  $\{\text{S}\text{F}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$ , 所以  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  也是  $\{\text{S}\text{F}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (生物)。

②  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  是  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  的上级,  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (云雀)

雀)是  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  的下级。

③  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  是  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  的构件,  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (云雀)是  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$ , 所以  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  也是  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (云雀)的构件。

④  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  有羽毛、无牙齿,  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (啄木鸟)是  $\{\text{R}\text{E}\text{R}\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$ , 所以  $\{\text{E}\text{O}\text{J}\text{H}\text{S}\text{J}\}$  (啄木鸟)有羽毛、无牙齿。

## 五、展望与思考

基于语义网络的方法进行语义知识的描述从本质上来讲是一种借助形式化模型,实现计算机对实际自然语言的语义分析和理解。语义网络不仅关注概念与概念之间动态的横向组合关系,同时也关注由概念联想形成的静态的纵向聚合关系,既能够解释事件中概念之间的语义组合形式和组合语义的生成,且网络结构又形象地刻画了自然语言文本中词语之间的运作机制和语言表达模式。本文以语义网络的研究理论、构建方法及在NLP中的实践应用为基础,详细探讨了在人工智能迅速发展的背景下彝语语义网络表示的关键问题和可能的应用,以真正实现分层次地、深入地、形式化地、精密地挖掘彝语文本中的语义信息。本文建议,在彝语语义网络表示的研究基础上,再结合大规模真实语言文本,下一步需要构造可有效应用于彝文信息处理的宝贵知识资源库,全面而深入地揭示各种语义知识,以便于计算机能够对彝语语义进行分析与计算。这无论是对彝语的词义消歧、机器翻译、信息检索等领域的自然语言处理问题,还是计算机辅助教学、语言学本体研究都有着极为重要的应用价值。

### 注释:

① 本文所说的“激活”即计算机通过调用语义知识充分理解自然语言,参见鲁川(2010)。

### 参考文献:

- [1] 袁毓林,曹宏.“语义网—本体知识—知识图谱”和语言研究[J].汉语学报,2021(1):17.
- [2] NIRENBURG S, RASKINV. Ontological Semantics (Language, Speech, and Communication) [M]. Cambridge: The MIT Press, 2004.
- [3] 刘海涛. 汉语语义网络的统计特性[J]. 科技通报, 2009(14): 2060.
- [4] 王成平. 彝语语言语料资源数据库的设计与共享的实现[J]. 中文信息学报, 2016(1): 129-132.
- [5] WANG Chengping. Research on design and sharing of yi language corpus resources database based on syntactic rules[J]. Solid State Technology, 2020(5): 10618-10629.
- [6] 鲁川. 汉语语法的意合网络[M]. 北京: 商务印书馆, 2001: 48.
- [7] 刘颖. 计算语言学[M]. 北京: 清华大学出版社, 2002: 137-138.
- [8] 鲁川. 汉语的意合网络[J]. 语言文字应用, 1998(2): 82.

山川名等专有名词的翻译中,把“a bbat cie lo”译为音译的“阿吧才咯”加上泛指性上义词“山”的“阿吧才咯山”,同样把“nyie lo hat”译为“泥罗汉山”;在毕摩经书的翻译中,把“jjyp lup”意译为“护佑神”、把“gep fi”意译为“生育神”、把“qu yyr shyx yyr”意译为“财神”等。

总之,在彝语文名词术语的汉语文翻译过程中,只有在遵循译名的单一性、科学性、系统性、简明性、约定俗成性等名词术语的翻译原则下,灵活应用名词术语翻译的方法与技巧,才有可能使译名在学术理论上站得住脚、在现实应用中有生命力。

#### 注释:

- ① 毕摩,彝语音译,彝族社会中一种替人礼赞、祈祷、消灾、祭祀的祭师。
- ② 为方便编辑,本文例句中的彝文用规范彝文拼音转写。

#### 参考文献:

- [1] 标准化与信息分类编码研究所. 中华人民共和国国家标准术语工作原则与方法, GB/T10112—1999[J]. 北京: 中国标准出版社, 2000: 5-6.
- [2] 冯志伟. 现代术语学引论[M]. 北京: 商务印书馆, 2011: 1-12.
- [3] G·隆多. 术语学概论[M]. 北京: 科学出版社, 1985: 20-22.
- [4] 余华. 彝语翻译史概述[J]. 民族翻译, 2015(4): 54-60.
- [5] 李旭练, 余华. 政治文献民族语文翻译工作的实践经验与发展路径[J]. 民族翻译, 2019(1): 43-48.
- [6] 余华. 论汉彝翻译中无等植物词汇的翻译[J]. 民族翻译, 2020(3): 35-40.

(上接第 17 页)

- [9] 王飞, 易锦竹, 谭新. 基于本体语义网络的语言理解模型[J]. 计算机科学, 2018, 45(6A): 102.
- [10] 鲁川. 知识工程语言学[M]. 北京: 清华大学出版社, 2010: 138-139.
- [11] 鲁川. 智能计算机的知识表示和汉语的语义研究[J]. 语文建设, 1992(11): 30-32.
- [12] 王俊, 高炜. 基于多元语义网络的民族信息资源库构建研究[J]. 苏州科技大学学报(自然科学版), 2018, 35(3): 75.
- [13] 谭莹, 张进, 夏立新. 语义网络发展历程与现状研究[J]. 情报、信息与共享, 2019(6): 107.
- [14] 冯志伟. 自然语言处理综论[M]. 北京: 电子工业出版社, 2005: 497.
- [15] 袁毓林. 怎样利用语言知识资源进行语义理解和常识推理[J]. 中文信息学报, 2018(12): 12-13.

## 四、结语

根据社会现实需要开展彝语文名词术语汉语文翻译的理论研究, 从而推动这项工作科学发展, 使之更好地为群众服务、为社会建设服务、为党和国家服务, 是彝语文翻译界应尽的责任。当然, 彝语文名词术语汉语文翻译工作是一项涉及面较广的系统工程, 其涉及名词术语的理论研究、收集整理与翻译, 以及组织协调工作及其经费等事项。对此, 要做好这项工作, 需要有关部门及彝语文翻译界的专家学者群策群力、与时俱进, 进行长期的努力。