

doi:10.16104/j.issn.1673-1883.2022.02.002

基于特征结构理论的彝语主谓谓语句结构标注与实践研究

田宣宣^{a,b}

(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:语义分析是当前计算语言学研究的一个重要任务,也是促进自然语言处理迈向新的阶段的主要挑战。在彝语言文字信息处理中,对彝语语句的语义标注是亟须解决的问题之一。研究选用特征结构理论框架,以彝语主谓谓语句的语义标注为切入点进行初步探索,用形式化手段表示彝语主谓谓语句的十一种语义关系类型,并归纳整理出彝语主谓谓语句的特征结构模型。研究表明,特征结构理论标注效率和标注精度较高,可以比较全面的描述彝语的复杂语义信息,为探索适合彝语语言事实的语义标注策略提供了一个可参考的解决方案。

关键词:彝语;主谓谓语句;特征结构;语义标注

中图分类号:H217 **文献标志码:**A **文章编号:**1673-1883(2022)02-0006-07

Structure Annotation and Practical Study of Subject Predicate Sentence in Yi Language Based on Feature Structure Theory

TIAN Xuanxuan^{a,b}

(a.Key Labortory of Ethnic Languge and Character Information Processing;b.College of Language and Literature, Southwest Mizu University, Chengdu, Sichuan 610041, China)

Abstract: Semantic analysis is not only an important task of computational linguistics, but also a major challenge to promote natural language processing to a new stage. In the Yi language character information processing, the semantic annotation of Yi language sentences is one of the urgent problems to be solved. This paper selects the theoretical framework of characteristic structure, starts from a preliminary exploration with the semantic annotation of Yi subject predicate sentences, uses formalized means to represent eleven semantic relationship types of Yi subject predicate sentences, and also summarizes the characteristic structure model of subject predicate sentences in Yi language. The research shows that the feature structure theory has high annotation efficiency and accuracy, and can comprehensively describe the complex semantic information of Yi language, which provides a reference solution for exploring the semantic annotation strategy for the facts of Yi language.

Keywords: Yi language; subject predicate sentence; characteristic structure; semantic annotation

一、引言

语义分析是指获取包括字、词、句子、篇章等各级语言单位的语义信息。在自然语言处理领域,语义分析主要包含两类任务:一是语义结构分析,旨在探讨语义基本单位及其组合规律;二是语义内容分析,即识别这些语言单位所描述的实体、事件及各种关系,包括实体间关系、事件与实体间关系和

事件间关系等^[1]。自步入信息化时代,如何从原始数据出发对其语义结构和语义内容进行规范和标注,并结合预训练的语言模型,实现有效、深入、自动的语义分析,是适应更为深入地研究语言和自然语言处理系统的根本要求。于是,“计算机语义标注”和“计算机语义理解”等概念接踵而来,大规模真实文本处理迫切需要建立一套适合本民族语言的语义标注体系,使其以一种精度高、形式化、可计

收稿日期:2022-01-09

基金项目:教育部项目彝语方言综合资源数据库的建设研究(17YJA740051);中央高校基本科研业务费专项基金项目基于汉彝映射关系的彝语方言资源智能化保护与开发研究(ZYN2022026);四川省2021-2023年高等教育人才培养质量和教学改革项目(JG2021-440);四川省民族语言文字信息处理重点实验室建设项目(2021PTJS32)。

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

算的方式呈现出来。

彝语属于汉藏语系藏缅语族彝语支,共有六大方言区,是彝族人民思想交流、信息传播和文化沟通的重要工具。彝语言文字信息处理是用计算机对彝语语言的音、形、义等信息进行处理,即对字、词、句、篇章的输入、输出、识别、分析、理解、生成等的操作与加工^[2]。目前,彝语言文字信息处理的理论和技术在字、词层面的分析比较成熟和完善,语义层面对知识资源的挖掘不断深入,并在应用开发中逐渐达到实用化水平。然而,随着彝语信息化进程的迅速发展,彝语自然语言处理对语义知识的需求越来越迫切,面临着对真实文本的语义进行深层次标注和分析的问题,这是制约彝语语义资源信息数据库建设和计算机语义理解的主要难点之一,阻碍着计算机彝文信息处理向智能化、深层次分析的方向发展。因此,亟须探索一种适合彝语独特特点的语义描写机制,完善计算机彝文信息处理技术的语义标注模型和理论,基于此面向真实文本构建大规模彝语语义标注资源,实现彝语语义知识描写的形式化和规范化,为彝语言文字信息处理中的语义消歧、语义知识抽取、信息检索、自动问答、文本蕴含及机器翻译等领域提供语言学知识和数据资源。

特征结构是描述某些语言学实体局部信息的一种表示方法,也是用于把信息约束加到可以接受该信息的语言学实体的一种方法^[3],对于彝语自然语言标注和处理具有重要的启发意义。本文尝试借助特征结构理论对彝语主谓谓语句进行语义标注,分析彝语主谓谓语句的内部语义关系和语义标注模型,探讨适合彝语语言事实的语义表示形式和有效的语义分析策略,为面向信息处理的大规模彝语语义资源建设提供一个全新的视角。

二、彝语主谓谓语句的特征结构标注思路

(一) 理论支撑

1. 主谓谓语句

在语言学界,主谓结构谓语的格式是汉语里最常见最重要的现象之一^[4],由此构成的“名₁-名₂-动”这种陈述关系层层套叠的句式,一般都称为“主谓谓语句”^[5]。它的一般语形可以表示为:S1+S2+V/A,其中,S1指全句的主语,也称为“大主语”,S2+V/A是整个句子的谓语,被称作“大谓语”,S2充当主谓短语的主语,也称作“小主语”,V/A是主谓短语的谓语,被称作“小谓语”。主谓谓语句的研究成果十分丰富,焦点主要集中在对其范围、句法特点

与功能、定价模式的分析上。由于主谓谓语句的句子成分之间存在着复杂的语义关系,对其语义类型的划分也是一项重要的研究内容。朱德熙^[6]在《语法讲义》中将主谓谓语句分出了七种语义类型;黄伯荣、廖序东、陈昌来等^[7]根据大主语和小主语之间的语义关系对汉语主谓谓语句进行了比较详细的分类;袁毓林、张旺熹、吴中伟等一些学者也从不同角度对主谓谓语句展开了较为全面的语义分析^[8-10]。总体来说,学界对主谓谓语句这种特殊语言现象的描写研究比较透彻,但随着信息处理技术的发展,这些成果对于应用于计算机理解自然语言还存在相当的困难,于是,一些学者对这类语句进行了句法和语义的形式化研究。缪建明、张全^[11]尝试在HNC理论下,从计算角度形成主谓谓语句特有的处理规则和专家处理系统;姚从军、罗丹^[12]借助组合范畴语法实现了汉语主谓谓语句的句法和语义的推演;陈禹^[13]基于“基线/加工”理论模型分析了主谓谓语句的语言组构模型等研究为计算机处理主谓谓语句提供了丰富的理论支持。

彝语中也存在着大量的主谓谓语句,如《凉山彝语语法》和《现代彝语概论》中将该类句式定义为一种以主谓词组作谓语的语句,谓语主要是说明或描写主语的^[14-15];《现代彝语》强调主谓词组做谓语具有陈述作用^[16];另外,陈国光、胡素华等也对彝语的主谓谓语句进行了实例分析。该类句式形式比较灵活,语义关系复杂,具有特殊的表意功能,给彝语言文字信息处理研究带来了较大的困难。因此,如何从信息处理的角度对彝语主谓谓语句进行语义标注和形式化分析,是面向计算机彝文信息处理中语义资源建设的不可缺少的一项基础性工作。

2. 特征结构理论

在语言学理论和自然语言处理领域,国内外学者都对“特征结构”进行过讨论。“特征”最初是语音学中用来描述语音的方法论^[17]。早期,布拉格学派把音位作为一组区别性特征来分析,“+”表示具有某个特征,“-”表示不具有某个特征。每个音位由若干区别性特征组成,如汉语音位/p/由[+双唇、+清音、+送气]等特征结合而成,其他音位至少有一个特征与/p/相区别,由此看来,特征是音位学分析中的最小单位。雅可布逊等人把声学语音学的成果应用到音系学分析中,提出了区别性特征的理论,认为任何语言的语音中都存在若干最小区别性,并根据发音状况把语音的区别归纳为十二对特征,这些特征旨在揭示音位对比,本质上都具有偶值性质。通过这一套区别特征和一套音系学规则,

可以对各种语言的语音系统作出分析归纳和描写说明^[18]。乔姆斯基和哈勒对十二对特征进行了增减^[19],之后这些区别性特征在不同的语言中有不同的增减变化,一些学者也用不同的特征去分析同一种语言。

20 世纪 50 年代,句法学中有人尝试用特征来描述句法结构。Chomsky^[20]在倡导的理论中进一步推广使用了“特征”一词,形式句法理论(如 LFG 和 GPSG)采用的复杂特征集类似于特征结构的描述机制,对语言实体进行精确的描述和自然语言的最终建模。冯志伟^[3]详细介绍了特征结构和把特征结构的信息内容结合起来的合一运算,指出特征结构是特征-值偶对的集合,其中特征是来自某个有限集合的不可分析的原子符号,而特征的值或者是原子符号,或者是特征结构,此外他还提出,特征结构可以类型化,通过类型化的特征结构把约束放在给定特征结构能够取值的类型上,也可以把它们组织到类型层级体系中,以捕捉一般化的东西。Witt 等^[21]展示了如何用 TEI 标记集来表示特征结构,以作为语言注释资源的元表示格式,认为特征结构是当今坚持不同语法理论的语言学家之间的一种常见的表示形式,用 TEI 标记集来表示特征结构可以为后续算法处理开辟新的可能性,是语言数据可持续性研究和保存工作的一部分。

20 世纪 80 年代,汉语语法研究领域开始用特征进行语义分析,提出了“语义特征”的概念。学界普遍引用和接受的是陆俭明的观点,认为语义特征着眼于分析概括同一句式的各个实例中处于同一关键位置上的词所共有的语义特征,以说明某些句式的独特特点^[22]。他还指出语言研究要满足“数字化”要求,以便于计算机存储、计算,那么语言和语言研究成果的各个层面、各种单位、各种规则的性质都必须用“特征”来加以呈现^[23]。2007 年武汉大学语言与信息研究中心尝试用特征结构理论进行语义标注的研究工作。2014 年《特征结构及其汉语语义资源建设》一书提出的特征结构语义描述模型和建构的大规模语义资源库在语义研究、自然语言处理等方面得到了广泛应用。特征结构理论是一种实用的语义分析策略。一般来说,一个短语或句子可表示成一个由实体、特征和特征值组成的三元组集合,这个集合称为该短语或句子的特征结构,找出短语或句子中词与词之间的语义关系并完整地描述出有语义关系的词语对,是特征结构的根本任务^[24]。特征结构有三种属性:第一,它允许语义的多重关联和交叉关联,即一个词可以和多个词

产生语义联系;第二,它允许递归和嵌套,即允许短语或句子的内部结构作为一个整体和其他词语产生意义关联;第三,它可以用特征结构图来表示,是一种用概念和语义关系来表达知识的无向图。

(二) 语料资源

本研究选用的语料均源自西南民族大学民族语言文字信息处理研究中心历年来完成的国家自然科学基金项目、国家社会科学基金项目、教育部项目等所收集、整理的彝语语料,且主要以喜德语音为标准音,以圣乍话为基础方言。根据研究需要本文对语料进行调整,筛选出 186 个彝语主谓谓语句,其中,大主语、小主语和小谓语等组成部分之间的语义关系比较复杂。本文并不试图判定彝语主谓谓语句究竟有多少语义关系类型,而是探索用特征结构理论分析该句式的可行性,研究面向彝语自然语言处理的语义表示方法。对这 186 个语句进行语义关系划分,大致有十一种类型,如表 1 所示。

表 1 彝语主谓谓语句语义关系分类表

| 序号 | 类型 | 代表例句 |
|----|--|---|
| 1 | S1 是受事, S2 是施事, 全句的语义关系是: 受事 施事——动作 | [1] 𐄂 𐄃 𐄄 𐄅 𐄆 𐄇 𐄈 𐄉 𐄊 𐄋 𐄌 𐄍 𐄎 𐄏 𐄐 𐄑 𐄒 𐄓 𐄔 𐄕 𐄖 𐄗 𐄘 𐄙 𐄚 𐄛 𐄜 𐄝 𐄞 𐄟 𐄠 𐄡 𐄢 𐄣 𐄤 𐄥 𐄦 𐄧 𐄨 𐄩 𐄪 𐄫 𐄬 𐄭 𐄮 𐄯 𐄰 𐄱 𐄲 𐄳 𐄴 𐄵 𐄶 𐄷 𐄸 𐄹 𐄺 𐄻 𐄼 𐄽 𐄾 𐄿 𐅀 𐅁 𐅂 𐅃 𐅄 𐅅 𐅆 𐅇 𐅈 𐅉 𐅊 𐅋 𐅌 𐅍 𐅎 𐅏 𐅐 𐅑 𐅒 𐅓 𐅔 𐅕 𐅖 𐅗 𐅘 𐅙 𐅚 𐅛 𐅜 𐅝 𐅞 𐅟 𐅠 𐅡 𐅢 𐅣 𐅤 𐅥 𐅦 𐅧 𐅨 𐅩 𐅪 𐅫 𐅬 𐅭 𐅮 𐅯 𐅰 𐅱 𐅲 𐅳 𐅴 𐅵 𐅶 𐅷 𐅸 𐅹 𐅺 𐅻 𐅼 𐅽 𐅾 𐅿 𐆀 𐆁 𐆂 𐆃 𐆄 𐆅 𐆆 𐆇 𐆈 𐆉 𐆊 𐆋 𐆌 𐆍 𐆎 𐆏 𐆐 𐆑 𐆒 𐆓 𐆔 𐆕 𐆖 𐆗 𐆘 𐆙 𐆚 𐆛 𐆜 𐆝 𐆞 𐆟 𐆠 𐆡 𐆢 𐆣 𐆤 𐆥 𐆦 𐆧 𐆨 𐆩 𐆪 𐆫 𐆬 𐆭 𐆮 𐆯 𐆰 𐆱 𐆲 𐆳 𐆴 𐆵 𐆶 𐆷 𐆸 𐆹 𐆺 𐆻 𐆼 𐆽 𐆾 𐆿 𐇀 𐇁 𐇂 𐇃 𐇄 𐇅 𐇆 𐇇 𐇈 𐇉 𐇊 𐇋 𐇌 𐇍 𐇎 𐇏 𐇐 𐇑 𐇒 𐇓 𐇔 𐇕 𐇖 𐇗 𐇘 𐇙 𐇚 𐇛 𐇜 𐇝 𐇞 𐇟 𐇠 𐇡 𐇢 𐇣 𐇤 𐇥 𐇦 𐇧 𐇨 𐇩 𐇪 𐇫 𐇬 𐇭 𐇮 𐇯 𐇰 𐇱 𐇲 𐇳 𐇴 𐇵 𐇶 𐇷 𐇸 𐇹 𐇺 𐇻 𐇼 𐇽 𐇾 𐇿 𐈀 𐈁 𐈂 𐈃 𐈄 𐈅 𐈆 𐈇 𐈈 𐈉 𐈊 𐈋 𐈌 𐈍 𐈎 𐈏 𐈐 𐈑 𐈒 𐈓 𐈔 𐈕 𐈖 𐈗 𐈘 𐈙 𐈚 𐈛 𐈜 𐈝 𐈞 𐈟 𐈠 𐈡 𐈢 𐈣 𐈤 𐈥 𐈦 𐈧 𐈨 𐈩 𐈪 𐈫 𐈬 𐈭 𐈮 𐈯 𐈰 𐈱 𐈲 𐈳 𐈴 𐈵 𐈶 𐈷 𐈸 𐈹 𐈺 𐈻 𐈼 𐈽 𐈾 𐈿 𐉀 𐉁 𐉂 𐉃 𐉄 𐉅 𐉆 𐉇 𐉈 𐉉 𐉊 𐉋 𐉌 𐉍 𐉎 𐉏 𐉐 𐉑 𐉒 𐉓 𐉔 𐉕 𐉖 𐉗 𐉘 𐉙 𐉚 𐉛 𐉜 𐉝 𐉞 𐉟 𐉠 𐉡 𐉢 𐉣 𐉤 𐉥 𐉦 𐉧 𐉨 𐉩 𐉪 𐉫 𐉬 𐉭 𐉮 𐉯 𐉰 𐉱 𐉲 𐉳 𐉴 𐉵 𐉶 𐉷 𐉸 𐉹 𐉺 𐉻 𐉼 𐉽 𐉾 𐉿 𐊀 𐊁 𐊂 𐊃 𐊄 𐊅 𐊆 𐊇 𐊈 𐊉 𐊊 𐊋 𐊌 𐊍 𐊎 𐊏 𐊐 𐊑 𐊒 𐊓 𐊔 𐊕 𐊖 𐊗 𐊘 𐊙 𐊚 𐊛 𐊜 𐊝 𐊞 𐊟 𐊠 𐊡 𐊢 𐊣 𐊤 𐊥 𐊦 𐊧 𐊨 𐊩 𐊪 𐊫 𐊬 𐊭 𐊮 𐊯 𐊰 𐊱 𐊲 𐊳 𐊴 𐊵 𐊶 𐊷 𐊸 𐊹 𐊺 𐊻 𐊼 𐊽 𐊾 𐊿 𐋀 𐋁 𐋂 𐋃 𐋄 𐋅 𐋆 𐋇 𐋈 𐋉 𐋊 𐋋 𐋌 𐋍 𐋎 𐋏 𐋐 𐋑 𐋒 𐋓 𐋔 𐋕 𐋖 𐋗 𐋘 𐋙 𐋚 𐋛 𐋜 𐋝 𐋞 𐋟 𐋠 𐋡 𐋢 𐋣 𐋤 𐋥 𐋦 𐋧 𐋨 𐋩 𐋪 𐋫 𐋬 𐋭 𐋮 𐋯 𐋰 𐋱 𐋲 𐋳 𐋴 𐋵 𐋶 𐋷 𐋸 𐋹 𐋺 𐋻 𐋼 𐋽 𐋾 𐋿 𐌀 𐌁 𐌂 𐌃 𐌄 𐌅 𐌆 𐌇 𐌈 𐌉 𐌊 𐌋 𐌌 𐌍 𐌎 𐌏 𐌐 𐌑 𐌒 𐌓 𐌔 𐌕 𐌖 𐌗 𐌘 𐌙 𐌚 𐌛 𐌜 𐌝 𐌞 𐌟 𐌠 𐌡 𐌢 𐌣 𐌤 𐌥 𐌦 𐌧 𐌨 𐌩 𐌪 𐌫 𐌬 𐌭 𐌮 𐌯 𐌰 𐌱 𐌲 𐌳 𐌴 𐌵 𐌶 𐌷 𐌸 𐌹 𐌺 𐌻 𐌼 𐌽 𐌾 𐌿 𐍀 𐍁 𐍂 𐍃 𐍄 𐍅 𐍆 𐍇 𐍈 𐍉 𐍊 𐍋 𐍌 𐍍 𐍎 𐍏 𐍐 𐍑 𐍒 𐍓 𐍔 𐍕 𐍖 𐍗 𐍘 𐍙 𐍚 𐍛 𐍜 𐍝 𐍞 𐍟 𐍠 𐍡 𐍢 𐍣 𐍤 𐍥 𐍦 𐍧 𐍨 𐍩 𐍪 𐍫 𐍬 𐍭 𐍮 𐍯 𐍰 𐍱 𐍲 𐍳 𐍴 𐍵 𐍶 𐍷 𐍸 𐍹 𐍺 𐍻 𐍼 𐍽 𐍾 𐍿 𐎀 𐎁 𐎂 𐎃 𐎄 𐎅 𐎆 𐎇 𐎈 𐎉 𐎊 𐎋 𐎌 𐎍 𐎎 𐎏 𐎐 𐎑 𐎒 𐎓 𐎔 𐎕 𐎖 𐎗 𐎘 𐎙 𐎚 𐎛 𐎜 𐎝 𐎞 𐎟 𐎠 𐎡 𐎢 𐎣 𐎤 𐎥 𐎦 𐎧 𐎨 𐎩 𐎪 𐎫 𐎬 𐎭 𐎮 𐎯 𐎰 𐎱 𐎲 𐎳 𐎴 𐎵 𐎶 𐎷 𐎸 𐎹 𐎺 𐎻 𐎼 𐎽 𐎾 𐎿 𐏀 𐏁 𐏂 𐏃 𐏄 𐏅 𐏆 𐏇 𐏈 𐏉 𐏊 𐏋 𐏌 𐏍 𐏎 𐏏 𐏐 𐏑 𐏒 𐏓 𐏔 𐏕 𐏖 𐏗 𐏘 𐏙 𐏚 𐏛 𐏜 𐏝 𐏞 𐏟 𐏠 𐏡 𐏢 𐏣 𐏤 𐏥 𐏦 𐏧 𐏨 𐏩 𐏪 𐏫 𐏬 𐏭 𐏮 𐏯 𐏰 𐏱 𐏲 𐏳 𐏴 𐏵 𐏶 𐏷 𐏸 𐏹 𐏺 𐏻 𐏼 𐏽 𐏾 𐏿 𐐀 𐐁 𐐂 𐐃 𐐄 𐐅 𐐆 𐐇 𐐈 𐐉 𐐊 𐐋 𐐌 𐐍 𐐎 𐐏 𐐐 𐐑 𐐒 𐐓 𐐔 𐐕 𐐖 𐐗 𐐘 𐐙 𐐚 𐐛 𐐜 𐐝 𐐞 𐐟 𐐠 𐐡 𐐢 𐐣 𐐤 𐐥 𐐦 𐐧 𐐨 𐐩 𐐪 𐐫 𐐬 𐐭 𐐮 𐐯 𐐰 𐐱 𐐲 𐐳 𐐴 𐐵 𐐶 𐐷 𐐸 𐐹 𐐺 𐐻 𐐼 𐐽 𐐾 𐐿 𐑀 𐑁 𐑂 𐑃 𐑄 𐑅 𐑆 𐑇 𐑈 𐑉 𐑊 𐑋 𐑌 𐑍 𐑎 𐑏 𐑐 𐑑 𐑒 𐑓 𐑔 𐑕 𐑖 𐑗 𐑘 𐑙 𐑚 𐑛 𐑜 𐑝 𐑞 𐑟 𐑠 𐑡 𐑢 𐑣 𐑤 𐑥 𐑦 𐑧 𐑨 𐑩 𐑪 𐑫 𐑬 𐑭 𐑮 𐑯 𐑰 𐑱 𐑲 𐑳 𐑴 𐑵 𐑶 𐑷 𐑸 𐑹 𐑺 𐑻 𐑼 𐑽 𐑾 𐑿 𐒀 𐒁 𐒂 𐒃 𐒄 𐒅 𐒆 𐒇 𐒈 𐒉 𐒊 𐒋 𐒌 𐒍 𐒎 𐒏 𐒐 𐒑 𐒒 𐒓 𐒔 𐒕 𐒖 𐒗 𐒘 𐒙 𐒚 𐒛 𐒜 𐒝 𐒞 𐒟 𐒠 𐒡 𐒢 𐒣 𐒤 𐒥 𐒦 𐒧 𐒨 𐒩 𐒪 𐒫 𐒬 𐒭 𐒮 𐒯 𐒰 𐒱 𐒲 𐒳 𐒴 𐒵 𐒶 𐒷 𐒸 𐒹 𐒺 𐒻 𐒼 𐒽 𐒾 𐒿 𐓀 𐓁 𐓂 𐓃 𐓄 𐓅 𐓆 𐓇 𐓈 𐓉 𐓊 𐓋 𐓌 𐓍 𐓎 𐓏 𐓐 𐓑 𐓒 𐓓 𐓔 𐓕 𐓖 𐓗 𐓘 𐓙 𐓚 𐓛 𐓜 𐓝 𐓞 𐓟 𐓠 𐓡 𐓢 𐓣 𐓤 𐓥 𐓦 𐓧 𐓨 𐓩 𐓪 𐓫 𐓬 𐓭 𐓮 𐓯 𐓰 𐓱 𐓲 𐓳 𐓴 𐓵 𐓶 𐓷 𐓸 𐓹 𐓺 𐓻 𐓼 𐓽 𐓾 𐓿 𐔀 𐔁 𐔂 𐔃 𐔄 𐔅 𐔆 𐔇 𐔈 𐔉 𐔊 𐔋 𐔌 𐔍 𐔎 𐔏 𐔐 𐔑 𐔒 𐔓 𐔔 𐔕 𐔖 𐔗 𐔘 𐔙 𐔚 𐔛 𐔜 𐔝 𐔞 𐔟 𐔠 𐔡 𐔢 𐔣 𐔤 𐔥 𐔦 𐔧 𐔨 𐔩 𐔪 𐔫 𐔬 𐔭 𐔮 𐔯 𐔰 𐔱 𐔲 𐔳 𐔴 𐔵 𐔶 𐔷 𐔸 𐔹 𐔺 𐔻 𐔼 𐔽 𐔾 𐔿 𐕀 𐕁 𐕂 𐕃 𐕄 𐕅 𐕆 𐕇 𐕈 𐕉 𐕊 𐕋 𐕌 𐕍 𐕎 𐕏 𐕐 𐕑 𐕒 𐕓 𐕔 𐕕 𐕖 𐕗 𐕘 𐕙 𐕚 𐕛 𐕜 𐕝 𐕞 𐕟 𐕠 𐕡 𐕢 𐕣 𐕤 𐕥 𐕦 𐕧 𐕨 𐕩 𐕪 𐕫 𐕬 𐕭 𐕮 𐕯 𐕰 𐕱 𐕲 𐕳 𐕴 𐕵 𐕶 𐕷 𐕸 𐕹 𐕺 𐕻 𐕼 𐕽 𐕾 𐕿 𐖀 𐖁 𐖂 𐖃 𐖄 𐖅 𐖆 𐖇 𐖈 𐖉 𐖊 𐖋 𐖌 𐖍 𐖎 𐖏 𐖐 𐖑 𐖒 𐖓 𐖔 𐖕 𐖖 𐖗 𐖘 𐖙 𐖚 𐖛 𐖜 𐖝 𐖞 𐖟 𐖠 𐖡 𐖢 𐖣 𐖤 𐖥 𐖦 𐖧 𐖨 𐖩 𐖪 𐖫 𐖬 𐖭 𐖮 𐖯 𐖰 𐖱 𐖲 𐖳 𐖴 𐖵 𐖶 𐖷 𐖸 𐖹 𐖺 𐖻 𐖼 𐖽 𐖾 𐖿 𐗀 𐗁 𐗂 𐗃 𐗄 𐗅 𐗆 𐗇 𐗈 𐗉 𐗊 𐗋 𐗌 𐗍 𐗎 𐗏 𐗐 𐗑 𐗒 𐗓 𐗔 𐗕 𐗖 𐗗 𐗘 𐗙 𐗚 𐗛 𐗜 𐗝 𐗞 𐗟 𐗠 𐗡 𐗢 𐗣 𐗤 𐗥 𐗦 𐗧 𐗨 𐗩 𐗪 𐗫 𐗬 𐗭 𐗮 𐗯 𐗰 𐗱 𐗲 𐗳 𐗴 𐗵 𐗶 𐗷 𐗸 𐗹 𐗺 𐗻 𐗼 𐗽 𐗾 𐗿 𐘀 𐘁 𐘂 𐘃 𐘄 𐘅 𐘆 𐘇 𐘈 𐘉 𐘊 𐘋 𐘌 𐘍 𐘎 𐘏 𐘐 𐘑 𐘒 𐘓 𐘔 𐘕 𐘖 𐘗 𐘘 𐘙 𐘚 𐘛 𐘜 𐘝 𐘞 𐘟 𐘠 𐘡 𐘢 𐘣 𐘤 𐘥 𐘦 𐘧 𐘨 𐘩 𐘪 𐘫 𐘬 𐘭 𐘮 𐘯 𐘰 𐘱 𐘲 𐘳 𐘴 𐘵 𐘶 𐘷 𐘸 𐘹 𐘺 𐘻 𐘼 𐘽 𐘾 𐘿 𐙀 𐙁 𐙂 𐙃 𐙄 𐙅 𐙆 𐙇 𐙈 𐙉 𐙊 𐙋 𐙌 𐙍 𐙎 𐙏 𐙐 𐙑 𐙒 𐙓 𐙔 𐙕 𐙖 𐙗 𐙘 𐙙 𐙚 𐙛 𐙜 𐙝 𐙞 𐙟 𐙠 𐙡 𐙢 𐙣 𐙤 𐙥 𐙦 𐙧 𐙨 𐙩 𐙪 𐙫 𐙬 𐙭 𐙮 𐙯 𐙰 𐙱 𐙲 𐙳 𐙴 𐙵 𐙶 𐙷 𐙸 𐙹 𐙺 𐙻 𐙼 𐙽 𐙾 𐙿 𐚀 𐚁 𐚂 𐚃 𐚄 𐚅 𐚆 𐚇 𐚈 𐚉 𐚊 𐚋 𐚌 𐚍 𐚎 𐚏 𐚐 𐚑 𐚒 𐚓 𐚔 𐚕 𐚖 𐚗 𐚘 𐚙 𐚚 𐚛 𐚜 𐚝 𐚞 𐚟 𐚠 𐚡 𐚢 𐚣 𐚤 𐚥 𐚦 𐚧 𐚨 𐚩 𐚪 𐚫 𐚬 𐚭 𐚮 𐚯 𐚰 𐚱 𐚲 𐚳 𐚴 𐚵 𐚶 𐚷 𐚸 𐚹 𐚺 𐚻 𐚼 𐚽 𐚾 𐚿 𐛀 𐛁 𐛂 𐛃 𐛄 𐛅 𐛆 𐛇 𐛈 𐛉 𐛊 𐛋 𐛌 𐛍 𐛎 𐛏 𐛐 𐛑 𐛒 𐛓 𐛔 𐛕 𐛖 𐛗 𐛘 𐛙 𐛚 𐛛 𐛜 𐛝 𐛞 𐛟 𐛠 𐛡 𐛢 𐛣 𐛤 𐛥 𐛦 𐛧 𐛨 𐛩 𐛪 𐛫 𐛬 𐛭 𐛮 𐛯 𐛰 𐛱 𐛲 𐛳 𐛴 𐛵 𐛶 𐛷 𐛸 𐛹 𐛺 𐛻 𐛼 𐛽 𐛾 𐛿 𐜀 𐜁 𐜂 𐜃 𐜄 𐜅 𐜆 𐜇 𐜈 𐜉 𐜊 𐜋 𐜌 𐜍 𐜎 𐜏 𐜐 𐜑 𐜒 𐜓 𐜔 𐜕 𐜖 𐜗 𐜘 𐜙 𐜚 𐜛 𐜜 𐜝 𐜞 𐜟 𐜠 𐜡 𐜢 𐜣 𐜤 𐜥 𐜦 𐜧 𐜨 𐜩 𐜪 𐜫 𐜬 𐜭 𐜮 𐜯 𐜰 𐜱 𐜲 𐜳 𐜴 𐜵 𐜶 𐜷 𐜸 𐜹 𐜺 𐜻 𐜼 𐜽 𐜾 𐜿 𐝀 𐝁 𐝂 𐝃 𐝄 𐝅 𐝆 𐝇 𐝈 𐝉 𐝊 𐝋 𐝌 𐝍 𐝎 𐝏 𐝐 𐝑 𐝒 𐝓 𐝔 𐝕 𐝖 𐝗 𐝘 𐝙 𐝚 𐝛 𐝜 𐝝 𐝞 𐝟 𐝠 𐝡 𐝢 𐝣 𐝤 𐝥 𐝦 𐝧 𐝨 𐝩 𐝪 𐝫 𐝬 𐝭 𐝮 𐝯 𐝰 𐝱 𐝲 𐝳 𐝴 𐝵 𐝶 𐝷 𐝸 𐝹 𐝺 𐝻 𐝼 𐝽 𐝾 𐝿 𐞀 𐞁 𐞂 𐞃 𐞄 𐞅 𐞆 𐞇 𐞈 𐞉 𐞊 𐞋 𐞌 𐞍 𐞎 𐞏 𐞐 𐞑 𐞒 𐞓 𐞔 𐞕 𐞖 𐞗 𐞘 𐞙 𐞚 𐞛 𐞜 𐞝 𐞞 𐞟 𐞠 𐞡 𐞢 𐞣 𐞤 𐞥 𐞦 𐞧 𐞨 𐞩 𐞪 𐞫 𐞬 𐞭 𐞮 𐞯 𐞰 𐞱 𐞲 𐞳 𐞴 𐞵 𐞶 𐞷 𐞸 𐞹 𐞺 𐞻 𐞼 𐞽 𐞾 𐞿 𐟀 𐟁 𐟂 𐟃 𐟄 𐟅 𐟆 𐟇 𐟈 𐟉 𐟊 𐟋 𐟌 𐟍 𐟎 𐟏 𐟐 𐟑 𐟒 𐟓 𐟔 𐟕 𐟖 𐟗 𐟘 𐟙 𐟚 𐟛 𐟜 𐟝 𐟞 𐟟 𐟠 𐟡 𐟢 𐟣 𐟤 𐟥 𐟦 𐟧 𐟨 𐟩 𐟪 𐟫 𐟬 𐟭 𐟮 𐟯 𐟰 𐟱 𐟲 𐟳 𐟴 𐟵 𐟶 𐟷 𐟸 𐟹 𐟺 𐟻 𐟼 𐟽 𐟾 𐟿 𐠀 𐠁 𐠂 𐠃 𐠄 𐠅 𐠆 𐠇 𐠈 𐠉 𐠊 𐠋 𐠌 𐠍 𐠎 𐠏 𐠐 𐠑 𐠒 𐠓 𐠔 𐠕 𐠖 𐠗 𐠘 𐠙 𐠚 𐠛 𐠜 𐠝 𐠞 𐠟 𐠠 𐠡 𐠢 𐠣 𐠤 𐠥 𐠦 𐠧 𐠨 𐠩 𐠪 𐠫 𐠬 𐠭 𐠮 𐠯 𐠰 𐠱 𐠲 𐠳 𐠴 𐠵 𐠶 𐠷 𐠸 𐠹 𐠺 𐠻 𐠼 𐠽 𐠾 𐠿 𐡀 𐡁 𐡂 𐡃 𐡄 𐡅 𐡆 𐡇 𐡈 𐡉 𐡊 𐡋 𐡌 𐡍 𐡎 𐡏 𐡐 𐡑 𐡒 𐡓 𐡔 𐡕 𐡖 𐡗 𐡘 𐡙 𐡚 𐡛 𐡜 𐡝 𐡞 𐡟 𐡠 𐡡 𐡢 𐡣 𐡤 𐡥 𐡦 𐡧 𐡨 𐡩 𐡪 𐡫 𐡬 𐡭 𐡮 𐡯 𐡰 𐡱 𐡲 𐡳 𐡴 𐡵 𐡶 𐡷 𐡸 𐡹 𐡺 𐡻 𐡼 𐡽 𐡾 𐡿 𐢀 𐢁 𐢂 𐢃 𐢄 𐢅 𐢆 𐢇 𐢈 𐢉 𐢊 𐢋 𐢌 𐢍 𐢎 𐢏 𐢐 𐢑 𐢒 𐢓 𐢔 𐢕 𐢖 𐢗 𐢘 𐢙 𐢚 𐢛 𐢜 𐢝 𐢞 𐢟 𐢠 𐢡 𐢢 𐢣 𐢤 𐢥 𐢦 𐢧 𐢨 𐢩 𐢪 𐢫 𐢬 𐢭 𐢮 𐢯 𐢰 𐢱 𐢲 𐢳 𐢴 𐢵 𐢶 𐢷 𐢸 𐢹 𐢺 𐢻 𐢼 𐢽 𐢾 𐢿 𐣀 𐣁 𐣂 𐣃 𐣄 𐣅 𐣆 𐣇 𐣈 𐣉 𐣊 𐣋 𐣌 𐣍 𐣎 𐣏 𐣐 𐣑 𐣒 𐣓 𐣔 𐣕 𐣖 𐣗 𐣘 𐣙 𐣚 𐣛 𐣜 𐣝 𐣞 𐣟 𐣠 𐣡 𐣢 𐣣 𐣤 𐣥 𐣦 𐣧 𐣨 𐣩 𐣪 𐣫 𐣬 𐣭 𐣮 𐣯 𐣰 𐣱 𐣲 𐣳 𐣴 𐣵 𐣶 𐣷 𐣸 𐣹 𐣺 𐣻 𐣼 𐣽 𐣾 𐣿 𐤀 𐤁 𐤂 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒 𐤓 𐤔 𐤕 𐤖 𐤗 𐤘 𐤙 𐤚 𐤛 𐤜 𐤝 𐤞 𐤟 𐤠 𐤡 𐤢 𐤣 𐤤 𐤥 𐤦 𐤧 𐤨 𐤩 𐤪 𐤫 𐤬 𐤭 𐤮 𐤯 𐤰 𐤱 𐤲 𐤳 𐤴 𐤵 𐤶 𐤷 𐤸 𐤹 𐤺 𐤻 𐤼 𐤽 𐤾 𐤿 𐥀 𐥁 𐥂 𐥃 𐥄 𐥅 𐥆 𐥇 𐥈 𐥉 𐥊 𐥋 𐥌 𐥍 𐥎 𐥏 𐥐 𐥑 𐥒 𐥓 𐥔 𐥕 𐥖 𐥗 𐥘 𐥙 𐥚 𐥛 𐥜 𐥝 𐥞 𐥟 𐥠 𐥡 𐥢 𐥣 𐥤 𐥥 𐥦 𐥧 𐥨 𐥩 𐥪 𐥫 𐥬 𐥭 𐥮 𐥯 𐥰 𐥱 𐥲 𐥳 𐥴 𐥵 𐥶 𐥷 𐥸 𐥹 𐥺 𐥻 𐥼 𐥽 𐥾 𐥿 𐦀 𐦁 𐦂 𐦃 𐦄 𐦅 𐦆 𐦇 𐦈 𐦉 𐦊 𐦋 𐦌 𐦍 𐦎 𐦏 𐦐 𐦑 𐦒 𐦓 𐦔 𐦕 𐦖 𐦗 𐦘 𐦙 𐦚 𐦛 𐦜 𐦝 𐦞 𐦟 𐦠 𐦡 𐦢 𐦣 𐦤 |

研究对象进行抽象的方法^[25]。特征结构用来刻画句子的句法结构和语义特征,形式上可以将各种语义特征用三元组描述,即[实体,特征,特征值],例如:

[1] 𠄎 𠄎 𠄎 (红车子)可表示为:[𠄎 𠄎, 𠄎 𠄎, 𠄎]

[2] 𠄎 𠄎 𠄎 𠄎 𠄎 (她说我们上课)可表示为:

[𠄎, , 𠄎]; [𠄎, , 𠄎 𠄎 𠄎 𠄎]; [𠄎 𠄎 𠄎, 𠄎 𠄎]

从例[1]可以看出,实体是“𠄎 𠄎(车子)”,特征值是“𠄎(红)”,特征“𠄎 𠄎(颜色)”并未出现;在例[2]中,“𠄎(她)”为“𠄎(说)”的特征值;“𠄎 𠄎 𠄎 𠄎(我们上课)”是“𠄎(说)”的又一个特征值,且“𠄎 𠄎 𠄎 𠄎(我们上课)”本身为一个特征结构,“𠄎 𠄎 𠄎 𠄎(我们)”是实体“𠄎 𠄎 𠄎 𠄎(上课)”的特征值。

分析所得特征的三元组表征方式具有以下特点:1)便于计算机处理和表示;2)一对语义关联对应一个特征三元组;3)不存在游离于关系之外的成分,有语义关系的词都要标注;4)对词与词之间语义知识的描述直观、自然、完整;5)三元组中,特征和特征值都可为实体;特征词可以不出现,需根据具体语言问题设计特征词表;特征值本身可以成为一个特征结构^[25]。

2. 特征结构图表示法

在计算语言学中,广泛使用的一种形式模型是图,图的主要元素是节点和边。一个特征三元组可看作两个“点”和连接它们的“边”,其中的“节点”表示实体或特征值,“边”表示特征。特征一定是某个节点的特征,这个节点就作为特征拥有者,另一个节点就作为特征值^[26]。于是一个特征三元组集合可表示为一个特征结构图,如图 1 所示。

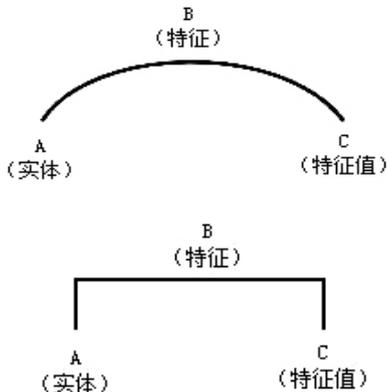


图 1 特征结构图

图 1 两个图同形,表现方式不同,反映了实体 A 和特征值 C 之间的语义关系,实线表示实体 A 的特征 B 的值是 C。例如:

[3] 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 (我有一些不同想法。)

它的特征结构如图 2 所示。

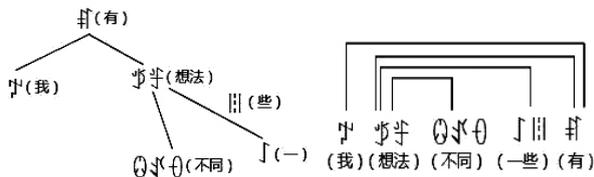


图 2 例[14]的特征结构图

特征结构图具有以下特点:1)特征结构图是一种不对称的无向图,动态描述语言成分的组合过程;2)各种语义关系之间可以出现交叉、循环、多边和回路现象;3)每个节点都为句子或短语中的词;4)图示是一种抽象概念的形式模型表示,是构成树库的基本元素;5)语义信息丰富,为计算机语义分析解决语义知识表达问题。

三、彝语主谓谓语句的特征结构标注实例分析

(一) 彝语主谓谓语句特征结构标注举例

如上所述,研究确定了彝语主谓谓语句的十一种语义关系类型。下文主要依据特征结构理论对各种类型的主谓谓语句语料进行语义标注,然后从标注结果中选取 11 个句子为考察对象,细致描述其特征结构三元组和特征结构图。分析结果如下:

(1) 第一种类型的特征结构标注分析

例[1]: 𠄎 𠄎 𠄎 𠄎 𠄎 (这个字你认识。)

句中大主语“𠄎 𠄎 𠄎 𠄎(这个字)”是受事,小主语“𠄎(你)”是施事,“𠄎(认识)”为“𠄎(你)”发出的动作。该句的原型句为“𠄎 𠄎 𠄎 𠄎 𠄎(你认识这个字)”。原型句中的宾语变成了主谓谓语句中的大主语,主语和谓语变成了主谓谓语句中的小主语和小谓语,其特征结构三元组和特征结构图如表 2 所示。

表 2 例[1]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|-------------|-------|
| [𠄎 𠄎, 𠄎, 𠄎] | |
| [𠄎, , 𠄎] | |
| [𠄎, , 𠄎 𠄎] | |

(2) 第二种类型的特征结构标注分析

例[2]: 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 𠄎 (我彝语汉语都懂。)

句中大主语“𠄎(我)”是施事,小主语“𠄎 𠄎 𠄎 𠄎(彝语汉语)”是受事,“𠄎(懂)”为“𠄎

(我)”发出的动作,其特征结构三元组和特征结构图如表 3 所示。

表 3 例[2]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|--|-------|
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |

(3)第三种类型的特征结构标注分析

例[3]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(我的父亲身体很好。)

句中大谓语“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (身体很好)”陈述说明大主语“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (父亲)”。“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (身体)”和“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (父亲)”有关,是“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (父亲)”的一部分;“ ㄉㄨㄛ^1 (好)”和“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (身体)”有关联,和“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (父亲)”无直接联系。如表 4 所示。

表 4 例[3]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |

(4)第四种类型的特征结构标注分析

例[4]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(小王我告诉他了。)

大主语“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (小王)”是话题,大谓语是一个主宾谓句,其宾语“ ㄉㄨㄛ^1 (他)”代指大主语,其特征结构三元组和特征结构图如表 5 所示。

表 5 例[4]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |

(5)第五种类型的特征结构标注分析

例[5]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(这件事情我有些不同想法。)

大谓语是个完整的主宾谓句,大主语“ ㄉㄨㄛ^1 (事情)”前面可以加上介词“对”“对于”等,且大主语和小宾语有语义关系,其特征结构三元组和特征结

构图如表 6 所示。

表 6 例[5]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1] | |

(6)第六种类型的特征结构标注分析

例[6]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(今天天气很好。)

句中大主语是表时间的名词或短语,大谓语是对大主语的陈述说明,其特征结构三元组和特征结构图如表 7 所示。

表 7 例[6]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [ㄉㄨㄛ^1 , $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |

(7)第七种类型的特征结构标注分析

例[7]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(这个城市我没有来过。)

句中大主语是表处所的名词或名词短语,大谓语是陈述说明大主语,其特征结构三元组和特征结构图如表 8 所示。

表 8 例[7]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$, ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |
| [ㄉㄨㄛ^1 , ㄉㄨㄛ^1 , ㄉㄨㄛ^1] | |

(8)第八种类型的特征结构标注分析

例[8]: $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ 。(这块布我用来做衣服了。)

句中大主语“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (布)”是小谓语“ ㄉㄨㄛ^1 (做)”的材料,大谓语是一个完整的主宾谓句。大主语“ $\text{ㄉㄨㄛ}^1 \text{ㄉㄨㄛ}^1$ (布)”和小谓语“ ㄉㄨㄛ^1 (做)”用介词“ ㄉㄨㄛ^1 (用)”来联系,其特征结构三元组和特征结构图如表 9 所示。

表 9 例[8]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|------------------------------------|-------|
| [$\text{ㄩ}, \text{ㄩ}, \text{ㄩ}$] | |

(9) 第九种类型的特征结构标注分析

例 [9]: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(这人说话粗鲁。)

该类型中小主语由动词或动宾短语充当。该句大主语“ ㄩ (人)”和小主语“ $\text{ㄩ} \text{ㄩ}$ (说话)”构成主谓关系,其特征结构三元组和特征结构图如表 10 所示。

表 10 例[9]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|------------------------------------|-------|
| [$\text{ㄩ}, \text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ} \text{ㄩ}, \text{ㄩ}$] | |

(10) 第十种类型的特征结构标注分析

例 [10]: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(一件礼物他都不买。)

该句的原型句为“ $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ ”,原型句中的宾语变成了主谓谓语句的大主语,谓语变成了主谓谓语句的小谓语,其特征结构三元组和特征结构图如表 11 所示。

(11) 第十一种类型的特征结构标注分析

表 11 例[10]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄩ} \text{ㄩ}, \text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ} \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ}$] | |

例 [11]: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(土豆一斤三块钱。)

该句语义较复杂,“ ㄩ (斤)”是实体“ $\text{ㄩ} \text{ㄩ}$ (土豆)”的特征,值为“ ㄩ (一)”;“ $\text{ㄩ} \text{ㄩ}$ (三块)”是土豆的价格,“ $\text{ㄩ} \text{ㄩ}$ (价格)”被隐藏了;“ $\text{ㄩ} \text{ㄩ}$ (三块)”也与“ $\text{ㄩ} \text{ㄩ}$ (一斤)”存在语义联系,其特征结构三元组和特征结构图如表 12 所示。

表 12 例[11]的两种标注结果

| 特征结构三元组 | 特征结构图 |
|---|-------|
| [$\text{ㄩ}, \text{ㄩ}, \text{ㄩ}$] | |
| [$\text{ㄩ} \text{ㄩ} (\text{ㄩ} \text{ㄩ}), (\text{ㄩ} \text{ㄩ}), \text{ㄩ} \text{ㄩ}$] | |
| [$\text{ㄩ}, \text{ㄩ}, \text{ㄩ}$] | |

(二) 彝语主谓谓语句的特征结构模型

进行语言信息处理研究,其中一项至关重要的基础工作就是建立自然语言的“语言模型”,以便于计算机能对其进行相应的分析与计算^[27]。根据对彝语主谓谓语句的特征结构分析结果可以发现,在十一种彝语主谓谓语句中,大主语、小主语和小谓语之间的语义表征存在共性,因此,一些类型的彝语主谓谓语句具有相同的语义关系表现形式,论文研究经归纳概括,整理出了彝语主谓谓语句的四类语义模型,如表 13 所示。

表 13 彝语主谓谓语句特征结构模型

| 类型 | 特征结构图 | 语义分析 |
|---------------------------|-------|---|
| 大谓语陈述说明大主语: S1+S2+V/A | | S1 是实体, S2 是 S1 的特征, V/A 是 S1 的特征值。 如: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(这个姑娘唱歌好听。) |
| 宾语前置句首或位于主语和宾语之间: S1+S2+V | | S1 是 V 的施事或受事, 其可改写为: S2+S1+V。 如: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(一个人我都不认识。) |
| 复指: S1+S2+S3+V | | 大谓语是完整的主宾谓句, 宾语 S3 复指大主语 S1, V 和 S3、S1 都有语义关系。 如: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(这个孩子同学都心疼他。) |
| 大主语前面可以加介词: S1+S2+S3+V | | 大谓语是一个完整的主宾谓句, 大主语 S1 与宾语 S3 有语义关系。 如: $\text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ} \text{ㄩ}$ 。(这个问题我没有意见。) |

四、结论与展望

彝语信息处理需要有什么样的语义知识?应借助哪种理论实现彝语语言知识资源的形式化表达?这是彝语及其信息化、智能化研究当前面临的重要问题,本文基于特征结构理论,尝试将特征结构三元组和特征结构图的语义标注方法应用到彝语主谓谓语句中,并根据标注结果建立彝语主谓谓语句的语义表示和分析模型。实践结果表明,该

理论可以比较完整地将彝语句子中所有具有语义关系的词联结起来,其直观的句子形式模型描述方式蕴含丰富的语义信息,是存储知识的基础,是语言分析和理解从语义到形式的基点,能够为研究有效的彝语语义标注策略和彝语的形式化表示提供一个可参考的解决方案,同时为彝语语言资源知识库的建设和应用研究提供一个理论框架和可行性思路,对推动计算机彝文信息处理向智能化方向发展奠定了良好的资源基础。

参考文献:

- [1] 姬东鸿.语义分析若干前沿问题[J].长江学术,2020(2).
- [2] 王成平,沙马拉毅.彝语语言文字信息处理技术的现况分析与发展前景[J].西南民族大学学报(人文社会科学版),2011(2).
- [3] 冯志伟.自然语言处理综论[M].北京:电子工业出版社,2005.
- [4] 朱德熙.语法答问[M].北京:商务印书馆,1985.
- [5] 陆俭明.“施-受动”主谓谓语句“功能-认知”研究[J].中国语文,2020(4).
- [6] 朱德熙.语法讲义[M].北京:商务印书馆,2003.
- [7] 黄伯荣,廖序东.现代汉语[M].北京:高等教育出版社,2017.
- [8] 袁毓林.话题化及相关的语法化过程[J].中国语文,1996(4).
- [9] 张旺熹.主谓谓语句的语义模式[J].世界汉语教学,1993(3).
- [10] 吴中伟.主谓谓语句 NP(VP AP)语义结构分析[J].语言研究,1996(1).
- [11] 缪建明,张全.现代汉语主谓谓语句的 HNC 处理研究[J].计算机科学,2006(12).
- [12] 姚从军,罗丹.面向信息处理的汉语主谓谓语句的组合范畴语法分析[J].中国社会科学院研究生院学报,2019(2).
- [13] 陈禹.主谓谓语句的“基线/加工”模式[J].现代外语,2019(4).
- [14] 张余蓉,蔡莹.凉山彝语语法[M].北京:民族出版社,1995.
- [15] 王秀英.现代彝语概论[M].成都:四川民族出版社,2010.
- [16] 丁椿寿,朱文旭.现代彝语[M].北京:中央民族学院出版社,1991.
- [17] ANDERSON S.R. Phonology in the Twentieth Century[M].Cambridge:Cambridge University Press,1995.
- [18] 石锋.从塞音的清浊看区别性特征的表现[J].南开学报,1988(1).
- [19] CHOMSKY N, HALLE M.Sound patterns of English[M]. New York: Harper and Row, 1968.
- [20] CHOMSKY N. Aspects of the Theory of Syntax[M].Cambridge:MIT Press,1965.
- [21] ANDREAS W, GEORG R, ERHARD H, et all. SusTEInability of linguistic resources through feature structures[J]. Literary and Linguistic Computing, 2009,24(3).
- [22] 陆俭明.语义特征分析在汉语语法研究中的运用[J].汉语学习,1991(1).
- [23] 陆俭明.语言研究要与未来接轨[J].语言战略研究,2021(6).
- [24] 陈波.特征结构及其汉语语义资源建设[M].武汉:武汉大学出版社,2014:36-42.
- [25] 刘海涛.依存语法的理论与实践[M].北京:科学出版社,2009.
- [26] 陈波,姬东鸿,吕晨.基于特征结构的汉语主谓谓语句语义标注研究[J].中文信息学报 2012(3).
- [27] KLEIN D, MANNING C. Corpus-based Induction of Syntactic Structure: Models of Dependency and Constituency[C]. In Proceedings of ACL,2004.