

基于情节本体的叙事性文本语义结构化表示方法研究*

宋宁远 王晓光

摘要 叙事性文本是对现实世界的记录,也是人文学者的研究对象与基础,对叙事性文本的阐释更是人文学术研究的基本任务。随着数字人文的发展,人文学者对远距离阅读与策略型阅读的需求与日俱增。作为构成叙事性文本的基本单元,事件和情节是文本解读的重点。为了更好地适应人文学者阅读需求的转变,建构叙事性文本智慧数据,支撑数字人文研究,以情节为视角,以事件为基础,对广泛存在的叙事性文本进行深层次语义表示与结构化处理显得尤为重要。但现阶段较为成熟的文本表示方法更适用于结构相对清晰、统一的科学文本,不能很好地表示叙事性文本中存在的复杂叙事结构,因此,以知识表示与知识组织为出发点,构建符合叙事性文本特征的语义结构化表示方法具有十分重要的意义。本文在梳理叙事学基本理论与概念,以及事件知识表示、事件本体与叙事本体等理论研究的基础上,对叙事文本中存在的事件、事件间关系进行语义建模与表示,同时完善了对情节的定义,构建了基于事件的情节本体 EBPO,提出了叙事性文本的“层次—网络”结构模型,用以实现对叙事性文本的语义结构化表示。初步标注的实验结果表明,借助情节本体及结构模型,可以实现对叙事性文本的语义结构化处理。图 8。表 4。参考文献 61。

关键词 叙事性文本 情节本体 语义结构 知识表示 数字人文

分类号 G254

Semantic Structured Representation Method of Narrative Text Based on Plot Ontology

SONG Ningyuan & WANG Xiaoguang

ABSTRACT

Narration is the record, deduction and re-creation of the development of the human world. Narrative text is a kind of rich semantic text, which uses symbolic systems such as words and images to represent real or fictional events and plots. Because of the different purposes, styles and genres of the text, it has various kinds and rich contents. The interpretation of narrative texts has always been the core and basic work of humanists. With the rise of digital humanities, a new research paradigm has emerged in the field of Humanities and Social Sciences; with more emphasis on data-driven research methods, the way of understanding narrative texts has changed, and new reading methods and research needs of narrative texts,

* 本文系教育部人文社科重点研究基地重大项目“大数据资源语义表示与组织”(编号:16JJD870002)的研究成果之一。(This article is an outcome of the major project “Semantic Representation and Organization of Big Data Resources”(No.16JJD870002) supported by Foundation of Key Research Institute of Humanities and Social Science at Universities.)

通信作者:王晓光,Email:whu_wxg@126.com,ORCID:0000-0003-1284-7164(Correspondence should be addressed to WANG Xiaoguang,Email:whu_wxg@126.com,ORCID:0000-0003-1284-7164)

including strategic reading and distant reading, have emerged. In order to fully meet this demand, structuring, semanticizing and standardizing narrative text content and constructing narrative text data infrastructure adapted to the digital humanities research paradigm have gradually become the key to future work.

As a special type of text, the complexity of narrative text lies in its emphasis on events as the basic component and the core structure. Events and plots all have abundant correlations. Therefore, narrative texts are dynamic and complex. This brings challenges to the full expression of narrative text features, the description of narrative structure, and the realization of semantic structured representation for narrative text.

In order to solve the problem, this study started from the basic theory of narratology, summarizes and collates the related statements about events, plots and narratives. Firstly, the characteristics of narrative texts were summarized, and an event knowledge representation scheme for narrative texts was proposed. Followed by the definition of plot, the ontology EBPO (Event Based Plot Ontology, EBPO) was designed and developed. The ontology defines 10 classes and 29 properties, which can express plots and events more clearly. Based on this ontology, a general process of deep semantic annotation oriented to the characteristics of narrative texts was proposed, and a preliminary deep semantic annotation experiment was carried out. In the semantic annotation experiment, five classical novel fragments were selected, such as *Pride and Prejudice*, *Moon and Sixpence*, *Living*, and so on. Natural language processing tool GATE 8.5.1 was used for annotating. With the results of semantic annotation, this paper proposed a hierarchical-network model for narrative texts. Taking *Pride and Prejudice*, *Moon and Sixpence* as examples, the annotation results were reorganized and correlated, and the semantic structured representation of narrative texts was realized.

The results show that the semantic structured representation of narrative text based on plot ontology can achieve more accurate definition on the events and plots. Bases on EBPO ontology and annotation ontology, the large-scale, high-quality, rich semantic narrative text corpus could be built, so the machine-aided event retrieval and reading methods including plot review, character relationship overview, plot development path can be realized. Future research will focus on the definition of plot, narrative mode, narrative style and other issues in order to achieve more accurate content representation of narrative text. 8 figs. 4 tabs. 61 refs.

KEY WORDS

Narrative texts. Plot ontology. Semantic structure. Knowledge representation. Digital humanities.

0 引言

叙事是用语言、图像等符号系统表现一件或一系列真实或虚构的事件与情节。叙事性文本是叙事的文字表示,同时也是人类对世界发展的记录、演绎及再创作。叙事性文本表现样式丰富,风格多样,种类繁多,包括小说、民间故事、回忆录等。对这些文本进行阐释历来是人文学者的基础工作之一。随着信息技术的不断

发展,数字人文逐渐成为人文社会科学领域的发展方向与趋势。在数字人文环境下,人文学者对叙事文本的阅读行为与使用策略发生了变化,产生了远距离阅读(distant reading)与策略型阅读(strategic reading)的需求^[1];除了对特定文本的理解之外,人文学者也开始关注诸如人物关系、情节发展线索等问题^[2];相关研究范式也逐渐由关注语篇及文献(document centric)向关注文献内的实体(entity centric)转变^[3];情节作为构成叙事性文本的基础,已经成为叙事性

文本研究的热点。因此,为了适应数字人文新的研究范式,实现叙事性文本的深度语义标注,建设高质量、大规模、机器可理解的叙事性文本语料库,促使叙事文本向智慧数据的转变^[4],进而满足事件抽取、情节检索、机器辅助的文本理解和宏观视角下的文本远距离阅读等数字人文研究新需求,有必要在建设情节本体的基础上,对叙事文本进行语义结构化表示。

现阶段,文本语义结构化表示在科学语篇情境下取得了较为丰富的成果,诞生了大量适用于不同任务的标签集^[5]、本体^[6]及语义标注方法^[7]。相较于结构清晰、观点明确、体例统一的科学文本,叙事性文本具有内容表达的动态性、叙事结构的复杂性、叙事模式的多样性等特点,因而缺乏与之相适应的语义结构化表示方法,这就给叙事性文本的机器理解带来了挑战,制约了叙事文本内容数据的处理与开发,无法更好地支撑数字人文研究。因此,深入理解叙事性文本内容结构,提出适用于叙事性文本的语义结构化表示方案就显得十分重要。

叙事性文本的复杂性在于强调以事件为基本组成,以由事件构成的情节为核心结构,事件、情节间均具有较为丰富的语义关联关系。因此,为了充分表示叙事性文本特征,描述叙事结构,首先需要对事件与事件之间的关系进行语义建模,在对情节进行定义和形式化表征与建模的基础上,设计面向叙事性文本的语义结构化表示方案,从而为事件检索、远距离阅读等提供数据基础,满足数字人文研究范式下的新型需求。

1 相关研究

1.1 叙事理论与叙事性文本

(1) 叙事基本理论

叙事理论是兴起于西方的一种文学理论,因对文学作品强大的阐释性而逐渐成为文学研究的重点^[8]。在《民间故事的形态学》^[9]中,普罗普对民间故事的叙事结构特征进行了归纳。

罗兰·巴特提出了叙事结构的三个层次,包括功能、动作与叙事^[10]。托多罗夫^[11]将叙事作品分解为故事、序列、主题句与词类。

除了早期对叙事结构的探讨,不同学者也针对叙事的组成要素提出了不同观点。艾布拉姆斯认为叙事包含了事件、人物,以及人物语言和行为^[12]。罗吉·福勒^[13]指出叙事是指详细叙述一系列事实、事件以及事件间关系的过程。日奈特^[14]认为叙事着重表现行动和事件,强调时间性和戏剧性。由此可以看出,事件是构成叙事的基础。叙事不仅包括事件、任务、背景等要素,而且包括“它们之间的各种关系”。例如,E.M.福斯特^{[15]1-21}提出了时间序列在构成叙事中的重要作用,认为叙事是对一系列按时间顺序排列的事件的叙述。另一种观点^[16]认为叙事是对事件按照逻辑序列的组织,故事的发展遵循一定的逻辑秩序。

(2) 情节概念与基本结构

以上研究均以事件为视角探讨了叙事的结构。除此之外,另有一些研究引入了“情节”的概念,指出情节是由事件序列组成,对叙事的构成具有重要作用。这些研究重点区分了情节、事件与叙事的关系与区别。胡亚敏^[17]指出情节是时间的形式系列或语义序列,是故事结构中的主干,也是任务、环境的支撑点。杰拉德·普林斯^{[18]63-71}把事件区分为状态性事件和行动性事件,并在此基础上给出了核心叙事及情节的定义。福斯特^{[15]72-74}指出了情节与故事的区别,认为情节是小说的一种较高层次的结构。吴琪^[19]、方珊^[20]将情节理解为处在“因果—事件”关系中的众多细节之和。谭光辉^[21]认为情节的基础是事件,情节是对事件用叙述话语进行处理的结果,是事件的时间化和因果化。赵毅恒^[22]认为情节是叙述文本中序列性组合的事件。

总体来看,学界对于情节的定义缺乏较为统一的标准,主流意见认为情节由事件序列组成,对情节的定义与揭示可以从理解事件与事件间关系的角度入手。这些观点为本文构建面

向叙事性文本的情节本体提供了参考。

(3) 叙事性文本特征

叙事性文本是叙事的语言及文字表达,在叙事作品中占有重要地位。叙事学理论中对叙事作品的分析存在明显的“故事”(story)和“话语”(discourse)的二元对立^[23],里蒙·凯南^[24]认为“叙事虚构作品”包含三个要素,“一是用语言性质的媒介将叙事信息由叙事者传递给接受者;二是描述了一系列事件;三是作品是虚构性质的”。除此之外,也有学者认为叙事作品由故事及叙事话语组成,并指出叙事话语是对叙事的符号表现^[25]。

综合以上观点,根据叙事的基本组成,可以归纳出叙事性文本的语义结构特点:①构成叙事性文本的基本语义单位为事件,根据组成元素的不同,事件又具有不同的类型;②不同事件之间通过时间及语义上的关系进行关联;③事件及事件间的关系构成了情节,进而由不同情节组织成叙事性文本,情节体现了叙事文本的叙事性特征。因此,为了实现对叙事性文本的语义结构化表示,需要对事件、事件间关系、情节等概念之间的关系进行准确定义与表示。

1.2 事件与事件知识表示

情节由事件组成,对于情节的描述需要深入理解事件的意义与组成,并在事件知识表示的基础上,完成对情节的规范化表征。对于事件的理解,哲学^[26]、语言学^[27,28]、认知科学^[29]等领域均给出了相应解释。在计算机及信息科学领域,主流研究将事件作为知识的基本形式与知识表示的基本框架,借此关联与事件发展过程相关的要素,更加突显了事件的动态性^[30]。

近年来,事件知识表示的研究重点关注事件的结构、要素及关系等。针对事件定义及要素组成,Nelson^[31]认为事件由行为人、行动等要素组成。仲兆满等人^[32]认为事件由动作、对象、时间、环境、断言及语言表现等六元素构成。schema.org^[33]也对事件尤其是文学事件(literary event)进行了规范化定义,包括日期(date)、起

始时间(start date)、参与者(attendees)、地点(location)等属性。在事件间关系研究方面,主流研究将事件间关系分为时间关系^[34]与因果关系^[35]两大类。事件间因果关系常用于事件推理^[36],有研究依据新闻事件^[37]、生物医学^[38]中的事件间因果关系构建了相应的事件序列。

在此基础上,学者多通过本体的形式,围绕特定任务实现对事件知识的形式化表征^[39]。其中,面向通用领域的事件本体以Event Ontology^[40]及Simple Event Model(SEM)^[41]为代表,对事件进行了一般描述与表征。LODE(Linking Description of Event)本体^[42]以实现事件关联发布为目的,定义了包括事件主体(agent)、客体(object)、时间、地点等要素。Corda等^[43]针对历史事件及历史文本,提出了历史事件本体(History event ontology),对事件推理的方法、规则等进行了约束。除此之外,还有面向学术领域^[44]、语义分析^[45]、人物建模^[46]等特定使用场景的事件本体,均对事件的一般组成要素进行了定义^[47]。

综合现有研究可以发现,组成事件的要素一般包括动作、时空环境、角色等。其中,动作是事件的核心,也是区分不同类型事件的主要依据。不同事件之间存在时间与语义关系,是构成事件网络的基础。

1.3 叙事性文本语义建模方法

现阶段叙事性文本语义建模多在事件知识表示的基础上,通过表征事件间语义关系,构建事件序列进而实现对叙事性文本的表示。故事本体(Stories ontology)^[48]定义了故事(story)、事件槽(event slot)、槽(slot)、子故事(substory)等,同时定义了事件序列(eventList),借此关联不同事件,但未能进一步定义事件间存在的具体语义关系。Damiano^[49]提出了一个叙事与动作本体(narrative & action),核心由动态要素类(dynamics)(动作、过程和状态)和实体类(entity)(角色、对象)组成,并定义了事件的情境信息。该本体描述了叙事对象的静态特征,

并给出了事件的状态属性,但同样没有解决叙事对象间的关联语义,无法更好地表征事件序列与情节特点。Nakasone 等^[50]采用修辞结构理论定义事件间存在的语义关系,实现对叙事性文本的语义建模,但却忽略了叙事性文本中天然存在的时序关系。BBC 的 Storyline Ontology^[51]侧重于新闻事件的组织,通过跟随关系 (follows) 对时序关系进行表征,但该模型对于叙事的表达能力较弱,不能较好地解释并表征叙事性文本的结构^[52]。

分析现有研究可以发现,叙事本体未能充分考虑叙事学中对于情节的解释,无法准确定义并描述情节,对于叙事性文本中普遍存在的语义关系考虑不足,因而不能较好地表征叙事逻辑。

2 面向叙事性文本的事件知识表示

根据叙事学基础理论,本研究认为叙事性文本是由一系列事件、情节及事件间关系组成,为了构建叙事性文本语义结构化表示方法,首先需要事件及事件间关系等进行准确定义。

2.1 事件的构成要素

与新闻事件、交通事件等真实世界内发生的即时性事件不同,叙事性文本中的事件更强调叙事性、戏剧性。为了突出叙事性文本的特征,本研究以叙事学相关研究为基础,以情节 (plot) 为核心结构,以事件为基本单元进行建模。因此本研究首先对叙事性文本、情节及事件进行表示。

作为组成叙事性文本的最基本单元,事件由动作 (action)、角色 (role)、情境 (context) 三大要素构成,表示特定时空环境下发生的由若干角色参与的特定动作。其中,动作是事件的核心元素,表示事件发展的变化过程,具备不同的动作类型 (action type)。角色表示事件的参与对象,通常包括动作发生的主体 (subject) 及作用客体 (object),分别表示动作的发起者及承受者,可以为人物、机构、团体等。情境主要描述事件发生的时空信息,包括时间 (time) 和空间 (space),时间为事件发生时的时间信息,可以为绝对时间或相对时间,包括起始时间、终止时间等属性;空间为事件发生时的位置信息,在叙事性文本中体现为场景、场所等。图 1 表示情节、事件及事件构成要素的结构关系。

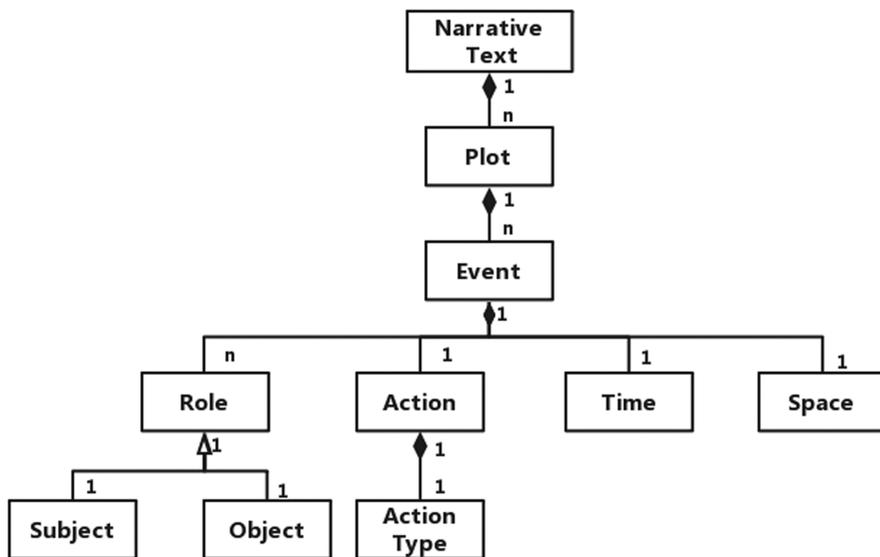


图 1 情节、事件及事件构成要素的结构关系

2.2 事件的类型

在叙事性文本中,事件因构成要素的不同而具有不同的状态和不同的类型。对事件类型的定义与揭示,一方面可以用来表示事件的不同状态,另一方面也可用来揭示事件的不同语义功能。语言学与叙事学的相关研究对事件类型进行了划分,如高彦梅^[53]⁸⁰⁻⁸⁶将语篇内的事件分为物质事件、心理事件、关系事件、言语事件、行为事件与存在事件六大类;普林斯^[18]⁶³⁻⁷¹将事件定义为状态型和行动型两大类,并从语言结构上给出区分两类事件的依据。

本文以此为基础,参照 Kaneiwa^[54]对事件种类的划分,将事件按照状态的不同划分为动作型事件(active event)和状态型事件(static event)两大类。同时,笔者认为两类事件的差异具体体现在构成要素的不同。其中,动作型事件通常用以对动作过程进行刻画,动作要素为必须要素,通常包括动作发起的主体及客体。状态型事件通常用以表示动作发生的前因、后果及相关情境,侧重于对时空环境及动作客体的描写,较少涉及对动作要素的叙述。因此,行动型事件及状态型事件可以进行如下表示:

Active Event <Role(Subject Object), Action, Time, Space>;

Static Event <Role(Object), Time, Space>

2.3 事件间的关系

事件间的关系较为复杂,Zacks^[26]、托马舍夫斯基^[55]等人认为事件间的关系通常包括时间关系与因果关系两大类。高彦梅^[53]⁹¹⁻¹¹¹在总结韩礼德、朗埃克等人研究成果的基础上,从语言学的角度将事件间关系分为时间关系、空间关系、联合关系、选择关系、起因—条件关系、解释关系、比较关系和投射关系。笔者认为,在叙事性文本中,事件间最主要的关系包括时序关系与语义关系两大类。其中,时序关系用以表示事件在时间维度上的先后次序,语义关系用以表示关联事件的逻辑。这两类关系共同作用于事件关联,是构成叙事情节的核心。

(1) 时序关系

时序关系是事件间的基本关系,决定事件发生的先后顺序及事件序列的发展方向。事件间的时序关系由时间要素决定,参照 W3C 的时间本体(time ontology)^[56]及 Allen^[57]关于事件时序逻辑结构的定义,笔者认为事件之间的时序关系包括 before、during 等 7 种,具体解释如表 1 所示。

表 1 事件间的时序关系

关系	定义
Before	表示事件发生的先后顺序,该事件的发生时间在另一事件之前。
During	表示事件发生的时间段在另一事件的发生时间段内。
Equals	表示两个事件的发生时间完全相同。
Finishes	表示该事件与另一事件同时结束。
Meets	表示该事件结束时,另一事件同时开始。
Overlaps	表示两个事件的发生时间具有重叠。
Starts	表示该事件与另一事件开始时间相同。

(2) 语义关系

在叙事性文本环境下,不同事件之间除了

时序关系外,还存在着一定的语义关系,表示事件间的逻辑结构与因果关系。Nakasone 曾借助

修辞结构理论(Rhetorical Structure Theory, RST)来构建文本叙事过程中不同事件间的基本语义关系^[50]。修辞结构理论由 Mann 等人^[58]提出,认为文本是由相互关联的文本块(text span)组成,并对文本块间的语义关系进行了解释说明,该理论可以较好地解释事件间的相互关系。Mann 共定义了 23 种不同的修辞关系,并初步分为 9 种类型,包括证据与证明关系(Evidence and justify),因果关系(relations of cause),背景、环境与阐释关系(background, circumstance and elaboration),动机关系(motivation),重述和总结关系(restatement and summary)等。除此之外,PDTB 2.0(Penn Discourse Tree Bank 2.0)定义了四种类型的语篇关系^[59],分别是时序关系(temporal)、因果关系(contingency)、对比关系

(comparison)与扩展关系(extension)等。Prasad^[60]在综合已有理论及研究的基础上,统一了语篇语义关系,设计了包括 20 种核心语篇语义关系在内的 ISO 标准(ISO 24617-8)。

本研究以修辞结构理论为基础,参考 PDTB 2.0、ISO 24617-8 等标准,面向叙事性文本的特征,以事件作为叙事性文本的基本组成结构,初步将事件间语义关系分为并列关系、起因—条件关系、解释关系、环境关系等四大类 11 种关系。其中,对比关系表示事件之前存在的转折关系,起因—条件关系主要表示事件间的因果逻辑关系,解释关系主要表示前后事件间一个事件对另一个事件进行解释说明的情形,环境关系侧重阐释事件发生的背景等信息,具体关系及其解释如表 2 所示。

表 2 事件间语义关系

类型	关系	定义	参考
对比关系	Contrast(对比)	两个事件之间存在的转折关系	RST、ISO 24617-8
起因—条件关系	Reason(因果)	当前事件的发生为另一事件发生的原因	RST、ISO 24617-8、PDTB 2.0
	Condition(条件)	当前事件作为另一个事件的存在条件	PDTB 2.0
	Motivation(动机)	当前事件直接导致了另一个事件的发展	RST
	Result(结果)	当前事件的发生为另一事件发生的结果	PDTB 2.0
解释关系	Interpretation(解释)	当前事件对另一事件的发展进行了解释说明	
	Summary(总结)	当前事件是对已发生的系列事件进行的总结	
	Evidence(证明)	当前事件为另一事件提供具体示例,进一步强化事件主旨	RST
	Restatement(重释)	两个事件从不同的角度描述了同一情况	ISO 24617-8
环境关系	Background(背景)	当前事件描述了另一事件发生的背景	RST
	Circumstance(环境)	当前事件描述了另一事件发生的情境	RST

3 基于事件的情节知识表示与本体建模

情节是文本叙事性的集中体现,也是构成

叙事性文本的核心结构。因此,清晰描述叙事性文本结构的关键在于对情节的语义表示与建模。情节由事件及其时序与语义关系构成,为了构建情节本体,实现对情节的知识表示,需要

在事件知识表示的基础上,以事件网络模型为基础,完善对情节的定义。

3.1 事件网络模型

事件网络模型是在对事件、事件类型、事件间关系等进行准确定义与描述的基础上,以不同事件为节点、以事件间的语义关系为边构建而成的有向图模型,其方向可以用事件的时序关系表示,同时也可以表示事件间的语义关系。

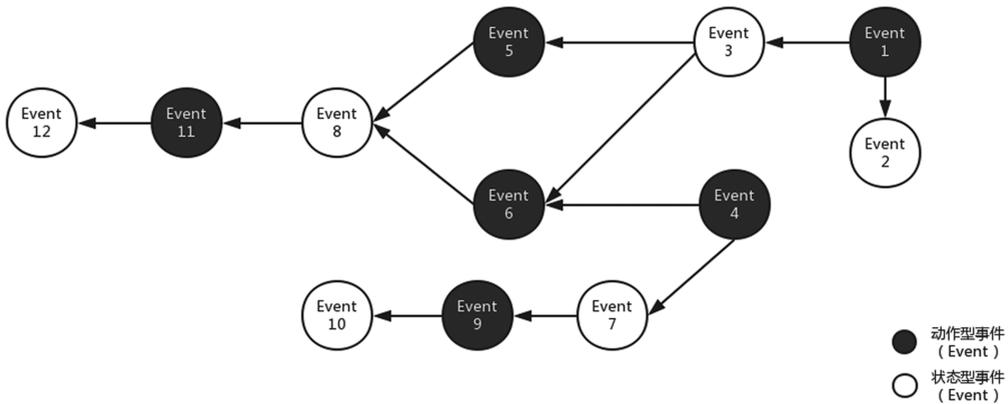


图2 事件网络模型

3.2 情节定义及语义表示

情节由若干事件组成,可以理解为具备不同语义及时序关系的事件序列。因此,情节可以用事件及事件间关系进行表示:

$$P < E, R >;$$

其中,E为事件,R为事件间关系。

情节具有极强的扩展性,为了实现对小说文本的结构化表示,需要对情节的组成进行清晰的界定。普林斯^{[18]82-92}将核心叙事(kernel narratives)定义为:全部叙事讲述 $n (n \geq 2)$ 个事件,而且事件局面(situation)或状态(state)的改变不超过一个。本文在这一理论的基础上,参考叙事性本体中关于情节的定义,将情节特征归纳为:①组成情节的最基本元素为事件;②情节必须包括两个以上的事件;③一个完整的最小粒度情节由事件序列构成,涉及状态的变化,

借助事件网络模型可以更加清晰地表现叙事性文本的语义结构特征。

图2表示事件网络的结构,以不同事件为节点,并对动作型事件与状态型事件进行了区分,节点间连线的方向表明事件的时序关系。例如,动作型事件Event1发生在状态型事件Event2与Event3的时间之前,可以用时序关系before表示。

一般应包含两种类型的事件,多数以状态型事件为结尾;④情节具有不同粒度,不同粒度的情节之间存在包含关系。若干子情节及情节之间的关联关系可以共同组成一个父情节。

因此,情节应当由一组行动性事件、状态性事件及事件间的关联关系组成。综合以上研究,情节可以具体表示为:

$$\text{Plot} < E_a, E_s, R_t, R_s >;$$

其中, E_a 为行动型事件, E_s 为状态型事件, R_t 为事件间的时序关系, R_s 为事件间的语义关系。

借助此定义,利用事件网络模型,可以较为清晰地对情节进行结构化表示。在事件网络模型图中,情节可以理解为具备一定结构的子图,且具有一定的语义属性。在图2的基础上,结合本文对情节的定义,可以在事件网络模型的基础上对情节进行表示,见图3。

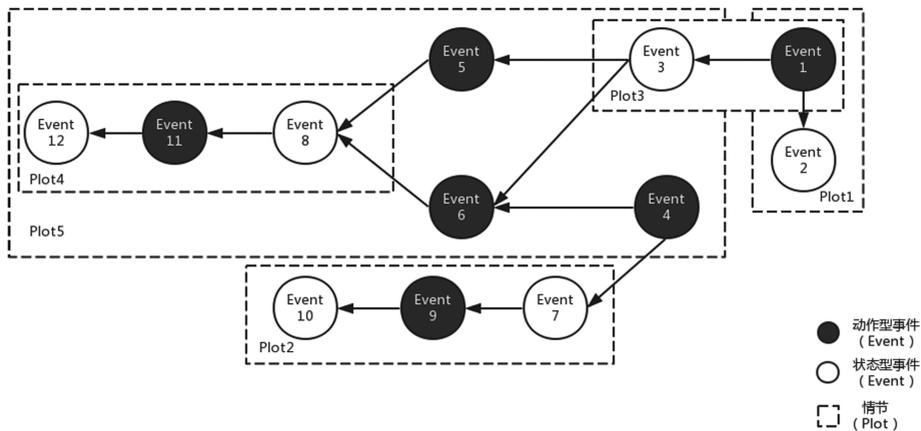


图3 基于事件网络模型的情节表示

据图3所示,情节是具有一定语义及时间关系的事件序列。情节与情节之间可以具有包含(Plot5包含Plot4)、重叠(Plot3与Plot5,Plot1与Plot3)、并发(Plot5与Plot2)等关系,这体现了叙事性文本中情节的扩展性以及叙事结构的多样性。此外,相同的事件也可以出现在不同的情节中,比如,动作型事件Event1分别引起了状态型事件Event2与Event3,构成了两个重叠的情节Plot1与Plot3。

3.3 情节本体建模

根据对事件及情节的定义,本文使用protégé5.2本体开发工具,设计了基于事件的情节本体EBPO(Event Based Plot Ontology)。该本体复用了CIDOC-CRM^[61]中的时间实体(E2 Temporal Entity)、W3C地理本体的空间信息(Spatial Thing),分别用以表示事件的时间(Time)、空间(Space)要素。该本体定义了事件间的语义与时序关系,及事件元素间的关系。EBPO本体的主要结构如图4所示。

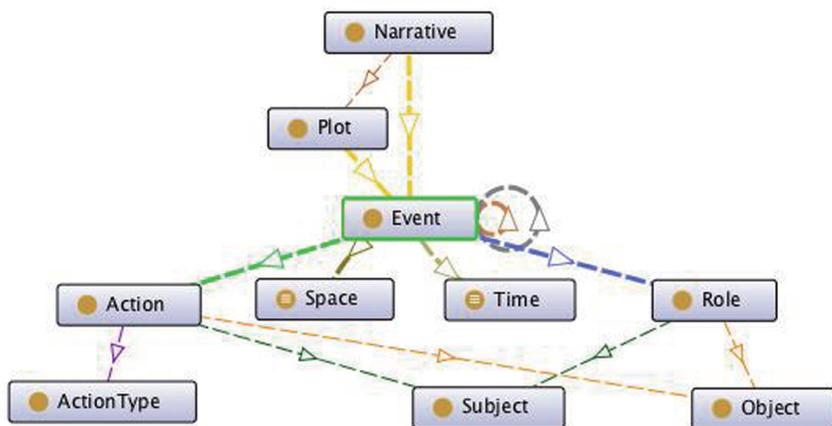


图4 情节本体EBPO的结构

由图 4 可以看出,该本体共定义了 10 个核心类,分别是叙述(Narrative)、情节(Plot)、事件(Event)、动作(Action)、空间(Space)、时间(Time)、角色(Role)、动作类型(Action Type)、主体(Subject)、客体(Object)。其中,动作类型、

主体、客体分别为动作及角色的子类。情节由事件构成,事件之间通过语义及时序两类关系进行关联。此外,该本体中还定义了包括事件间关系、事件内各元素关系在内的 31 种对象属性,其中核心对象属性如表 3 所示。

表 3 EBPO 本体核心对象属性

核心属性	领域	范围	说明
before、after、equals、during、contains、finishes、finishedBy、meets、metBy、overlaps、overlapedBy、starts、startedBy	Event	Event	事件类之间的时序关系
background、cause、justify、condition、interpretation、summary、contrast	Event	Event	事件类之间的语义关系
hasAction	Event	Action	事件的动作要素
hasType	Action	ActionType	事件动作要素所具有的类型
hasEvent	Narrative、Plot	Event	叙事及情节中所包含的事件
hasObject	Role、Action	Object	角色、动作的客体
hasSubject	Role、Action	Subject	角色、动作的主体
hasRole	Event	Role	事件中包含的角色要素
hasPlot	Narrative	Plot	一个叙事所包含的情节
hasTime	Event	Time	事件中包含的时间要素
hasPlace	Event	Place	事件中包含的空间要素

4 叙事性文本语义结构化表示方法

4.1 基于情节本体的叙事性文本语义标注

在情节本体的基础上,实现对叙事性文本语义结构化表示,首先需要对叙事性文本进行语义标注(semantic annotation)。语义标注以本体、受控词表等为基础,通过文本分析、概念识别、关系抽取等揭示文本内容的语义特征,是实现文本由非结构化向结构化、半结构化转变的关键,也是构建叙事性文本智慧数据的核心。在叙事性文本语境下,基于情节本体的叙事性文本语义标注流程如图 5 所示。

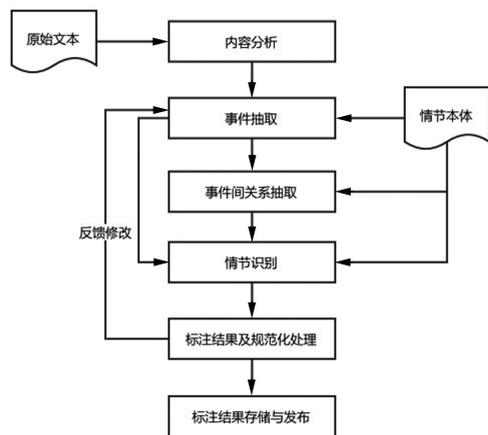


图 5 基于情节本体的叙事性文本语义标注流程

在此基础上,本研究随机抽取选择了5段小说文本片段作为标注实验样本,遵循叙事性文本语义标注流程,进行标注实验。选择GATE8.5.1作为标注工具,并通过调用插件Language: Chinese实现中文标注。同时使用GATE内嵌的自然语言处理工具ANNIE对所选样本进行命名实体识别,以在一定程度上实现机器辅助的叙事性文本半自动语义标注。标注样本及标注结果基本统计结果如表4所示。

表4 标注样本及标注结果基本统计

样本	情节	事件	角色
《傲慢与偏见》	3	11	3
《月亮与六便士》	2	10	6
《白鹿原》	5	20	8
《平凡的世界》	3	13	4
《活着》	4	14	2

4.2 叙事性文本“层次—网络”模型

叙事性文本具备一定的结构特征,以事件为核心,在叙事性文本语义标注与关系识别的基础上,叙事性文本可以表示为由事件及事件序列相互关联的事件网络,事件网络的节点和边可以通过事件及事件间的时序及语义关系来表示,因此叙事性文本具有一定的网络结构特征。在事件网络的基础上,不同粒度的情节之间存在着包含与被包含关系,凸显了叙事性文本的层次结构特征。除了包含关系之外,不同粒度的情节之间仍然具有一定的网络结构特征。据此可以将叙事性文本的结构理解为“层次—网络”模型,如图6所示。

4.3 叙事性文本语义结构化表示应用案例

在对所选叙事性文本片段进行语义标注的基础上,本研究参照叙事性文本“层次—网络”模型对标注结果进行了重新组织与关联,以实

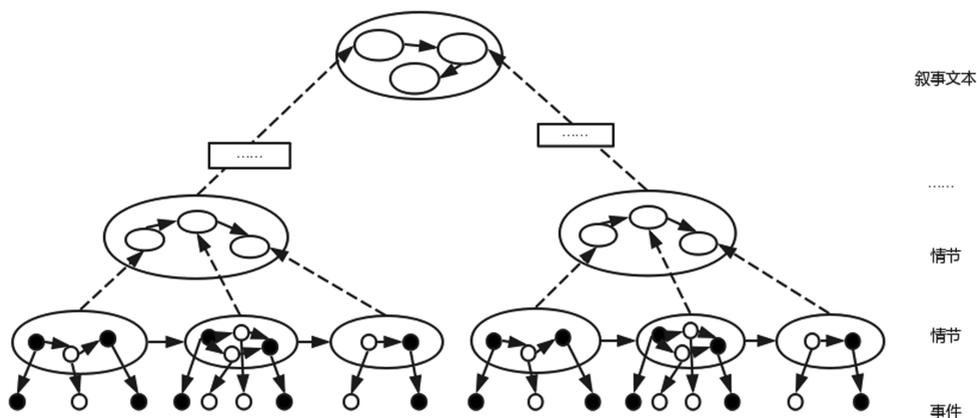


图6 叙事性文本“层次—网络”模型

现对叙事性文本的语义结构化表示。本文选取《傲慢与偏见》《月亮与六便士》中的文本片段为案例进行介绍。

(1) 案例一

…伊丽莎白拿出姐姐写来的信,一封封细读起来。没想到这时候达西先生竟然又来了!他匆匆忙忙地解释着,听菲茨威廉上校说她脸

色不好就来看看,伊丽莎白冷淡地回应着,达西自觉没趣,便在屋里踱来踱去。几分钟后,他突然转过身,走到伊丽莎白跟前激动地说:“我实在克制不住,我感觉再也压抑不住。请允许我告诉你,我非常爱你,深深地爱慕着。”

伊丽莎白惊讶得无法形容,脸蛋儿通红双眼瞪得圆溜溜。达西以为她想继续往下听呢,

立即倾诉了自己的一片深情,并承认因为伊丽莎白出身低微总令他产生矛盾心理,可是他爱得太过于强烈,最终还是克制不住灼热的情感,希望美丽的姑娘能接受他的求婚。

生性傲慢的达西在说这些话时,脸上仍然流露出一种稳操胜券的神气,伊丽莎白看得明明白白。待这位求婚者一讲完,她便尽量让自己恢复平静,一字一句地说道:“抱歉,达西先生,不管是否能接受,我都不能表示自己的感激之情。你这种爱慕之情表露得过于勉强,而且

盛气凌人,我只能给你带去痛苦。不过,这样一来,你倒很容易克制那份感情了。”达西正倚着壁炉,两眼直瞪瞪地盯着她,心里羞愤无比。他竭力装出镇定自若的样子,好一会儿才说:“我真荣幸,能够得到这样的回答!可以请教一下吗,我怎会遭到如此无礼的拒绝?”(选自简·奥斯汀《傲慢与偏见》)

在对案例一内容进行结构化处理后,可以得出相应的语义结构化表示结果,具体如图7所示。

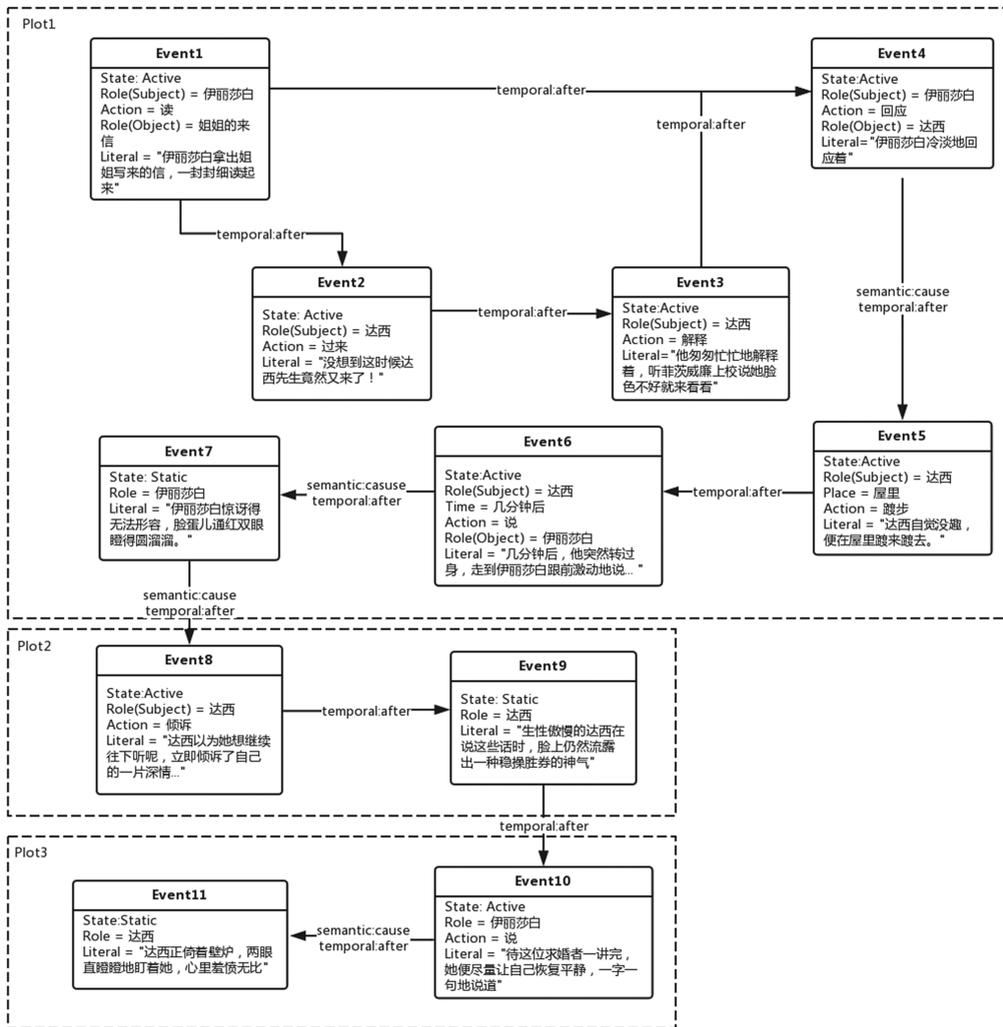


图7 叙事性文本结构化表示实例一

从图7可以得出,本次标注实验从文本中共计提取出11个事件。其中事件7、事件9、事件11为状态性事件,分别描述了角色伊丽莎白和达西的状态特点,可以理解为一系列行动性事件后造成的结果。以这三个状态性事件为情节边界,可以区分出三个情节,其中情节1叙述了“达西向伊丽莎白表白”,情节2叙述了“达西向伊丽莎白继续表白”,情节3叙述了“伊丽莎白拒绝达西”。

(2) 案例二

……在她们走出去以后,思特里克兰德把门关上,走到桌子的另一头,在皇家法律顾问和

那位政府官员中间坐下来。他又一次把红葡萄酒传过来,给客人递雪茄。皇家法律顾问称赞酒很好,思特里克兰德告诉我们他是从什么地方买来的。我们开始谈论起酿酒同烟草来。皇家法律顾问给大家说了他正在审理的一个案件,上校谈起打马球的事。我没有什么事好说,所以只是坐在那里,装作很有礼貌地津津有味地听着别人谈话的样子。因为我知道这些人谁都和我无关,所以就从容地仔细打量起思特里克兰德来。他比我想象中的……(选自毛姆《月亮与六便士》)

案例二语义结构化表示结果如图8所示:

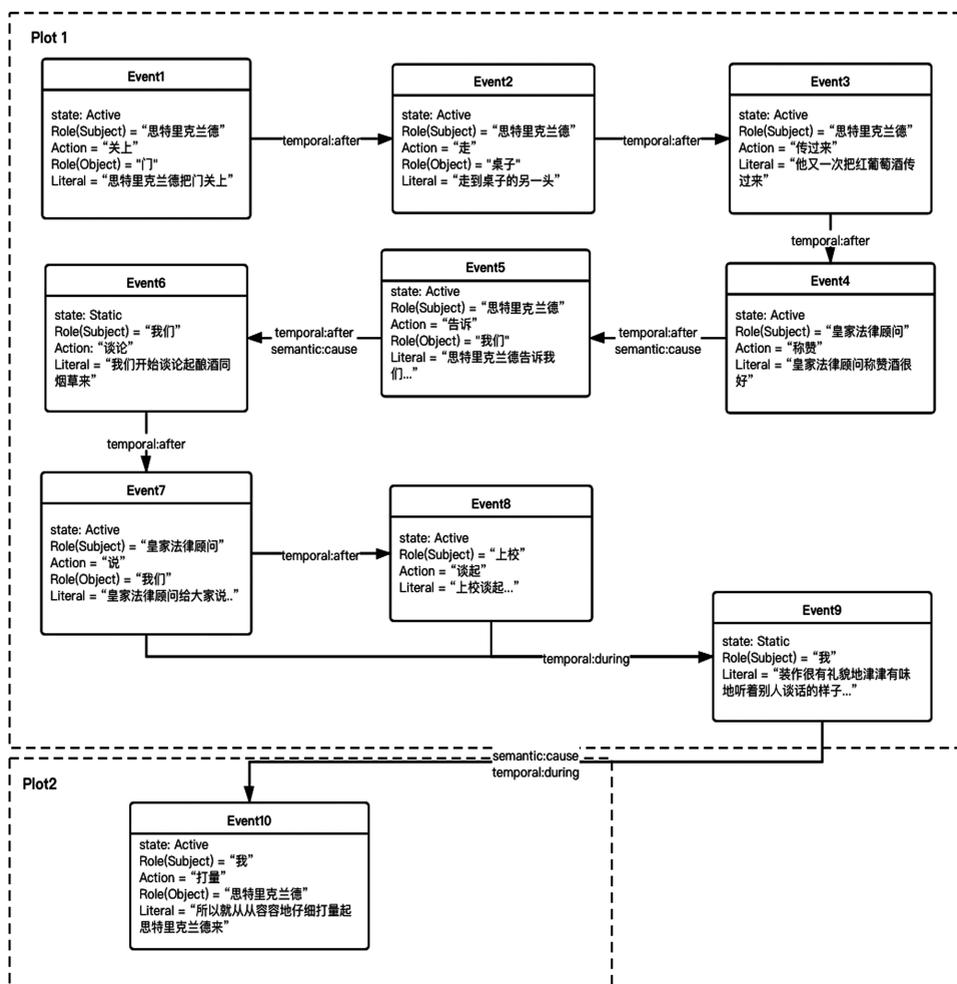


图8 叙事性文本结构化表示实例二

据图 8 所示,本次标注实验共提取事件 10 个,其中事件 9 为状态性事件,描述了“我”的状态特点。据此将该文本段划分为两个情节。情节 1 侧重介绍交谈的场景及动作,情节 2 描述“我”打量思特里克兰德。

根据标注结果可以发现,基于情节本体的叙事性文本语义结构化表示方法可以在准确表示事件、区分事件类型的基础上,实现较为精确的情节定义。同时,该标注结果也显示,叙事性文本具备一定的“层次—网络”结构特征,可以选择事件网络作为文本内容及知识的组织方式。

5 总结

本文提出了一种针对叙事性文本的语义结构化表示与知识组织方式,该方式将事件作为组成叙事性文本的最基本单元,在对事件及事件间关系进行语义建模的基础上,通过对情节的定义,设计并开发了情节本体 EBPO,并根据“层次—网络”模型关联了不同文本内容,共同组成具备叙事性特征的文本。本文提出的理论方法可以借助 TEI 等文本结构化标注方案,利用 EBPO 本体及标注本体,结合事件抽取、语义消歧、事件关联与推理等技术方法,实现对叙事性文本内容的细粒度、富语义、结构化标注,并根据事件网络实现叙事性文本的内容关联组织。这种知识组织方式可以在一定程度上实现计算机辅助的情节回顾、人物关系概览、情节发展路径呈现等基于事件的检索与浏览,将更加符合数字人文环境下人文学者的远距离阅读与策略型阅读的需求。

值得注意的是,随着自然语言处理方法的不断发展,命名实体识别、情感分析等文本处理技术日趋完善,这给叙事性文本语义化表示提供了基础,并可以在一定程度上实现由机器辅助的叙事性文本半自动语义标注。但相较

于概念实体,现有技术的事件、情节等富语义文本内容的准确表征与自动标注等方面还存在不足,未来研究仍需要在建设大规模、高质量语料库的基础上,实现对相关算法的优化与改进。本研究从基础理论出发,在探究叙事性文本内容组织方式的基础上,提出了一套较为适用的叙事性文本表示方法,可为建设高质量语料库提供帮助,也使利用深度学习技术实现叙事性文本的自动语义标注与关联组织成为了可能。

关于情节的定义与表征是深入理解叙事性文本的关键,也是对叙事性文本结构化表示与组织的核心工作。如何准确定义情节、区分事件与情节、发现情节与事件的关系,历来是叙事学、语言学乃至认知心理学等领域的研究重点与难点。本文设计了一种按照事件状态及事件要素组成的情节表征方法与模型,能够在一定程度上标识情节边界,实现情节划分。但情节本身带有强烈的主观认识,并受到文本体裁、风格等多种因素影响,语义特征较为复杂。这就给构建情节本体带来了困难与挑战,未来仍需要借鉴多领域理论与研究成果,更为精确地定义情节、情节间语义关系与情节的属性特点,实现对情节的语义表征。

受作者及体裁的影响,叙事性文本存在诸如倒叙、插叙、乱叙等不同叙事特点及叙事模式,这对事件与情节之间的关系表示提出了挑战。本研究设计的叙事性文本事件网络知识表示方法,能够在一定程度上揭示叙事性文本的叙事特点,但该方法仍属于对叙事性文本的通用型建模方法,尚不能清晰定义与表示不同类型的叙事结构与叙事风格。如何在本研究的基础上,结合不同类型的叙事风格、叙事体裁与特定语境,面向不同类型用户的个性化信息需求,更为准确地表征叙事性文本,提出契合不同叙事风格的语义关联方式,以适应不同种类的叙事特点,是未来研究中的重点。

参考文献

- [1] 胡悦融, 马青, 刘佳派, 等. 数字人文背景下“远距离可视化阅读”探析[J]. 图书馆论坛, 2017, 37(2): 1-9. (Hu Yuerong, Ma Qing, Liu Jiapai, et al. A review of visual distant reading in digital humanities[J]. Library Tribune, 2017, 37(2): 1-9.)
- [2] Edmond J. How scholars read now: when the signal is the noise[J/OL]. Digital Humanities Quarterly, 2018, 12(1) [2019-12-23]. <http://www.digitalhumanities.org/dhq/vol/12/2/000388/000388.html>.
- [3] Giunchiglia F, Dutta B, Maltese V. From knowledge organization to knowledge representation[J]. Knowledge Organization, 2014, 41(1): 44-56.
- [4] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018, 44(1): 17-34. (Zeng Marcia Lei, Wang Xiaoguang, Fan Wei. Smart data from libraries, archives and museums and its role in the digital humanity researches[J]. Journal of Library Science in China, 2018, 44(1): 17-34.)
- [5] 王晓光, 刘怡丹. 图书表示模型 BITS 及其对中文科技图书的适用性分析[J]. 数字图书馆论坛, 2016(8): 36-42. (Wang Xiaoguang, Liu Yidan. Research on structures of BITS and its suitability to Chinese STM books [J]. Digital Library Forum, 2016(8): 36-42.)
- [6] 薛家秀, 欧石燕. 科学论文篇章结构建模与解析研究进展[J]. 图书与情报, 2019(2): 120-132. (Xue Jiaxiu, Ou Shiyen. Research progress on discourse structure modelling and discourse parsing of scientific articles [J]. Library and Information, 2019(2): 120-132.)
- [7] Ciancarini P, Di Iorio A, Nuzzolese A G, et al. Semantic annotation of scholarly documents and citations[C]// Congress of the Italian Association for Artificial Intelligence. Springer, Cham, 2013: 336-347.
- [8] 程锡麟. 叙事理论概述[J]. 外语研究, 2002(3): 10-15. (Cheng Xilin. General review of narrative theory[J]. Foreign Languages Research, 2002(3): 10-15.)
- [9] 唐伟胜. 范式与层面: 国外叙事学研究综述——兼评国内叙事学研究现状[J]. 外国语, 2003(5): 60-66. (Tang Weisheng. Levels and paradigms of narratological study: with a brief comment on narratological study in China[J]. Journal of Foreign Languages, 2003(5): 60-66.)
- [10] Barthes R, Duisit L. An introduction to the structural analysis of narrative[J]. New Literary History, 1975, 6(2): 237-272.
- [11] 尚必武, 胡全生. 经典、后经典、后经典之后——试论叙事学的范畴与走向[J]. 当代外国文学, 2007(3): 120-128. (Shang Biwu, Hu Quansheng. Classics, post-classics, after the post-classics: on the category and trend of narratology[J]. Contemporary Foreign Literature, 2007(3): 120-128.)
- [12] Abrams M H, Harpham G. A glossary of literary terms[M]. Boston: Cengage Learning, 2011: 50-52.
- [13] 罗吉·福勒. 现代西方文学批评术语词典[M]. 袁德成, 译. 成都: 四川人民出版社, 1987: 172. (Fowler R. A dictionary of modern critical terms[M]. Yuan Decheng, trans. Chengdu: Sichuan People's Publishing House, 1987: 172.)
- [14] Gerard G. Narrative discourse: an essay in method[M]. Ithaca: Cornell University Press, 1984: 23-24.
- [15] E.M.福斯特. 小说面面观[M]. 冯涛, 译. 北京: 人民文学出版社, 2009. (Forster E M. Aspects of the novel [M]. Feng Tao, trans. Beijing: People's Literature Publishing House, 2009.)

- [16] 华莱士·马丁. 当代叙事学[M]. 伍晓明, 译. 第二版. 北京: 北京大学出版社, 2005: 83-95. (Martin W. Recent theories of narrative[M]. Wu Xiaoming, trans. Second Edition. Beijing: Peking University Press, 2005: 83-95.)
- [17] 胡亚敏. 叙事学[M]. 武汉: 华中师范大学出版社, 1994: 119-129. (Hu Yamin. Narratology[M]. Wuhan: Central China Normal University Press, 1994: 119-129.)
- [18] 杰拉德·普林斯. 叙事学: 叙事的形式与功能[M]. 徐强, 译. 北京: 中国人民大学出版社, 2013. (Prince G. Narratology: the form and functioning of narrative[M]. Xu Qiang, trans. Beijing: China Renmin University Press, 2013.)
- [19] 吴琪. 叙事学视野下的小说情节[D]. 成都: 四川大学, 2003. (Wu Qi. A study of plot in the perspective of narratology[D]. Chengdu: Sichuan University, 2003.)
- [20] 方珊. 形式主义文论[M]. 济南: 山东教育出版社, 1999: 89-99. (Fang Shan. Literary theory of form-structure[M]. Jinan: Shandong Education Press, 1999: 89-99.)
- [21] 谭光辉. 论情节类型: 时间与因果的复杂组合方式[J]. 文艺争鸣, 2016(7): 135-141. (Tan Guanghui. On the type of plot: the combination of time and causality[J]. Literary and Artistic Contention, 2016(7): 135-141.)
- [22] 赵毅衡. “叙述转向”之后: 广义叙述学的可能性与必要性[J]. 江西社会科学, 2008(9): 31-41. (Zhao Yiheng. After the “Narrative Turn”: the possibility and necessity of generalized narratology[J]. Jiangxi Social Sciences, 2008(9): 31-41.)
- [23] 谢龙新. 经典“叙事”概念: 外延, 内涵及其超越[J]. 湖北师范学院学报(哲学社会科学版), 2010(5): 24-29. (Xie Longxin. Classical narrative concept: extension and meaning[J]. Journal of Hubei Normal University (Philosophy and Social Science), 2010(5): 24-29.)
- [24] 凌晨光. 作为叙事虚构作品的小说的特性[J]. 东岳论丛, 2003, 24(6): 111-113. (Ling Chenguang. The characteristics of novels as narrative fictions[J]. Dong Yue Tribune, 2003, 24(6): 111-113.)
- [25] 周靖波. 影视叙事文本特性初探[J]. 现代传播: 北京广播学院学报, 2003(3): 67-72. (Zhou Jingbo. A probe into the characteristics of movie narrative texts[J]. Modern Communication: Journal of Beijing Broadcasting Institute, 2003(3): 67-72.)
- [26] Kim J. Events as property exemplifications[M]//Brand M, Walton D. Action theory. Dordrecht: Springer, 1976: 159-177.
- [27] Timberlake A. Aspect, tense, mood[J]. Language Typology and Syntactic Description, 2007, 3: 280-333.
- [28] Chang J. Event structure and argument linking in Chinese[J]. Language and Linguistics, 2003, 4(2): 317-351.
- [29] Speer N K, Zacks J M, Reynolds J R. Human brain activity time-locked to narrative event boundaries[J]. Psychological Science, 2007, 18(5): 449-455.
- [30] 周文, 刘宗田, 孔庆苹. 基于事件的知识处理研究综述[J]. 计算机科学, 2008, 35(2): 160-162. (Zhou Wen, Liu Zongtian, Kong Qingping. A survey of event-based knowledge processing[J]. Computer Science, 2008, 35(2): 160-162.)
- [31] Nelson K, Gruendel J. Event knowledge: structure and function in development[M]// Event knowledge: structure

- and function in development. Erlbaum, 1986.
- [32] 仲兆满, 刘宗田, 李存华. 事件本体模型及事件类排序[J]. 北京大学学报(自然科学版), 2013, 49(2): 234-240. (Zhong Zhaoman, Liu Zongtian, Li Cunhua. Event ontology model and event class ranking[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2013, 49(2): 234-240.)
- [33] Event in Schema.org[EB/OL]. [2018-10-19]. <https://schema.org/Event>.
- [34] Allen J F, Ferguson G. Actions and events in interval temporal logic[J]. Journal of Logic and Computation, 1994, 4(5): 531-579.
- [35] Li M, Chen J, Chen T, et al. Probability for disaster chains in emergencies[J]. Journal of Tsinghua University Science and Technology, 2010, 50(8): 1173-1177.
- [36] 仲兆满, 刘宗田, 周文, 等. 事件关系表示模型[J]. 中文信息学报, 2009, 23(6): 56-60. (Zhong Zhaoman, Liu Zongtian, Zhou Wen, et al. The model of event relation representation[J]. Journal of Chinese Information Processing, 2009, 23(6): 56-60.)
- [37] 王佳琪, 张均胜, 乔晓东. 基于文献的科研事件表示与语义链接研究[J]. 数据分析与知识发现, 2018, 2(5): 32-39. (Wang Jiaqi, Zhang Junsheng, Qiao Xiaodong. Analyzing representation and semantic links of scientific research events[J]. Data Analysis and Knowledge Discovery, 2018, 2(5): 32-39.)
- [38] Kerlin B, Cooley B C, Isermann B H, et al. Cause-effect relation between hyperfibrinogenemia and vascular disease[J]. Blood, 2004, 103(5): 1728-1734.
- [39] 刘宗田, 黄美丽, 周文, 等. 面向事件的本体研究[J]. 计算机科学, 2009, 36(11): 189-192. (Liu Zongtian, Huang Meili, Zhou Wen, et al. Research on event-oriented ontology model[J]. Computer Science, 2009, 36(11): 189-192.)
- [40] Event Ontology[EB/OL]. [2018-10-17]. <http://motools.sourceforge.net/event/event.html>.
- [41] Hage W R V, Malaisé V, Segers R, et al. Design and use of the Simple Event Model (SEM)[J]. Web Semantics Science Services & Agents on the World Wide Web, 2011, 9(2): 128-136.
- [42] Linking description of event [EB/OL]. [2018-10-17]. <http://linkedevents.org/ontology>.
- [43] Corda I, Bennett B, Dimitrova V. A logical model of an event ontology for exploring connections in historical domains[C]//Workshop on Detection, Representation and Exploitation of Events in Semantic Web (Derive 2011), Tenth International Semantic Web Conference (ISWC). 2011.
- [44] Jeong S, Kim H G. SEDE: an ontology for scholarly event description[J]. Journal of Information Science, 2010, 36(2): 209-227.
- [45] Teymourian K, Paschke A. Towards semantic event processing[C]//Proceedings of the Third ACM International Conference on Distributed Event-Based Systems. ACM, 2009: 29.
- [46] Han Y J, Park S Y, Park S B, et al. Reconstruction of people information based on an event ontology[C]//Proceedings of 2007 International Conference on Natural Language Processing and Knowledge Engineering. IEEE, 2007: 446-451.
- [47] 张旭洁, 刘宗田, 刘炜, 等. 事件与事件本体模型研究综述[J]. 计算机工程, 2013, 39(9): 303-307. (Zhang Xujie, Liu Zongtian, Liu Wei, et al. Survey of event and event ontology model research[J]. Computer

- Engineering, 2013, 39(9): 303-307.)
- [48] Stories ontology [EB/OL]. [2018-10-17]. <https://bartoc.org/en/node/18293>.
- [49] Damiano R, Lieto A. Ontological representations of narratives: a case study on stories and actions[C]//Satellite workshop of the 35th Meeting of the Cognitive Science Society CogSci. 2013.
- [50] Nakasone A, Ishizuka M. ISRST: an interest based storytelling model using rhetorical relations [C]//International Conference on Technologies for E-Learning and Digital Entertainment. Springer-Verlag, 2007: 324-335.
- [51] BBC storyline ontology [EB/OL]. [2018-10-17]. <https://www.bbc.co.uk/ontologies/storyline>.
- [52] Winer D. Review of ontology based storytelling devices[J]. Lecture Notes in Computer Science, 2014, 8002: 394-405.
- [53] 高彦梅. 语篇语义框架研究[M]. 北京: 北京大学出版社, 2015.(Gao Yanmei. Constructing semantic frames in discourse[M]. Beijing: Peking University Press, 2015.)
- [54] Kaneiwa K, Iwazume M, Fukuda K. An upper ontology for event classifications and relations[C]//Proceedings of 20th Australian Joint Conference on Artificial Intelligence, Gold Coast, Australia. DBLP, 2007:394-403.
- [55] 赵毅衡. 论底本:叙述如何分层[J]. 文艺研究, 2013(1):5-15.(Zhao Yiheng. The theory on dibon: how to stratify narrative[J]. Literature & Art Studies, 2013(1): 5-15.)
- [56] Time ontology in owl [EB/OL]. [2018-10-16]. <https://www.w3.org/TR/owl-time>.
- [57] Allen J F. Towards a general theory of action and time[J]. Artificial Intelligence, 1984, 23(2): 123-154.
- [58] Mann W C, Thompson S A. Rhetorical structure theory: description and construction of text structures[M]//Kempen G. Natural language generation. Springer Netherlands, 1987:85-95.
- [59] Prasad R, Dinesh N, Lee A, et al. Attribution and its annotation in the Penn Discourse TreeBank[J]. TAL, 2006, 47(2): 43-64.
- [60] Prasad R, Bunt H. Semantic relations in discourse;the current state of ISO 24617-8[C]//Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11). 2015.
- [61] Doerr M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata[J]. Ai Magazine, 2003, 24(3):75-92.

宋宁远 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

王晓光 武汉大学信息管理学院教授。湖北 武汉 430072。

(收稿日期:2019-01-03;修回日期:2019-06-12)