Introduction and Overview

Introduction

Contemporary societies are exposed to a continuous flow of information, producing vast amounts of documents at different points in time. Assume that someone needs a comprehensive report of the *events* that occurred in the 2020 Minneapolis protests (e.g., When did they start? Why did they start? Who was involved? Who was in favor/against? How did they end?). An attentive reader will have to (1) collect documents from different sources over a period of time and order them chronologically; (2) select the relevant events and participants (i.e., who, what, and where), deduplicating repetitions at in- and cross-document level; (3) reconstruct the chronological order of the events (i.e., when); (4) connect the relevant events and participants in a coherent way (i.e., why and how); (5) distinguish between important and peripheral information; and (6) distinguish between reporting of events and perspectives on events (i.e., who thinks/feels what).

Available technologies struggle to solve these issues. News aggregation systems can easily monitor the burst and the development of a topic in terms of the quantity of documents published, but they fail in providing content-based analysis. People have to read the documents and a single, stand-alone coherent report manually.

Current natural language processing (NLP) systems also fail when documents' coherence and cross-document connections play a major role in the extraction of information, as in this case. They can identify complex information, but they lack a method combining it into a unitary and coherent message. Steps in this direction have been conducted with the development of entailment recognition tasks (Dagan et al., 2013; Bowman et al., 2015), end-of-story prediction tasks (Mostafazadeh et al., 2016, 2017), narrative

chains (Chambers and Jurafsky, 2008, 2009), and script embeddings (Regneri et al., 2010; Pichotta and Mooney, 2014; Rudinger et al., 2015; Modi, 2016; Bisk et al., 2019), but they are still limited and in their infancy.

It has been suggested that what actually makes us human is our ability to tell stories and create narratives (Boyd, 2009; Gottschall, 2012). Narratives represent an evolutionary asset of mankind that allows us to make sense of our experiences and reality, identify explanatory patterns, and build models for reasoning and decision making (Boyd, 2009). People can easily refer to changes in the world, identify their participants, distinguish relevant information, and have expectations of what can happen next. The human ability to find and create narratives allows us to deal with huge streams of fragmented information on a daily basis. Such narrative structures are at the heart of information sharing, as is exemplified by the structure of news articles. Developing systems and approaches that can represent, understand, and generate narratives would lead to more user-friendly and "intelligent" systems that could effectively help humans in their everyday navigation across information and contribute to the reduction of the negative effects of information overload.

Narrowing down the generic expression of "what happens in a narrative," we can see that at least three big notions are involved: events, representing the "what"; their participants, representing the "who and where"; and a set of relations that connect the previous elements together and contribute to the development and coherence of narratives, at both local and global levels.

A somewhat simplified picture that emerges is that of stories as sequences of events. This is actually true for many stories and in particular for stories emerging from news articles. An additional dimension highly connected to the production of stories is that of a moral compass. Many stories, regardless of the specific media that realize them (i.e., a painting, a novel, a play, a song, among others), contain a point or a moral. Aesop's *Fables*, a collection of fables from ancient Greece, were originally addressed to an adult public, covering different themes (e.g., religion, politics, societal issues), and all have an ethical dimension.[1] As a more provocative statement, all religious textbooks (e.g., the Bible, the Koran, the Vedas, among others) can be see as stories with a moral. Others include changes of fortune, insights and growth of personality, or depictions of personalities. Handling such aspects and dimensions of stories is an open challenge for NLP systems, because multiple complex aspects (e.g., coherence of the document, characters' intentions and personalities, moral insights) have to be modeled and dealt with at the same time.

[1] All stories end with an ethical summary (e.g., "This story teaches us that [. . .]") of what one learns from each story.

This volume, on the other hand, focuses on methods and approaches that address the "simple" event-sequence type of story. In the following paragraphs, we present a (short) overview of the components and notions that are involved in this type of story to help the readers to navigate through the upcoming chapters.

Event Structure: Microlevel of Analysis

Event structures have been at the heart of linguistics, philosophy, and artificial intelligence. The dialogue across these disciplines to better understand what events are and how to automatically work with them has been continuous and it is still ongoing. Much work has been conducted in this area, which we call microlevel of analysis of events, and consensus has been reached across disciplines and communities on some aspects.

From an ontological point of view, events are considered objects that can occupy portions of time and space. There is also a general consensus to define events as things that occur or happen or hold true. Using a more appropriate ontological terminology, and especially of upper ontologies (Niles and Pease, 2001; Gangemi et al., 2002), events are "perdurants"; i.e., concepts that must be defined dependent on time. An entity's continuity of existence in time has a pivotal role in conceptual classification of human knowledge systems (Huang, 2016).

Human perception, action, language, and thought manifest a commitment to events. It is quite easy to refer to events, either directly (example (1a)) or by means of anaphoric pronouns (example (1b)); to quantify over them; or to modify them through adverbial phrases (example (1c)).

(1) a. *John had a* ***walk***.
b. *John* ***walked***$_i$. ***It***$_i$ *was nice.*
c. *John* ***walked***$_i$ *in the park with Mary.*

The investigation of adverbial modification has led to the formulation of a program to systematically capture and account for appropriate entailment and representation of propositions involving event expressions. From example (1c) it is quite intuitive to infer that the complex event of John walking in the park with Mary entails the simpler event of John walking. Because events can be represented as predicates expressing a, *n*-ary relation between entities, it becomes very difficult to explain why example (1c) that expresses a ternary relation (*John*, *in the park*, *with Mary*) also entails a different unary relation (*John*; Kenny, 2003). The solution of this "riddle" impinges

on three subjects (logic, linguistics, and ontology) and has been presented (not without problems) by Davidson (1969). It requires a reification process (i.e., the introduction of new entities in the domain of discourse) of the events as if they were individuals, thus allowing quantification and reference. Predicates of natural language predicate denote events; thus, they require an additional event argument as part of their argument structures (Bach, 1986; Dowty, 1989; Parsons, 1990). This operation has opened new perspectives in the treatment of events and the study of event semantics. For instance, it strengthens the ontological analysis of events as spatiotemporal entities and, at the same time, allows the extension of the event argument to any predicate (Higginbotham, 1985).

When it comes to the linguistic realizations of events, the introduction of the event argument allows a comprehensive analysis of events without being restricted to any specific parts of speech. Verbs, of course, represent the preferred and unmarked realization of events, but any other part of speech that instantiates a predicate may be considered an event. An interesting applicative case of this vision is represented by the TimeML Annotation Guidelines for English (Saurí et al., 2006), which exhaustively illustrates all possible linguistic realizations of events, ranging from verbs to nouns, adjectives, and predicative constructions with prepositional phrases.

As events pertain to reality, it has been noted that their linguistic realizations convey meaning about the temporal properties of events. Work in this area has investigated both when events happen (Reichenbach, 1947) and their intrinsic temporal properties.[2] These matters focus on the lexical semantic properties of events and are also known as "lexical aspect" or *aktionsarten* (German for "kinds of action").

The seminal work of Vendler (1967) focused on verbs and laid down a four-class typology of types of actions: activities, states, accomplishments, and achievements. The distinctions across these classes is based on a set of parameters involving the change undergone during the time the event occurs (dynamicity), the presence of a natural ending point for the event (telicity), and the temporal span the event occupies (duration). States have no internal structure or change during the time span in which they hold (or are true). In a sentence like "*Marc knows English*" there is no difference between the knowing of English by Marc at different moments in time, nor must the knowing of English reach a natural ending point for it to be valid. Activities are events that consist of successive actions over a period of time; they have

[2] Very recently, attention has also been focused on a more neglected aspect of events: space. Interested readers are referred to Pustejovsky (2013)

an internal change and a duration but no natural end point. In "*John walks in the park*," the event is perceived as valid whether John walks for 10 minutes or an hour. Accomplishments are events with a duration and a necessary end point. In "*John drew a circle*" the event requires some time in order to reach a "climax," or its natural end point. Failure to do so will result in a nonvalid event: "*John is drawing a circle*," pictures John in the act of drawing a circle and it entails that the circle has not been finished yet. Achievements are events with an instantaneous end point and no duration. In "*John arrived*," the event is valid only when John is actually at his destination; i.e., the event has reached its climax but there is no extension of the event over time.

The validity of this classification has been proved by means of batteries of linguistic tests (e.g., compatibility with temporal adverbials introduced by *for* or *in*; compatibility with progressive form, among others). Furthermore, different organizations and refinements have been proposed over time by different authors (Pustejovsky, 1991; Krifka, 1992; Tenny, 1994; Lenci and Bertinetto, 2000; Verkuyl, 2013). Overall, following Bach (1986), we can characterize a global class of *eventualities* whose major distinction is along the dynamicity parameter, differentiating between states and events, with the latter further distinguished by different subtypes.

Recent work in computational linguistics (Setzer, 2001; Pustejovsky et al., 2003a; Linguistic Data Consortium [LDC], 2005) has adopted the term "event" as a general umbrella expression to include both states and events. At the same time, different classifications of events have been proposed. An extensive overview of the different annotation schemes for events shows that none of them adopts a typology based on the temporal properties of events. Event classes are based on either semantic descriptions of the events (e.g., "Attack"; "Transportation", among others) as in LDC (2005) and Mitamura et al. (2015) or on syntactic-semantic properties of the events that can be further used for reasoning (Pustejovsky et al., 2003a) or dealing with temporal relations (O'Gorman et al., 2016).

A challenging aspect of events concerns how their meaning must be represented to account for their internal structure. A large body of literature has developed the idea that event meaning can be analyzed into its structural components. Decompositional models (Jackendoff, 1983; Dowty, 1989; Levin and Rappaport Hovav, 1995) have promoted an analysis of event meaning in their compositional primitive elements. An accomplishment event such as "*John drew a circle*" could be decomposed into a fixed set of semantic primitives: CAUSE(John, BECOME(DRAWN(y))). In these works, there is a consensus around the idea that complex events, such as accomplishments, are composed of an inner and an outer event, where the outer event expresses

causation and agency and the inner event expresses telicity and change of state. Limitations of primitive-based approaches are known and correspond to a risk of proliferation of the primitives themselves. A different approach was proposed by Pustejovsky (1991): rather than decomposing event structures into primitives, events are decomposed into subevents (i.e., substructure of events), each with predicative content, together with rules that govern event composition (see also Chapter 1).

Because events can be represented as predicates, one of their functions is that of expressing a relation with respect to the elements to which they apply. Elements to which event predicates apply are referred to as event participants, and they contribute to the definition of event structures. Different proposals and approaches to effectively modeling event structures have been made in the literature. Besides some inherent differences, theses models are largely concerned not only with identifying the event participants but also in assigning roles to the participants within the event. In other words, the (linguistic) investigation of event structures aims at understanding the "meanings of arguments (e.g., subjects and objects) in the linguistic coding of events and their structural expressions in sentences" (Marantz, 2013, p. 152).

Most meaning representations use a lexically based approach that assumes that the lexical semantics of a verb determines the complements that occur with it in a clause. Recent developments in the generative semantics tradition place the construction of meaning at the interface between syntax and semantic interpretation, with the syntactic configurations determining event structure interpretations.

Access to the event structure has important consequences from a computational linguistics (CL)/NLP perspective. On the one hand, it appears that CL/NLP has used the patterns mapping the semantic arguments of events into syntactic structures for both the creation of language resources (e.g., VerbNet, PropBank, among others) and the development of systems. On the other hand, access to the event structure represents a first step for the development of more advanced systems for understanding narratives. Knowing the actors involved in an event and, most important, their roles and the semantics associated with them may help the development of systems that will allow some form of reasoning and (possibly) prevent wrong inferences.

Finally, recent work has started to investigate "*where do events come from?*" By introducing refinements on the interplay between events and relationships, Guarino and Guizzardi (2016) proposed to distinguish events from scenes. Scenes are described as things that happen in a specific spatiotemporal

region; i.e., perdurants that are the object of a unitary perception act.[3] A scene always has some time duration and it occurs in a certain place. In this analysis, events are described as relationships emerging from scenes through a *focusing process*. The focus here is to be interpreted as the relationships that the different participants have with respect to the event (i.e., the event structure): "[s]o we can distinguish some *core participants*, and others that are not involved at all in the event, except in a very indirect way" (Guarino and Guizzardi, 2016, p. 246). By refining the individuation principle that applies to these notions, different scenes must have different spatiotemporal locations, whereas different events may share the same spatiotemporal location.

Connecting Events: Macrolevel of Analysis

As soon as we move outside the boundaries of single sentence analysis of events and their structures, it becomes even more evident how events are complex hubs of information. As a matter of fact, sentences hardly exist or are interpreted in isolation. Context, being the actual context of utterance or the previous discourse history, plays a major role.

When investigating events in discourse, multiple aspects are at play. For instance, the event structure may be influenced or modified by the context of occurrence. In this section, we will overview three aspects that involve event–event relations and that play a crucial role for the development of robust narrative understanding systems: coreference relations, temporal relations, and finally, causal relations.

Davidson (1969) already used anaphoric relations in support of his proposal of an event argument. However, being able to exactly identify when two events are coreferential – i.e., they denote the same event mention – is not an easy task. Strict definitions of event coreference will require an exact match across different features, such as (i) the event description; i.e., what has happened; (ii) the event participants; i.e., who is involved and their roles; and (iii) the spatiotemporal location; i.e., when and where the event mention occurred. As Lu and Ng (2018) clearly illustrated, an end-to-end system for event coreference resolution requires addressing the following subtasks: (1) extract the event mentions (event trigger) and their associated participants, (2) extract the entity mentions (including time and place) and determine

[3] It is easy to draw parallels between this definition of scene and frames in frame semantics (Fillmore, 1968, 1982) or that of scripts (Schank and Abelson, 1977).

those that are coreferential, and (3) determine which events are coreferential by matching event triggers and event participants. A challenging aspect of event coreference is the potentially high variation of linguistic realizations of an event (i.e., verbs, nouns, adjectives, prepositional phrases, pronouns). In addition to cross–part of speech relations, challenges are represented by lexical semantic relations such as similarity, hyperonymy/hyponymy, among others, that can be used to refer to event mentions. This variability of realizations combined with the fact that events can relate to each other in different ways makes the task of event coreference even more difficult. In particular, this means that events can corefer fully (identity) or partially (quasi-identity; Poesio et al., 2004; Hovy et al., 2013). Among the set of relations that give rise to semi-identity is the subevent relation (Araki et al., 2014; Araki and Mitamura, 2015).[4] Event coreference resolution is an important building block for properly understanding narratives and extracting storylines both from a single document and when combining multiple documents.

Temporal relation extraction, more than event coreference, is an essential task for the extraction and generation of storylines. When reading a text, we interpret sentences in succession, and at the same time we fit them into a temporal structure. This temporal structure is responsible for allowing us to make inferences on the chronological order of events (i.e., a timeline reflecting the actual temporal order in which events might have occurred), even in the absence of explicit information in the text. In this case, the set of devices that natural languages have at their disposal is varied and larger than those used to express events. These include grammatical devices such as tense and aspect (grammatical aspect), use of discourse connectives such as prepositions and conjunctions, the structure of the discourse itself, and, finally, our own event knowledge. Their combination is responsible for generating what we perceive as a temporal structure of events.

An important aspect concerns the distinction between the actual temporal structure, the chronological order of events as they happen in the world, and the narrative conventions that different text genres adopt when presenting a story. For instance, in the case of news articles, the order of presentation of events is influenced by their perceived news value, which can actually be independent of their chronology. This makes the extraction of such relations from these types of texts even more complex. Additional relevant aspects concern the granularity of the temporal relations used to chronologically connect events. TimeML (Pustejovsky et al., 2003b) qualifies as the most advanced proposal

[4] For a recent overview of corpora and systems on this topic, readers are referred to Lu and Ng (2018).

for annotating temporal relations in texts and subsequently using the annotated data to train systems to extract timelines. The task is far from being solved due to numerous pending issues, including the number and types of temporal relations to be used and the sparseness of the annotation between potential event pairs. A more comprehensive overview of the task and solutions in the perspective of storyline extraction is presented in Chapter 4.

Causality plays a central role in the making of stories. Sequences of events ordered in time do not make up a story per se unless there is some sort of "explanation" or logical connections among them. To clarify, consider the tradition of the annals (*annales*), a literary genre very popular in ancient Rome and in the Middle Ages, devoted to chronological records of events. These, like any other timeline, present events exactly in the order in which they happened but do not make connections or infer explanations as to the reasons why an event happened. Causality, on the other hand, is the essential glue that explains why certain things happened and contributes to the making of a coherent story. The addition of causal relations across pairs of events can be envisioned as an aiding cleaning process of timelines in the generation of storylines: only events that are coherently connected will be maintained and presented to the users. Chapter 5 introduces an overview of the role of causal relations for storyline extraction, showing shortcomings of existing approaches and suggesting directions for future work.

One of the goals of this short overview is to show how much of the attention of scholars in linguistics, NLP, and artificial intelligence has mainly focused on the microlevels of analysis of events. Although there is an abundance of work on analysis of language outside the sentence boundary and in the context of larger and more complex units, such as text or discourse, when it comes to events the macrolevel of analysis appears still in its infancy. The complexity of the phenomena at stake (coreference, chronological order, causal inference, commonsense knowledge) are such that available systems as well as theoretical frameworks all present inadequate solutions. It is the goal of this volume to present both a perspective and the state-of-the-art viable approaches that aim at connecting into a homogeneous framework these two levels of analysis. Only by this means can we move toward new and more viable solutions for improving automatic methods for narrative understanding.

Part One: The Foundational Components of Storylines

The first six chapters of this volume introduce the foundational elements of a storyline, as well as state-of-the-art solutions.

Pustejovsky's chapter (Chapter 1) provides an excellent entry point for one of the key components of stories, namely, events and their structure. The chapter offers a thorough overview of research conducted on event structure representations by taking into account different perspectives including linguistics, NLP, and artificial intelligence. The dialogue that it creates across these perspectives will help the reader to appreciate the complexity of the problem and the proposed solutions, highlighting differences and commonalities. It is exactly on the commonalities of the different approaches that a unifying perspective of event semantics is outlined and proposed. An innovative distinction is proposed between surface events, namely, denoted by verbal predicates, and latent event structures; i.e., the finer-grained subeventual representations of events. The chapter offers an example of how to integrate the subeventual event representations into a (lexical) resource designed to model atomic event representations (i.e., VerbNet; Kipper et al., 2008). The distinction between surface and latent (or deep) event structures of sentences and texts is a step forward toward the development of a general computational theory of event structure, with a common vocabulary for events and their relations that may enable reasoning at multiple levels.

Bonial and colleagues (Chapter 2) dive deep into existing ontologies and lexicons for representing atomic event structures. Different approaches and perspectives involved in the development of these useful tools are presented, allowing the reader to also become familiar with the different theoretical backgrounds that have informed such efforts (of particular interest is the section dedicated to semantic role labeling), but it also demonstrates the variety of ways and extents to which lexical resources and ontologies can be integrated and thus enrich each other. A result of this integration effort is the rich event ontology (REO). REO offers a a novel hierarchy of event concepts capable of linking resources for atomic event structure representations (e.g., VerbNet and FrameNet). An immediate advantage of this integration is the expansion of the conceptual coverage of atomic events, thus facilitating deeper reasoning about events. Additionally, REO extends event representations with typical temporal and causal relations between events, a key information that could be integrated as prior event knowledge in systems for the extraction of storylines.

Models of events structure are essential in order to develop a robust storyline extraction framework. By taking the participants in the stories (and thus in the events) as the organizing criterion for the identification of storylines, the required model of event structure needs to express exactly what participants do and what happens to each of them in an event mention. Such an approach is presented by Croft and colleagues (Chapter 3). They propose a model of event decomposition that takes into account time, causation, and qualitative