Survey on Narrative Structure: from Linguistic Theories to Automatic Extraction Approaches

Aman Berhe* — Camille Guinaudeau* — Claude Barras**

* *Université Paris-Saclay, CNRS-LISN*
** *Vocapia Research*

ABSTRACT. *Narration is an essential element in the transmission of written and oral stories and corresponds to both who is telling the story and how the story is told. Philosophers and structuralists have analyzed and defined different narrative structures. Recently, researchers in Machine Learning (ML) or Natural Language Processing (NLP) have been particularly interested on the extraction and understanding of narrative structure in different collections. On this work, we present a survey on theories, research and techniques around narrative structure: from linguistic theories to automatic approaches used for the extraction and analysis of narrative structure in different multimedia collections and annotation tools.*

KEYWORDS: *Narratives, Narratives structure, Multimedia collection.*

RÉSUMÉ. *La narration est un élément essentiel à la transmission d'histoires écrites et orales et correspond à la fois à la personne qui raconte l'histoire et à comment l'histoire est racontée. Philosophes et structuralistes ont analysé et défini différentes structures narratives existantes. Récemment, des chercheurs en apprentissage automatique ou en traitement automatique des langues se sont particulièrement intéressés à l'extraction et à la compréhension de la structure narrative dans divers contenus. Dans ce travail, nous présentons une revue des théories, recherches et techniques autour de la structure narrative. Cette description recouvre à la fois les théories linguistiques et les approches automatiques utilisées pour l'extraction et l'analyse de la structure narrative dans divers contenus multimédias ainsi que les outils d'annotation mis en place pour développer des corpus.*

MOTS-CLÉS : *Narration, Structures narratives, Ressources multimédias.*

1. Introduction

Narrative is a way to tell an information or a story from a particular point of view. It is used to tell stories, facts or scientific results in the form of texts, audios and videos for the purpose of entertainment, education or history preservation. The way narratives progress gradually is referred to as narrative structure. Narratives are audience interactive as they include high level themes related to deep human emotions and create a strong connection and motivation among the audience. This connection is also established through a dramatic element that helps to capture the audience, the narrative-hook, that is a core point of the structure and progress of the narrative. Onega and Landa (2014) defined narrative as the semiotic representation of a series of events meaningfully connected in a temporal and casual way. They further classified films, plays, comic strips, novels, newsreels, diaries, chronicles and treatises of geological history as narratives. Movies, TV series, fictional books, audio recordings that focus on telling a story follow a complex sequence of steps to mesmerize audiences from the start to the end of the intended story. These sequences of steps are referred to "narrative structure".

Since the 19th century, some renowned philosophers (Lucas, 1968), formalists (Propp, 1958) or structuralists (Leach, 1970) have studied narratives and their structural development in an intensive manner, from the point of view of literature. Their work has been the basis for the development of the study of storytelling in general and narrative structure specifically. The importance of narrative is undoubtedly clear on conveying information, entertaining, history preservation or documenting, starting from the pictorial stone scripture of the Stone Age period until the most sophisticated storytelling tools of this current era. That is why the literature-wise study of narratives and their structure was vastly explored.

Narratology (Genette, 1983; Onega and Landa, 2014) is, etymologically, the science of narratives as defined by the structuralists. The concept of narratology has been evolving through the years and now studies the narrative aspects of literary and non literary genres, such as poems, films, drama, history or advertisement. Narratives come in different mediums, linguistic narrative (history, novel, short stories) or audiovisual narratives (films, TV series, TV shows). Each medium allows for a specific presentation of the stories, different points of view, various degrees of narratorial intrusiveness and different time-handling techniques. Consequently, each narrative medium requires an analytical approach to narrative structures.

Narrative structure can be used for the reorganization of a huge collection of multimedia contents (Berhe *et al.*, 2020). Hence, automatic methods should take advantage of the narrative structure found in multimedia contents for better accessibility, management and understanding of these collections. On the other hand, automatic analysis, understanding and extraction of narrative structure may benefit writers of short skits, comedies or folktales to structure their writings and capture their audience. Additionally, it is an important asset for producing long and continuous episodes without contradictions or continuity problems for sequel film makers and TV show producers.

Recommendation systems, more specifically entertainment recommendation systems, may also benefit from the extraction of narrative structures with regard to recommending contents according to narrative structures rather than just meta-data and content similarity. Hence, the giant video content platforms, such as YouTube, Netflix, etc., may use narrative structure for better retrieval and searching. Furthermore, the gaming industry can apply narrative structures on programming different steps and situations that are connected narrative-wise, for example in video games (Vargas, 2017). The education system may also utilize narrative structures to present contents of different courses so that it can be entertaining and engaging for children and teachers at different levels of studies. Finally, social media posts and tweets are also used to identify narratives and understand events that happened and are going to happen in the future (Brogan, 2015; Sadler, 2018).

Computational narrative is the research domain that involves Artificial Intelligence (AI), Machine Learning (ML) or Natural Language Processing (NLP) approaches for automatic extraction or analysis of narratives. The representation, analysis, extraction and manipulation of narrative structure according to the existing narrative theories and narrative structures have been studied. The advancement of NLP research and results in many areas has inspired researchers to continue working on narrative structure understanding using automatic methods. In order to elaborate and evaluate the proposed algorithms, annotations and visualization tools have a vital role. Narratives have been annotated manually and using annotation tools (Finlayson, 2013). Visualization tools made annotation and understanding of narratives less difficult considering automatic methods.

In this survey, we investigated published articles and books that focus on the automatic understanding of narrative and their structures. Besides, available databases and annotations that focus on automatic approaches are described. The papers presented in this work focus on narrative content extraction, analysis and understanding. To our knowledge, this survey is one of its kind. There are surveys (Finlayson, 2013) on short narrative texts, such as folktales, and their focus was on morphologies and linguistic characteristics. However, our survey covers all modalities (textual, audio and visual) of narrative content, and the investigation starts from the narrative theories and gradually develops into the state-of-the-art ML and NLP algorithms used for the understanding of narratives and their structures. Therefore, the survey discusses work from the basic narrative theories until the development of automatic methods for the understanding of narratives and their structure. Hence, we believe this work can pave the way for future studies in automatic processing of narrative contents and facilitate the road with already available methods and resources, and included missing points that can motivate the advancement of research towards narrative structure understanding, extraction, analysis from a machine learning and natural language processing point of view.

The paper is organized in the following manner. Section 2 briefly discusses the history of narrative theories and the evolution of narrative structures' definition. Section 3 presents features utilized to automatically understand narratives and their structures, based on different modalities. Section 4 describes the available annotated datasets and visualization tools concerning narratives and their structure. Section 5 details algorithms and computational models applied on narrative structure. Finally, Section 6 concludes the paper and recommends possible future research on narrative structure.

2. Short history of narratives

Narratology has been dominated by structuralist approaches since the 1990s, and has been developed into a variety of theories, concepts, and analytical procedures. The term "narratology" was introduced in the structuralist study of narratives by Tzvetan Todorov in 1969. Schmid (2010) believed that narrativity can be identified by two distinct concepts. The first one is the classical narrative theory, long before the term "narratology" was first used, and the second one is the structuralist concept of narrative. Genette (1988) developed a theory of narratological poetics that may be used to address the entire creation of narrative processes in use. Structuralism was further shaped by Lévi-Strauss (1958) who claimed that myths found in various cultures can be interpreted in terms of their repetitive structure, which leads to the study and formulation of narrative structures.

In Aristotle's approach [1], a narrative is classified into three main parts which are the beginning, the middle, and the end (Lucas, 1968; Whalley *et al.*, 1997). The beginning is where the characters and main settings are introduced. In the middle, the conflict starts and the protagonists get acquainted with the problem. At the end, the problem is solved and the life of the protagonists goes back to normal. Many narrative content writers follow Aristotle's three-stage structure in different mediums of narrative, including Hollywood movies (Field, 2009).

Propp (1958) focused on repeated plot elements in his studies of the morphology of folktales, which he called "functions", and their associated character roles. He defined function as "an act of a character, defined from the point of view of its significance for the course of the action". Each function involves a set of characters who filled certain roles, the *dramatis personae* of the morphology. He identified seven *dramatis personae* classes: Hero, Villain, Princess, Dispatcher, Donor, Helper and False Hero. Propp (1958) identified 31 elements of stories that can be categorized into four spheres, namely the introduction, the body of the story, the donor sequence (the sequence of actions by a provider to the hero) and the hero's return, in his studies. This categorization was first developed by Todorov and Weinstein (1969) who state that there are five steps that most narrative stories or plots follow. These are equilibrium (starting the story where the life of characters are normal), disruption (the life of a character or characters is disrupted), realization (characters are informed about the situation and chaos occurs), restored order (characters resolve the disruption) and equilibrium again (equilibrium is restored, new equilibrium).

1. Aristotle's *Poetics*, 347-342 B.C., is a little collection of lecture notes, yet for many centuries it served as the foundation of narrative theory.

Novelist Freytag (1872) developed a narrative pyramid as a description of the narrative structure in fictions. In dramatic narratives, he proposed a dramatic structure containing five parts (Exposition, Rising Action, Climax, Falling Action and Denouement), which were also shared by Todorov. Many films and dramatic fictions use Freytag's pyramid of dramatic sentiments to present narratives of any kind.

Comparatively, many narrative theories and structures share at least one common point which originates from Aristotle's theory. But Todorov and Propp shared most of the steps in their narrative theory and structure, even if they differed on the story and the content of the narrative. They also agreed with Aristotle's structure in a more general way. The beginning step in Aristotle is equivalent to introduction and equilibrium in Propp's and Todorov's narrative theory, respectively. The middle step in Aristotle is equivalent to the body of the story, the donor sequence step, in Propp's theory, and disruption and realization steps, in Todorov's. Finally, the end stage of Aristotle's theory is equivalent to the hero's return and new equilibrium steps, in Propp and Todorov respectively.

In his narrative approach, Lévi-Strauss found out, through his studies of hundreds of myths, that we, as human beings, make sense of the world or the people, or events, as binary opposites (Lévi-Strauss, 1958). He indicated that binary opposites are the center of narratives, so that narratives are organized around the conflict between such opposites (e.g. good vs evil, man vs woman, peace vs war, wisdom vs ignorance, etc.).

Some researchers took advantage of the contents of the narratives and tried to find structure within the contents. For example, Cohn (2013) applied narrative grammar to verbal discourse and movies, claiming that "the narrative structure orders information into a particular pacing, from which a reader can extract a sequence's meaning—both the objects that appear across panels and the events they engage in". Bordwell (2013), Berger (1997), and Chatman (1980), on their side, have proposed to split the narratives depending on the structure and the contents. Bordwell (2013) and Chatman (1980) divided the narrative into *histoire* and *discours*, which literally mean "story" and "discourse", or plot, respectively. "Story" is the content of the narrative. It can also be described as the raw material of dramatic actions which is made up of events, characters, entities, etc. Plot (sometimes called "discourse") is the way a story is presented. Berger (1997) divided the narrative into "fabula" and "syuzhet" which are equivalent to story and plot respectively. Van Dijk (1981) considered episodes to be semantic units of discourse and characterized an episode of a discourse as a specific "sequence of propositions". Furthermore, he specified such a sequence must be coherent according to conditions of textual coherence.

Most of the above narrative theories were established from text books as stories or fairy tales. Research has been done based on these structures and morphologies in the literature domain. Many modern writers used Campbell's theory (Campbell, 2008) of mythological structure of the journey of a hero. Screen and story writers—for movies, theatres or TV series for example—have also different ways of writing the narrative that goes on through the media pieces by pieces. The most common techniques followed by film makers (mainly in Hollywood) are three-act (III-act) and five-act (V-act) structures (Field, 2009) formed by decomposing the concept of Lucas (1958), Todorov and Weinstein (1969), and Freytag (1872). An act, as defined by McKee (1997), is a "series of sequences that peaks in a climatic scene which causes a major reversal of values, more powerful in its impact than any previous sequence or scene". All the theories and approaches on narrative structure, for both textual and audiovisual contents, can be summarized in Figure 1. Hauge (2011) proposed a structure with six stages known as "Michael Hauge's Six Stage Plot Structure" that is based on the three-act (III-act) narrative structure.

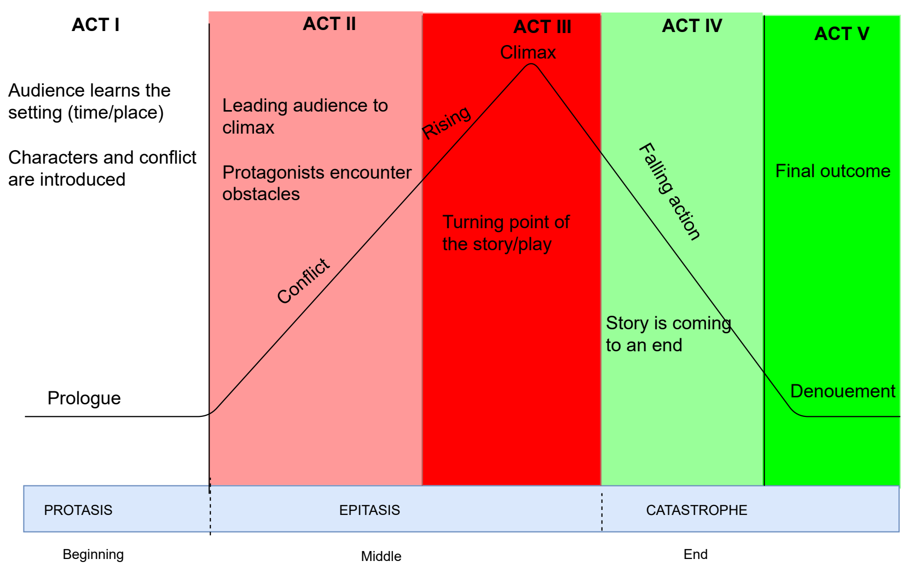

Figure 1. *Narrative structure summary adapted from Freytag's (1872) pyramid.*

Figure 1 divides a narrative into five parts, the V-act structure, but also embeds the three-part structure (III-act). Act I and II in the V-act structure are equivalent to act I of the III-act structure. Act III in the V-act structure is act II in the III-act structure. Act IV and V, in the V-act structure, are equivalent to the act III of III-act structure. The III-act structure (Whalley *et al.*, 1997) is mostly a replica of Aristotle's narrative theory and the V-act structure is coined from Freytag's pyramid (Freytag, 1872).

3. Modalities for analysis and extraction of narrative (structures)

Since the 1970s, narratives and storytelling have been investigated using scientific methods in the field of artificial intelligence for understanding and evaluating human cognition theories (Vargas, 2017; Finlayson, 2012; Andersen and Slator, 1990). The field of computational narrative links the daily human activities (narratives) and the computing world (machines computations) by analyzing and modeling narratives, narrative understanding and machine-readable representations of narratives with the purpose of enabling computers to tell a story.

The understanding and extraction of narrative structure is a difficult problem for computers due to the complex nature of the different inputs that constitute a narrative and different ways of constructing a narrative structure. Most of the research that focus on the extraction of narratives and their structures use simple stories or folktales in the form of text input. Recent work, however, also focuses on feature films, video shots and long and progressive videos (i.e. TV series). In the context of multimedia collections, narratives can also be carried by several modalities. Researchers have been dealing with different modalities, used jointly or independently, to extract narratives and their structure.

In the following subsections, textual and audiovisual features are presented. Subsection 3.1 presents the works that dealt with the textual modalities of narratives and the features used and Subsection 3.2 introduces audiovisual modalities used for automatic computation of narrative contents.

3.1. *Textual modality*

Most of the research on the extraction of narratives and their structures used simple stories or folktales in the form of text input, based on Vladimir Propp's influential theory of the structure of folktale plots. Finlayson (2012), Valls-Vargas *et al.* (2014), Bod *et al.* (2018) and Malec (2010) focused their study on narrative discourse of Propp's folktales, and Elson (2012) concentrated his work on a set of fables as well as longer texts including literary fiction and epic poetry.

Finlayson (2012), in his work on Russian folktales, uses semantic level descriptions, known as "event timeline", and abstracts them to the next higher level, structures, such as Villainy, Struggle, Victory, and Reward. Event timeline corresponds to the order of the events as expressed in the text, which is different from their temporal order within the story world. It includes straightforward things that happen ("He fought the dragon."), times and dates ("He fought the dragon on Tuesday."), aspectual constructions ("He began to fight the dragon.") and subordinate expressions ("He promised he would fight the dragon, but never did."). The author believed that each of these types of occurrences has a particular meaning in the timeline of the story world. He used specifically Propp's morphology and proposed an Analogical Story Merging (ASM) algorithm that extract plot patterns. The algorithm is discussed in Section 5 in detail.

Malec (2010) uses a semantic markup language, PftML, to automatically analyze and parse folk narratives. The language proposed by Malec (2001) is similar to an XML grammar and enables to decompose tales into Propp's folktales functions.

Lô *et al.* (2020) apply machine learning and natural language processing approaches to identify, analyze and generate West African and West European folktales by using semantic and syntactic coherence of the narratives. The authors use explicit West African features such as culture-specific protagonists, other characters or objects to appear in the generated texts. They utilize word embeddings and term frequency-inverse document frequency (TF-IDF) to build a character-level and a word-level recurrent neural model. Additionally, they search for recurring patterns that could indicate the existence of a distinctive narrative structure between West African and Western European folktales. To do so, they employ linguistic markers to divide the folktales following the III-act structure and classify the obtained segments.

Concerning literary fictions, in order to model the narrative structure of fables and literary fictions, Elson (2012) extracts textual features vectors based on word distance, type of punctuation and character mention and constructs conversational networks that are then analyzed with regard to literary theories.

Valls-Vargas *et al.* (2014) extract entities from a narrative text and compute a feature vector using both linguistic information and external knowledge (WordNet and ConceptNet) for each extracted entity. The features are computed from the parse tree of a sentence where an entity was found, the sub-tree representing the entity, the leaves of the sub-tree (i.e., word-level tokens with POS tags) and the dependency lists that contain a reference to any node in the entity's sub-tree. The features are automatically extracted thanks to *Voz*[2], a system that explores techniques for automatic extraction of narrative information from text.

Film scripts are also among the studied sources of textual modality for narratives. A lot of them, including several genres, are available and openly accessible in different websites and fan pages, e.g. IMSDb.[3] Murtagh *et al.* (2009) study the narrative structure of the *Casablanca* film script and of six episodes of *CSI* (*Crime Scene Investigation*) TV series. They show that film scripts are important to analyze and understand narrative structures based on the theories of McKee (1997) on principles of screenwriting. To do so, they take all words into account in semi-structured texts that were extracted from film scripts.

Finally, Kim and Monroy-Hernandez (2016) and Barbieri (2007) use the theory of narratives to do task-specific application which are based on textual information. Kim and Monroy-Hernandez (2016) extract social media content based on narrative theory. For that, they annotate the beginning, middle and end of an event according to a narrative structure and then automatically cluster the text into parts based on the sentences of the event. Similarly, in social-media narrative content, Tekiroğlu *et al.* worked on generating counter-narrative (CN) against online hate speech (HS). CN is a non-aggressive response that offers feedback through fact-bound arguments to withstand HS. To this end, Tekiroğlu *et al.* (2020) build a dataset based on quality decomposed into two terms, conformity and diversity (lexical and semantic).

2. https://sites.google.com/site/josepvalls/home/voz.
3. The Internet Movie Script Database, www.imsdb.com.

3.2. *Audiovisual modalities*

The video and audio of a film or a TV series (or any type of multimedia document) are essential and core information narrators. Little research is conducted on audiovisual data for the analysis, extraction, understanding and description of narrative structure from TV series or other multimedia collection.

Nonetheless, some studies focus on the computation of narrative structure using videos of Hollywood movies. Zhao and Ge (2010), for example, work on the calculation of a structure-model for Hollywood movies, applying film-making rules and film grammars to their data. Other works investigate narrative structure in educational videos in order to help and motivate students (Dorai *et al.*, 2003; Phung *et al.*, 2002). Dorai *et al.* (2003) built a hierarchical structure that decomposes the video into sections based on the contents presented to the students. Their method exploits variations in color distributions in frames to separate sections containing some form of text from narrator content. In their study, Phung *et al.* (2002) use three narrative structure parts (narration, conversation/discussion and linkage sections) in the domain of educational videos.

TV series or sequential TV shows can be treated as a large collection of meaningful episodes, scenes or events. They can also be seen as a collection of short pieces of logical narrative units ordered chronologically or casually. The narratives in this kind of large collection can vary in type and length, but this collection has at least one main narrative that goes on from the beginning to the end of the series that can be dynamically captured by creating links between the smallest logical story unit, mostly known as a scene (Bost, 2016; Tapaswi *et al.*, 2014; McKee, 1997; Zhao and Ge, 2010). Narrative units are the smallest portions of a narrative content or story (Berhe *et al.*, 2019; Ercolessi *et al.*, 2011).

As mentioned in Section 2, narrative structure is made up of two components, the plot and the story. Story refers to the raw material of dramatic action and answers the story questions like who, what and where. It also corresponds to the description of settings, characters and events. Plot refers to how the story is told. It is a sequence of events that drives the story forward from beginning to end. It is more concerned with the characters and their interaction and answers the questions like how and when actions/events have occurred. Some researchers (Bost *et al.*, 2016; Tapaswi *et al.*, 2014; Ercolessi *et al.*, 2012; Berhe *et al.*, 2020) rely on such narrative elements (characters interactions, events, entities) to perform narrative structure extraction on TV programs.

Entities are important element of narratives. They refer to the mentions of places, names and organizations. Piskorski *et al.* (2020) use entities to extract events where the target entities have participated or been mentioned for a structured news data.

Events can be used to connect narratives and form a structure of the narrative. They can come in different granularity, for example in a scene, in an episode (mostly in books, films, TV series) or a phrase (mostly in folktales). Chambers and Jurafsky

(2009) propose to learn narrative schema, coherent sequences or sets of events using unsupervised techniques. To do so, they extract chains of events to obtain narratives from a document (Chambers and Jurafsky, 2009; Chambers and Jurafsky, 2008). Similarly, Regneri *et al.* (2010) and Finlayson (2012) propose to learn event scripts from lists of actions using multiple sequence alignment techniques.

Narrative structure extraction can rely on different kind of features coming from the various modalities of the document under consideration: linguistic or syntactic features, audio information (pitch, rythm) or visual clues (shots, movement). In the next section, annotations, available dataset and visualization tools for the different narrative contents are discussed.

4. Annotations and visualization tools for narrative structure

In modeling techniques for narratives understanding, annotation is a necessary step to represent the story from a text to machine-readable format and evaluate the automatic extraction methods developed.

Table 1 presents datasets specifically used for narrative content analysis, extraction and understanding and discussed in this survey. It describes the name of the datasets (some of them do not have a name), the annotation type, size (number of narrative units in the dataset), number of words (number of scenes for video datasets), modality, language and the availability of the dataset (where "available" refers to the online availability of the dataset and "NA" refers to "not available"). Furthermore, some research works have also presented a summary of datasets for narrative contents (see, for example, Doukhan *et al.* (2015)).

Different annotation schemas and environments have been proposed. Some of the most famous and reliable annotation environments are the Story Workbench by (Finlayson, 2008) (2008) and the Scheherazade system by Elson (Elson, 2012). Both dealt with the annotation of folktales and short narrative texts. Finlayson worked on 15 Russian folktales, a subset of Propp's original tales, annotated for 18 aspects of meaning by 12 annotators using the Story Workbench, a general text-annotation tool developed for this work (Finlayson, 2012). Each aspect was annotated by two annotators with an inter-annotator agreement between 0.7 and 0.8. Finally, Sloetjes and Wittenburg (2008) provide a multimedia annotation tool, ELAN, which makes it possible to annotate multiple category of annotations on the same multimedia documents.

The annotation of narrative documents, particularly multimedia narratives, is a very time-consuming task. Li *et al.* (2018) and Eisenberg and Finlayson (2019) have worked on the annotation of narrative elements of short stories in two different ways. Li *et al.* (2018) produce a guideline to annotate directly the narrative structure based on Freytag's (1872) pyramid, and Eisenberg and Finlayson (2019) provide a guideline for narrative characteristics annotation to collect human judgments on narrative characteristics.

Dataset	Annotation	Size	# of words or scenes (*)	Modality	Lang.	Access
GV-LEx (Doukhan *et al.*, 2015)	Lexical and structural	101	66,935	text/speech	Fr	online
Propp's folktales (Bod *et al.*, 2018)	Lexical and structural	450	-	text	En	NA
(Lô *et al.*, 2020)	West African & Western European tales	742	406,403	text	En	online
French tales (Garcia-Fernandez *et al.*, 2014)	Narrative classification	107	85,600	text	Fr	online
Russian folktales	Linguistic	15	18,862	text	En	online
Movies (Guha *et al.*, 2015)	ACT boundary	9	27*	video	En	NA
ScriptBase (Gorinski and Lapata, 2015)	Summaries, loglines & taglines	1,276	-	text	En	NA
NarrativeQA (Kočiskỳ *et al.*, 2018)	Question answering	1,572	15,406	text	En	online
FSD (Liu *et al.*, n.d.)	Stories for each scene	60	1,569*	video	En	online
TRI-POD (Papalampidi *et al.*, 2019)	Turning points of screenplay	99	54,600	text	En	online
CSI (Fremann *et al.*, 2018)	Entities	39	>500,000	text/video	En	online
Casablanca (Murtagh *et al.*, 2009)	Structural	1	77*	text	En	NA

Table 1. *Summaries of available datasets regarding narrative content. "*" shows the number of scenes available in a dataset when the number of words are not present and the dataset is described only by scenes.*

Garcia-Fernandez *et al.* (2014) propose the digitization and annotation of a tales corpus from a narrative point of view (only the French tales corpus is available) and classify it according to the Aarne & Thompson (1961) narrative classification of folktales. Doukhan *et al.* (2015) provided GV-LEx, a corpus of French folktales annotated using textual and audio modalities, during their study of the relationships between the textual structures of tales and speech prosody. They annotated 89 text and 12 speech corpora with the targeted application of an expressive text-to-speech synthesis system embedded in a humanoid robot. They performed lexical level (extended definitions of enumerations, time, place and person named entities, and part of speech (PoS) tags) and supra-lexical level (the segmentation of tales into a sequence of episodes, the localization and attribution of direct quotations, together with tale protagonists co-references) annotations. Lô *et al.* (2020) provide two corpora of West African and Western European folktales that are used in three experiments on cross-cultural folktales analysis. They collected a total of 742 English narratives, 252 of which were West African, and the other 490 Western European, to predict the next words that continues the narrative based on an input seed from the narratives in the corpora. The West African folktales were written by authors from Anglophone West African countries such as Gambia, Ghana, and Nigeria, while the Western European folktales were from countries such as Netherlands, Germany, France, and the UK (Lô *et al.*, 2020). Finlayson (2013) presents a survey of corpora for the advancement of scientific understanding of narrative. He identifies 167 unique text collections (155 with some sort of annotation) that could be considered a “corpus” and contained 17 different broad types of narratives, 5 different modalities, and approximately 42 different types of annotations. From the annotations presented, the most common and complex are events, named entities, and roles annotations.

Concerning the annotation of multimedia documents, Ercolessi *et al.* (2011), Bost (2016) and Liu *et al.* (2020) have annotated several seasons of TV series with scene boundaries. Bost annotated 5 seasons of *Game of Thrones*, 2 seasons of *Breaking Bad* and 1 season of *House of Cards*. Ercolessi *et al.* annotated *Buffy The Vampire Slayer* and *Mac and Alice*. Liu *et al.* collected 60 episodes of *The Flintstones* TV series (which are composed of 1,569 scenes) and annotated the dataset into story. To this end, 105 undergraduate engineering students in data science were invited to annotate the scene labels and each student annotated 4 episodes. They have provided that dataset as Flintstones Scene Dataset (*FSD*).[4] Liu *et al.* constructed the dataset on the assumption of the "three-act" structure (see Figure 1). Furthermore, Tapaswi *et al.* (2014) annotated face tracks, shots and scene boundaries, script-to-video alignment in *The Big Bang Theory* TV series, and some story-line in *Game of Thrones*. Papalampidi *et al.* (2019) developed the TuRnIng POint Dataset (TRI-POD)[5] composed of 99 annotated screenplays. Their work focused on identifying turning points of screen plays based on textual information. Similarly, Fremann *et al.* (2018) built a dataset of

4. Flintstones Scene Dataset (FSD) is available at https://github.com/llafcode/The_FSD_dataset.git.
5. https://github.com/ppapalampidi/TRIPOD.

episodes of *Crime Scene Investigation* TV series [6] for natural language understanding. The dataset is composed of 39 episodes (seasons 1-5) with screenplays and entities annotations (perpetrator/s in a crime scene). For these annotations, they hired three post-graduate annotators that were not regular fans of the TV series.

Movies have also been used as main sources of audiovisual and linguistic analysis. Guha *et al.* (2015) annotated 9 movies according to the III-act narrative structure. They used film experts to annotate 2 act boundaries as they believe that accurate detection of act boundaries requires knowledge of screenwriting and narrative structure. The dataset ScriptBase (Gorinski and Lapata, 2015) compiles a collection of 1,276 movie scripts (from 1909 to 2013) divided into 23 genres; each movie is on average accompanied by 3 user summaries, 3 loglines (one-sentence summary of a movie), and 3 taglines (short snippets used to promote a movie). Kočiskỳ *et al.* (2018) developed a dataset, NarrativeQA, of stories on books (collected from project Gutenberg [7]) and movie scripts based on question answering using summaries. NarrativeQA is composed of 1,572 stories, evenly split between books and scripts, and 46,765 question-answer pairs. Lewis *et al.* (2017) collected a large scale dataset of 10,945 subtitles files associated with movies metadata that were pre-processed so subtitles contain only linguistic information. Finally, the last movie dataset was proposed by Murtagh *et al.* (2009). It is composed of the *Casablanca* film script divided into 77 successive scenes. The source text for the 77 scenes, containing in total 6,710 words, including metadata, varies between 5 and 1,017 words.

In order to visualize important information or evaluate automatic systems, visualization and annotation tools were proposed: StoryFlow, StoryGraph, Yarn, MovieGraph and StoryCake. StoryFlow (Liu *et al.*, 2013), StoryGraph (Tapaswi *et al.*, 2014) and Yarn (Padia *et al.*, n.d.) were developed to visualize a succession of events in a narrative using merging and diverging timelines, with the temporal continuity of these events in mind and less concern about the exactitude of their temporal placement. MovieGraph (Vicol *et al.*, 2018) proposed a graph-based visualization of a video clip for the annotation and visualization of social situations in a movie clip. Kim *et al.* (2017) developed a visualization technique to explore non-linear narratives in movies. They introduced Story Explorer, an interactive tool that visualizes narrative patterns of a movie via portraying events of a story in chronological order. Story Explorer displays a story curve together with information such as characters and settings. Finally, StoryCake (Qiang *et al.*, 2017) proposed a hierarchical plot visualization method according to the story elements and the hierarchical relationships of entities. Table 2 describes annotation and visualization tools used in narrative contents.

The available datasets and their annotations discussed above are used in different tasks of automatic narrative content extraction, analysis and understanding. In the next section, the algorithms designed to obtain the narrative structure and better processing of narratives from various types of narrative contents are described.

---

6. CSI dataset is available at https://github.com/EdinburghNLP/csi-corpus.
7. http://www.gutenberg.org/.

Tool	Purpose	Source	Description
StoryFlow	Visualization	(Liu *et al.*, 2013)	Visualization of sequences of events
StoryGraph	Visualization	(Tapaswi *et al.*, 2014)	Visualization of sequences of events
Yarn	Visualization	(Padia *et al.*, n.d.)	Visualization of sequences of events
MovieGraph	Visualization & Annotation	(Vicol *et al.*, 2018)	Visualization and annotation of social situations in a movie clip
Story Explorer	Visualization	(Kim *et al.*, 2017)	Interactive display of a story curve with different narrative elements information
StoryCake	Visualization	(Qiang *et al.*, 2017)	Plot visualization of a story using entity relationships
Story-Workbench	Annotation	(Finlayson, 2008)	Annotation of folktales and short narrative texts for many linguistic aspects of narratives
Scheherazade	Annotation	(Elson, 2012)	Annotation of folktales and short narrative texts for many linguistic aspects of narratives
ELAN	Annotation	(Sloetjes and Wittenburg, 2008)	Annotation of multiple categories on the same multimedia document

Table 2. *Summary of available annotation and visualization tools used in narrative documents.*

Finally, few research papers have used recent natural language processing (NLP) methods, such as attention mechanisms, for the understanding of narrative content. For example, Lô *et al.* (2020) train a bag of words (BoW), an LSTM (long-short term memory neural network), and other classifiers in order to identify, analyze, and generate West African folktales for better extraction of the organization of folktales according to the III-Act narrative structure.

5. Methods of automatic narrative structure extraction

Natural language Processing (NLP) has been vastly investigated for the understanding and extraction of important information from textual documents. Understanding and analysis of narratives and their structure has been dealt according to multiple tasks using different machine learning and natural language processing approaches.

Table 3 summarizes the main tasks and describes the algorithms, type of narrative content and modalities used in the papers. The following subsections discuss the main

Task	Algorithm	Narrative	Modality
Representation (2)	Co-occurrence, bootstrapping	Fables	Text
Semantic analysis (3)	Correspondence analysis		Text
Content Extraction (7)	PMI, CRF, decision-tree, NER, graph analysis	Movies, speech, folktales	Text, Speech, Video
Event chains (4)	PMI, NER	Folktales	Text
Morphology (Propp's) (2)	ASM	Folktales	Text
Decomposition & linking (10)	Community clustering, recursive algorithms	Movies & TV shows	Video
Hyperlinking (10)	NER, classification	Social media, folktales, movies	Text, Video

Table 3. *Summary of tasks on narratives and algorithms used. Numbers in the parenthesis, in the task column, refers to the number of articles included in this work that use the task on narrative contents.*

methods (some of the methods overlap with each other) used to understand and extract narratives and their structure.

5.1. *Key narrative elements extraction*

Narrative elements of narrative contents constitute an important information. One way of using narrative elements for a better understanding of narratives is the creation of a character network, i.e. a graph that illustrates the interaction of the characters, as suggested in the survey of fictional character networks (Labatut and Bost, 2019). The extraction of social networks (Agarwal and Rambow, 2010; Elson, 2012; Bost *et al.*, 2016) explaining the connection of characters to understand the story between them showed promising results. Valls-Vargas *et al.* (2017) build a graph which captures the narrative entities, such as characters, organization and places, and in turn depict the story between the entities, referred to as story graph. Similarly, Bost *et al.* (2016) take advantage of the plot properties of narratives in TV series to construct a character network and represent the dynamics of the characters.

Content extraction techniques have been used to extract important features of events, stories or entities (Chen *et al.*, 2015; Tapaswi *et al.*, 2015; Yu *et al.*, 2016; Arnulphy *et al.*, 2015; Ghannay *et al.*, 2018). As an example, Tapaswi *et al.* (2015) work on the alignment of plot synopsis to video to provide story-based retrieval from videos. To do so, they consider shots and sentences as atomic units and extract named entities from the text and person identification from the video to create alignments between synopsis and sets of shots (scenes). Arnulphy *et al.* (2015) work on event extraction from textual documents in the TimeML [8] challenges for the French and English languages. They used event descriptors to assign every word to a label that indicates whether it is an event or not by using conditional random field (CRF) and decision-tree based algorithms. Ghannay *et al.* (2018) use an end-to-end entity recognition (NER) approach for a slot filling task which is a semantic concept extraction in speech, in the framework of a human/machine spoken dialog dedicated to hotel booking.

Chambers and Jurafsky (2009), Vargas (2017) and Finlayson (2012) describe in their research works the importance of event chains, that is to say, connections of events found in a narrative. Valls-Vargas *et al.* (2014) extract entities from events and compute a feature-vector for each of them using linguistic information and external knowledge. They propose the *Continuous Jaccard* measure to estimate the similarity between the extracted entities to decide if they represent characters or not. Chambers and Jurafsky (2008) extract narrative event chains from raw texts through a three-step process. First, they learn basic information about the narrative chain (the protagonists and constituent sub-events). Then, they use the Pointwise Mutual Information (PMI) measure to compute how often events share grammatical arguments. Finally, they build a global narrative score such that all events in the chain provide feedback on the event. In other words, they find the next most likely event to occur, given all narrative events in a document, by maximizing the PMI score. Chambers and Jurafsky (2009) learn narrative schemas, coherent sequences or sets of events (e.g. arrested (Police, Suspect), convicted (Judge, Suspect)) using unsupervised techniques based on co-referencing arguments in chains of verbs. Finally, Reagan *et al.* (2016) use plot sequences or event sequences to construct the story arcs of English fiction books.

Finlayson (2012) introduce the Analogical Story Merging (ASM) algorithm, based on Bayesian model merging, a machine learning technique for learning regular grammars, in order to learn Propp's morphology. To do so, he takes descriptions at the semantic level and abstracts them to a higher level, i.e. structures such as Villainy, Struggle, Victory and Reward. Model merging used to derive a regular grammar is the foundation for ASM with two key differences: filtering and analogical mapping. The filtering process constructs another model from the final merged model from which all states that do no meet certain criteria are removed. The states that survive this cutting become the alphabet, or for Propp's morphology, the functions. In the work of Finlayson (2012), it is shown that the ASM algorithm learns a big part of Propp's theory of folktales structure.

5.2. *Decomposition*

Another approach to extract narrative (structure) consists in the decomposition of the stories contained in the narrative into story-lines and the re-connection of the decomposed stories. For example, Park *et al.* (2012) work on detecting some story lines,

8. ISO-TimeML is an International Standard for time and event markup, and annotation.

organized around characters, from narratives of a movie. They propose a Character-net that can represent the relationships between characters using dialogues, and a method that can extract the sequences via clustering communities of characters based on heuristic algorithm. Many research works (Li *et al.*, 2001; Zhao and Ge, 2010; Adams *et al.*, 2002; Guha *et al.*, 2015) also decompose movies into acts using computational methods for better understanding of the narrative act boundaries and the semantics of a narrative in movies. Adams *et al.* (2002) study film grammar and decomposition of a movie with the goal of automatically locating dramatic events and section boundaries. In their work, they are able to reconstruct the dramatic development of films, focusing on the filmmakers' point of view. Film grammars or Hollywood film-making strategies can work on full movies and standalone episodes of TV series. They use the attributes of motion and shot length to define and compute a measure of the tempo of a movie. They applied Deriche's recursive filtering algorithm to detect the edges of a movie to locate significant tempo pace changes. Similarly, Guha *et al.* (2015) deconstruct a movie into a III-act structure (act I (exposition), act II (conflict) and act III (resolution)) thanks to a popular movie grammar, followed by most screen writers. They detect the act boundaries based on the knowledge of film grammar and features from three modalities (text, video, music). They address the problem of automatically detecting the III-act narrative structure in movies in an unsupervised manner. They cast the problem as one dimensional edge detection in the story intensity curve P and used probability distribution to find the act boundaries. Finally, Lee *et al.* (2021) decompose narrative multimedia plots into story-lines based on the estimation of the personality of the characters. They estimate a character's personality from the average length of dialogues and the ratio of out-degree for in-degree in a graph of characters relationships.

5.3. *Hyperlinking*

When the narrative document (movie, fiction, TV series or TV shows) is quite large, it becomes very complicated to extract its narrative. Character interactions and stories that flow through, from the beginning till the end, are intertwined. Therefore, in order to better understand the narratives and their structure from a large collection, documents need to be reorganized in a more sensible way and in smaller narrative units such as scenes. This can be done by creating links between the narrative units according to a narrative point of view. Many researchers try to link multimedia documents, coined by the term multimedia hyperlinking (Bois *et al.*, 2017a; Bois *et al.*, 2017c; Budnik *et al.*, 2018; Chaturvedi *et al.*, 2018).

Multimedia hyperlinking is a way to navigate videos inside a collection by jumping from one video to another, using different techniques. Bois *et al.* (2015); Ordelman *et al.* (2015); Kim and Monroy-Hernandez (2016), etc. design some linking categories or typologies for multimedia hyperlinking and build graphs to easily explore news by following links that lead to similar news. Kim and Monroy-Hernandez (2016) use narrative theory as a framework to identify the links between social media contents. They first identify and fill narrative gaps in a social media record, then they link content to narrative categories with respect to storytelling roles. Ordelman *et al.* (2015) present a video-hyperlinking method based on named entity identification. They investigate an unconstrained, multimodal perspective on the identification of anchor points, and a perspective based on the detection of entities from existing metadata and/or automatic audiovisual analysis.

Graph based hyperlinking methods have also shown promising results (Vicol *et al.*, 2018; Li *et al.*, 2017; Valls-Vargas *et al.*, 2017; Bois *et al.*, 2017b; Ercolessi *et al.*, 2012). Valls-Vargas *et al.* (2017) and Vicol *et al.* (2018) produce graphs of stories using narrative elements such as entities. To do so, Valls-Vargas *et al.* (2017) use co-referenced entities, entities and the role of the entity to build their story graph from short textual documents (Russian folktales in English). They extract entities and classify their roles and finally build the graph with the help of a common sense knowledge database. On their side, Vicol *et al.* (2018) propose a method for querying videos and text with graphs, and show that: first, their graphs contain rich and sufficient information to summarize and localize each scene and second, subgraphs allow them to describe situations at an abstract level and retrieve multiple semantically relevant situations. Their graphs capture people's interactions, emotions and motivations which must be inferred from a combination of visual cues and dialog. Vicol *et al.* (2018) use 8 different node types (such as characters, relationship, topics, etc.) while constructing the graph and prepare the Moviegraph dataset. Bois *et al.* (2017b) generate links between news documents surrounding a specific event. They propose a set of intuitive properties that a graph should exhibit to be explorable and used nearest neighbor approaches to create the graph.

Goyal *et al.* (2010) explore NLP techniques to automatically generate plot unit representation. To this end, they develop a tool that produces plot unit representations, known as AESOP, and use it to affect projection rules in order to connect situations with the respective characters. They identify affect states and map the affect states onto characters in a story using "projection rules". They use co-occurrence with Evil/Kind Agent patterns, and bootstrapping over conjunctions of verbs.

5.4. *Clustering*

Long and progressive multimedia contents, such as TV series, TV shows or documentaries have interesting intertwined narratives and they follow different narrative structure. In TV series, Ercolessi *et al.* (2012), Bost *et al.* (2016) and Chaturvedi *et al.* (2018) linked scenes using the concept of multimedia hyperlinking and used these links to tie different videos together and recreate one whole narrative. Chaturvedi *et al.* (2018) identify instances of similar narratives from a collection of narrative texts of movies. They found correspondences between narratives in terms of plot events and resemblances between characters and their social relationships. They coin the term story-kernel to quantify the correspondence similarity. Ercolessi *et al.* (2012) apply clustering methods for plot de-interlacing. They aim at grouping semantically related scenes into stories or sub-stories of a TV series episode (*Ally McBeal* and *Malcolm in The Middle*). They cluster scenes using traditional agglomerative clustering and graph based community detection algorithm, known as Louvain (Blondel *et al.*, 2008), to group scenes of the TV series.

Hierarchical clustering is one of the machine learning algorithms used to capture the semantics of narrative documents. For example, Murtagh *et al.* (2009) use hierarchical clustering through a sequence of agglomerations of successive scenes or temporal segments or intervals of successive scenes. They take into account the sequential nature of the scenes, ensured through the requirement that agglomerations must be adjacent. The scenes are compared through Correspondence Analysis using euclidean embeddings of film scripts. Murtagh *et al.* (2009) take all words into account in the semi-structured texts that composed film scripts. Each scene is cross-tabulated by the set of all words so that, in this cross-tabulation table, at the intersection of scene i and word j, the value corresponds to a presence (1) or absence (0) value.

6. Conclusion

Narrative structure is an important pattern found in narratives contents (text, audio, video) and AI, ML and NLP approaches have been used to extract, analyze and represent it. In this work, we have presented different kinds of narrative theories and structures, starting from Aristotle. Most of the structures can be summarized using the III-Act (Three stage) more general structure and the V-Act (Five stage) more specific one, see Figure 1. Folktales have been the main source of studies on narratives and their structures using computational methods. However, due to the fast and ever growing multimedia contents research have also been dealing with films, TV shows or TV series, recently. To this end, research relies on textual and audiovisual modalities as well as narrative elements (characters, entities or events) extracted from these modalities.

Machine learning and NLP algorithms, such as clustering, correspondence analysis and deep learning are used for the analysis and understanding of narrative and their structures in different narrative content modalities (textual and audiovisual). Recently, pre-trained deep learning models (including language models) are applied to represent and extract narratives for different purposes.

Annotation and visualization tools have been developed to alleviate the problem of the lack of annotated corpora for narrative structure analysis. Visualization tools can also provide easier overviews of narratives for a better understanding and representation of very long and complicated narratives. However, there is no standard dataset for narrative structure besides the Propp's for folktales (Finlayson, 2013). In order to accelerate the research on narrative structure extraction or analysis, the release of a (multimodal) corpora is, therefore, necessary. Annotation guidelines presented should be standardized so that every one could annotate any narrative content.

Concerning applications, the domains of audiovisual content analysis, summarization, event extraction and recommendation could benefit greatly from the extraction of narrative structure as a narrative structure could quickly and effectively represent a content. Therefore narrative elements could be used for the task of multimedia summarization, video recommendation and production of narrative contents (either multi-modal or mono-modal).

In our studies, only few research papers use deep learning approaches and the current advancement of NLP algorithms and pre-trained models, such as BERT (Devlin *et al.*, 2019) and GPT (Floridi and Chiriatti, 2020). A recent work on the detection of the most salient scenes within a set of semantically related scenes, Berhe *et al.* (2020), shows that time distributed LTSM produce promising results. We believe that current advancements of NLP may greatly help the future works on narrative content extraction, analysis and understanding. Furthermore, state-of-the-art neural network architectures, such as transformers, that have remarkable advancement on NLP, could take narrative content understanding to the next level.

7. References

Adams B., Dorai C., Venkatesh S., "Toward Automatic Extraction of Expressive Elements from Motion Pictures: Tempo", *IEEE Transactions on Multimedia*, vol. 4, n° 4, p. 472-481, 2002.

Agarwal A., Rambow O., "Automatic detection and classification of social events", *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, p. 1024-1034, 2010.

Andersen S., Slator B. M., "Requiem for a theory: the 'story grammar'story", *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 2, n° 3, p. 253-275, 1990.

Arnulphy B., Claveau V., Tannier X., Vilnat A., "Supervised Machine Learning Techniques to Detect TimeML Events in French and English", *Proceedings of the 2015 International Conference on Applications of Natural Language to Information Systems*, p. 19-32, 2015.

Barbieri M., Automatic summarization of narrative video, PhD thesis, Eindhoven University, 2007.

Berger A. A., *Narratives in Popular Culture, Media, and Everyday Life*, Sage, 1997.

Berhe A., Barras C., Guinaudeau C., "Video Scene Segmentation of TV Series Using Multimodal Neural Features", *Series-International Journal of TV Serial Narratives*, vol. 5, n° 1, p. 59-68, 2019.

Berhe A., Guinaudeau C., Barras C., "Scene Linking Annotation and Automatic Scene Characterization in TV Series.", *Proceedings of the 2020 Text2Story Workshop at European Conference on Information Retrieval*, p. 47-53, 2020.

Blondel V. D., Guillaume J.-L., Lambiotte R., Lefebvre E., "Fast unfolding of communities in large networks", *Journal of statistical mechanics: theory and experiment*, vol. 2008, n° 10, p. P10008, 2008.

Bod R., Fisseni B., Kurji A., Löwe B., "Objectivity and Reproducibility of Proppian Narrative Annotations", *Proceedings of the 3rd Workshop on Computational Models of Narrative*, p. 15-19, 2018.

Bois R., Gravier G., Jamet E., Morin E., Robert M., Sébillot P., "Linking multimedia content for efficient news browsing", *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, p. 301-307, 2017a.

Bois R., Gravier G., Jamet E., Robert M., Morin E., Sébillot P., "Language-based construction of explorable news graphs for journalists", *Proceedings of the Workshop on Natural Language Processing meets Journalism in Empirical Methods in Natural Language Processing*, p. 31-36, 2017b.

Bois R., Gravier G., Sébillot P., Morin E., "Vers une typologie de liens entre contenus journalistiques", *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles*, p. 515-521, 2015.

Bois R., Vukotić V., Simon A.-R., Sicre R., Raymond C., Sébillot P., Gravier G., "Exploiting multimodality in video hyperlinking to improve target diversity", *Proceedings of the International Conference on Multimedia Modeling*, p. 185-197, 2017c.

Bordwell D., *Narration in the fiction film*, Routledge, 2013.

Bost X., A Storytelling Machine?: Automatic Video Summarization: the Case of TV Series, PhD thesis, Université d'Avignon, 2016.

Bost X., Labatut V., Gueye S., Linares G., "Narrative smoothing: dynamic conversational network for the analysis of TV series plots", *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, p. 1111-1118, 2016.

Brogan M. K., "How Twitter Is Changing Narrative Storytelling: a Case Study of the Boston Marathon Bombings", *Elon Journal of Undergraduate Research in Communications*, vol. 6, n^{o} 1, p. 28-47, 2015.

Budnik M., Demirdelen M., Gravier G., "A study on multimodal video hyperlinking with visual aggregation", *Proceedings of the 2018 IEEE International Conference on Multimedia and Expo*, p. 1-6, 2018.

Campbell J., *The Hero with a Thousand Faces*, New World Library, 2008.

Chambers N., Jurafsky D., "Unsupervised learning of narrative event chains", *Proceedings of the 2008 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p. 789-797, 2008.

Chambers N., Jurafsky D., "Unsupervised learning of narrative schemas and their participants", *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, p. 602-610, 2009.

Chatman S. B., *Story and Discourse: Narrative Structure in Fiction and Film*, Cornell University, 1980.

Chaturvedi S., Srivastava S., Roth D., "Where have i heard this story before? identifying narrative similarity in movie remakes", *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p. 673-678, 2018.

Chen Y., Xu L., Liu K., Zeng D., Zhao J., "Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks", *Proceedings of the Annual Meeting of the Association for*

Computational Linguistics and the International Joint Conference on Natural Language Processing, p. 167-176, 2015.

Cohn N., “Visual narrative structure”, *Cognitive science*, vol. 37, nº 3, p. 413-452, 2013.

Devlin J., Chang M.-W., Lee K., Toutanova K., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, *Proceedings of the 2019 Meeting of the North American Chapter of the Association for Computational Linguistics*, p. 4171-4186, 2019.

Dorai C., Oria V., Neelavalli V., “Structuralizing Educational Videos Based on Presentation Content”, *Proceedings of the 2003 IEEE International Conference on Image Processing*, p. II-1029, 2003.

Doukhan D., Rosset S., Rilliard A., d’Alessandro C., Adda-Decker M., “The GV-LEx Corpus of Tales in French”, *Language Resources and Evaluation*, vol. 49, nº 3, p. 521-547, 2015.

Eisenberg J., Finlayson M. A., “Annotation Guideline No. 1: Cover Sheet for Narrative Boundaries Annotation Guide”, *Journal of Cultural Analytics*, vol. 4, nº 3, p. 11199, 2019.

Elson D. K., *Modeling narrative discourse*, Columbia University, 2012.

Ercolessi P., Bredin H., Sénac C., Joly P., “Segmenting TV series into scenes using speaker diarization”, *Proceedings of the Workshop on Image Analysis for Multimedia Interactive Services*, p. 13-15, 2011.

Ercolessi P., Sénac C., Bredin H., “Toward plot de-interlacing in tv series using scenes clustering”, *Proceedings of the 10th international workshop on content-based multimedia indexing*, p. 1-6, 2012.

Field S., *Selling a Screenplay: The Screenwriter’s Guide to Hollywood*, Delta, 2009.

Finlayson M. A., “Collecting Semantics in the Wild: The Story Workbench.”, *Proceedings of the AAAI Fall Symposium: Naturally-Inspired Artificial Intelligence*, p. 46-53, 2008.

Finlayson M. A., Learning Narrative Structure from Annotated Folktales, PhD thesis, Massachusetts Institute of Technology, 2012.

Finlayson M. A., “A Survey of Corpora in Computational and Cognitive Narrative Science”, *Sprache Und Datenverarbeitung*, vol. 37, nº 1–2, p. 113–141, 2013.

Floridi L., Chiriatti M., “GPT-3: Its Nature, Scope, Limits, and Consequences”, *Minds and Machines*, vol. 30, nº 4, p. 681-694, 2020.

Frermann L., Cohen S. B., Lapata M., “Whodunnit? Crime Drama as a Case for Natural Language Understanding”, *Transactions of the Association for Computational Linguistics*, vol. 6, nº 1, p. 1-15, 2018.

Freytag G., *Die Technik des Dramas*, Autorenhaus Verlag, 1872.

Garcia-Fernandez A., Ligozat A.-L., Vilnat A., “Construction and Annotation of a French Folkstale Corpus”, *Proceedings of the 2014 International Conference on Language Resources and Evaluation*, p. 2430-2435, 2014.

Genette G., *Narrative discourse: An essay in method*, vol. 3, Cornell University Press, 1983.

Genette G., *Narrative Discourse Revisited*, Cornell University Press, 1988.

Ghannay S., Caubrière A., Estève Y., Camelin N., Simonnet E., Laurent A., Morin E., “End-To-End Named Entity and Semantic Concept Extraction from Speech”, *Proceedings of the 2018 IEEE Spoken Language Technology Workshop*, p. 692-699, 2018.

Gorinski P., Lapata M., "Movie Script Summarization as Graph-Based Scene Extraction", *Proceedings of the Annual Meeting of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, p. 1066-1076, 2015.

Goyal A., Riloff E., Daumé III H., "Automatically Producing Plot Unit Representations for Narrative Text", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, p. 77-86, 2010.

Guha T., Kumar N., Narayanan S. S., Smith S. L., "Computationally Deconstructing Movie Narratives: an Informatics Approach", *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 2264-2268, 2015.

Hauge M., *Writing Screenplays That Sell*, Bloomsbury Publishing, 2011.

Kim J., Monroy-Hernandez A., "Storia: Summarizing social media content based on narrative theory using crowdsourcing", *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, p. 1018-1027, 2016.

Kim N. W., Bach B., Im H., Schriber S., Gross M., Pfister H., "Visualizing Nonlinear Narratives with Story Curves", *Proceedings of the 2017 IEEE Transactions on Visualization and Computer Graphics*, p. 595-604, 2017.

Kočiskỳ T., Schwarz J., Blunsom P., Dyer C., Hermann K. M., Melis G., Grefenstette E., "The Narrativeqa Reading Comprehension Challenge", *Transactions of The Association for Computational Linguistics*, vol. 6, n° 1, p. 317-328, 2018.

Labatut V., Bost X., "Extraction and Analysis of Fictional Character Networks: A Survey", *ACM Computing Surveys*, vol. 52, n° 5, p. 1-40, 2019.

Leach E. R., *Claude Lévi-Strauss*, Viking Press, 1970.

Lee O.-J., You E.-S., Kim J.-T., "Plot Structure Decomposition in Narrative Multimedia by Analyzing Personalities of Fictional Characters", *Applied Sciences*, vol. 11, n° 4, p. 1645, 2021.

Lévi-Strauss C., *Anthropologie Structurale*, Plon Paris, 1958.

Lewis R. J., Grizzard M., Lea S., Ilijev D., Choi J.-A., Müsse L., O'Connor G., "Large-Scale Patterns of Entertainment Gratifications in Linguistic Content of US Films", *Communication Studies*, vol. 68, n° 4, p. 422-438, 2017.

Li B., Cardier B., Wang T., Metze F., "Annotating High-Level Structures of Short Stories and Personal Anecdotes", *Proceedings of the 11th International Conference on Language Resources and Evaluation*, p. 1-7, 2018.

Li R., Tapaswi M., Liao R., Jia J., Urtasun R., Fidler S., "Situation Recognition with Graph Neural Networks", *Proceedings of the 2017 IEEE International Conference on Computer Vision*, p. 4173-4182, 2017.

Li Y., Ming W., Kuo C. J., "Semantic Video Content Abstraction Based on Multiple Cues", *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo*, p. 159-159, 2001.

Liu C., Shmilovici A., Last M., "Towards Story-Based Classification of Movie Scenes", *PloS one*, vol. 15, n° 2, p. e0228579, n.d.

Liu S., Wu Y., Wei E., Liu M., Liu Y., "StoryFlow: Tracking The Evolution of Stories", *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, p. 2436-2445, 2013.

Lô G., de Boer V., van Aart C. J., "Exploring West African folk narrative texts using machine learning", *Information*, vol. 11, n° 5, p. 236, 2020.

Lucas D. W., "Aristotle Poetics", *The Classical Review*, 1968.

Malec S., "Autopropp: Toward the Automatic Markup, Classification, and Annotation of Russian Magic Tales", *Proceedings of the 2010 International Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, p. 68-74, 2010.

Malec S. A., "Proppian Structural Analysis and XML Modeling", *Computers, Literature and Philology*, vol. 4, n° 1, p. 68-74, 2001.

McKee R., *Story: Substance, Structure, Style and the Principles of Screenwriting*, HarperCollins Publishers, 1997.

Murtagh F., Ganz A., McKie S., "The structure of narrative: the case of film scripts", *Pattern Recognition*, vol. 42, n° 2, p. 302-312, 2009.

Onega S., Landa J. A. G., *Narratology: an Introduction*, Routledge, 2014.

Ordelman R., Aly R., Eskevich M., Huet B., Jones G. J., "Convenient discovery of archived video using audiovisual hyperlinking", *Proceedings of the 3rd Edition Workshop on Speech, Language & Audio in Multimedia*, p. 23-26, 2015.

Padia K., Bandara K. H., Healey C. G., "A System for Generating Storyline Visualizations Using Hierarchical Task Network Planning", *Computer Graphics*, n.d.

Papalampidi P., Keller F., Lapata M., "Movie Plot Analysis via Turning Point Identification", *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, p. 1707-1717, 2019.

Park S.-B., Oh K.-J., Jo G.-S., "Social network analysis in a movie using character-net", *Multimedia Tools and Applications*, vol. 2, n° 59, p. 601-627, 2012.

Phung Q. D., Dorai C., Venkatesh S., "Narrative structure analysis with education and training videos for e-learning", *Proceedings of the 16th International Conference on Pattern Recognition*, p. 835-838, 2002.

Piskorski J., Zavarella V., Atkinson M., Verile M., "Timelines: Entity-Centric Event Extraction from Online News", *Proceedings of the 2020 Text2Story Workshop at European Conference on Information Retrieval*, p. 105-114, 2020.

Propp V., *Morphology of the Folktale*, University of Texas Press, 1958.

Qiang L., Bingjie C., Haibo Z., "Storytelling by the Storycake Visualization", *Visual Computer*, vol. 33, n° 10, p. 1241-1252, 2017.

Reagan A. J., Mitchell L., Kiley D., Danforth C. M., Dodds P. S., "The Emotional Arcs of Stories are Dominated by Six Basic Shapes", *European Physical Journal Data Science*, vol. 5, n° 1, p. 1-12, 2016.

Regneri M., Koller A., Pinkal M., "Learning script knowledge with web experiments", *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, p. 979-988, 2010.

Sadler N., "Narrative and Interpretation on Twitter: Reading Tweets by Telling Stories", *New Media and Society*, vol. 20, n° 9, p. 3266-3282, 2018.

Schmid W., *Narratology: an introduction*, Walter de Gruyter, 2010.

Sloetjes H., Wittenburg P., "Annotation by Category-ELAN and ISO DCR", *Proceedings of the 2008 International Conference on Language Resources and Evaluation*, p. 816-820, 2008.

Tapaswi M., Bäuml M., Stiefelhagen R., "StoryGraphs: Visualizing Character Interactions as a Timeline", *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, p. 827-834, 2014.

Tapaswi M., Bäuml M., Stiefelhagen R., "Aligning Plot Synopses to Videos for Story-Based Retrieval", *International Journal of Multimedia Information Retrieval*, vol. 4, n^{o} 1, p. 3-16, 2015.

Tekiroğlu S. S., Chung Y.-L., Guerini M., "Generating Counter Narratives Against Online Hate Speech: Data and Strategies", *Proceedings of the 2020 Annual Meeting of the Association for Computational Linguistics*, p. 1177-1190, 2020.

Todorov T., Weinstein A., "Structural analysis of narrative", *NOVEL: A forum on fiction*, vol. 3, p. 70-76, 1969.

Valls-Vargas J., Ontanón S., Zhu J., "Toward automatic character identification in unannotated narrative text", *Proceedings of the 7th intelligent narrative technologies workshop*, p. 38-44, 2014.

Valls-Vargas J., Zhu J., Ontañón S., "Towards Automatically Extracting Story Graphs from Natural Language Stories", *Workshops of the 31st AAAI Conference on Artificial Intelligence*, 2017.

Van Dijk T., "Episodes as Units of Discourse Analysis. Analyzing Discourse: Text and Talk", 1981.

Vargas J. V., *Narrative Information Extraction with Non-Linear Natural Language Processing Pipelines*, 2017.

Vicol P., Tapaswi M., Castrejon L., Fidler S., "Moviegraphs: Towards Understanding Human-Centric Situations from Videos", *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, p. 8581-8590, 2018.

Whalley G., Baxter J., Atherton P. *et al.*, "Aristotle's Poetics: Translated and with a Commentary By George Whalley", *Dramatic Theory and Criticism (DTC): Greeks to Grotowski*, vol. 26, n^{o} 1, p. 36-37, 1997.

Yu H., Zhang S., Morency L.-P., "Unsupervised Text Recap Extraction for TV Series", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, p. 1797-1806, 2016.

Zhao Z., Ge X., "A computable structure model for hollywood film", *Proceedings of the 2010 IEEE International Conference on Image Processing*, p. 877-880, 2010.