CEUR-WS.org/Vol-2723/short50.pdf

Computing Narrative

Ben Miller[a], Julie S. Park[a]

[a]Emory University, 537 S. Kilgo Cirlce, Atlanta, GA 30322, USA

Abstract

Presuming it is possible to produce an analytic model of narrative representable in the architecture of computation, other affordances of digital media should be adaptable to empirical, quantitative inquiry of stories and their structure. *Computing Narrative* is a brief description of some projects attempting those representations and the implications thereof for critical inquiry of stories and their structure.

Keywords

Computational models of narrative, Natural Language Processing, Literary studies

1. Representing Stories with Data

How do current attempts to computationally model and generate narrative offer a landscape of cultural and technical practices in which to situate the problematic transformation of stories and storytelling into data? Furthermore, how might that transformation alter our relationship to culture, critical inquiry, evidence in the humanities, and our expectations of narrative?

Presuming it is possible to produce an analytic model of narrative representable in the architecture of computation, one sensitive to work on narratology ranging from the structural slot-grammar of Propp [35] to the theorizing of human experientiality by Fludernik [10], other affordances of digital media should be adaptable to empirical, quantitative inquiry of stories and their structure. Earlier techniques in information science have made strings, and then keywords, entities, gestures, sounds, maps, and networks searchable; now, subjective cognitive frameworks such as narrative are being rendered as procedures and data in creative contexts, such as games, and in critical contexts, such as by story generation and annotation systems. These systems, ranging from platforms to games to annotation schema to bots, are enabling alternative, cultural and analytic practices to flourish in concert with computation. Scholars like Grigar and Barber [16], Montfort [31], and Hayles [19] have long advocated the value of deep critical engagement in both these systems and their outputs.

What typifies narrative-generating projects and systems is that they can be seen to represent theoretical formalisms, such as focalization [13], with information architectures, such as directed acyclic graphs [39]. This translation of the cognitive structures of storytelling into software, a process better called a transduction as it relates to the conversion of information from one medium to another, holds whether we consider systems based on rules that shuffle manually-constructed snippets of text, or on generative models that build stories by predicting words in a sequence. In their survey of natural language generation, Gatt and Krahmer

CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands
b.j.miller@emory.edu (B. Miller); julie.sohyun.park@emory.edu (J.S. Park)
http://benmiller.mit.edu/ (B. Miller)
0000-0002-5170-6850 (B. Miller)
© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

2018 merge these categories, differentiating instead on two other factors. First, they consider whether the system is text-to-text or data-to-text, meaning does it take as its input elements from pre-existing stories, such as lines from personal narratives, or data from real or simulated events, such as weather data or sports scores. Second, they ask whether the system includes some type of overarching planning function, commonly described as a drama manager [3]. The majority of their work considers the broader problem of language generation, rather than narrative generation. However, when they do address narrative generation, most specifically with reference to [25] it is on terms similar to what we describe above: “a database of entities,” a graph describing the schema of events that relate one to the next in the manner of a story, and a function that dispenses the elements of the database according to the learned and programmed rules of what makes a story. By the conclusion of their survey, Gatt and Krahmer put forward that there is little real distinction between the text-to-text and data-to-text approaches. What is left, we would argue, is the simpler framework of a dictionary, and a function.

Consider, as an example of a creative approach to narrative generation, Nick Montfort’s interactive fiction library for Python, *Curveship* [30]. It represents aspects of focalization, such as tense, with language drawn from dictionaries and called via functions, such as in the snippet of code below [29].

```
def acknowledge(tense_er, tense_rs, _, discourse, __):
    'Produces a rather empty utterance when there is nothing to represent.'
    template = 'nothing special [happen/1/v]'
    para = Paragraph(discourse.spin['template_filter'], [template])
    para.set(discourse.spin['narrator'], discourse.spin['narratee'],
             tense_er, tense_rs, discourse.spin['progressive'])
    return [para]
```

In this example, each verb is conjugated by a function to manifest a focalization strategy. The strategy invoked here is the discourse order, “progressive,” where events follow one after the next in chronological order. That order is called by the function, “discourse.spin,” on lines 4, 5, and 6. *Curveship* is reflective of other experiments in software that enable the writing of interactive fiction such as *Inform 7* [32], which predates it, or *Ink* [20], which came later. These types of direct, scripting-based tools let authors control aspects of the narration and provide ways to dynamically but deliberately alter foundational aspects of discourse, such as tense, person, mood, and distance.

These experiments with which to craft stories allow for revelations about how aspects of focalization influence the effect of narration on readers. These revelations, and their relationship to narrative structure, rely on the capacity of software architectural features. Such features mainly include dictionaries of words and their equivalences known as synonym rings, and functions that can be run. Sequentially, these features allow for the tools to instantiate rules that generate grammatically comprehensible permutations, such that the “nothing special [happen/1/v]” above can become, “nothing special happened / happens / is happening,” depending upon the desired focalization.

Researchers can glean many observations from the output of these tools and their inner workings. One fundamental lesson is confirmation that the abstraction and reduction of narrative to systems of rules and formalisms can be executed to yield recognizable, compelling stories. By measures such as critical interest, as with *Facade* [24], by community recognition, as with *Violet* [11], or by commercial success, as with *Middle-earth: Shadow of Mordor* [28] or *Crusader Kings III* [33], procedural generation of story and story elements is a fixture in contemporary culture. This transduction of conceptual categories, like narrative order, to software, has consequences.

2. Narrative Empiricism

Experiments like *Curveship* in narrative generation based on the abstraction and reduction of narrative to code and data have enabled static and dynamic narrative generating machines. These machines range from Talespin [26] to Wordsmith [46] to the OpenAI/GPT-2 based AI Dungeon 2 [45]. In turn, these experiments and their generation of narrative in statistical form, provide an empirical perspective on cultural phenomena. However, it is this empiricism, a kind of return to formalism, that seems to engender a split amongst humanists. At one end of the spectrum we could situate work like [5], where 2,389 participants were asked to evaluate the stability and affect of story retellings, or [27] where 227 raters were asked to evaluate the degree of certainty of a speaker relative to the kind of event about which they were providing testimony. These studies function as an empirical work drawing upon fields like social psychology to help ensure methodological validity. They, and many others, treat narrative as decomposable data, asking readers to function as evaluators who operationalize aspects of stories. A story, in empirical studies like these, might be represented by a continuous value called "surprisingness," evaluated on a 7-point scale [5], or by a cluster number denoting the intersection of two continua, certainty-to-uncertainty, and event-to-negation of event [27].

This empirical work undertaken by digital humanists, computer scientists, and cognitive scientists alike, parallels work in other domains. With experiments in expressive computation, such as were the subject of Harrigan and Wardrip-Fruin [18], there first came theoretical formalisms, followed by top-down structural models based on expertise and theory, and then, when computing power and cheap storage made available large corpora, a pivot towards data-mining. This pivot affects what the authors consider as ends of a multidimensional spectrum – the more empirical work described above at one end, and the more performative work that unfolds with reader interaction at the opposite. For example, at this end of the spectrum, one might find physical works like the shuffle book, *Composition No. 1*[41], where narrative structure seems to function more as an emergent category dependent upon a reader's reading. One still finds the granular markers of mood and distance, such as tense and person, but the larger framing elements of narrative, such as the markers that convey a story's degree of fictionality through the embedding of narration in a dream, are left more to chance. One could also find work that begins with a prompt, then proceeds to describe twisty passages through large textual datasets, e.g. the aforementioned *AI Dungeon 2*. That piece of software relies on a complex language model where words, phrases, and chunks of text are statistically linked through stronger and weaker linkages across 1,600 dimensions to the 5GB of text in what is known as the GPT-2 (Generative Pretrained Transformer 2) model. The updated, GPT-3 model draws on even more text and provides even more dimensions of connectivity, but the underlying architecture is identical.

As seen from just these examples, work on the computational modeling of narrative ranges across media from text to video to games, and proceeds under many avenues: script and frame-based approaches to infer common story patterns [6, 43], hybrid approaches blending top-down models and manual annotation with machine learning to focus on the impact of features such as focalization [4, 9], syntactic approaches to build connections between story grammars and linguistic ones [23, 1] , and quantification-based approaches to apply lessons from statistics to story structure analysis [21, 2]. These approaches indicate growing recognition under many names, including but not limited to distant reading, cultural analytics, and quantitative literary studies. These various experiments bear some resemblances from their methods aimed to elicit structure from narrative, an approach derived from data mining. Yet their remaining terminological and methodological diversity suggests that research in to story structure occurs across disciplines, that the results are unsatisfactory, and also that there is much debate as to how to understand and represent the structure of stories.

3. Narrative Inference

Narrative, like most objects studied by humanists, is a subjective, culturally-dependent phenomenon [38, 17]. As such, attempts to model it fall victim to the problem of validation in the humanities; each reading of a narrative serves as one of many possible readings by one of many possible readers. In short, to paraphrase Piper [34], even structural approaches to the study of narrative emerged from a discipline not subject to a need for quantitative proof of validity or invalidity. Critical scaffolding was at one point in time sufficient. But, as we move increasingly towards generative systems, towards generative cultural objects, empirical measures of agreement on the meaning of these cultural objects seem to substitute for measures of validity. In the above mentioned study by Chambers and Jurafsky [6], a valid response is one that repeats. By this, we mean that the induction of a schema or script, a commonly occurring pattern of actions within a story, relies at present on the mining of stories from a segment of a large linguistic data corpus, the approximately 6GB of the New York Times component of the *English Gigaword Corpus* [15]. What remains from data mining approaches such as schema or script induction is a kind of story common denominator: the events most likely to be written about in the way they are most likely to be described.

Even prior to the age of data mining, when narratologists relied on the descriptive strength, groundedness, and the utility of a theory, the validity of a theory often meant ignoring the pragmatics of culture, of a particular reader. While frameworks like Propp's can be applied in contexts beyond the Russian Folk Tale, it was work by folklorists like Colby [7] and Dundes [8] that dramatically emphasized how structurally distinct storytelling structures could be.

With predictive language models like Radford et al. [36], readers find stories that emerge as something like symptoms from preexisting large corpora. This emergence is something that present developers of these models with, "concerns about malicious applications" [36]. In essence, a reader provides these language-model based systems with prompting text. Based upon the dense web of connections present within the model, a series of textual snippets are provided. If the model is coherent, sufficiently large, and the connections modeled meaningfully, the snippets are related and what emerges is something that can be read like a story. For example, the snippet below is from *AI Dungeon 2* and contains three parts: generated text from a prior prompt, a new user-entered prompt, and text generated in response to that new prompt.

```
Your job as a security robot was to protect shipments
from being stolen by criminals or terrorists. However,
these days there are no more thieves and terrorists.
The world has become too dangerous for anyone but the
most skilled mercenaries to survive.
```

```
> You go for a walk.
It's been awhile since you've done much physical
activity. Your body aches all the time and you don't
feel like going out for a jog right now. You need to
get some fresh air and exercise.
```

The tool seems to present a coherent world, a milieu that incorporates a set of related objects presented with appropriate language: robots, protection, thieves, mercenaries, survival. However, there is only linguistic continuity, not logical continuity. When responding to the prompt, "You go for a walk," the engine responds paradoxically, both in favor of and against exercise. Currently, these types of generative algorithms fall into two categories. Those without drama managers, like *AI Dungeon 2*, and those with, like the experiments of Gervás et al. [14].

4. Narrative Emergence

What differentiates these two types is more than just the underlying technology. The technology reflects broader theoretical differences and a persistent question that is itself broader than one related to narrative. Namely, in the case of narrative generation, is the logic of the story reflected in the structure of the language, or is it a set of extrinsic structures which the language of a story prompts in the manner of an indexical reference? To frame the question more concretely, do narrative generation systems need to explicitly model the narrative structure, or does modeling the language of stories carry forward the necessary coherence for story generation?

For researchers in the humanities looking to apply computational methods for the analysis of narrative, this question is fundamental. It suggests that the conceptual framework of narrative, while describable in structural terms, may be an emergent phenomena – a function of the text, rather than a feature.

5. Making Connections

Our position, is that empirical studies of narrative, and the transduction of narrative features into software architectures enable more informed critiques and meta-analyses of stories. In the survey above, we described a number of storytelling systems. Let's return here to three: the software experiment with focalizations, *Curveship*; *Durruti*[44], an analytic tool that extends a throughline of CS research in natural language generation from the early architectures of Schank and Abelson [42] past the schemas of Chambers and Jurafsky [6]; and *AI Dungeon 2*. What we contend, based on these examples, is that the categorical distinctions upon which theoretical approaches to narrative rely, such as appear throughout the work of Gennette [13], Fludernik [10], Ryan [40], and others, are principally useful as analytic tools for manual, close reading. They are descriptive, top-down formalisms that require close reading to identify the distinctions they put forth. Some of the distinctions upon which they rely that would seem straightforward, such as between prolepsis, or events being related ahead of their occurrence, and chronicle, or events being related in order of their occurrence. However, research shows they are still incredibly challenging interpretations for human annotators. A recent, well-designed study [22], yielded an inter-annotator agreement score of 0.24 for annotations of these features in modernist novels, a score the author noted is normally considered to be, "quite low." Empirical research on features of narrative reveal how complex and potentially subjective these conceptual categories may be.

The challenge of human readers identifying elements of narratological frameworks is distinct from the challenge machine readers face. However, current narrative analytic systems avoid much of this challenge. Instead, most still proxy position in the text for any attempt at position within the discourse [21]: it is easier to know which line a system is on, than where that line resides within the time of the storyworld. More complex approaches, as with the proposal by [37], are still entirely reliant on human annotators. Effectively, computational analysis of narrative at scale often retreats from a consideration of narrative to instead focus on more concrete textual features. Again, this retreat conveys the conceptual complexity of the elements of narrative and point towards narrative being more of an emergent phenomenon than an intrinsic one.

We write this not to critique these projects, or the theoretical frameworks upon which they rely, but to indicate how difficult a task they attempt. Even framing structures, a feature that typically precedes and concludes a story, thus having a fixed textual position, presents dramatic challenges. Signaling that a story begins as a dream, something that should be manifestly apparent given the necessity that it conveys to readers the fantastic nature of the events that follow, are challenging to identify computationally. This kind of discourse signaling relies, to borrow language from Ryan's work on possible worlds and earlier work on modal logic, on recognizing how significantly the textual story world departs from a shared reference world. However, instead of being able to compare from story world to reference world, analytic tools based on automatic inference of schema and scripts, like *Durruti*, rely on recognizing patterns from databases of preexisting text. They are, in essence, examples of the text-to-text generation described by [12], though the generated text is an interpretation, not a story. The inferential capacity of tools like these is that of pattern matching, and comparative, but not evaluative. Consider the snippet of code from *Durruti* responsible for building the semantic models.

```
conditioned_models = {}
for feature in combos_search:
    #bubble up the scope of all features in for_export
    local_data = {f: for_export[f][feature] for f in for_export}
    local_data["feature_tuple"] = feature
    conditioned_models[feature] = PMI_Model(local_data)
```

As one can see on fifth line, *Durruti* represents events in the schema as feature tuples. These tuples are commonly occurring sets of subjects, verbs, and objects. To reiterate, the analysis generated by this tool is dependent upon the recognition of repeated patterns of actions, followed by the sequencing of those actions in the order in which they appear within the text, not the story. Using a set of functions, it builds for the critic a dictionary of event schemas for consultation and interpretation.

The limits these tools express are the necessary by-products of representing complex cognitive structures, such as temporality, mood, speed, and framing, with software architectures. We do not point this out to indicate that one is more complex than the other, but to say that the complexity of each leaves one underspecified in the other. The current architecture for representing stories in software – dictionaries, graphs, and tuples – fall short of clearly representing features of narrative not because they are insufficiently complex. Instead, they indicate that the categories themselves are subjective, challenging, and most importantly for our consideration here, emergent. That narrative would seem to be an emergent phenomenon, as it refuses to be well-behaved and readily annotated, implies that readers can accept and consume stories produced by theoretically imperfect systems. Critical, writerly, and commercial culture's embrace of generative storytelling indicates that these architectures produce stories worth reading.

References

[1] J. Araki et al. "Detecting Subevent Structure for Event Coreference Resolution." In: *LREC*. 2014, pp. 4553–4558.

[2] D. Bamman, T. Underwood, and N. A. Smith. "A bayesian mixed effects model of literary character". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2014, pp. 370–379.

[3] J. Bates. "Virtual reality, art, and entertainment". In: *Presence: Teleoperators & Virtual Environments* 1.1 (1992), pp. 133–138.

[4] R. Bod et al. "Objectivity and reproducibility of Proppian narrative annotations". In: *The Third Workshop on Computational Models of Narrative*. Ed. by M. A. Finlayson. 2012, pp. 17–21.

[5] F. Breithaupt et al. "Fact vs. affect in the telephone game: All levels of surprise are retold with high accuracy, even independently of facts". In: *Frontiers in psychology* 9 (2018), p. 2210.

[6] N. Chambers and D. Jurafsky. "Unsupervised learning of narrative schemas and their participants". In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 2009, pp. 602–610.

[7] B. N. Colby. "Cultural patterns in narrative". In: *Science* 151.3712 (1966), pp. 793–798.

[8] A. Dundes. *Sacred narrative: readings in the theory of myth*. Univ of California Press, 1984.

[9] M. A. Finlayson. "Inferring Propp's functions from semantically annotated text". In: *The Journal of American Folklore* 129.511 (2016), pp. 55–77.

[10] M. Fludernik. *An introduction to narratology*. Routledge, 2009.

[11] J. Freese. *Violet*. https://ifdb.tads.org/viewgame?id=4glrrfh7wrp9zz7b. 2008.

[12] A. Gatt and E. Krahmer. "Survey of the state of the art in natural language generation: Core tasks, applications and evaluation". In: *Journal of Artificial Intelligence Research* 61 (2018), pp. 65–170.

[13] G. Genette. *Narrative discourse: An essay in method*. Vol. 3. Cornell University Press, 1983.

[14] P. Gervás et al. "The long path to narrative generation". In: *IBM Journal of Research and Development* 63.1 (2019), pp. 8–1.

[15] D. Graff et al. "English gigaword". In: *Linguistic Data Consortium, Philadelphia* 4.1 (2003), p. 34.

[16] D. Grigar and J. Barber. "Defending your life in MOOspace: A report from the electronic edge". In: *High wired: On the design, use, and theory of educational MOOs* (1997), pp. 192–231.

[17] D. F. Harrell. "Shades of computational evocation and meaning: The GRIOT system and improvisational poetry generation". In: *Proceedings, sixth digital arts and culture conference*. Citeseer. 2005, pp. 133–143.

[18] P. Harrigan and N. Wardrip-Fruin. *Second person: Role-playing and story in games and playable media*. The MIT Press, 2007.

[19] N. K. Hayles. *My mother was a computer: Digital subjects and literary texts*. University of Chicago Press, 2010.

[20] Inkle. *ink*. https://github.com/inkle/ink. 2020.

[21] M. L. Jockers. *Macroanalysis: Digital methods and literary history*. University of Illinois Press, 2013.

[22] E. Kearns. "Annotating and quantifying narrative time disruptions in modernist and hypertext fiction". In: *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*. 2020, pp. 72–77.

[23] Z. Liu, T. Mitamura, and E. Hovy. "Evaluation algorithms for event nugget detection: A pilot study". In: *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. 2015, pp. 53–57.

[24] M. Mateas and A. Stern. "Façade: An experiment in building a fully-realized interactive drama". In: *Game developers conference*. Vol. 2. 2003, pp. 4–8.

[25] N. McIntyre and M. Lapata. "Learning to tell tales: A data-driven approach to story generation". In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 2009, pp. 217–225.

[26] J. Meehan. "The metanovel: Writing stories by computer (Tech. Rep. No. 74)". In: *New Haven, CT: Yale University, Department of Computer Science* (1976).

[27] B. Miller. "Reading Certainty: Evidence from a Large Study on NLP and Witness Testimony". In: *Proceedings of Digital Humanities 2020 (Ottawa, Canada; July 2020)*. Ottawa, Canada: Alliance of Digital Humanities Organizations, 2020.

[28] Monolith Productions. *Middle-earth: Shadow of Mordor*. Microsoft Windows, OS X, Linux, XBox, Playstation. 2014.

[29] N. Montfort. *Curveship*. https://https://github.com/cjb/curveship. 2010.

[30] N. Montfort. "Curveship: An interactive fiction system for interactive narrating". In: *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*. 2009, pp. 55–62.

[31] N. Montfort. *Twisty Little Passages: an approach to interactive fiction*. MIT Press, 2005.

[32] G. Nelson. *Inform 7*. http://inform7.com/. 2019.

[33] Paradox Development Studio. *Crusader Kings III*. Microsoft Windows, Linux. 2020.

[34] A. Piper. *Validation and Subjective Computing*. March 25, 2015.

[35] V. Propp. “Morphology of the Folktale. 1928”. In: *Trans. Svatava Pirkova-Jakobson. 2nd ed. Austin: U of Texas P* (1968).

[36] A. Radford et al. “Better language models and their implications”. In: *OpenAI Blog https://openai. com/blog/better-language-models* (2019).

[37] N. Reiter, M. Willand, and E. Gius. “A Shared Task for the Digital Humanities Chapter 1: Introduction to Annotation, Narrative Levels and Shared Tasks”. In: *Journal of Cultural Analytics* (2019), p. 11192.

[38] P. Ricoeur. *Time and Narrative, vol. 1, trans. Kathleen McLaughlin*. 2009.

[39] M. O. Riedl and R. M. Young. “From linear story generation to branching story graphs”. In: *IEEE Computer Graphics and Applications* 26.3 (2006), pp. 23–31.

[40] M.-L. Ryan. *Possible worlds, artificial intelligence, and narrative theory*. Indiana University Press, 1991.

[41] M. Saporta. *Composition no. 1: roman*. Éditions du Seuil, 1962.

[42] R. C. Schank and R. P. Abelson. “Scripts, plans, and knowledge”. In: *IJCAI*. Vol. 75. 1975, pp. 151–157.

[43] D. Simonson and A. Davis. “Interactions between narrative schemas and document categories”. In: *Proceedings of the First Workshop on Computing News Storylines*. 2015, pp. 1–10.

[44] D. Simonson and A. Davis. “NASTEA: Investigating narrative schemas through annotated entities”. In: *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*. 2016, pp. 57–66.

[45] N. Walton. *AI Dungeon 2*. Accessed: 20 July 2020. 2019. URL: https://www.aidungeon.io/.

[46] A. Wright. *Algorithmic authors*. 2015.