Mark A. Finlayson
Florida International University

A Survey of Corpora in Computational and Cognitive Narrative Science

The use of corpora, and especially annotated corpora, is key to advancing the scientific understanding of narrative. I seek here to answer several interrelated questions: what corpora have been used in computational and cognitive research on narrative to date? What are the general characteristics of these corpora, and are they adequate for the overall task of a scientific understanding of narrative? What types of narratives have been studied, what kinds of annotations have been applied to those narratives, and what gaps need to be filled? To answer these questions, I conducted a wide-ranging literature review, collecting a bibliography of over 2600 conference papers, journal articles, theses, reports, and books on the topic of the computational, cognitive, and (more generally) scientific[1] study of narrative. I surveyed this bibliography for the use of corpora, and identified 167 unique text collections (155 with some sort of annotation) that could be considered a "corpus". These corpora were used in more than 180 publications across twelve fields. This set contained seventeen different broad types of narratives, five different modalities, and approximately 42 different types of annotation. Of the types of annotations present, the most common complex is an annotation of Events, Named Entities, and Roles (~33 % of corpora) which indicates that a representation of the "*who does what to whom"* (i.e., the fabula) of narratives has been a key concern of researchers. Worryingly, I was able to identify only four corpora which provide enough information to capture a basic representation of fabula that were not annotated directly by their authors. This means that there is much work to be done in improving data collection standards in the field. Another observation was that the sense and meaning of words, narrative units, and other levels of the narratives is woefully under-represented. Finally, another important finding is that over half of the corpora identified have fewer than 20 texts, and many of these "narrative" corpora contain texts that are only barely "narrative" in nature. These observations lead to some clear recommendations for the directions of future research and resource-collection efforts.

The paper is organized as follows. First I discuss the need for and utility of corpora in the scientific study of narrative (§1). I then outline the method I used for surveying the literature (§2), which resulted in a comprehensive list of 64 workshops, symposia, conferences, and other meetings that have been held on narrative science in the past 30 years (Table 7 in the Appendix). I then proceed to a detailed analysis of the results of the survey, identifying the different types of narratives, modalities, annotations, annotation methods, and so forth that I found (§3). Because the type and extent of annotation is

1 For my meaning of the terms *annotation*, *narrative,* and *science,* see midway through §1.

critical to a corpus's utility, I discuss these in detail (§4). These observations lead me to identify several important gaps in the available corpora, and I will discuss what needs to be done to address these gaps (§5). I conclude with a concise list of the contributions of this work, and potential future directions (§6).

1 The Need for Corpora

Data is a key component of science. Without data we cannot connect our theories to the real world; we cannot support or reject our hypotheses; and we are significantly hobbled in our ability imagine new research directions. In narrative science, corpora are one of the most significant sources of data. A corpus can turn a study of narrative from example-driven, intuitive, anecdotal art to data-driven science. Corpora allow more objectivity in the study of narrative, by making exact the specific language examples that are being analyzed. Corpora allow repeatability and the comparison of results, by allowing researchers to re-run experiments with exactly the same inputs, and test competing hypotheses against the same data. The use of corpora is also a practical matter: collecting narrative artifacts allow focusing of resources, such as time, money, and expertise, which are scarce and rarely adequate for the needs at hand.

In the science of narrative, *annotated* corpora are especially important. First, though, some definitions. My use of the word *annotation* is the same as in corpus linguistics, in that it covers "any descriptive or analytic notations applied to raw language data" (Bird and Liberman 2001). Because here we are concerned with narratives, I can reformulate this as any descriptive or analytic notations applied to raw *narrative* data. Here I explicitly allow the possibility that the raw narratives "texts" are communicated to us in a non-linguistic or partially-linguistic format, say, movies, video, images, cartoons, paintings, game logs, etc. With these different modalities in mind, it is important to note that here I am using *narrative* in a relatively broad sense of the term, namely, any sequence of events that are locally coherent and connected, with clear chains of cause and effect concerning a set of agents and their goals and motivations. These narratives might be most properly called *action discourses* (see Finlayson & Corman, 2014, this volume), and may or may not have an extra "narratable quality" that makes it a story worth telling. Finally, my use of the word *science* and its variants is here in the strong sense Anglo-Saxon use of the term, namely, a field of study that follows, after some fashion, the Baconian scientific method to advance and test hypotheses.

Annotation is needed because, although we have access to the narrative "texts" themselves, narrative theories or experiments usually deal with levels of analysis hidden from view, and not easily calculated from the surface form of the texts. These hidden levels are the basic data against which our hypotheses will be tested, but our tools are generally inadequate to collecting that data. Narratives are complex cognitive and linguistic objects. They describe complex chains of events, with relationships between characters and objects that reflect the full complexity of human life. Narratives are often communicated via sophisticated forms of discourse, relying on deep knowledge of the world, society, and culture. Many of the phenomena that are interesting from the point of view of narrative science happen at a level beyond our ability to automatically analyze by machine. By way of example, suppose one wants to conduct a narrative experiment comparing the discourse structure of different textual narratives. One must clearly start with some representation of the discourse structure, but although automatic analyzers exist that will provide a discourse structure (e.g., Lin et al 2014), they are brittle, error-prone, and don't work in most languages other than English. Thus, if researchers seek clean, error-free data, they are required to perform annotation, usually via some mix of automatic methods and trained human annotators, to provide the initial data required. For other more complicated levels of analysis, we often do not even have any ability at all to do an automatic analysis of the data, requiring annotation to be performed completely manually.

Nevertheless, for all of their advantages, corpora present several dangers. The first is that the corpora used for particular tasks might be non-representative for the phenomena or domain at hand, leading to inaccurate conclusions. This danger can be mitigated when corpus developers are clear about when, where, how, and why they collected the texts they included in the resource. A second, more serious problem is that because corpora, especially well-annotated corpora, are difficult and expensive to create, researchers will often use an available corpus for their work, even if that corpus is less than ideal, or even inadequate, for their scientific needs. Because the researchers have an expensive hammer, provided by someone else, everything starts to look like a nail.

The third, and potentially most troublesome, danger of using corpora is that it can drive a field into a "local maxima" solution. This happens because, despite whatever the known flaws of a particular corpus, it is easier to use that corpus than build another one. Researchers then measure the accuracy of their results (or algorithms, models, or other scientific explanations or hypotheses) against that data, usually with metrics that allow their results to be compared to other work on the same corpora. This leads to optimization of research results to that specific corpus—*overfitting* in the parlance of machine learning. Researchers become more focused on increasing their performance on certain well-known metrics, than actually getting their results to generalize across a larger body of data. (See, e.g., Oepen, 2007).

With these advantages and dangers clearly in mind, I ask several key questions about the use of corpora in narrative science. First, *what corpora have been used in the field so far*? This is an important question for both the beginning and established researcher. By identifying corpora already extant, we can save ourselves a lot of work. Second, *how have these corpora helped or hindered the progress of narrative science*? Answering this question will give us clear idea of what advantages we have gained from the corpora in use, and what barriers those corpora have held up to progress. This will also give us insight into the major concerns of the field, regardless of whether or not those concerns are made explicit. Third, with a clear set of past resources and uses identified, we can proceed to ask the question: *what directions should corpus creation for narrative science take in the coming years?*

2 Survey Method

To answer these questions I conducted an extensive survey of the literature that deals with cognitive and computational issues of narrative. The survey consisted of three stages. First, I sought to identify a set of works that were a representative sample of the literature on the topic of the science of narrative (§2.1). Second, I reviewed the set for the use of corpora (§2.2). Third, for each of those corpora, I noted down a number of important characteristics, tabulating them for comparison (§2.3). This tabulated data forms the raw material from which the analyses in Sections 3 & 4 were derived.

2.1 Finding Articles on Narrative Science

The goal of the survey's first stage was to identify a large and hopefully representative set of works on the topic from which corpora could be extracted.

I identified 2,601 total works including workshop and conference papers, journal articles, books, reports, and theses. Despite the size of the set, it is almost certainly incomplete, and as of this writing I continue to add to the list. To assemble the set, I began with two well-known resources: Erik Mueller's webpage listing Story Understanding Resources (Mueller 2014), which lists a number of narrative-science-related papers and meetings (16 of the latter), and the IRIS Bibliography on Interactive Storytelling (Cavazza, Champagnat et al. 2009, IRIS 2014), which contains 1,881 publications as of this writing. I collected the articles listed in these resources and filtered out works not related to narrative. This left approximately 1,000 works. I next read citation lists of those works to identify articles and meetings missing from the set. I did this for approximately 800 works, which allowed me to identify approximately 48 additional meetings and 1,600 additional works. I stopped collecting when reading the citation lists rarely produced a new work not already in the set.

There were two criteria for inclusion of a work in the set. First, a work was included if its main topic contained either the word "narrative" or "story", and the general approach of the research reported had a scientific approach, especially if it was either cognitive or computational. This criterion was important for curtailing the search to a manageable size, and excluding the extremely large number of articles that discuss narrative from a non-scientific point of view. Second, a work was included if it was cited by a significant number of articles already in the set (ten or more). This had the effect of including seminal works from other fields that themselves were not cognitive or computational in nature. Good examples here would be Propp's seminal work on the structure of Russian fairy tales (Propp 1968), introductions to narratological theory (Bal 2009), or influential tracts on narrative-related phenomena (Bordwell 1985).

2.2 Finding Articles Describing Corpora

With the large set of works in hand, I then skimmed the contents of those works to identify those that used a narrative corpus. This consisted of reading at least the title and abstract of every work, and often closely skimming the remainder of the work's text. If it was still unclear, I performed a closer read to determine whether a corpus had been used. For the purposes of this survey, if a work met any of the following criteria, I marked it as having used a narrative corpus:

1. The work explicitly described the creation of a resource called a "corpus", which consisted of a set of stories, with annotations of some sort, which was released for general use. This covers what we classically would conceive of as an "annotated corpus".
2. The work described research on a set of stories that were specifically gathered together, with or without annotations, to do an experiment, and those stories were identified clearly enough, or captured in such a way, that they that could conceivably be packaged up and distributed independently.
3. The work described research in which a set of stories was generated during an experiment (such as playing a narrative game, or a run of an interactive story system), and these stories were captured for later analysis.
4. The work described the analysis of a one or more artificially constructed stories used to illustrate a particular technique, method, or theoretical construct, and these stories were identified or actually presented in the paper.

Excluded from this analysis were the following:

1. Works that used corpora that clearly did not contain stories, or used corpora which probably contained stories but the stories were not clearly marked.
2. Works that used a set of documents or other artifacts (such as commonsense rules, museum artwork, or places on a map) that might be used to generate or build stories, but were not themselves stories.
3. Works that described an experiment that generated a set of stories for analysis, but those stories were not captured, and so could not conceivably be released for general consumption.

To make these criteria clearer, consider two examples of near misses, namely, works that at first glance look like they might have a corpus relevant to this study, but actually did not. First consider the work by Michel and collaborators titled "Quantitative Analysis of Culture Using Millions of Digitized Books", published in Science (Michel, Shen et al. 2011). In this work, the authors performed an informatics analysis of approximately 4 % of all books printed between the years 1800 and 2000, comprising approximately five million volumes. This work would seem to fit the criteria: it definitely used a corpus, and that corpus almost certainly contained stories. However, the work was excluded for at least two reasons. First, although the work might be seen as relevant to narrative in the large, the work was not specifically focused on narrative, but rather on “culturomics”. Second, although the corpus almost certainly contained narratives (in the form of novels, shorts stories, plays, scripts, and so on) there was no attempt to separate narrative from non-narrative text. Thus the corpus is not a corpus of narratives, and significant work would need to be done to figure out which texts were actually narratives and which were not.

A second example of a near miss is one much closer to the concerns of narrative science, namely, work on the generation of dilemma-based narratives (Barber and Kudenko 2007). In this work, the authors describe a system that automatically generates interactive stories. They describe the system architecture and rules that define the domain of the system. They also describe an evaluation in which eight players were asked to use the system and report their experiences. Here there is no doubt that the work is relevant to narrative science. Again, however, it fails the test for use of a narrative corpus for at least two reasons. First, the artifacts laid out in the paper are not narratives themselves, but rather domain rules that constrain the generation of narratives. Second, although the authors surely generate a set of narratives in the course of their experiments, those narratives were not captured in any way, and so could not be released to other scientists to use.

2.3 Coding Features of the Corpora

In the survey’s second stage I identified over 180 works that used corpora in the research. In the third stage, I carefully read each paper mentioning the use of a corpus, and noted down thirteen features of those corpora. These features were chosen based on the information that I have found useful in dealing with corpora in the past. Sometimes corpora were mentioned in several different works; these corpora were only counted once. The features captured were:

1. Name of the corpus, if any
2. Year the corpus was first described
3. The canonical reference for the corpus, plus any additional references that mention use of the corpus
4. The original stated purpose of the corpus
5. The field, roughly, to which the work’s reporting the corpus belong
6. URL of the corpus if it is available online
7. Modalities of the narratives in the corpus
8. Whether or not the description in the work is sufficient for exact identification of the texts contained in the corpus
9. Whether or not the description in the work is sufficient for replication of the annotation technique
10. The number of stories in the corpus, or my best guess based on descriptions in the work

11. The number of words in the corpus, as best as could be determined. This was left blank if it could not be estimated in any reasonable way
12. The different layers of annotation applied to the corpus
13. The number of different layers of annotation applied to the corpus that are relevant to constructing a representation of the fabula (see §4.1).

Originally I had a 14th feature, namely, the language in which the stories were communicated. However, every corpus that used written or spoken language used English. I identified 167 unique corpora across the whole bibliography. These corpora contained approximately 17.2 million stories. 112 of the corpora (66 %) reported word counts or described their contents in such a way that word counts could be estimated. The total number of words reported was approximately 2.1 billion. Removing the five largest corpora (each with a million or more stories), there are 583,361 stories comprising at least 22.7 million words. Figure 1 shows a histogram of corpus sizes. Fully 69 corpora (41 %) have fewer than ten stories. Over half of the set (85 corpora, 51 %) have fewer than 20 stories.

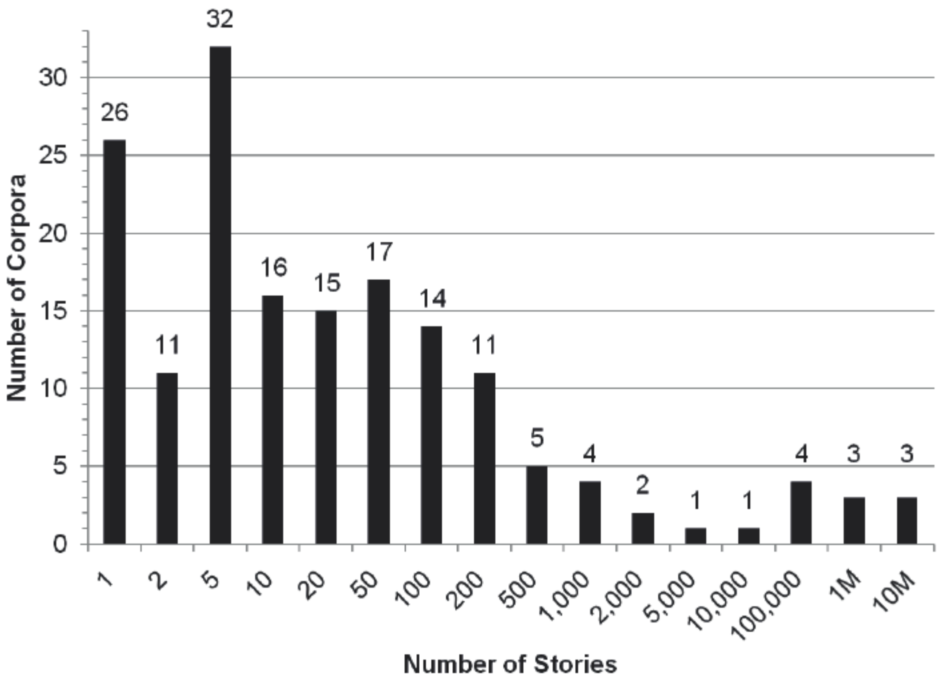

Figure 1: Histogram of the number of corpora against the number of stories. The total number of corpora is 167. Each bin represents the number of corpora that have the labeled number of stories, up to the amount of the next bin.

2.4 Availability of the Data

The data collected for this survey could potentially be quite valuable to students and researchers working on narrative science. Therefore, I am releasing the raw data of the survey in the form of two files, available for download from the web as supplemental materials[2]. The first is the raw bibliographic data of the survey, containing 2601 citations. The second file is a spreadsheet with the coded features of each corpus, plus the analyses that underlie the next two sections.

3 Survey Analysis

I begin my analysis by examining the range of values found among the 167 corpora along four key dimensions: field of study, modality, text type, and annotation style. A detailed analysis of the types of annotation found across the corpora may be found in Section 4.

3.1 Fields

For each corpus I noted down the fields of study most closely aligned with the topic of the work in which the corpus was described. These assignments were somewhat *ad hoc*, as field boundaries are by no means well defined, and I approached not as an unassailable classification but rather as a rough guide to help orient us to the data.

Field	#	%	
Artificial Intelligence	76	46 %	
Psychology	29	17 %	
Computer Science	24	14 %	
Computational Linguistics	16	10 %	
Cognitive Science	12	7 %	
Literary theory	3	2 %	**13 Corpora below here (8 %)**
Communications	2	1 %	
Management Science	2	1 %	
Marketing	2	1 %	
Narratology	2	1 %	
Anthropology	1	1 %	
Folkloristics	1	1 %	
Total	**170**		

Table 1: Distribution of corpora amongst fields. The total is greater than 167 because some corpora fall into multiple fields.

2 A comparison table and a bibliographic database are available at http://dspace.mit.edu/handle/1721.1/92563; a typeset version of the bibliography including the features used for comparison can be accessed at http://www.uvrr.de/tl_files/Supplemente/corp_finlayson.pdf.

Here I made my best guess based on my understanding of the work of the authors, the publication venue, and the stated goals of the corpus. I identified twelve distinct fields with works featuring narrative corpora in my bibliography, as shown in Table 1.

As can be seen, Artificial Intelligence tops the list, with almost 50 % of all the corpora. This is not surprising, given the starting seeds for the bibliographic search, the constraints governing inclusion of copora in the list (i.e., restriction to computational or cognitive concerns), and the general preoccupation of AI with story understanding. Additionally, *Computational Linguistics* may perhaps be thought of as a subfield of AI, further increasing that share. Perhaps the other observation to come from this data is that the survey is almost certainly incomplete: the almost non-existent counts for fields such as Narratology, Literary Theory, and Folkloristics, which almost certainly have many more annotated corpora available, indicates that the survey did not capture all relevant fields.

3.2 Modalities

I also tabulated the modalities found amongst corpora in the survey, shown in Table 2. By "modality", I mean whether the stories in the corpus were communicated as texts, images, video, etc: the sensory channel(s) in which the stories exist.

Text was by far the most common modality. This is not surprising, given the number of stories in textual form, the ease with which computers can store and manipulate text, and the focus on text processing in Artificial Intelligence and related disciplines.

Game Logs was a type of unusual modality that was unexpected. Four corpora used data structures for games that outlined a narrative, either as a recording or as a template (Orkin, Smith et al. 2010, Doran and Parberry 2011, Ha, Rowe et al. 2011, Lee, Mott et al. 2011). It somewhat debatable if these corpora meet the criterion outlined above for being "narratives" themselves, rather than materials used to build narratives. But in my judgment they seemed sufficiently narrative-like to warrant inclusion.

Modality	**#**	**%**
Text	153	92 %
Video	13	8 %
Image	7	4 %
Speech	5	3 %
Game Logs	4	2 %
Total	**183**	

Table 2: Distribution of the corpora amongst modalities. The total is greater than 167 because some corpora use multiple modalities.

3.3 Text Types

Perhaps the most interesting tabulation that I performed, aside from the types annotation layers used, was the grouping of types of text. For each corpus identified, I read the descriptions provided in the papers and wrote down what types of texts the corpus contained. When this list was complete, I did a manual clustering into groups, focusing first on keyword similarity, and then looking carefully at odd descriptions of text types to see if they fit into an existing group. This analysis produced 17 different types of narrative, shown in Table 3.

By volume the most common text types were web logs, a.k.a. blogs, which added up to over fifteen million stories and billions of words. Additional large sets were simple scenarios (approximately one million stories) and newswire text (approximately one million stories and one billion words).

By far the most common text type, by corpus count, was what I will call "simple scenarios" (46 corpora, 28 %). These are extremely simple descriptions of situations, actions, or interactions. I present three select examples to illustrate the type.

Ex. 1. There was a viper and a weasel. The weasel caught the viper. The viper bit it. It died. The viper went away. (Knapp 2009)

	Text Type	**#**	**%**
1	Simple Scenarios	46	28 %
2	Fables	28	17 %
3	Folktales	21	13 %
4	Short Stories	13	8 %
5	News Articles	10	6 %
6	Novels	9	5 %
7	Films, Commercials, Documentaries	8	5 %
8	Summaries of Movies or Plays	6	4 %
9	Blogs/Weblogs	6	4 %
10	Games or Generated Narratives	6	4 %
11	Historical Stories, Biography	5	3 %
12	Personal Narratives (e.g., Diaries)	4	2 %
13	Cartoons, Comics	4	2 %
14	Play or Movie Scripts	4	2 %
15	Legal Cases, Fact Patterns	3	2 %
16	Speeches & Jokes	2	1 %
17	Hyperstories, Nonlinear Narratives	1	1 %
	Total	**176**	

Table 3: Distribution of corpora amongst text types. The total is greater than 167 because some corpora use multiple modalities.

In this paper, Knapp was developing a conversational system that drew upon a collection of stories that notionally corresponded to an experience base. These stories involved a small number of characters and actions, and were generally just a handful of words long.

Although extremely simple and non-literary, Knapp's stories do have a bit of narrative structure involving things like climaxes and reversals of fortune. These would seem to meet fall into most people's general idea of a story. In contrast, take the next example:

Ex. 2. Andrew and Roxanne decided to go to the park. The weather was sunny. Andrew ran through the park. He threw the Frisbee to Roxanne. (St. John 1992)

St. John had a commonsense model of the world that was used to generate a million or so simple situations like those shown in Example 2. These scenarios are story-like in that they have characters and events in a coherent and related sequence, but not story like in that they have no higher-level narrative structure or literary worth (see the discussion of Level II narratives in Finlayson & Corman, 2014, this volume). And even further down the chain toward "non-story" items are scenarios like these:

Ex. 3. Locate broom. Run broom across floor to gather dust into a pile. Place dust pile in trash can. (Smith and Arnold 2009)

In this paper Smith and Arnold attempt to learn commonsense plans about the world from stimuli provide by users via the web, which they call "stories", such as those found in Example 3. These "stories" read more like lists of instructions, and probably would not pass a "narrative test", namely, if a large number of people were asked to judge whether or not these items were stories.

After simple scenarios, the second and third most common text types by corpus count are Fables and Folktales, respectively. Together they account for as much as simple scenarios, around 30%. These artifacts are clearly stories, even when the fables are created by experimenters directly, or adapted from existing folktale or fable collections.

3.4 Annotation Styles

I also tabulated the styles of annotation. Here I determined if the corpus had been annotated by the authors themselves, by independent annotators, by experimental subjects, by automatic analyzers, by a process that also simultaneously generated the narratives, or if the annotations were given in the original material. This breakdown is given in Table 4. There were 155 annotated corpora, with 32 corpora being annotated by an unchecked or non-human generated annotation. The key observation of this data is that a majority of corpora involve author-generated annotations. I will discuss problems with this in Section 5.3.

4 Annotation Analysis

Because narrative corpora have their greatest utility for narrative science when they have been annotated for various relevant features, I spent a significant amount of effort cataloging and categorizing the different types of annotations that were applied to the corpora identified in the survey. This classification was done by me personally, and not by my research assistants. Although, there were large swaths of commonality between types of annotation layers, finding correspondences is a tricky process requiring careful attention to the form, content, and purpose of the annotation layers.

Annotation Style	**#**	**%**	
Authors	92	59 %	
Annotators	27	17 %	
Subjects	16	10 %	
Automatic	16	10 %	25 corpora (15 %) between these three styles, jointly labeled “Other”
Generated	5	3 %	
Given	4	3 %	
Unannotated	12	8 %	
Total:	**204**		

Table 4: Distribution of corpora amongst annotation styles. The total is greater than 167 because some corpora use multiple styles.

By way of example, take two example corpora: a corpus of two simple scenarios (about 20 words) expressed in a logic encoding described by Mueller (2007) and the corpus of over one million news articles described by Chambers and Jurafsky (2009), where the texts were run through an automatic NLP tool chain. To the untrained eye, these annotation suites might seem completely distinct. On closer inspection, however, we find that, although the representational formalisms are significantly different, both corpora sought to capture similar sorts of semantics from the texts. Mueller’s formalism was the event calculus, which encoded actions performed in the story. Each action was assigned a type, had participants, and was pinned to a particular time point on the timeline. The participants were uniform across the whole story. Chambers and Jurafsky did standard semantic role labeling (Palmer, Kingsbury et al. 2005), ran their text through a co-reference resolver, and then through a relationship detection system that sought to impose a partial order on the role-labeled verbs. The both can be described as being concerned with the following sorts of information:

1. Events, where an event is either defined as an action of a character or a verb
2. Referring Expressions and Coreference, where characters or objects in the story are reified and their mentions are either explicitly (through links) or implicitly (via uniform naming) grouped together
3. Roles, where agents, patients, or other roles for events are identified, and their fillers marked.
4. Time, in which events are ordered, fully or partially, on a timeline

Mueller’s study additional is concerned with Cause, where causal linkages between actions are encoded in the logic representation.

I used state-of-the-art NLP tasks, such as semantic role labeling, co-reference resolution, word sense disambiguation, and so forth, as a rough guide for classification of annotation layers. Using this procedure I cataloged, classified, and counted the annotation layers for each corpus. In this analysis I ignored some common annotations, in particular, syntactic analyses that can be done mostly automatically and almost completely accurately by machine for fluent English text: tokens, part of speech tagging, lemmatization, sentence boundary detection, and CFG parses I all considered to be "default" annotations that could be automatically applied to any corpus with little effort.

This analysis produced a list of 23 different layers of annotation, with a 24th category "Other" that contained nineteen unique layers not found in any other corpus. Six layers had sub-type layers, which were annotated on enough corpora to comprise their own distinct grouping. For example, some corpora, when encoding referring expression and coreference, only did so for the main characters (or other named entities) present, not for all things mentioned. This strategy was usually pursued as a way of lessening the annotation workload. Table 5 shows the full list of annotation layers (except for the one-off layers). Figures 2 & 3 show the corpora from the survey charted against the number of stories and the number of words. Note that 55 corpora (33%) were not described in enough detail to estimate word counts, and so these are excluded from Figure 3. In addition to the 24 layer types identified, I also grouped the layers into four broad categories: Fabula-relevant, Narrative Features, Other Common Layers, and One-Off Layers. Detailed explanations of these groups, the different layers, and what they capture are in the following subsections.

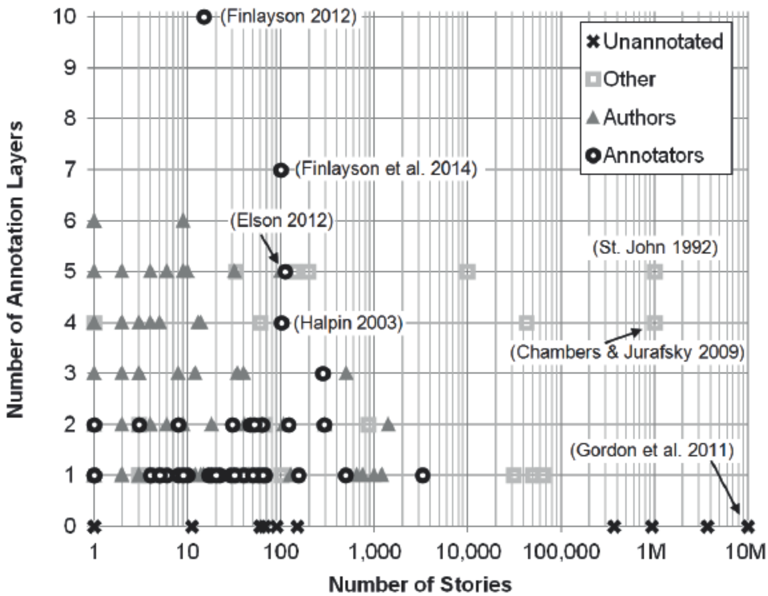

Figure 2: Number of annotation layers for each corpus, against the number of stories, categorized by annotation style. See §3.4 for an explanation of the annotation styles. The notable corpora mentioned by name here are described in Table 6

	Layer Type	#	%	Layer Subtype	#	%
1	Events	56	34 %			
2	Coreference	55	33 %	for Named Entities only	44	26 %
3						
4	Event Roles	32	19 %			
5	Cause	28	17 %			
6	Narrative Features	27	16 %	Suspense-Related Features	6	4 %
7	Other	19	12 %			
8	Beliefs, Desires, Intensions	18	11 %			
9	Narrative Structures	19	11 %	Archetype markings only	2	1 %
10	Relationships	14	8 %	Social relationships only	3	2 %
11	Time	11	7 %			
11	Emotions	10	6 %			
12	Spatial Relationships	10	6 %			
13	Properties / Attributes	9	5 %	on Named Entities only	5	3 %
14	Questions &Answers	9	5 %			
15	Senses	7	4 %	on Named Entities or Verbs only	2	1 %
16	Discourse Structure	6	4 %			
17	Topics	4	2 %			
18	Argumentation Structure	3	2 %			
19	Dialogue	3	2 %			
20	Images	3	2 %			
21	Metadata	3	2 %			
22	Gestures	2	1 %			
23	Summaries	2	1 %			
24	Videos	2	1 %			

Table 5: Distribution of corpora amongst annotation layers. The left-hand list gives the main annotation layers, including the "Other" category. Many corpora have more than one layer. The right-hand list shows layers subtypes that are counted as part of their superordinate category, but have their own distinct identity

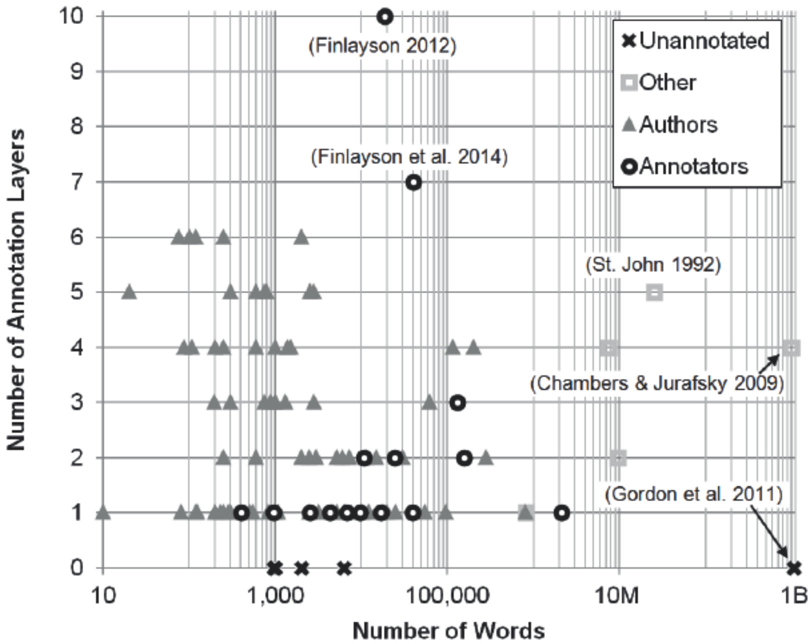

Figure 3: Number of annotation layers for each corpus, against the number of words, categorized by annotation style. Not all corpora are represented in this chart, as word counts could not be estimated for all corpora. See §3.4 for an explanation of the annotation styles. The notable corpora mentioned by name here are described in Table 6.

4.1 Capturing the Fabula: "Who Does What to Whom"

The first broad group of annotation layers comprises those layers that are relevant to the representation of the fabula of the narratives. These include layers that mark events, entities (e.g., coreference or referring expressions), event roles, links between events (such as cause or time), word meaning, and static relationships between characters. Figure 4 shows a scatter plot of corpora sizes (in number of stories) against the number of fabula-relevant annotation layers present in the corpus.

Of these layers, by far the most common are the annotation of events and the annotation of coreference—roughly corresponding to the atomic elements of the plot and the characters of the story. Approximately one-third of the corpora have both an event and coreference annotation.

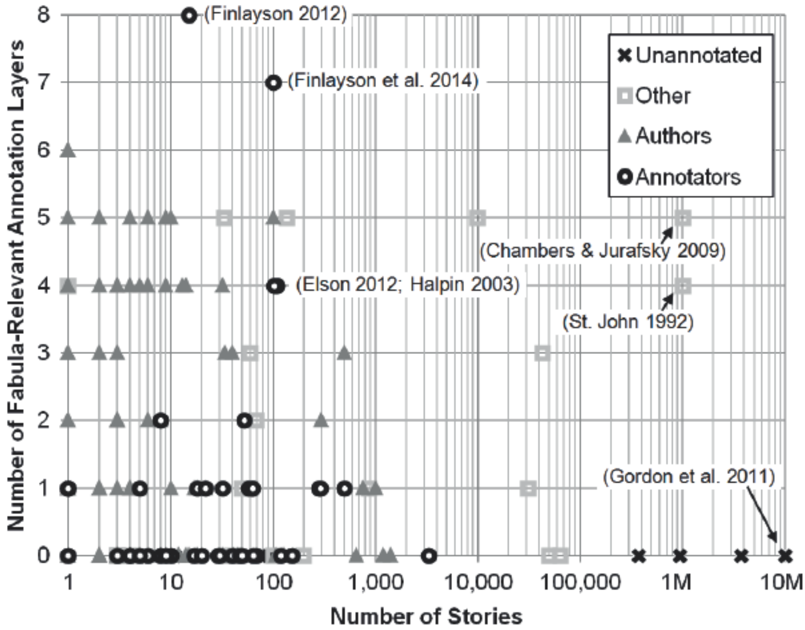

Figure 4: Number of annotation layers relevant to the fabula for each corpus, against the number of stories, categorized by annotation style. See §3.4 for an explanation of the annotation styles. The notable corpora mentioned by name here are described in Table 6.

Closely associated with these annotations are roles and cause, coming in at just less than 20 % of all the corpora each. Roles connect events to objects (the coreference clusters), and causal links connect events to each other, chaining them up in cause-and-effect sequences.

Beliefs, Desires, and Intensions (BDIs) are another relatively common type of annotation, found in much of the work using planning for narrative generation or narrative understanding (e.g., Charles, Lozano et al. 2003, Riedl and Young 2010, Bahamón and Young 2013). Planning frameworks not only encode events and their temporal or causal relationships, but they also encode the goals, plans, intensions, beliefs, and desires of characters in the story, to better represent both our understanding and to generate realistic narratives.

Temporal relationships are also of great relevance to the fabula: annotation layers that fall into this category capture strict or partial orderings of events on a timeline, or relationships between individual events or happenings in the story world.

Relationships other than cause and time are also relevant to understanding the fabula of a story. Here I include such things such as familial or stative relationships, attributes and properties, and spatial relationships.

Finally, the semantic sense of items in other representational layers, be they words, events, actors, or things, is critical to an accurate representation of the fabula. Surprisingly, very few corpora (seven corpora, or 4 %) contained any serious annotation of sense.

4.2 Narrative Features

There is a second main group of annotation layers, that of layers relevant to narratives as narratives *per se*. This distinction mirrors the one made in (Finlayson & Corman, 2014, this volume), between Level I and Level II narratives. Level I narratives are plain action discourses, but Level II narratives are action discourses with additional narrative structure. The first of these types of annotation layers are direct narrative features, which seek to capture some binary or scalar property of texts that is closely related to their narrative nature. These include, for example:

- conflict metrics (Ware, Young et al. 2012)
- point of view (Wiebe 1994)
- novelty (Peinado, Francisco et al. 2010)
- genre/voice features (Garzotto, Herrero et al. 2010)
- marking of whether or not a passage can be considered a "story" (Ceran, Karad et al. 2012)

A clear subtype of this class is the marking of narrative features relevant to suspense in the narrative, which was found on six corpora.

The second type of narrative-relevant annotation layer was direct annotation of narrative structure. This include things such as:

- Plot unit analyses (Lehnert 1981)
- Proppian functional analyses (Lendvai, Declerck et al. 2010, Finlayson 2012)
- Narrative content units (Passonneau, Goodkind et al. 2007)

The third type of narrative-relevant annotation is the marking of emotion, in particular, the emotion that a particular word, phrase, statement or passage might raise in a reader of the text.

Finally, discourse structure was often a concern. Discourse structure—by which I mean a PDTB (Prasad *et al.* 2014) or RST (Mann & Thompson 1987) analysis, or something similar—is relevant to the narrative itself because it tries to represent how the fabula is transformed into the actual text, and how these parts interlock and support one another.

4.3 Other Common Annotation Layers

The third major group of annotation layers mostly concerned connecting narratives to practical application. These included the following types:

A number of corpora were annotated with questions and answers, in an attempt to understand, for example, the inferences that a reader might make when reading a narrative, or the types questions that would naturally arise from the information presented.

A related annotation type was summaries, where narratives would be associated with summarized versions of themselves, usually for testing summarization algorithms or narrative understanding systems.

Related to senses is the concept of *topics*, where portions of narratives, or the whole narratives themselves, were marked for its relevance to a particular topic. These annotation were usually relevant to information retrieval applications

There is a small but robust community in computational narrative that works on the use of narrative in law, and thus often finds useful annotation of argumentation structure, which lays out argumentation assumptions, implications, and dependencies.

Those interested in interactive narrative systems and training computers to speak properly with humans often sought annotations of dialogue in narratives. This found form as association of dialogue with the speaker, or annotation of features of the dialogue such as to whom it was address or the general purpose of the utterance (speech acts).

There is another community that is interested in the intersection of narrative with visual representations, be those representations static, such as photographs, cartoons, or other images, or dynamics, such as videos or storytelling gestures.

Finally, some fields are concerned not only with the contents of the narratives themselves, but by whom they were told, where they were written down, in what era, and so forth. This sort of metadata is useful to Folklorists, for example, who are interested in tale provenance and evolution (Broadwell and Tangherlini 2012).

4.4 Miscellaneous Annotation Layers

There were nineteen other annotation layers that only occurred once in the entire survey, and did not seem to fit well with any other already-established annotation layer type. These layers ran the gamut, including things such as:

- encodings of world knowledge (Frank, Koppen et al. 2008)
- force dynamics (Ontañón and Zhu 2010)
- camera shots (Jhala and Young 2009)
- strategies (Burke and Kass 1995)
- dimensions of psychological models, such as the event indexing model or situation model (Zwaan, Langston et al. 1995, Zacks, Speer et al. 2009)
- internal structure of referring expressions (Hervás and Finlayson 2010)

A full list of other annotation layers, and their associated citations, can be found in the supplemental materials for this paper.

4.5 Notable Corpora

There are several notable corpora indicated in Figures 2, 3, and 4. These corpora are listed in the next table, although with key features. They are notable either because large in terms of number of words or texts, or are quite deeply annotated.

Reference	Description	# Texts	# Words	# Layers	Anno. Style
Chambers and Juraf-sky 2009	Google Gigaword newspaper article corpus	~1M	~1B	4	Automatic
Elson 2012	DramaBank: fairy tales, news articles, epic poetry, plus others	110	?	5	Annotated
Finlayson, Halverson et al. 2014	Islamist extremist stories	100	40,480	7	Annotated
Finlayson 2012	Russian folktales analyzed by Vladimir Propp	15	18862	10	Annotated
Gordon, Bejan et al. 2011	Stories extracted from blog posts on the internet	~10M	>1B	0	-
Halpin 2003	Children's stories written in the classroom	102	~20k	4	Automatic
Orkin, Smith et al. 2010	Data generated by people playing The Restaurant Game (game logs)	9882	n/a	4	Generated
St. John 1992	Simple generated scenarios	1M	~25M	4	Generated

Table 6: Corpora notable for their size or depth of annotation.

5 Gaps

The detailed survey reported here leads me to observe several major gaps in the corpora available for the science of narrative. Each gap begets a recommendation for future researchers when conducting corpus-driven narrative science.

5.1 The Non-Narrative Nature of Narrative Corpora

The most serious gap identified is that many of the narrative corpora used contain texts which might not be considered "narratives" at all. Fully 46 of the corpora discovered (28 %) consist of what I describe as "simple scenarios", some examples of which are given in Section 3.3. If narrative at all, these texts usually only fall into what I would call "Level I" narratives (Finlayson 2014, this volume), which consist of action discourses on a specific situation and coherent topic. To literary theorists, narratologists, and even sophisticated readers, these simple scenarios hardly rise to the level of narratives in any interesting sense of the word.

To be fair, one major research problem in this area is understanding the fabula. Researchers struggle to achieve a basic understanding of *who is doing what to whom* in the stories their systems read, and also struggle to make effective use of this information, connecting it to inferences and conclusions via a commonsense understanding of the world. Therefore, it is reasonable to expect that these researchers would naturally bias toward fabula-focused stories, trying to cut out confounding factors (such as “Level II” narrative complexity).

If one includes in one’s count the next two largest text types, *fables* and *folktales*, the count of corpora reaches 95, or 58 % of the total. While fables and folktales are most certainly narratives by any measure, they are still narratively relatively simple. What all of this means is that the narrative nature of much of the material that underlies computational and cognitive narrative science is narratively light or, at worst, suspect.

This less-than-narrative character of these corpora is not without cost. One problem seems to be that it widens a divide between humanists interested in sophisticated Level II narratives and scientists and computationalists blocked by barriers at Level I. This makes these large Level I corpora useful to mainly to scientists and computationalists, reducing the opportunities for collaboration and joint insight with humanists.

This leads me to a first recommendation, which is that corpus builders should make every effort to move away from Level I type narratives to more sophisticated Level II narratives, that is, texts with significant narrative complexity. Corpus builders must of course still attend to the immediate needs of the research for which they are building the corpus. But they should not immediately drop to the lowest common denominator of simple Level I narratives. That approach results in corpora which are useful for only a narrow range of research. Instead, researchers should spend as much effort as they can afford, up front, to identify narratives that are not only useful to their own purposes, but also will be of interest to those work at all levels of the narrative spectrum.

5.2 A Focus on the Fabula, and Lack of Narrative Features Standard

As was noted in Section 4.1, and hand-in-hand with the non-narrativity of the identified corpora, approximately a third of the corpora contain an annotation of the core of the fabula, namely: events and coreference. Approximately ⅔ of those (20 % of the total) contain cause and role information. This serves as an admirable foundation for future work with these corpora, and points the way forward to a core set of annotation layers that all corpus builders should seek to include in their resources.

But there are two major gaps here. First, there is a surprising lack of focus on meaning and sense at all levels of annotation. At the lower level of meaning, at the word and event level, only seven corpora (4 %) have any real attempt to annotate sense! (And two of these corpora only annotate those senses on named entities or verbs.) This means that while the core structure of the fabula is well represented, we have almost no idea what the words themselves mean. Structure is important in narrative, no doubt, but narratives are rich with meaning, and their purpose is the communication of particular meanings. If we don’t annotate this meaning, then we are missing much of what narrative is all about.

At a higher level, there is even less attention to the “meaning” of narratives as concerns humanists or literary theorists. Here we might seek, for example, meaning relevant to its impact on people, on emotions, on what we learn, or on what we believe. Only a handful of corpora actually attempt to capture these layers. We see ten corpora annotated with emotion-related information (6%), and one corpus annotated with moral information (Dehghani, Tomai et al. 2008). But beyond this, the meaning relevant to those who study narratives in their natural context is lost.

This leads to a second, related point, in that while there is rough agreement on the sorts of things that must be annotated to capture the core of the fabula, there is far from any consensus on what features should be annotated to capture the core "narrative-specific" features of texts, such as narrativity (Abbott 2014), conflict or suspense, or eventfulness (Huhn 2014; Labov 2006),. Here I think of things like narrative identity, distance, and point of view, or the relationship between narrative time and discourse time, or narrative levels, or encodings of audience response (e.g., emotion).

The recommendation here is for researchers to make efforts to move beyond the fabula, and include in their analysis and data collection layers that touch on information other than just fabula structure. First, there is a surprising surfeit of sense-related information, importantly at the word level, and interestingly at the higher, social and individual meaning level. Second, at the very least, we need to start a discussion in the field about what core narrative features we need to really represent the rough approach of the narrative Here, an initial proposal has come from Mani (2013), where he has surveyed quite a bit of narratological theory and laid out a proposal for annotation of a variety of narrative features that are found across works in the area.

5.3 Author-Provided Annotations

One of the biggest problems in the narrative corpora identified in this survey is the prevalence of author-provided annotations. Almost 60 % (92 corpora) of the resources identified were annotated either partially or fully by the authors. This continuing reliance on annotation by authors themselves is disappointing and worrisome in an age of sophisticated natural language processing technology and deep understanding of the methods, techniques, and metrics for building reliable, useful annotated corpora. Author-provided annotations, while understandable from a cost and labor point of view, are damaging from a scientific point of view. I am not saying that author-provided annotations do not have their place (indeed, they are often key to humanist approaches, and for developing initially adequate analytical categories for later annotation). But for the production of a corpus that is to be used as data to support or reject a hypothesis, author-provided annotations break the principle of double-blind science, where the instruments for the collection of the data themselves are not invested in the outcome of experiment. For a field like physics or chemistry, the instruments for data collection are inanimate; for language-related sciences like the study of narrative, the data are the annotations, and the instruments are usually people. When the data are produced by people, they should be produced by independent annotators specifically trained to a reliable, documented standard. Not doing so not only undermines the results of the specific studies supported by such data, but undermines the field as a whole: it prevents us from shaking off our "toy demonstration" roots from the early days of AI, and becoming scientists in the full fledged used of the term. Researchers themselves should not be doing annotation, as it makes it easy to "bake in the answer" to the data (accidentally or not). The recommendation for this gap is clear: researchers must spend the extra effort to hire, train, and engage independent annotators to produce their data. Anything less becomes a major chink in the armor of the work.

5.4 The Focus on Text

Less of a problem, and more of a consistent bias to be aware of, is the overwhelming focus on textual narratives. 152 of the corpora (92 %) used the textual modality. This is understandable for several reasons, in particular: the ease with which text can be input and read out from computers; the significant amount of technology available for processing and understanding text; and the large number of narratives available in text form. However, narratives are told across a number of different modalities, and it should be a goal of the field to develop analysis tools for, and to test our theories on, these differently-moded narratives. Examples of alternate modalities represented in this survey include video (Mbengue 2013), image (Lee et al. 2012), speech (Tomai et al. 2011), and game logs (Orkin et al. 2010), but many other modalities are possible.

5.5 Practical Problems

In addition to the higher-level concerns presented in the previous sections, there are a number of practical problems with many of the corpora, which lessens their utility for scientific work.

First, 42 corpora (25 %) with textual data don't report their results in such a way that the word counts may be estimated. An even larger fraction (approximately 50 %) don't report exact word counts in the paper, forcing downstream users to estimate the counts from clues. Word counts are a key piece of information for textual corpora, and should always be reported. Ideally, not only the total word count for the corpus, but information on the distribution of word counts among documents or stories should also be included.

Second, fully 69 corpora (41 %) have fewer than ten stories. 51 % (85) have fewer than 20 stories. This means that a large fraction of scientific work on narrative has been done on narrative corpora that are quite small. This raises grave questions about the scability and generalizability of the results in those works.

Third, few corpora or narrative sets are released in downloadable format. I found only four corpora in my survey that included a URL in published work from which the corpus could be directly downloaded, even though 114 corpora (68 %) were created in the year 2000 or later. Moreover, only 89 corpora (53 %) report enough detail to actually identify the stories used in the work. The cause of these problems may be oversight, lack of a facility to properly archive such data, or copyright issues. Regardless, this failure to release corpora to the community means that later scientific work is severely hampered in taking advantage of these resources.

6 Contributions & Future Work

Narrative corpora are a key component for the advancement of the science of narrative. In this work, I have made three contributions that support the use of narrative corpora for future work in this area.

First, I have conducted the first large-scale survey of narrative corpora used in the computational and cognitive study of narrative. This includes the identification of a large bibliography of 2601 works along with the identification of over 180 works that use 167 sets of stories that could be called "corpora".

Second, I have identified a range of dimensions for narrative corpora, tabulating the values of each corpus along those dimensions. These include key dimensions such number of stories, number of words, number and types of annotation layers, and so forth. This survey gives us the first relatively representative view of the use of narrative corpora in computational and cognitive narrative research.

Finally, I have identified several major problems in the use of narrative corpora, and proposed recommendations for future work to address these gaps. These gaps include:

1. The non-narrative nature of many "narrative" corpora
2. The focus on fabula structure at the expense of fabula meaning or more narratively interesting features
3. The overwhelming reliance on annotations created by the authors themselves
4. The overwhelming focus on narratives in text form, neglecting other modalities
5. A number of practical problems, such as not reporting key information for the corpora, focusing on too-small corpora, and not making the data available.

By attending to these gaps, and the recommendations for future work that naturally follow from them, narrative researchers will not only improve their own work, but also advance the field, allowing us to truly bring the study of narrative into a scientific frame of mind.

7 Acknowledgements

The preparation of this article was supported by the U.S. Defense Advanced Research Project Agency (DARPA) under contract number D12AP00210. The views expressed here are solely mine, and do not necessarily reflect those of DARPA, the U.S. military, or the U.S. government. I thank my research assistants Benjamin Frank, Brandon Irving, and Sharon Mozgai for their invaluable help in locating and organizing the references used in this study.

References

Abbott, H. Porter. (2008). The Cambridge introduction to narrative (2nd ed.). Cambridge, UK: Cambridge University Press.

Bahamón, J. C. and R. M. Young (2013). CB-POCL: A Choice-Based Algorithm for Character Personality in Planning-based Narrative Generation. Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13). M. A. Finlayson, B. Fisseni, B. Löwe and J. C. Meister. Dagstuhl, Germany, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 32: 4-23.

Bal, M. (2009). Narratology: Introduction to the Theory of Narrative (3rd ed.). Toronto, Canada, University of Toronto Press.

Barber, H. and D. Kudenko (2007). Interactive Generation of Dilemma-based Narratives. Proceedings of Narrative AI and Intelligent Serious Games for Education Symposium. AISB'07, Newcastle University, Newcastle upon Tyne, UK.

Bird, S. and M. Liberman (2001). "A formal framework for linguistic annotation." Speech Communication 33(1-2): 23-60.

Bordwell, D. (1985). Narration in the fiction film. Madison, WI, USA, The University of Wisconsin Press.

Broadwell, P. M. and T. R. Tangherlini (2012). TrollFinder: Geo-Semantic Exploration of a Very Large Corpus of Danish Folklore. Third Workshop on Computational Models of Narrative (CMN). M. A. Finlayson. Istanbul, Turkey, European Language Resources Association (ELRA): 52-59.

Burke, R. and A. Kass (1995). "Supporting Learning Through Active Retrieval of Video Stories." Journal of Expert Systems with Applications 9(3): 361-378.

Cavazza, M., R. Champagnat, R. Leonardi and T. I. Consortium (2009). The IRIS Network of Excellence: Future Directions in Interactive Storytelling. Proceedings of the Second Joint International Conference on Interactive Digital Storytelling (ICIDS 2009). I. A. Iurgel, N. Zagalo and P. Petta. Guimarães, Portugal, Springer: 8-13.

Ceran, B., R. Karad, S. Corman and H. Davulcu (2012). A Hybrid Model and Memory Based Story Classifier. Third Workshop on Computational Models of Narrative (CMN). M. A. Finlayson. Istanbul, Turkey, European Language Resources Association (ELRA): 60-64.

Chambers, N. and D. Jurafsky (2009). Unsupervised Learning of Narrative Schemas and their Participants. 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. Suntec, Singapore: 602-610.

Charles, F., M. Lozano, S. J. Mead, A. F. Bisquerra and M. Cavazza (2003). Planning Formalisms and Authoring in Interactive Storytelling. Proceedings of Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2003). S. Göbel, N. Braun, U. Spierling, J. Dechau and H. Diener. Darmstadt, Germany, Fraunhofer IRB Verlag: 216-225.

Dehghani, M., E. Tomai, K. Forbus and M. Klenk (2008). An Integrated Reasoning Approach to Moral Decision-Making. Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI). Chicago, IL, USA: 1280-1286.

Doran, J. and I. Parberry (2011). A Prototype Quest Generator Based on a Structural Analysis of Quests from Four MMORPGs. Proceedings of the Second International Workshop on Procedural Content Generation in Games (PCGames 2011). Bordeaux, France, ACM: Article No. 1.

Elson, D. K. (2012). DramaBank: Annotating Agency in Narrative Discourse. Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC). Istanbul, Turkey: 2813-2819.

Finlayson, M. A. (2012). Learning Narrative Structure from Annotated Folktales. Ph.D. Dissertation, Massachusetts Institute of Technology.

Finlayson, M. A., J. R. Halverson and S. R. Corman (2014). The N2 Corpus: A Semantically Annotated Corpus of Islamist Extremist Stories. Proceedings of the Language Resources and Evaluation Conference (LREC). Reykjavík, Iceland: 896-902.

Frank, S. L., M. Koppen, L. G. M. Noordman and W. Vonk (2008). "World Knowledge in Computational Models of Discourse Comprehension." Discourse Processes 45(6): 429-463.

Garzotto, F., E. Herrero and F. Salgueiro (2010). One Tool-Many Paradigm: Creativity and Regularity in Youngsters' Hyperstories. Proceedings of the Third Joint International Conference on Interactive Digital Storytelling (ICIDS 2010). R. Aylett, M. Y. Lim, S. Louchart, P. Petta and M. Riedl. Edinburgh, UK, Springer: 44-49.

Gordon, A. S., C. A. Bejan and K. Sagae (2011). Commonsense causal reasoning using millions of personal stories. Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI '11). San Francisco, CA, USA: 1180-1185.

Ha, E. Y., J. P. Rowe, B. W. Mott and J. C. Lester (2011). Goal Recognition with Markov Logic Networks for Player-Adaptive Games. Proceedings of the Seventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-11). Stanford, CA, USA, AAAI Press: 32-39.

Halpin, H. R. (2003). The Plots of Children and Machines: The Statistical and Symbolic Semantic Analysis of Narratives. MS Master of Science Thesis, University of Edinburgh.

Hervás, R. and M. A. Finlayson (2010). The Prevalence of Descriptive Referring Expressions in News and Narrative. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Short Papers (ACL 2010). Uppsala, Sweden: 49-54.

Hühn, Peter: "Event and Eventfulness". In: Hühn, Peter et al. (eds.): The Living Handbook of Narratology. Hamburg: Hamburg University. http://www.lhn.uni-hamburg.de/article/event-and-eventfulness [view date:19 May 2014]

IRIS (2014). IRIS Bibliography on Interactive Storytelling. http://iris.ofai.at:7777/iris_db/.

Jhala, A. and R. M. Young (2009). Comparing Effects of Different Cinematic Visualization Strategies on Viewer Comprehension. Proceedings of the Second Joint International Conference on Interactive Digital Storytelling (ICIDS 2009). I. A. Iurgel, N. Zagalo and P. Petta. Guimarães, Portugal, Springer: 26-37.

Knapp, D. (2009). Using Analogies across Narratives to drive Dialogue. Proceedings of the 2nd Workshop on Intelligent Narrative Technologies (INT2). S. Louchart, M. Mehta and D. L. Roberts. AAAI Spring Symposium Series, Palo Alto, CA, USA, AAAI Press. 2: 67-74.

Labov, W. (2006). Narrative pre-construction. Narrative Inquiry, 16 (1), 37–45.

Lee, C., S. Muresan, et al. (2012). Computational Analysis of Referring Expressions in Narratives of Picture Books. 1st Workshop on Computational Linguistics for Literature (CLfL 2012): 1-7.

Lee, S. Y., B. W. Mott and J. C. Lester (2011). Director Agent Intervention Strategies for Interactive Narrative Environments. Proceedings of the Fourth Joint International Conference on Interactive Digital Storytelling (ICIDS 2011). M. Si, D. Thue, E. André et al. Vancouver, Canada, Springer: 140-151.

Lehnert, W. G. (1981). "Plot Units and Narrative Summarization." Cognitive Science 5(4): 293-331.

Lendvai, P., T. Declerck, S. Darányi, P. Gervás, R. Hervás, S. Malec and F. Peinado (2010). Integration of Linguistic Markup into Semantic Models of Folk Narratives: The Fairy Tale Use Case. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC). Valetta, Malta, European Language Resources Association (ELRA): 1996-2001.

Lin, Ziheng, H.T. Ng and M.-Y. Kan (2014). A PDTB-Styled End-to-End Discourse Parser. Natural Language Engineering, 20, pp 151-184.

Mani, I. (2013). Computational Modeling of Narrative, Morgan & Claypool Publishers.

Mann, William C. and Sandra A. Thompson (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. Text 8 (3): 243-281.

Mbengue, C., D. A. Gómez Jáuregui, et al. (2013). CONTACT: A Multimodal Corpus for Studying Expressive Styles and Informing the Design of Individualized Virtual Narrators. Proceedings of the 6th INT Workshop, Boston, MA, USA: 86-88.

Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, T. G. B. Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak and E. L. Aiden (2011). "Quantitative Analysis of Culture Using Millions of Digitized Books." Science 331(6014): 176-182.

Mueller, E. T. (2007). Understanding Goal-Based Stories through Model Finding and Planning. Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies. B. S. Magerko, M. O. Riedl, B. Loyall, R. M. Young and M. Mateas. Arlington, VA, USA, AAAI Press. 1: 95-101.

Mueller, E. T. (2014). Story Understanding Resources. http://xenia media mit.edu/~mueller/storyund/storyres html.

Oepen, Stephan (2007). Beyond the Science of the Wall Stree Journal. Unified Linguisitic Annotation Workshop, Bergen, Norway.

Ontañón, S. and J. Zhu (2010). Story and Text Generation through Computational Analogy in the Riu System. Proceedings of the Sixth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-10). Stanford, CA, USA, AAAI Press: 51-56.

Orkin, J., T. Smith, H. Reckman and D. Roy (2010). Semi-Automatic Task Recognition for Interactive Narratives with EAT & RUN. Proceedings of the Intelligent Narrative Technologies III Workshop (INT3). A. Jhala, M. O. Riedl and D. L. Roberts. 5th

International Conference on the Foundations of Digital Games Conference (FDG 2010), Monterey, CA, USA, ACM Press: Article No. 3.

Palmer, M., P. Kingsbury and D. Gildea (2005). "The Proposition Bank: An annotated corpus of semantic roles." Computational Linguistics 31(1): 71-105.

Passonneau, R. J., A. Goodkind and E. T. Levy (2007). Annotation of Children's Oral Narrations: Modeling Emergent Narrative Skills for Computational Applications. Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2007). Key West, FL, USA, AAAI: 253-258.

Peinado, F., V. Francisco, R. Hervás and P. Gervás (2010). "Assessing the Novelty of Computer-Generated Narratives Using Empirical Metrics." Minds and Machines 20(4): 565-588.

Prasad, Rashmi, B. Webber & A. Joshi (2014) Reflections on the Penn Discourse TreeBank, Comparable Corpora and Complementary Annotation. Computational Lingustics. doi:10.1162/COLI_a_00204

Propp, V. (1968). The Morphology of the Folktale (2nd ed.). Austin, University of Texas Press.

Riedl, M. O. and R. M. Young (2010). "Narrative Planning: Balancing Plot and Character." Journal of Artificial Intelligence Research 39: 217-268.

Smith, D. and K. Arnold (2009). Learning Hierarchical Plans by Reading Simple English Narratives. Common Sense and Intelligent User Interfaces: Story Understanding and Generation for Context-Aware Interface Design. C. Havasi, H. Lieberman and E. T. Mueller. International Conference on Intelligent User Interfaces (IUI), Sanibel Island, FL, Association for Computing Machinery: 16-19.

St. John, M. F. (1992). "The Story Gestalt - A Model of Knowledge-Intensive Processes in Text Comprehension." Cognitive Science 16(2): 271-306.

Tomai, E., Thapa, L., Gordon, A. S., & Kang, S.-H. (2011). Causality in Hundreds of Narratives of the Same Events. Papers from the 2011 AIIDE Workshop, Intelligent Narrative Technologies IV (INT 4), Stanford, CA, USA: 77-83.

Ware, S. G., R. M. Young, B. E. Harrison and D. L. Roberts (2012). Four Quantitative Metrics Describing Narrative Conflict. Proceedings of the Fifth Joint International Conference on Interactive Digital Storytelling (ICIDS 2012). D. Oyarzun, F. Peinado, R. M. Young, A. Elizalde and G. Méndez. San Sebastián, Spain, Springer: 18-29.

Wiebe, J. M. (1994). "Tracking Point of View in Narrative." Computational Linguistics 20(2): 233-287.

Zacks, J. M., N. K. Speer and J. R. Reynolds (2009). "Segmentation in Reading and Film Comprehension." Journal of Experimental Psychology: General 138(2): 307-327.

Zwaan, R. A., M. C. Langston and A. C. Graesser (1995). "The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model." Psychological Science 6(5): 292-297.

Appendix: Comprehensive List of Narrative Science Meetings

	Year	Abbr	Name	Location	# Papers
1	1987	IGEL 1987	Empirical Study of Literature and Media	Siegen, Germany	-
2	1990	IGEL 1990	Empirical Study of Literature and Media	Amsterdam	-
3	1992	IGEL 1992	Empirical Study of Literature and Media	Memphis	-
4	1994	IGEL 1994	Empirical Study of Literature and Media	Budapest	-
5	1995	-	Interactive Story Systems	Palo Alto	32
6	1996	IGEL 1996	Empirical Study of Literature and Media	Nakoda, Canada	-
7	1998	IGEL 1998	Empirical Study of Literature and Media	Utrecht	-
8	1999	-	Interactive Frictions: An International Conference on Interactive Narrative	Los Angeles	-
9	1999	-	Creative Language: Humour and Stories	Sussex	12
10	1999	-	Narrative Intelligence	N. Falmouth, MA	28
11	2000	IGEL 2000	Empirical Study of Literature and Media	Toronto	-
12	2000	NILE 2000	Narrative in Interactive Learning Environments	Edinburgh	11
13	2001	ICVS 2001	Virtual Storytelling	Avignon	24
14	2002	-	Storytelling in Collaborative Virtual Environments	Bonn	12
15	2002	IGEL 2002	Empirical Study of Literature and Media	Pecs, Hungary	-
16	2002	NILE 2002	Narrative in Interactive Learning Environments	Edinburgh	9
17	2003	ICVS 2003	Virtual Storytelling	Toulouse	30
18	2003	TIDSE 2003	Technologies for Interactive Digital Storytelling	Darmstadt	41
19	2004	IGEL 2004	Empirical Study of Literature and Media	Edmonton	-
20	2004	NILE 2004	Narrative in Interactive Learning Environments	Edinburgh	15
21	2004	SRMC'04	Story Representation, Mechanism, and Context	New York	13
22	2004	TIDSE 2004	Technologies for Interactive Digital Storytelling	Darmstadt	41
23	2005	-	Narrative Learning Environments	Amsterdam	8
24	2005	ICVS 2005	Virtual Storytelling	Strasbourg	33
25	2006	IGEL 2006	Empirical Study of Literature and Media	Munich	-
26	2006	NILE 2006	Narrative in Interactive Learning Environments	Edinburgh	11
27	2006	TIDSE 2006	Technologies for Interactive Digital Storytelling	Darmstadt	37
28	2007	ICVS 2007	Virtual Storytelling	Saint Malo	22
29	2007	INT	Intelligent Narrative Technologies	Arlington, VA	28
30	2008	ITIS 2008	Integrating Technologies for Interactive Stories	Playa del Carmen	6
31	2008	ICIDS 2008	Interactive Digital Storytelling	Erfurt	43
32	2008	IGEL 2008	Empirical Study of Literature and Media	Memphis	-
33	2008	NILE 2008	Narrative in Interactive Learning Environments	Edinburgh	13
34	2008	SRMC'08	Story Representation, Mechanism, and Context	Vancouver	9
35	2009	-	DARPA Experience-based Narr. Memory Systems	Arlington, VA	-

	Year	Abbr	Name	Location	# Papers
36	2009	-	Story Understanding and Generation for Context-Aware Interface Design	Sanibel Island, FL	8
37	2009	CMN'09	Computational Models of Narrative	Beverly, MA	-
38	2009	ICIDS 2009	Interactive Digital Storytelling	Guimaraes	50
39	2009	INT2	Intelligent Narrative Technologies	Palo Alto	21
40	2010	-	DARPA Negotiation Across Cultures	Arlington, VA	-
41	2010	CMN'10	Computational Models of Narrative	Arlington, VA	17
42	2010	ICIDS 2010	Interactive Digital Storytelling	Edinburg	50
43	2010	IGEL 2010	Empirical Study of Literature and Media	Utrecht	-
44	2010	INT3	Intelligent Narrative Technologies	Monterey	17
45	2011	ICIDS 2011	Interactive Digital Storytelling	Vancouver	53
46	2011	INT4	Intelligent Narrative Technologies	Palo Alto	19
47	2011		DARPA STORyNET: Stories, Neuroscience and Experimental Technologies	Charlottesville	-
48	2011	-	DARPA Narrative Networks (N2): The Neurobiology of Narratives	San Francisco	-
49	2011	-	DARPA Narrative Networks (N2): Modeling, Simulating and Sensing Narrative Influence	Arlington, VA	-
50	2012	CLFL 2012	Computational Linguistics for Literature	Montreal	14
51	2012	CMN'12	Computational Models of Narrative	Istanbul	30
52	2012	HT 2012	Narrative Connections	Milwaukee	3
53	2012	ICIDS 2012	Interactive Digital Storytelling	San Sebastian	23
54	2012	IGEL 2012	Empirical Study of Literature and Media	Montreal	-
55	2012	INT5	Intelligent Narrative Technologies	Palo Alto	12
56	2013	CLFL 2013	Computational Linguistics for Literature	Atlanta	10
57	2013	CMN'13	Computational Models of Narrative	Hamburg	29
58	2013	ICIDS 2013	Interactive Digital Storytelling	Istanbul	35
59	2013	INT6	Intelligent Narrative Technologies	Boston	20
60	2014	IGEL 2014	Empirical Study of Literature and Media	Turin	-
61	2014	CMN'14	Computational Models of Narrative	Quebec City	tbd
62	2014	CLFL 2014	Computational Linguistics for Literature	Gothenburg	tbd
63	2014	ICIDS 2014	Interactive Digital Storytelling	Singapore	tbd
64	2014	INT7	Intelligent Narrative Technologies	Milwaukee	tbd
				Total Papers:	**889**

Table 7: A list of 64 conferences, workshops, symposia, and other meetings that have been held on the computational or cognitive science of narrative in the past 30 years. Meetings were included in this list only if they exclusively focused on narrative science.The last five conferences have not been held at the time of this writing. Meetings with a dash in the rightmost column lack published papers, instead featuring only talks, posters, or abstracts.