The Third Workshop on Computational Models of Narrative

(CMN'12)

Workshop Programme

Saturday, 26 May 2012

12:30	Welcome and Introduction, *M.A. Finlayson*
13:00	**Invited Keynote**: Crowd Sourcing Narrative Logic: Towards a Computational Narratology with CLÉA, *J.C. Meister*
	Session I: Representation
14:00	Toward Sequencing "Narrative DNA": Tale Types, Motif Strings and Memetic Pathways, *S. Darányi, P. Wittek, L. Forró*
14:20	Computational Models of Narratives as Structured Associations of Formalized Elementary Events, *G.P. Zarri*
14:35	Objectivity and Reproducibility of Proppian Narrative Annotations, *R. Bod, B. Fisseni, A. Kurji, B. Löwe*
14:50	An Experiment to Determine whether Clustering will Reveal Mythemes, *R. Lang, J.G. Mersch*
15:00	In Search of an Appropriate Abstraction Level for Motif Annotations, *F. Karsdorp, P. van Kranenburg, T. Meder, D. Trieschnigg, A. van den Bosch*
15:15	Understanding Objects in Online Museum Collections by Means of Narratives, *C. van den Akker, M. van Erp, L. Aroyo, R. Segers, L. van der Meij, S. Lêgene and G. Schreiber*
15:30	**Best Student Paper on a Cognitive Science Topic**: Indexter: A Computational Model of the Event-Indexing Situation Model for Characterizing Narratives, *R.E. Cardona-Rivera, B.A. Cassell, S.G. Ware, R.M. Young*
15:50	Coffee Break
16:30	Towards Finding the Fundamental Unit of Narrative: A Proposal for the Narreme, *A. Baikadi, R.E. Cardona-Rivera*
16:40	People, Places and Emotions: Visually Representing Historical Context in Oral Testimonies, *A.T. Chen, A. Yoon, R. Shaw*
	Session II: Corpora
16:55	TrollFinder: Geo-Semantic Exploration of a Very Large Corpus of Danish Folklore, *P.M. Broadwell, T.R. Tangherlini*
17:15	A Hybrid Model and Memory Based Story Classifier, *B. Ceran, R. Karad, S. Corman, H. Davulcu*
17:30	A Crowd-Sourced Collection of Narratives for Studying Conflict, *R. Swanson and A. Jhala*
17:50	Towards a Culturally-Rich Shared Narrative Corpus: Suggestions for the Inclusion of Culturally Diverse Narrative Genres, *V. Romero, J. Niehaus, P. Weyhrauch, J. Pfautz, S.N. Rielly*
18:00	Towards a Digital Resource for African Folktales, *D.O. Ninan and O.A. Odejobi*
18:15	Formal Models of Western Films for Interactive Narrative Technologies, *B. Magerko, B. O'Neill*
18:35	End

Workshop Programme

(cont…)

Sunday, 26 May 2012

	Session III: Similarity
9:00	Detecting Story Analogies from Annotations of Time, Action and Agency, *D.K. Elson*
9:20	Story Comparison via Simultaneous Matching and Alignment, *M.P. Fay*
9:35	Similarity of Narratives, *L. Michael*
9:55	Which Dimensions of Narratives are Relevant for Human Judgments of Story Equivalence?, *B. Fisseni, B. Löwe*
10:10	Story Retrieval and Comparison using Concept Patterns, *C.E. Krakauer, P.H. Winston*
10:30	Coffee Break
	Session IV: Generation
11:00	From the Fleece of Fact to Narrative Yarns: A Computational Model of Composition, *P. Gervás*
11:20	Is this a DAG that I see before me?: An Onomasiological Approach to Narrative Analysis and Generation, *M. Levison, G. Lessard*
11:40	Automatically Learning to Tell Stories about Social Situations from the Crowd, *B. Li, S. Lee-Urban, D.S. Appling, M.O. Riedl*
12:00	Prototyping the Use of Plot Curves to Guide Story Generation, *C. León, P. Gervás*
12:10	Simulating Plot: Towards a Generative Model of Narrative Structure, *G.A. Sack*
12:30	Lunch
	Session V: Persuasion
14:30	A Choice-Based Model of Character Personality in Narrative, *J.C. Bahamon, R.M. Young*
14:45	Persuasive Precedents, *F. Bex, T. Bench-Capon, B. Verheij*
15:00	Integrating Argumentation, Narrative and Probability in Legal Evidence, *B. Verheij*
15:10	Arguments as Narratives, *A. Wyner*
15:20	Towards a Computational Model of Narrative Persuasion: A Broad Perspective, *J. Niehaus, V. Romera, J. Pfautz, S.N. Reilly, R. Gerrig, P. Weyhrauch*
15:30	Discussion
16:00	Coffee Break
16:30	Discussion
18:00	End

Editor

Mark A. Finlayson	Massachusetts Institute of Technology, USA

Organizing Committee

Mark A. Finlayson (Chair)	Massachusetts Institute of Technology, USA
Pablo Gervás	Universidad Complutense de Madrid, Spain
Deniz Yuret	Koç University, Turkey
Floris Bex	University of Dundee, UK

Programme Committee

Steve Corman	Arizona State University, USA
Barbara Dancygier	University of British Columbia, Canada
Hasan Davulcu	Arizona State University, USA
David K. Elson	Google, USA
Matthew P. Fay	Massachusetts Institute of Technology, USA
Andrew Gordon	Institute for Creative Technologies, USA
Benedikt Löwe	University of Amsterdam, the Netherlands
Livia Polanyi	LDM Associates, USA
Emmett Tomai	University of Texas-Pan American, USA
Bart Verheij	University of Groningen, the Netherlands
Patrick H. Winston	Massachusetts Institute of Technology, USA
R. Michael Young	North Carolina State University, USA

Table of Contents

Invited Keynote

Author Index

Preface

This workshop is third in a series dedicated to advancing a nascent field: the computationally-grounded science of narrative. The past decade has seen a resurgence of interest in a formal understanding of the phenomenon of narrative. Since 1999 there have been over fifteen conferences, symposia, and workshops (including those in this series) focusing on applying computational and experimental techniques to narrative. The field is strongly interdisciplinary: it has engaged researchers across the humanities, social sciences, cognitive sciences, and computer sciences. With this building momentum the coming years promise great advances in our understanding of the fundamental nature of narrative and its place in human cognition and society.

We call this workshop series *Computational Models of Narrative* because we believe that a true science of narrative must adhere to the principle espoused by Herbert Simon in his book *The Sciences of the Artificial*: that without computational modeling, the science of a complex human phenomenon such as narrative will never be successful. To our mind this expands the workshop's purview beyond the limited body of effort that directly incorporates computer simulation. It gives us a broad mandate to include a great deal of cognitive, linguistic, neurobiological, social scientific, and literary work: indeed, any research where the researchers have successfully applied their field's unique insights to narrative in a way that is compatible with a computational frame of mind. We seek work whose results are thought out carefully enough, and specified precisely enough, that they could eventually inform computational modeling of narrative.

In keeping with interdisciplinary nature of the field, we have made an explicit decision to move around between different communities so as to enhance engagement, cross-pollination, and visibility (at least for the first handful of workshops). Last year we were hosted by the Association for the Advancement of Artificial Intelligence (AAAI). This year we are hosted by the Language Resources and Evaluation Conference (LREC), which is solidly placed in the computational linguistics community. For the next meeting we will likely set our sights on a conference in cognitive science or neuroscience; after that, probably the humanities or the social sciences. In recognition of a critical blocker, the special focus of this edition of the workshop is the identification, collection, and construction of shared resources and corpora that support the computational modeling of narrative. LREC, therefore, was a perfect venue.

This year we are pleased to be giving a best paper award: Best student paper on a cognitive science topic. The award goes to Mr. Rogelio E. Cardona-Rivera, who is a Ph.D. student in computer science at North Carolina State University, for his paper titled "Indexter: A Computational Model of the Event-Indexing Situation Model for Characterizing Narratives," co-authored with Bradley A. Cassell, Stephen G. Ware, and R. Michael Young. This paper award is part of our effort to reach out and engage the diverse communities that are relevant to the computational science of narrative.

We thank our sponsors for their support of the workshop. Their continued interest and generosity allowed us to offer a number of travel grants this year, as well as bring in our invited speaker. They include: Robb Wilcox at the Office of Naval Research Global; Ivy Estabrooke of the Human Social Cultural and Behavioral Sciences Program at the Office of Naval Research; and Bill Casebeer at the Defense Advanced Research Projects Agency. The Cognitive Science Society provided support to offer the best student paper award, in the form of a check and complimentary membership in the society for the next year.

Mark A. Finlayson
Workshop Chair
Cambridge, Massachusetts

Invited Keynote: Crowd Sourcing Narrative Logic: Towards a Computational Narratology with CLÉA

Jan Christoph Meister

Department of Language, Literature and Media I, Faculty of Humanities
Universität Hamburg, Hamburg, Germany
jan-c-meister@uni-hamburg.de

Abstract

I discuss a collaborative, computer aided approach towards building and exploiting a shared resource that can aid further research into the history and development of narrative, as well as into its phenomenology and logic. The particular example illustrating this approach is a project called CLÉA, short for *Collaborative Literature Éxploration and Annotation*.

CLÉA, a Google *Digital Humanities Award* funded project[1], is a browser based annotation and text analysis environment which comprises three functional modules:

1. a working environment for highly unconstrained, non-deterministic collaborative markup of narratives;

2. a repository which manages and distributes object data (texts and corpora, imported among other from Google Books), and meta-data (stand-off markup files) as well as tag sets (consisting of TEI-XML compliant narratological tags) aggregated and generated by users;

3. a heuristic module, based on a machine learning component, that aims to identify and draw the human user's attention to hitherto unnoticed patterns and regularities in the phenomenology of human-made narratives.

In the field of narrative studies, whether philological or computational, CLÉA's crowd sourcing approach is novel in at least two regards: One, it aims at basing our analyses, theories and models of the phenomenon of narrative representation on collaborative corpus studies, rather than on the detailed, in-depth investigation of a small number of exemplary narratives that is undertaken by an individual researcher. Two, it augments human research activity by a computational heuristic that works bottom-up, through the statistical analysis of human generated meta-data, rather than top-down, i.e. by mapping a pre-defined abstract model or taxonomy of narrative onto the original object-data itself. In doing so, this heuristic acknowledges the crucial fact that narratives are always human constructs, and not real-world entities that can be found "out there" and on their own.

However, CLÉA amounts to more than merely facilitating a new praxis of analysing and modeling the symbolic processes and products that human cultures designate as 'narrative'. In a methodological perspective and with regard to the current transformation of the humanities in general, CLÉA also demonstrates the impact and potential relevance of the new scientific paradigm of the Digital Humanities. In a disciplinary perspective on the other hand, and in particular in that of narratology and of computational science, CLA may be considered as an example for an emerging inter-discipline, tentatively labeled by Inderjeet Mani and others as Computational Narratology.[2] Accordingly, I will pay equal attention to the demonstration of CLÉA as a concrete example, and to the reflexion of broader methodological and programmatic consequences. Adopting a computer aided, crowd sourcing based approach in the study and modelling of narrative, I believe, will do more than afford us new insights into the logic of narratives as objects of our research. It will also help us to understand better on which premises our reasoning about narrative is based, and which contingencies and constraints — disciplinary as well as cultural — we might have to take into account in our future work.

[1] CLÉA is the current, second phase development of the initial stand alone desktop application CATMA (Computer Assisted Textual Markup and Analysis) conceptualized and developed at Hamburg University from 2009 onward. In *heure*CLÉA, its third development phase which we plan to commence in 2013, we will extend system support for narratological markup and implement a robust machine learning based heuristic module. For more technical information on CATMA and CLÉA as well as downloads, see `http://www.catma.de`.

[2] See Mani's forthcoming article "Computational Narratology" in Hühn, Peter et al. (eds) *The living handbook of narratology*. Hamburg, Hamburg University Press. `http://hup.sub.uni-hamburg.de/lhn`

Toward Sequencing “Narrative DNA”: Tale Types, Motif Strings and Memetic Pathways

Sándor Darányi, Peter Wittek, László Forró†
Swedish School of Library and Information Science
University of Borås
50190 Borås, Allégatan 1, Sweden
†8220 Balatonalmádi, Remetevölgyi út 27, Hungary

E-mail: sandor.daranyi@hb.se, peterwittek@acm.org, salmonix@gmail.com

Abstract

The Aarne-Thompson-Uther Tale Type Catalog (ATU) is a bibliographic tool which uses metadata from tale content, called motifs, to define tale types as canonical motif sequences. The motifs themselves are listed in another bibliographic tool, the Aarne-Thompson Motif Index (AaTh). Tale types in ATU are defined in an abstracted fashion and can be processed like a corpus. We analyzed 219 types with 1202 motifs from the “Tales of magic” (types 300-749) segment to exemplify that motif sequences show signs of recombination in the storytelling process. Compared to chromosome mutations in genetics, we offer examples for insertion/deletion, duplication and, possibly, transposition, whereas the sample was not sufficient to find inverted motif strings as well. These initial findings encourage efforts to sequence motif strings like DNA in genetics, attempting to find for instance the longest common motif subsequences in tales. Expressing the network of motif connections by graphs suggests that tale plots as consolidated pathways of content help one memorize culturally engraved messages. We anticipate a connection between such networks and Waddington’s epigenetic landscape.

Keywords: tale type, motif, motif sequence, mutation, recombination, plot development, memetic pathway, epigenetic landscape

1. Introduction

Recently Darányi (2010) has analyzed the role of formulaity in oral and written narratives, and hinted at a parallel with sublanguages for indexing (Harris, 2002) also used in immunology (Harris *et al*., 1989) and bioinformatics (Leontis & Westhof, 2003). The similarity between these wildly different application domains goes back to the use of motifs. In the literary sense, a motif is an instance of a prominent yet little investigated content-bearing unit: an element that keeps recurring in an artifact – e.g. in film, music, but also in folklore or scientific texts – by means of which often a narrative theme is conveyed. As Uther notes, “Although the definitions of a tale type as a self-sufficient narrative, and of a motif as the smallest unit within such a narrative, have often been criticized for their imprecision, these are nevertheless useful terms to describe the relationships among a large number of narratives with different functional and formal attributes from a variety of ethnic groups, time periods, and genres. The general distinction of a motif as one of the elements of a tale (that is, a statement about an actor, an object, or an incident) is separated here from its content. In fact, a motif can be a combination of all three of these elements, for example, when a woman uses a magic gift to cause a change in the situation. “Motif” thus has a broad definition that enables it to be used as a basis for literary and ethnological research. It is a narrative unit, and as such is subject to a dynamic that determines with which other motifs it can be combined. Thus motifs constitute the basic building blocks of narratives” (Uther, 2004).

On the other hand in bioinformatics oftentimes the task is to compare a protein of unknown structure with its homologues of known 3-D structures based on the idea of motifs (Buhler & Tompa, 2002). The concept of a motif here refers to a Hidden Markov Model stating that e.g. in a deoxyribonucleic acid (DNA) sequence, amino acids such as arginine, leucine, cysteine and histidine, follow each other with certain probabilities. Based on such conceptual similarities between the two domains, Darányi and Forró (2012) postulate a parallel between coding textual and genetic information, pointing toward “narrative genomics” as a recombination theory of content variation. A related phrase, the concept of “narrative DNA” (i.e. recombinative narrative elements similar to DNA, a building block of life with the genetic instructions used in the development and functioning of all known living organisms) goes back to Bruce (1996), with the idea reinforced by Gill (2011).

This paper is structured as follows: Section 2 discusses related work, whereas Section 3 outlines text evolution as a recombination process. In Section 4 we briefly list the material and method used in this study, with the results in Section 5, their discussion and future work in Section 6, and our conclusions in Section 7.

2. Background considerations and related work

Here we continue to use metadata to exemplify our hypothesis. The metadata in case is the Arne-Thompson-Uther Tale Type Catalog (ATU), a classification and bibliography of international folk tales (Uther, 2004), an alphanumerical, basically decimal classification scheme describing tale types in seven major chapters (animal tales, tales of magic, religious tales, realistic tales (novelle), tales of the stupid ogre (giant, devil), anecdotes and jokes, and formula tales), with an extensive Appendix discussing discontinued types, changes in previous type numbers, new types, geographical and ethnic terms, a register of motifs exemplified in tale types, bibliography and abbreviations, additional references and a subject index.

The numbering of the tale types runs from 1 to 2399 (in fact, 2411). Individual type descriptions uniformly come with a number, a title, an abstract-like plot mostly tagged with motifs, known combinations with other types, technical remarks, and references to the most important literature on the type plus its variants in different cultures. At the same time, as the inclusion of some 250 new types in the Appendix indicates, tale typology is a comprehensive and large-scale field of study, but also unfinished business: not all motifs in the Aarne-Thompson Motif Index (AaTh; Thompson, 1955-58) were used to tag the types, difficulties of the definition of a motif imposed limitations on its usability in ATU, and considerations related to classification of narratives had to be observed as well.[1]

In the ATU, tale types are defined as canonical motif sequences such that motif string A constitutes type X, string B stands for type Y, etc. Also, it is important to note that tale types were not conceived in the void, rather they extract the essential characteristic features of a body of tales from all over the world. An example is an excerpt from Type 300 *The Dragon-Slayer: "*A youth acquires (e.g. by exchange) three wonderful dogs [B421, B312.2]. He comes to a town where people are mourning and learns that once a year a (seven-headed) dragon [B11.2.3.1] demands a virgin as a sacrifice [B11.10, S262]. In the current year, the king's daughter has been chosen to be sacrificed, and the king offers her as a prize to her rescuer [T68.1]. The youth goes to the appointed place. While waiting to fight with the dragon, he falls into a magic sleep [D1975], during which the princess twists a ring (ribbons) into his hair; only one of her falling tears can awaken him [D1978.2]."

Together with the AaTh, ATU is the standard reference tool for librarians and digital curators alike, although other manuals such as Jason (2000) also come handy as means of orientation. When using the ATU, it is regarded as a matter of fact that its descriptive units, motifs, constitute the highest level of abstraction, and there are no units of content above this. However, Darányi and Forró (2012) have recently shown that, contrary to expectations, motifs sometimes agglomerate into higher-order multiplets, some of them being even collocated, i.e. tale types as motif strings are not *entirely* unique and must have been persistent enough to be reused as building blocks of plots.

In the above study, the authors considered ATU as a text corpus and analysed its sub-section "Supernatural adversaries" (types 300-399) in particular and section "Tales of magic" (types 300-749) in general. The two subcorpora were scrutinized for multiple motif co-occurrences and visualized by the two-mode clustering of a bag-of-motifs matrix. Having excluded types not indexed by motifs at all, the first part of the experiment (300-399) worked with 52 tale types defined on the basis of 281 motifs, and the second part (300-745A) with 219 types and 1202 motifs, respectively. After ontology visualization leading to the above conclusion, their cautiously optimistic suggestion was that as the complete AaTh contains about 40.000 motifs, this could allow for the prevalence of robust motif sequences as a new kind of metadata, and enable the use of both single and chained motifs as tags for semantic markup. Secondly, they hypothesized that since only canonical sequences of tale functions (a limited set of action types used by another limited set of actors) are known to result in "valid", i.e. acceptable, Russian fairy tales (Propp, 1968), collocated motif strings might play the same role. Thus motif substrings could be exchanged between narratives in the course of text variation, and a simple model borrowed from genetics, four types of chromosome mutation, could exemplify narrative evolution as a recombination process.

3. Narrative element recombination

These ideas were of interest to us for two reasons. The first broad context was the perception of text variation as an evolutionary process, and the task of mapping evolving semantic content onto structures with both hierarchical and multivariate access. In this frame, the reason why some motif strings have evolved and survived relates to a kind of selection pressure in a cultural historical setting, yet to be modelled. To this end, ATU and AaTh as tools have pioneered and mastered the hierarchical approach to content description but are wanting in terms of being understood as multivariate products at the same time. This is a current deficiency that cannot be overlooked or neglected when it comes to any kind of their overhaul in and for a digital environment.

In other words, for modelling one needs descriptive units of content which can index the source material in its entirety, are both multivariate by nature and fit the hierarchical classification structure, plus flexible enough to evolve, i.e. become more and more enriched variants of the original standard classifications. Indexing by single text words or phrases plus by motifs is clearly not enough to meet this goal. On the other hand, the existence of persistent motif strings in multiple copies underlying several types indicates that more than one level of semantic metadata may pertain to the body of tales we want to index.

The other broad context is the parallel between the linguistic and the genetic code as vehicles of information transfer over time. Both use coded transfer mechanisms to transmit their messages, capture instructions to reproduce meaning from form (we regard context as form here); and in both, sequence plays an important role in the coding and decoding process.

Tale types as motif sequences follow the sublanguage approach to content representation, pioneered by Harris (2002). As pointed out by Darányi (2010), this domain-specific practice from the life sciences can be recognized in formal descriptions of narrative content, too. Below the similarities between their communication patterns allow for methodology import between the two

[1] Hans-Jörg Uther, personal communication (02-12-11).

domains:
(1) Content is sequential, coded by an alphabet and compiled based on the combinations of its elements, i.e. irrespective of their order on a basic observation level. This holds for nucleotides – the building blocks of nucleic acids such as DNA and RNA – and motifs, the building blocks of tale types alike;
(2) On a next level, adding grammar and moving over to permutations, sequences start to play a role. Canonical nucleotide sequences generate secondary and tertiary – in fact spatial – structures such as the famed double helix; canonical motif sequences may contribute to the evolution of tale types, themselves representatives of tale variants. Moreover, function sequences develop into fairy tale subtypes as shown by plot analysis (Propp, 1968), and canonical mytheme sequences constitute myths and mythologies (Lévi-Strauss, 1964-71; Maranda, 2001). In a sense, reading and understanding the genetic code and narratives alike demands the mastering of abstract grammars with their equally abstract vocabularies;
(3) As said the concept of motifs is widely used in bioinformatics. Motifs in this sense mean primary nucleotide sequences of functional importance for structure generation. Sequential motifs include structural and regulatory motifs, with different functionalities pertaining to them; we anticipate methodological undercurrents linking the two knowledge domains which need to be explored in more detail.
(4) Chromosome and story mutations may be more similar than thought previously. Chromosomal mutations produce changes in whole chromosomes (more than one gene), or in the number of chromosomes present, with the major types being (a) deletion – loss of part of a chromosome; (b) duplication – extra copies of a part of a chromosome; (c) inversion – reversal in the direction of a part of a chromosome; and (d) translocation – part of a chromosome breaks off and attaches to another one.
Whereas most mutations are neutral and have little or no impact on the functionality of the product, their adding up can dramatically affect the survival rate of the outcome, leading to new genotypes and phenotypes in the course of evolution. In the same vein, deletion and translocation could be standard tools in the narrative building toolkit; inversion is suggested to play a central role in the Bible (Christensen, 2003), and duplication is evident e.g. in the case of the Proppian narrative scheme where complete tale moves may be repeated several times or combined with one another by different embeddings (Propp, 1968). This indicates the need for a theory of text evolution as a series of narrative element recombinations, forming from simple to more complex structures by “mutation mechanisms”.

4. Material and method

From the sample of 219 tale types as in Darányi and Forró (2012), examples for mutation types were manually selected and disambiguated where more than one tale variant was coded by the same ATU number, plus a set with the same motif (L161) both in terminal and non-terminal positions was separated for network visualization. Until better tools become available and allow for more stringent procedures, we defined insertion and deletion as added or missing inlays within a sequence of motifs. Transposition was considered a single motif or motif string added after a marker. Duplication was regarded as string repetition, and inversion as a reversed motif string.

5. Results

Below we identify three out of the above four major mutation types in our metadata to show how different mechanisms may lead to tale element recombination.

5.1 Insertion and deletion

This type is inherent in e.g. ATU 545A *The Cat Castle*: [B211.1.8 / B422 / B421 / B435.1] - B581.1.2 - **N411.1.1** - F771.4.1 - D711 - **B582.1.**2, and 545B *Puss in Boots*: [B211.1.8 / B422 / B435.1 / B435.2 / B441.1] - [B580 / B581 / B582.1.1] - **K1917.3** - **K1952.1.**1 - [F771.4.1 / **K722**] - D711, where motifs separated by / refer to storytelling alternatives, e.g. both [B211.1.8 / B422 / B421 / B435.1] and [B211.1.8 / B422 / B435.1 / B435.2 / B441.1] represent helpful animals. B581.1.2 and [B580 / B581 / B582.1.1], respectively, stand for bringing luck; F771.4.1 is castle owned by ogre, and D711 means disenchantment by decapitation. Therefore the underlying joint storyline is “Helpful animal brings luck by defeating ogre, culminating in his own decapitation”. In the first plot, N411.1.1 (Cat as sole inheritance) and B582.1.2 (Animal wins husband for mistress) are insertions indicated by boldtype, whereas in the second, K1917.3 (Penniless wooer: helpful animal reports master wealthy and thus wins girl for him), K1952.1.1 (Poor boy said by helpful animal to be dispossessed prince (wealthy man) who has lost clothes while swimming (in shipwreck)), and K722 (Giant tricked into becoming mouse. Cat eats him up) appear as additions to the basic plot. Since in *The Cat Castle*, a poor girl finds a husband, whereas in *Puss in Boots*, a poor man marries a princess, i.e. we have the heroine and hero oriented variants of the same story, it is an open question whether additions or deletions have resulted in these variants.

5.2 Transposition

For transposition, we depart from the observation that in the sample, motif L161 (Lowly hero marries princess) occurred in 20 tale types (9 % of the 219 plots), and out of these, it was in 15 types in terminal position, i.e. the tale finished with the wedding, whereas in 5 cases the adventures continued.
Consider the story of Aladdin as an example. Its ATU summary goes like this: “A magician orders a (stupid) boy, Aladdin, to fetch a lamp for him out of a cave of treasures. The cave opens and closes by means of a magic ring [D1470.1.5]. Aladdin finds the lamp [D812.5, D840, D1470.1.16, D1421.1.5, D1662.2], but when he wants to leave the cave it does not open (the magician has closed it). When Aladdin rubs the magic ring (lamp) in despair, a helpful genie appears and leads him out. Aladdin reaches his mother's house and wishes for riches and a castle [D1131.1]. Both wishes are fulfilled by the genie (by another spirit who appears in the same way when the lamp or the ring is rubbed). Aladdin woos the princess, but her

father intends to marry her to another man (Aladdin marries the princess [L161]). The magician exchanges the old, magic lamp (which the princess had kept) for a new, worthless one [D860, D371.1]. He wishes himself to be transferred to Africa together with the princess and the castle [D2136.2]. Aladdin is imprisoned. He rubs the ring [D881] and the genie takes him to the castle where the princess is. She poisons the magician (Aladdin kills him). Aladdin takes the lamp again and uses it to return with the castle and the princess to his home." One can easily anticipate a tale variant which finishes with the wedding, so that a second, from somewhere else translocated plot could be concatenated. This is described as: 561 *Aladdin*: D1470.1.5 - [D812.5 / D840 / D1470.1.16 / D1421.1.5 / D1662.2] - D1131.1 - L161 - [**D860 / D371.1**] - **D2136.2** - **D881**.

The other three examples are as follows:

Type 502_1 *The Wild Man*: "A king catches a wild man (Iron John) and puts him into a cage, forbidding anyone to set him free. His son frees the prisoner because his ball rolls into the cage or because he feels pity for him. The prince is afraid of his father's anger and leaves home (his father drives him away to be killed or sends him to another king) along with a servant. On their way the servant persuades the prince to exchange clothes. The prince becomes a servant at the court of another king. At a tournament he appears unrecognized three times on a splendid horse [R222] which he received from the wild man and wins the hand of a princess. Or, he wins the princess because he has helped her father in war [L161]. Often the wild man is disenchanted [G671]. In some variants the prince works for a while at the wild man's house where he disobeys instructions (e.g. looks into a forbidden chamber [C611], cares for a horse although it is not allowed [B316]) and his hair turns to gold." As this is a tale whose initial situation is not formalized in terms of motifs, we summed up the plot of the first variant as R222 - L161 - **G671**. In its second variant, no wedding takes place, i.e. L161 is missing, hence that version was not considered for exemplification here. However, in the above variant, G671 as a new ending to the story suggests a possible transposition.

Type 400_1 *The Man on a Quest for His Lost Wife* is summed up in the ATU as follows: "This tale exists chiefly in three different forms: (1) A man in distress (impoverished fisherman, merchant) unwittingly promises his (unborn) son to the devil [S240]. When the boy is delivered to him later, the devil cannot use him because he is protected by magic [K218.2] (cf. Type 810). Thus the boy is cast out in the sea (river, desert). He arrives in a foreign country and finds a lonely castle where he meets a bewitched princess (maiden, fairy) in the form of a serpent (deer). He rescues her by enduring three nights of torture [D758.1]. They marry [F302, L161]. When he wants to visit his parents, his wife gives him a ring to carry him home [D1470.1.15], and she forbids him to call her to come to him [C31.6] (to boast of her beauty [C31.5]). At home he is induced (by his mother) to break the taboo. His wife appears [D2074. 2.3.1], takes the ring, and leaves him destitute. The man sets out in search of his wife [H1385.3]. On his way he meets three hermits (rulers of animal kingdoms, or moon, sun, and wind) whom he asks for directions [B221, H1232, H1235]. With the help of the third he arrives at the empire of his wife, or he pretends that he wants to help three giants who are fighting over magic objects (inheritance, booty). He steals the magic objects (magic sword [D1400.1.4], magic coat or hood [D1361.14], seven-league boots [D1521.1]) [D831, D832] (cf. Type 518). With their help he is able to overcome the obstacles on the way to his wife [D2121]. When he finds his wife, she is about to marry another man [N681]. He discloses his identity as her real husband.(2) Meeting the princess and disenchantment as in version (1); but the disenchantment is not complete. The princess wants to travel back to her own distant land. She asks her rescuer to wait for her at a certain time and place. She appears three times, but each time a servant (witch) has put her husband into a deep sleep from which he cannot be awakened [D1364.15, D1364.4.1, D1972]. The princess informs him (in a letter) how and where to find her (on the glass mountain). The man sets out to find her. Continued as in version (1). (3) A youth watches a flock of birds (swans, ducks, geese, doves) land on the shore. The birds take off their feather coats and become beautiful maidens [D361.1]. While they are bathing, the youth steals the feather coat of the most beautiful girl, who cannot leave with the others and thus must marry the youth [D721.2, B652.1]. Later, because of carelessness (of the man's mother), the maiden takes back her coat [D361.1.1] and flies away (together with her children). She tells the youth her destination in the otherworld (e.g. glass mountain). The man sets out in search of his wife (as in version 1)." As for the three variants, the formula of the first one is: S240 - K218.2 - D758.1 - [F302 / L161] - **D1470.1.15 - [C31.6 / C31.5] - D2074.2.3.1** - H1385.3 - B221 - H1232 - H1235 - D1400.1.4 / D1361.14 / D1521.1 - [D831 / D832] - D2121 - N681. The second one, 400_2, replaces the segment **D1470.1.15 - [C31.6 / C31.5] - D2074.2.3.1** by **D1364.15 - D1364.4.1 - D1972** which is regarded transposition for the time being, and repeats the rest of the string from H1385.3 to N681. The third variant, 400_3, mentions that the beautiful girl having lost her bird shape must marry her captor but does not index the story with L161, nonetheless after having replaced the beginning of the plot by **D361.1 - [D721.2 / B652.1] - D361.1.1**, i.e. bird shape lost and regained, it continues with H1385.3 to N681 as above.

Finally type 303 *The Twins or Blood-Brothers* tells the following story: "After having eaten a magic fish (apple, water) [T511.5.1, T511.1.1, T512], a woman gives birth to twins. (Cf. Type 705A.) Grateful animals accompany the grown-up brothers, or animals give them one or more of their young ones because the brothers did not kill them. (The brothers are given unusual animals; they win them or bring them up; in some variants, the animals are born at the same time as the brothers [T589.7.1].) Together with his animals, one of the brothers sets out. When the brothers separate, they agree upon a life token that gives a warning when one of them is in mortal danger and needs help: Water will become cloudy, a plant or a tree dry up, a knife stuck in a tree will grow rusty, etc. [E761]. The first brother frees a princess (three princesses) from a dragon (trolls), unmasks an impostor ("Red Knight") who pretended to be the princess's rescuer, and marries the princess [R111.1.3, K1932, H83, L161]. Cf. Type 300. Against a warning, the hero follows a light [G451] (is tempted by an animal). He falls into the power of a witch and is turned to stone [D231]. His twin brother is warned by the life token and sets forth in quest of him. The princess mistakes him for her husband, as the two brothers are very much alike [K1311.1]. At night the brother puts a naked sword in the bed between himself and his sister-in-law [T351]. Then he finds the witch, makes her remove the spell from his brother, and kills her. The first brother learns that the second has slept with his wife and kills him out of jealousy [N342.3]. Later on, when he asks his wife why she had put the sword in the bed, he realizes that his brother was innocent. The brother is resuscitated by magic means [B512] (water of life). In some variants, a youth saves the life of a raven (crane, eagle). As a reward he obtains magic objects. The youth defeats a sea monster, delivers three princesses, and marries the youngest of them." Its formula is: [T511.5.1 / T511.1.1 / T512] - T589.7.1 - E761 - R111.1.3 - K1932 - H83 - L161 - **G451- D231- K1311.1 - T351 - N342.3- B512.** We regard the segment in boldtype as a transposition but at the same time warn the reader that screening for the transposed chunks in the complete ATU was not possible for this paper, and therefore this part of our results remains a suggestion only (Table 1).

Finally, for the same reason, we were not able to isolate inversion, i.e. reversed motif string in our material.

[Table 1 comes approximately here]

5.3 Duplication

A good example for motif string duplication is type 700 *Thumbling*: "A childless couple wish for a child, however small he may be. They have a boy (by supernatural birth) the size of a thumb [F535.1]. Thumbling takes food to his father on the farm and drives the wagon (plow) by sitting in the horse's (ox's) ear [F535.1.1.1]. He allows himself to be sold to strangers and then runs away from them. He lets himself be sold to thieves and accompanies them while they steal. Thumbling either helps them or he betrays them by his shouting; he then robs the thieves. Cf. Type 1525E. He is swallowed by a cow [F911.3.1], talks from the cow's insides and reappears [F913] (in the sausage prepared from the intestines of the slaughtered cow [F535.1.1.8]). Someone takes the intestines (sausage) and, frightened by Thumbling's voice inside, throws them away. Thumbling is swallowed by a wolf (fox) who eats the intestines [F911.3.1]. He talks from the wolf's belly and the wolf becomes sick and frightens (warns) shepherds. The wolf dies (is killed) and Thumbling is rescued [F913], or he persuades the wolf to take him to his father's house [F535.1.1]." We notice that in the respective motif sequence, F535.1 - F535.1.1.1 - **F911.3.1 - F913** - F535.1.1.8 - **F911.3.1 - F913** - F535.1.1, the segment in boldtype is repeated twice.

It is interesting to compare *Thumbling* with the related type 333 *Little Red Riding Hood*: "A little girl, called "Red Riding Hood" because of her red cap, is sent to her grandmother who lives in the forest and is warned not to leave the path [J21.5]. On the way she meets a wolf. The wolf learns where the girl is going, hurries on ahead, and devours the grandmother (puts her blood in a glass and her flesh in a pot). He puts on her clothes and lies down in her bed. Red Riding Hood arrives at the grandmother's house. (She has to drink the blood, eat the flesh, and lie down in the bed.) Red Riding Hood doubts whether the wolf is her grandmother and asks him about his odd big ears [Z18.1], eyes, hands, and mouth. Finally the wolf eats Red Riding Hood [K2011]. A hunter kills the wolf and cuts open his belly. Red Riding Hood and the grandmother are rescued alive [F913]. They fill the wolf's belly with stones [Q426]; he is drowned or falls to his death." Its formula is J21.5 - Z18.1 - K2011 - F913 - Q426, that is, both tales contain the motif F913./Victims rescued from swallower's belly/ (Table 2). Representing now the two related tale types as a directed graph whose nodes stand for the motifs and whose edges are numbered according to tale types, we notice that motif duplication yields a loop (Fig. 1).

[Table 2 and Figures 1-2 come approximately here]

5.4 Plots as memetic pathways

For visual inspection we regarded the motif index of ATU as a description of a directed graph whose nodes are motifs from AaTH. A directed edge starts from motif A to motif B if there is at least one tale type in which motif A and motif B are subsequent motifs in this order. An edge is labelled by all the tale types in which such an order appears (Fig. 2).

We note in passing that tales and their variations have been created by thousands of individuals, which is also true for content on the World Wide Web. While individuals can impose order on the web at the local level, its true global organization is utterly unplanned, and high-level structure needs to be extracted a posteriori. If we consider the graph of the web where the nodes are websites and the directed edges are links between them, we may notice the presence of so-called hubs and authorities (Kleinberg, 1999). A hub is a page that points at many other pages, whereas an authority is a page that is linked in by many different hubs.

Google's PageRank algorithm followed this line of thought to evaluate websites and rank websites (Page *et al.*, 1999). Trying to establish a ranking of motifs, we attempted to find a similar structure in their network. Early results however remained inconclusive and indicate the absence of clear hubs and authorities in our limited sample. There are motifs with a high number of both incoming and outgoing edges, but no definite sinks or sources. Therefore a ranking will have to be based on centrality or the degree of a node.

This is illustrated in Fig. 2 where, even at this small scale, the motif network shows an interesting structure. For example K1932 makes an excellent dense centre which exemplifies that there are no real hubs or authorities, but common motifs that appear in different tale types and in different positions. A hub would mean a motif from which there is an extraordinary number of possible continuation in different tale types. We do not see this, therefore we

may believe that story lines follow a restricted number of possibilities (hence one can rightfully suspect a Hidden Markov Model). An authority or a sink in the graph would be a motif that gathers plot lines, many different tale types would end or go through the very same motif. We do not see this either. H1242 is similar to K1932.
Another interesting option is to depart from the engraving function of storylines. When repeated in the course of oral transmission, such as retelling, such canonical plots as tale types preserve themselves by being repeated a thousand times and resulting in as many variants. With the above graph representation convention, one is in a position to combine this engraving function and the now forking then intertwined nature of the web of plots with individual storylines as memetic pathways. Memetic refers here to memes, those hypothetical units of cultural heritage which, by analogy with genes, self-replicate to maintain themselves (Dawkins, 1976). In this somewhat lose analogy, self-replication errors in genes result in mutations whereas self-replication errors in memes lead to text variation.

6. Discussion and future work

Our ongoing experiments suggest that better algorithms will identify not only motif sequences, but will also yield visual representations of the major "narrative mutation" types. In other words we expect that by visual inspection of a network of memetic pathways, one will be able to tell apart more popular motifs from less used ones, plus spot characteristic narrative element recombinations underlying ATU.
Secondly, by considering plot direction as its gradient, we anticipate a connection between such pathways and Waddington's epigenetic landscape (1957). Brock explains the significance of this concept as follows: "Genes provide continuity and a degree of permanence, passing in predictable ways from parents to offspring, from cell to dividing cell. Genes can be detected and sequenced, their frequencies quantified. Much more elusive, though, are the effects of environment on genes. Remarkably, in 1932, at a time when genes were recognized as discrete heritable units but their structure and function unknown, Conrad Hal Waddington used the term 'epigenetics' to refer to the external manifestation of genetic activity. He presented the 'epigenetic landscape' as a way to visualize the forces affecting cell differentiation. In this model, marbles (cells) move varying ways down a landscape whose contour is affected by genes. Details within the contours are further defined by factors above ('epi-') the fixed genetic level, and these details determine the final resting state of differentiation for each cell type.
Whether epigenetic factors act above, below, before, or after the gene depends on the factor. More importantly, 'epigenetics' today commonly refers to changes that are heritable but do not involve changes in the DNA sequence. Specifically, these are changes that affect gene expression, without changing DNA sequence, which can be passed on at least one generation" (Brock, 2010). As far as we can tell, Waddington's original idea could model the interaction between motifs shaping a landscape from "below" in a tectonic sense, socio-historical constraints influencing it from "above", and plot development as the marbles rolling down the landscape, while its modern interpretation would possibly amount to different readings of the same storyline without alterations to its narrative structure.
The formal connection between memetic pathways and the epigenetic landscape is that two-dimensional (planar) graphs correspond to landscapes (Cantwell & Forman, 1993; Minor & Urban 2008).

7. Conclusions

To use the terminology of Dawkins (1976), we considered tale types as memetic sequences of motifs, i.e. semantic content with a memory engraving function. Carried out manually, an initial tale type screening on a small test sample indicated that insertions, deletions, repetitions and possible transpositions of single motifs or motif sequences in the sample metadata corpus were not unlike chromosome mutations in genetics.
To regard the development of sequential semantic content an evolutionary process will have to be addressed in more detail in a next paper. Just identifying common structure between tales, and variation in such structure is not sufficient to claim evidence for evolution though. The problem of handling text variation has been there since the 19th century, and regarding text variants as an evolutionary series goes back to Lévi-Strauss' Oidipus analysis (1958) and his consecutive research on the canonical formula of myth. Hence the genetic metaphor for storytelling is a clarification attempt to see if one can model the process to a better extent, and the term "evolution" was used in a loose sense, indicating some sort of directed progress, just like e.g. in cultural evolution. It is also clear that a fitness function will be crucial to prove our point but we focused on simpler parts of the proposed model at this time.
It remains to be seen if motif networks based on more material than our current sample will show the hubs and authorities structure of the web. Our current assumptions are based on the analysis of a much larger graph whose visualization for this paper ran into problems hence we regard this issue unresolved. However, in another paper we report about adding taxonomy-like information to see if a more explicit graph structure will result (Declerck *et al.*, 2012). Finally, to map natural language expressions in tales to motifs as higher order content indicators, i.e. actively incorporate features at the fine-grained, grammatical level of folk narratives remains a critical task (Lendvai *et al.*, 2010).

8. Acknowledgements

The authors are grateful to two unknown reviewers for their helpful critic, to Hans-Jörg Uther (Enzyklopädie des Märchens, Göttingen) and Theo Meder (Meertens Instituut, Amsterdam) for discussions on the subject, and Artem Kozmin (Russian State University of the Humanities, Moscow) for the digitized variant of the AaTh.

9. References

Brock, T. (2010). Nurturing the Concept of Epigenetics. Cayman Chemicals Technical Report 2150 (April

2010).
Bruce, W.K. (1996). Splicing the double helix: Narrative DNA and system assaults. Muncia: Ball State University.
Buhler, J., Tompa, M. (2002). Finding Motifs Using Random Projections. *Journal of Computational Biology* 9 (2), pp. 225-242.
Cantwell, M.D., Forman, R.T.T. (1993). Landscape graphs: Ecological modeling with graph theory to detect configurations common to diverse landscapes. *Landscape Ecology* 8(4), pp. 239--255.
Christensen, D. L. (2003). *The unity of the Bible: exploring the beauty and structure of the Bible.* Mahwah, N.J.: Paulist Press.
Darányi, S. (2010). Examples of Formulaity in Narratives and Scientific Communication. In *Proceedings of the 1st International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*. Szeged: University of Szeged, pp. 29-35.
Darányi, S. and Forró, L. (2012). Detecting Multiple Motif Co-occurrences in the Aarne-Thompson-Uther Tale Type Catalog: A Preliminary Survey. *Anales de Documentación*, In press.
Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
Declerck, T., Lendvai, P., Darányi, S. (2012). Multilingual and Semantic Extension of Folk Tale Catalogues. Accepted for DH-12.
Gill, J. (2011). An investigation of cultural complexity via memetics: Methodological rationale and its operationalisation. In: Sheffield Doctoral Conference, Sheffield, 19-20th April 2011. (Unpublished).
Harris, Z.S., Gottfried, M., Ryckman, T., Mattick, P., Daladier, A., Harris, T.N., Harris, S. (1989). *The form of information in science: analysis of an immunology sublanguage.* Dordrecht: Kluwer.
Harris, Z. S. (202). The structure of science information. *Journal of Biomedical Informatics* 35, pp. 215–221.
Jason, H. (2000). *Motif, Type and Genre. A Manual for Compilation of Indices & A Bibliography of Indices and Indexing*, Helsinki: Academia Scientiarum Fennica.
Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, pp. 604-632.
Lendvai, P., Declerck, T., Darányi, S., Gervás, P., Hervás, R., Malec, S., Peinado, F. (2010). Integration of Linguistic Markup into Semantic Models of Folk Narratives: The Fairy Tale Use Case. In: *Proceedings of the Seventh International conference on Language Resources and Evaluation*, Pages 1996-2001, Valetta, Malta, European Language Resources Association (ELRA), pp. 1996-2001.
Leontis, N.B., Westhof, E. (2003). Analysis of RNA motifs. *Current Opinion in Structural Biology*, 13, pp. 300-308.
Lévi-Strauss, C. (1958). The structural study of myth. In Sebeok, T.A. (Ed.): *Myth: A Symposium.*. Bloomington: Indiana University Press, pp. 50-66.
Lévi-Strauss, C.(1964-71). *Mythologiques I-IV.* Paris : Plon.
Maranda, P. (Ed.). (2001). *The double twist: from ethnography to morphodynamics.* Toronto: University of Toronto Press.
Minor, E.S., Urban, D.L. (2008). A Graph-Theory Framework for Evaluating Landscape Connectivity and Conservation Planning**.** *Conservation Biology*, 22 (2), pp. 297–307.
Page, L., Brin, S., Motwani, R., Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, 1999.
Propp, V.J. (1968). *Morphology of the folktale*. Austin: University of Texas Press.
Rodolfa, K.T. (2008). Inducing pluripotency. StemBook (Ed.), *The Stem Cell Research Community*, StemBook.
Thompson, S. (1955-58). *Motif-Index of Folk-Literature 1–6*. Bloomington: Indiana University Press.
Uther, H.J. (2004). *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson*, Part I, Helsinki: Academia Scientiarum Fennica.
Waddington, C.H. (1957). *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology* London: Allen & Unwin.

Computational Models of Narratives as Structured Association of Formalized Elementary Events

Gian Piero Zarri
Sorbonne University, LaLIC/STIH Laboratory,
Maison de la Recherche, 28 rue Serpente, 75006 Paris, France.
E-mail: zarri@noos.fr, gian_piero.zarri@paris-sorbonne.fr

Abstract

In this paper, we describe the conceptual tools that, in an NKRL context (NKRL = Narrative Knowledge Representation Language), allow us to obtain a (computer-suitable) description of full "narratives" as logically- and temporally-ordered streams of formalized "elementary events". After having introduced, first, the main principles underpinning NKRL, we describe in some detail the characteristics of the second order (reification-based) tools, like the "completive construction" and the "binding occurrences", which implement concretely the association of the NKRL-formalized elementary events. Examples concerning some recent applications of NKRL in different domains will be used in the paper to better explain the use of these tools.

Keywords: knowledge representation, narratives, events and elementary events.

1. Introduction

"Narrative" is a generic term that denotes a *logically- and temporally ordered sequence of "elementary events"* – a narrative can also consist of a single event. An elementary event describes in turn the specific *behaviours* (actions, processes…), *experiences* (situations, states…) etc., temporally and spatially constrained, that characterize some (not necessarily humans) entities or groups of entities. As opposed to "*fictional*" narratives (novels, poetry…), "*non-fictional*" narratives concern 'daily life' descriptions included in natural language (NL) documents as corporate reports, news stories, legal texts, medical records, etc., but also in multimedia supports like surveillance videos, actuality photos or eLearning sources. The real time description of the different moves (equivalent to *elementary events*) of a robot trying to pour a cup of coffee for an impaired elderly is also a (fictional/non-fictional) narrative. Because of the 'pervasiveness' of the *narrative resources* (Finlayson *et al.*, 2010), being able to represent and manage in a general, accurate, and effective way their *semantic/logical content* is both *conceptually relevant and economically important*.

These last years – thanks mainly to the contribution of the European Commission through several EC-financed projects – a modelling tool called NKRL ("Narrative Knowledge Representation Language") has been specified and implemented for representing and managing, in a normalised way, the 'meaning' of (mainly non-fictional) complex multimedia narrative sources, see Zarri (2009; 2011a; 2011b). NKRL is, at the same time:

- A knowledge representation system for describing in adequate detail and in computer-understandable format the content of (non-fictional) narrative information. One of the main characteristics of the language concerns *the addition of an "ontology of event" to the standard "ontology of concepts"*. The ontology of events consists of a hierarchical structure (HTemp, hierarchy of templates) of *n*-ary "*templates*": each of them corresponds in turn to the *formal description of a general class of elementary events* like "move a physical object", "be present in a place", "produce a service" etc. The 'concrete' events like "Yesterday, Lucy moved the wardrobe", "John lives in Paris" are obtained by *instantiating* the corresponding templates. The "*connectivity phenomena*" that assure the coherence of a whole narrative by linking together its constitutive "elementary events" are dealt with making use of "reification" mechanisms, see next Section.
- A system of powerful *reasoning* (inference) procedures. For example, the "*transformation rules*" try to replace some queries that failed with one or more different queries that are not '*strictly equivalent*' but only '*semantically close*' to the original ones (analogical reasoning). The 'hypothesis rules', in contrast, allow us to build up *causal-like explications of given events* according to common-sense schemata formed of several 'reasoning steps', to be validated through unification with the contents of a knowledge base. See, e.g., Zarri (2005) for more details about the main features of the NKRL rules.
- A *wholly implemented software environment* – in two versions, an SQL- and a file-supported one.

In this paper, we will describe first, Section 2, the general framework of second-order tools that allow us to collect the single elementary events into a whole narrative. Section 3 will present a detailed example; Section 4 supplies some information about other approaches used to deal with a narrative context. Section 5 will be a short "Conclusion".

2. Associating the elementary events

In NKRL, the templates – and their instances, called "*predicative occurrences*", which supply then the *formal description of the elementary events* – are represented according to the *n*-ary schema denoted by Eq. 1:

$(L_i \ (P_j \ (R_1 \ a_1) \ (R_2 \ a_2) \ \ldots \ (R_n \ a_n)))$. (1)

The meaning of the different symbols used in Eq. 1 is:

- L_i is the *symbolic label* identifying the particular n-*ary structure*, for example, the NKRL representation of a template, or the predicative occurrence corresponding to the specific elementary event relating that "John has given a book to Mary".
- P_j is the *conceptual predicate*, corresponding then, at *deep level*, to surface items in the style of "transfer of objects", "move", "give", etc.
- R_k is a generic *functional role*, see Zarri (2011c), i.e., an operator denoting the "*specific function*" of the *arguments of the predicate* ("John", "book" and "Mary" in the previous example) with respect to this predicate. In the example, "John" is linked to the MOVE (or GIVE etc.) conceptual predicate by a SUBJECT or AGENT functional role, "book" by an OBJECT or PATIENT role and "Mary" by a BENEFICIARY role.
- a_k denotes the NKRL representation of the *arguments of the predicate* (e.g., in the previous example, the individuals JOHN_, BOOK_1 and MARY_).

Note that each of the $(R_i \ a_i)$ cells of Eq. 1, *taken individually*, represents some sort of *binary relationship*. The main point to emphasize is, however, that the whole conceptual structure represented by Eq. 1 can be fragmented into binary units for practical purposes like the storing within a database, *but must be considered globally as a full* n-*ary structure whenever significant querying/inference operations must be envisaged about its whole 'meaning'*, see (Zarri 2009: 14-21) in this context.

To avoid both the typical ambiguities of natural language and the possible 'combinatorial explosion' problems – see the discussion in (Zarri 2009: 56-62) – both the (unique) *conceptual predicate* of Eq. 1 and the associated *functional roles* are "primitives" in NKRL. Predicates P_j pertain then to the set {BEHAVE, EXIST, EXPERIENCE, MOVE, OWN, PRODUCE, RECEIVE}, and the functional roles R_k to the set {SUBJ(ect), OBJ(ect), SOURCE, BEN(e)F(iciary), MODAL(ity), TOPIC, CONTEXT}, see the examples below. On the contrary, the a_i terms (the arguments of the predicate) in Eq. 1 *are not primitives* in NKRL and pertain to an "open", conventional *ontology of concepts* – called HClass, hierarchy of classes in NKRL – whose *low levels* must normally be updated (i.e., new concepts and their instances must be added) whenever a new application in a new domain has to be considered.

2.1 Connectivity phenomena and reification

We have already mentioned, in Section 1, those "*connectivity phenomena*" that, in natural language terms, are expressed through associative syntactic features like causality, goal, indirect speech, co-ordination, subordination etc. They constitute, then, *the 'surface evidence' of those deep semantic mechanisms that assure the logical coherence among the components (elementary events) of a specific narrative.*

In NKRL, the connectivity phenomena are dealt with making use of *second order structures* obtained from the *reification* of *generic* (i.e., not only predicative, see below) occurrences. Concretely, the reification is based on the use of the *symbolic labels denoting NKRL well-formed conceptual expressions*, in the style then of the L_i terms in Eq. 1 above. "Reification" is intended here – as usual in a Knowledge Representation context – as the possibility of *creating new objects ("first class citizens") out of already existing entities and to 'say something' about them without making reference to the original entities.*

2.2 Completive construction

A first example of reification mechanism is supplied by the so-called "*completive construction*". This consists in using *as filler of a functional role R_k (see Eq. 1) in a predicative occurrence pc_i (an instance of a template) the symbolic label L_j of another (generic) occurrence c_j. Only one* of the functional roles of pc_i can be filled with a symbolic label L_j: i.e., *only one* of the binary cells of Eq. 1 can then be denoted as $(R_k \ L_j)$. Moreover, only the OBJ, MODAL, TOPIC and CONTEXT functional roles of pc_i can accept as filler the symbolic label L_j of a generic occurrence c_j. "*Generic*" means that the c_j used as (implicit) filler of a functional role can correspond not only to *predicative occurrence*, but also to one of those *binding occurrences* introduced in 2.3 below. We can also note that, for implementation reasons, the label L_j used as filler is *prefixed* by a "sharp", "#", code. The general format of a "*completive construction filler*" corresponds then, actually, to #symbolic_label, see the example in Table 1. Note that symbolic_label is a regular "*concept*" of HClass, the 'standard' ontology of concepts of NKRL; symbolic_label has then as instances all the specific labels used to denote (generic) occurrences in a particular NKRL application.

An example of completive construction is given in Table 1, which represents a fragment of narrative concerning a recent application of NKRL techniques in the gas/oil domain, see Zarri (2011a). The fragment says: "On October 16, 2008, at 9h10, the field operator, denoted by the label INDIVIDUAL_PERSON_104, sends to the control activities leader (INDIVIDUAL_PERSON_102) a message confirming that valve VALVE_FCV401 is open". For clarity's sake, we reproduce also in Table 2 the template, Move:StructuredInformation, used to create the predicative occurrence virt2.c69 (an instance of this template) represented in the upper part of Table 1. We can observe immediately that the general structure of both the predicative occurrences of Table 1 and of the template of Table 2 corresponds well to the (*n*-ary) structure of Eq. 1 – see the equivalence among the symbolic labels L_j, virt2.c69, virt2.c70 and Move:StructuredInformation (4.42), the presence of a unique (conceptual) predicate, the (seven) functional roles introducing the arguments of the predicate, etc. Note that, in the templates, these arguments (the a_i terms of Eq.1) are concretely implemented, see Table 2, as *variables (var_i) and constraints on these variables***.**

Table 1: An example of completive construction

virt2.c69)	MOVE	SUBJ	INDIVIDUAL_PERSON_104: (GP1Z_COMPLEX)
		OBJ	#virt2.c70
		BENF	INDIVIDUAL_PERSON_102: (GP1Z_MAIN_CONTROL_ROOM)
		MODAL	vhf_audio_transmitter
		date-1:	2008-10-16/09:10
		date-2:	

Move:StructuredInformation (4.42)

INDIVIDUAL_PERSON_104 *sends to the control activities leader the message represented by the predicative occurrence* virt2.c70.

virt2.c70)	PRODUCE	SUBJ	INDIVIDUAL_PERSON_104: (GP1Z_COMPLEX)
		OBJ	(SPECIF assessment_ positive_)
		TOPIC	(SPECIF VALVE_FCV401 open_)
		date-1:	2008-10-16-08:41
		date-2:	2008-10-16-08:50

Produce:Assessment/Trial (6.32)

INDIVIDUAL_PERSON_104 *confirms that* VALVE_FCV401 *is open.*

The constraints are expressed as *HClass concepts or combinations of concepts* – as already stated, HClass is the *standard* ontology of concepts in NKRL. The two NKRL ontologies, HClass and HTemp, *interact then strictly*. When creating a *predicative occurrence* like virt2.c69 to represent a particular elementary event, the role fillers *must conform to the constraints of the father-template.* In virt2.c69, e.g., INDIVIDUAL_PERSON_102 and INDIVIDUAL_PERSON_104 are "*individuals*", *instances* of the HClass concept individual_person; this last is a specialization of human_being, specialization in turn of human_being_or_social_body, see the constraints on the arguments *var1*, associated with the SUBJ(ect) role, and *var6*, associated with the BEN(e)F(iciary) role, in the template of Table 2. GP1Z_COMPLEX and GP1Z_MAIN_CONTROL_ROOM are also individuals. The former is an instance of the concept let_down_station, a specialization of the HClass concept location_ – see the constraint on the variable *var2* in Table 1 – through a specialization chain of concepts that includes, among other things, oil/gas_processing_plant, industrial_premises and extended_location. The latter is an instance of the concept control_room, a specialization of location_ through, among other things, office/room_, building/area_component and extended_location. Note that, in the templates and predicative occurrences, *the "determiners/attributes" of the location type are implemented as "lists" of concepts/individuals that are linked to the corresponding arguments of the predicate through the "colon" operator, ":"*, see Tables 1 and 2. The elements of a template (as SOURCE, BENF etc. in Table 2) included in square brackets, "[ ]", are 'optional', i.e., they can be present or not in their instances. Two special operators, date-1 and date-2 – that can be assimilated to specific functional roles – are used to introduce, in the predicative occurrences (see Table 1), the *temporal information* associated with an elementary event. These operators do not appear in the formulation of templates given that these last, as *general classes* of elementary events, are to be considered as a-temporal. A description of the formal system used in NKRL for the representation and management of temporal information can be found, e.g., in Zarri (2009: 76-86, 194-201).

Table 2: An example of template

name : Move:StructuredInformation
father : Move:TransmitInformation
position : 4. 42
NL description : 'Transmit a Structured Information'

MOVE	SUBJ	*var1*: [(*var2*)]
	OBJ	*var3*
	[SOURCE	*var4*: [(*var5*)]]
	[BENF	*var6*: [(*var7*)]]
	[MODAL	*var8*]
	[TOPIC	*var9*]
	[CONTEXT	*var10*]
	{ [modulators], ≠abs }	

var1	=	*human_being_or_social_body*
var3	=	*symbolic_label*
var4	=	*human_being_or_social_body*
var6	=	*human_being_or_social_body*
var8	=	*electronic/media_product*, *information_support*, *service_*, *services_agency*, *transmission_medium*
var9	=	*sortal_concept*
var10	=	*situation_*, *symbolic_label>*
var2, *var5*, *var7*	=	*location_*

The function of the "attributive operator" SPECIF(ication) – employed in virt2.c70 to build up the "structured arguments" (expansions) used as fillers of the OBJ(ect) and TOPIC functional roles – consists in adding some 'properties' (positive_, open_) to the terms, concepts (assessment_) or individuals (VALVE_FCV401), which represent the 'head' of the SPECIF list. SPECIF(ication) = S is one of the *four operators that make up the AECS sub-language*. AECS includes also the disjunctive operator ALTERN(ative) = A, the distributive operator ENUM(eration) = E and the collective operator COORD(ination) = C. The interweaving of the four operators within an expansion is controlled by the so-called "*priority rule*", see Zarri (2009: 68-70).

Returning now to the completive construction, the association of *var3* with the constraint symbolic_label in the template Move:StructuredInformation of Table 2 makes it clear that, in the predicative occurrences derived from this template, the OBJ(ect) of the transfer is *the content of a full elementary event*, see Table 1. This modality of use of the "completive construction" to represent all sorts of situations corresponding to a "*transfer of information*" is particularly popular in an NKRL context.

2.3 Binding occurrences

A second, more general way of linking together NKRL elementary events within the scope of a full narrative consists in making use of "*binding occurrences*", i.e., lists labelled with specific "*binding operators*" Bn_i whose arguments arg_i are represented (reification) by symbolic labels L_j of (predicative or binding) c_j occurrences. The general expression of a binding occurrence bc_i is then:

$$(\mathrm{Lb}_k\ (\mathrm{Bn}_i\ \mathrm{L}_1\ \mathrm{L}_2\ \ldots\ \mathrm{L}_n))\ , \qquad (2)$$

where Lb_k is now the symbolic label identifying the whole (*autonomous*) binding structure. Eq. 2 is particularly important in an NKRL context given that it represents also *the general model of a full narrative as structured association of (formalized) elementary events.*

Unlike templates and predicative occurrences, binding occurrences are characterized by the absence of any predicate or functional role. The eight binding operators corresponding to Bn_i in Eq. 2 are: ALTERN, COORD, ENUM (see also the AECS operators in the previous sub-section), CAUSE (the 'strict causality' operator), REFER (the 'weak causality' operator), GOAL (the 'strict intentionality' operator), MOTIV(ation, the 'weak intentionality' operator), COND(ition), see Zarri (2009: 91-98) for more details. Some restrictions must be respected in order to set up *well formed* binding occurrences. For example while, for the binding occurrences of the ALTERN, COORD and ENUM type, *no restriction is imposed on the cardinality of the list* (i.e., on the possible number of arguments L_i), *only two arguments* L_m *and* L_n *are admitted* in the binding occurrences labelled with CAUSE, REFER, GOAL, MOTIV and COND. The binding occurrences that make use of these five binding operators are then simply of the type: $(\mathrm{Lb}_k\ (\mathrm{Bn}_i\ \mathrm{L}_m\ \mathrm{L}_n))$.

To supply now a simple example of use of the binding occurrence tools, let us suppose we should want to represent that: "On October 16th, 2008, at 8h26, the Control Room operator pushes the SEQ1_BUTTON *in order to* start the auxiliary lubrication pump M202". Two elementary events are involved here: they are represented by the two predicative occurrences virt2.c32 and virt2.c33 in Table 3 and denote, the first, the action of "pushing", and the second, the (possible) "start" of the pump. We must additionally introduce a *binding occurrence* virt2.c30 – labelled using the GOAL binding operator and involving only two arguments L_m and L_n, see above – to link together the conceptual labels virt2.c32 (L_m, the planning activity) and virt2.c33 (L_n, the intended result). The global meaning of virt2.c30 is then: "the activity described in virt2.c32 is focalised towards (GOAL) the realization of virt2.c33". Note also that, in agreement with the semantics of GOAL, see Zarri (2009: 71), virt2.c33, the 'result', is *characterized by the presence of an uncertainty attribute code*, "*****", to indicate that, at the moment of 'pushing', the real instantiation of a situation like 'pump running' cannot be categorically stated.

3. NKRL Modelling of a Full Narrative

The *second order structures* of NKRL, completive construction and binding occurrences, allow us to take correctly into account the *connectivity phenomena*; accordingly, they play also a crucial role in the *full modelling of complete narratives*. As an example, we supply in Table 4 the NKRL representation of a narrative proper to the context of the gas/oil application already mentioned: "On November 1st, 2008, at 10h15, the start-up procedure of the GP1Z turbine was stopped by the production activities leader, given that he had been informed by a field operator of the presence of an oil leakage concerning an auxiliary lubrication pump".

Table 3: Binding and predicative occurrences

virt2.c32)	BEHAVE	SUBJ	INDIVIDUAL_PERSON_102: (GP1Z_MAIN_CONTROL_ROOM)
		MODAL	button_pushing
		TOPIC	SEQ1_BUTTON
		date-1:	2008-10-16-08:26
		date-2:	
Behave:ActExplicitly (1.12)			
*virt2.c33)	MOVE	SUBJ	AUXILIARY_LUBRICATION_PUMP_M202: (idle_)
		OBJ	AUXILIARY_LUBRICATION_PUMP_M202: (running_)
		date-1:	2008-10-16-08:26
		date-2:	
Move:ForcedchangeofState (4.12)			
virt2.c30)	(GOAL virt2.c32 virt2.c33)		

The – *mandatory* – starting point for the creation of the NKRL model of a full narrative is the set up of a *binding occurrence showing the main topics dealt with in the narrative*. This 'upper level' binding occurrence is often, as in the present case, of the COORD(ination) type: we have estimated here, e.g., that the narrative was formed of *three independent but strictly connected items,* relating the first the causes of the turbine's stop, and giving information the second and the third about the jobs of the two involved people. But the upper level binding occurrence can be labelled using any of the eight binding operators introduced above. Having set up the top level of the conceptual representation, the different blocks listed in this binding occurrence are *successively expanded* and the corresponding elementary events suitably encoded.

Let us consider, e.g., the binding occurrence virt3.c6 that illustrates the *two – strictly associated*, COORD – reasons of the stop. The first is described in the *completive construction* formed by the implicit inclusion of virt3.c8, the 'message' signalling the leakage, as OBJ(ect) of the transmission of information between the two individuals mentioned in the predicative occurrence virt3.c7. Note that, thanks to the completive construction mechanism, the two occurrences virt3.c7/virt3.c8 perform as a unique conceptual unit: the insertion of #virt3.c8 within virt3.c6 concerns only the coherence controls of the software, and does not alter at all the cardinality (two) of the COORD's arguments. The *second* reason of the stop is described in virt3.c9: when the leakage is detected we can note, temporal modulator obs(serve), that the auxiliary pump is linked to the turbine – coupled_with is a specialization of binary_relational_property, specific term in HClass, through property_, of the high level non_sortal_concept.

Table 4: NKRL modeling of a gas/oil narrative

virt3.c1) (COORD virt3.c2 virt3.c3 virt3.c4)

The conceptual model of the narrative is formed of three components.

virt3.c2) (CAUSE virt3.c5 virt3.c6)

The first component consists of a CAUSE binding relationship.

```
virt3.c5)  PRODUCE  SUBJ     INDIVIDUAL_PERSON_102:
                             (GP1Z_MAIN_CONTROL_ROOM)
                    OBJ      activity_stop
                    TOPIC    (SPECIF turbine_startup GP1Z_TURBINE)
                    date-1:  1/11/2008/10:15, (1/11/2008/10:30)
                    date-2:
```

Produce:PerformTask/Activity (6.3)

On November 1st, 2008, INDIVIDUAL_PERSON_102 *ends the start-up of the* GP1Z_TURBINE.

virt3.c6) (COORD virt3.c7 #virt3.c8 virt3.c9)

The second term of the CAUSE *relationship consists of a* COORD *binding occurrence.*

```
virt3.c7)  MOVE  SUBJ     INDIVIDUAL_PERSON_104: (GP1Z_COMPLEX)
                 OBJ      #virt3.c8
                 BENF     INDIVIDUAL_PERSON_102:
                                 (GP1Z_MAIN_CONTROL_ROOM)
                 MODAL    vhf_audio_transmitter
                 date-1:  1/11/2008/10:15
                 date-2:
```

Move:StructuredInformation (4.42)

INDIVIDUAL_PERSON_104 *sends to* INDIVIDUAL_PERSON_102 *the message represented by the predicative occurrence* virt3.c8.

```
virt3.c8)  PRODUCE SUBJ  INDIVIDUAL_PERSON_104: (GP1Z_COMPLEX)
                    OBJ   detection_
                   TOPIC (SPECIF lubrication_oil_leakage
                            (SPECIF around_
                            AUXILIARY_LUBRICATION_PUMP_M202))
                   date-1: 1/11/2008/10:02
                   date-2: 1/11/2008/10:15
```

Produce:PerformTask/Activity (6.3)

INDIVIDUAL_PERSON_104 *has discovered the presence of lubrication oil leakage around the lubrication pump M202.*

```
virt3.c9)  OWN  SUBJ     AUXILIARY_LUBRICATION_PUMP_M202
                OBJ      property_
                TOPIC    (SPECIF coupled_with GP1Z_TURBINE)
                { obs }
                date-1:  1/11/2008/10:02
                date-2:
```

Own:CompoundProperty (5.42)

On November 1st, 2008, at 10h02, we can observe that the auxiliary lubrication pump is related to the GP1Z_TURBINE.

```
virt3.c3)  BEHAVE  SUBJ     INDIVIDUAL_PERSON_102:
                              (GP1Z_MAIN_CONTROL_ROOM)
                   MODAL    production_activities_leader
                   { obs }
                   date-1:  1/11/2008/10:15
                   date-2:
```

Behave:Role (1.11)

We can remark that INDIVIDUAL_PERSON_102 *fulfils the function of production activities leader.*

```
virt3.c4)  BEHAVE  SUBJ     INDIVIDUAL_PERSON_104: (GP1Z_COMPLEX)
                   MODAL    field_operator
                   { obs }
                   date-1:  1/11/2008/10:15
                   date-2:
```

Behave:Role (1.11)

We can remark that INDIVIDUAL_PERSON_104 *fulfils the function of field operator at the GP1Z complex.*

Note that "obs" – see (Zarri 2009: 71-75) for the specific NKRL "determiners" represented by the "modulators", and the "temporal modulators" like "obs(serve)" in particular – is used to indicate that the situation described in the associated predicative occurrence is true at the date stored in the date-1 block of the occurrence, without having, at this level, any possibility of giving information about the duration of the situation, which probably extends in time before and after the given date – see also the two 'job' occurrences virt3.c3 and virt3.c4.

Eventually, we can remark that the logical arrangement of a narrative (like that of Table 4) expressed in NKRL terms can always be represented as a tree structure, see Fig. 1. This remark is not new, and can be considered as valid in general independently from the formalization adopted see, e.g., the "story trees" of Mani and Pustejovsky (2004).

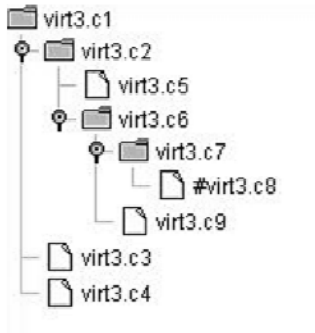


Figure 1: Tree structure corresponding to the narrative of Table 4

4. Related work

With respect to the practical solutions suggested for dealing with the *connectivity phenomena* from a Computer Science/Artificial Intelligence point of view, we can evoke here, first, some 'old', well-known proposals of Schankian inspirations evoking all sorts of scripts, scenarios, TAUs, MOPs etc., see, e.g., (Dyer, 1983; Kolodner, 1984). The SnePS (Semantic Network Processing System) proposal of Stuart Shapiro, see (Shapiro, 1979) pertains roughly to the same period. Going back to the fifties-sixties we can also note that, among the "*correlators*" introduced by Silvio Ceccato's in a Mechanical Translation (MT) context to represent both fictional and nonfictional narratives as *recursive networks of triadic structures* (Ceccato, 1961), some concerned coordination and subordination, apposition, subject-predicate relationships, etc.

Among the recent suggestions in a "narrative" domain, we can mention in particular some mechanisms used in a Conceptual Graph's framework for dealing with "contexts". John Sowa's Conceptual Graphs (CGs), see (Sowa, 1999), are based on a powerful graph-based representation scheme that can be used to represent *n*-ary relationships between complex objects in a global system. Contexts in CGs are dealt with making use of *second order (nested graphs) extensions* that represent *CGs' solution to the "connectivity phenomena" problem*. Nested graphs bear some resemblance to NKRL's constructs like completive construction and binding occurrences, as we can see from Sowa's analysis (Sowa, 1999: 485-486) of the sentence "Tom believes that Mary wants to marry a sailor". This is decomposed, as in NKRL, in a first part "Tom believes that..." in a completive construction style and a second "...Mary wants to marry..." where two elementary events signalled by the presence of the two predicates "want" and "marry" are linked together making use of a "binding occurrence"-like formal expression.

In a generic 'Linguistics/Computational Linguistics' framework, Episodic Logic, EL (Schubert and Hwang, 2000) is a 'Natural Language-like', highly formalized logical representation for narrative understanding allowing, among other things, the expression of sentence and predicate reification, of intensional predicates (corresponding to wanting, believing, making, etc.), of episodes, events, states of affairs, etc. "Episodes" can be explicitly related in terms of part-whole, temporal and causal relations. Some solutions for the connectivity phenomena management have also been put forward by the Discourse Representation Theory, DRT (Kamp and Reyle, 1993). DRT is a semantic theory developed for representing and computing *trans-sentential anaphora and other forms of text cohesion* see, e.g., the solution suggested (through, e.g., "embedding functions" similar to the context solutions proposed by Sowa, see above) for managing *all sort of context-related problems*. The "Text Meaning Representation" model, TRM (Nirenburg and Raskin, 2004) is part of an implemented theory of Natural Language processing, OntoSem (Ontological semantics). TRM presents some interesting similarities with NKRL, see a detailed analysis in (Zarri, 2009: 146-149).

5. Conclusion

In this paper, we have described the conceptual tools that, in an NKRL context, allow us to set-up a computational model of full "narratives" as logically- and temporally-ordered sets of formalized "elementary events". These tools are represented by unification-based, second order semantic/syntactic structures like the "completive construction" and the "binding occurrences".

As already stated in the "Introduction", NKRL is also a *fully operational environment*. Many successful, real-size experiments in many different domains (from "terrorism" to the "corporate", "cultural heritage" and "legal" domains, to the management of "storyboards" for the gas/oil industry, the implementation of assisted living procedures in a secure environment etc.) have proved the practical utility of this tool.

6. References

Ceccato, S., Ed. (1961). *Linguistic Analysis and Programming for Mechanical Translation* (Technical report RADC-TR-60-18). Milano: Feltrinelli.

Dyer, M.G. (1983). *In-Depth Understanding*. Cambridge (MA): The MIT Press.

Finlayson, M.A., Gervás, P. Mueller, E., Narayanan, S. and Winston P. (2010). Preface: Computational Models of Narrative. In: M.A. Finlayson *et al.* (Eds.), *Computational Models of Narratives – Papers from the AAAI Fall Symposium*. Menlo Park (CA): AAAI Press, pp. vii--viii.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation.* Dordrecht: Kluwer.

Kolodner, J.L. (1984). *Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model*. Hillsdale (NJ): Lawrence Erlbaum Associates,

Mani, I. and Pustejovsky, J. (2004). Temporal Discourse Models for Narrative Structure. In: *Proceedings of the ACL Workshop on Discourse Annotation.* Stroudsburg (PA): The Association for Computational Linguistics (ACL), pp. 57--64.

Nirenburg, S. and Raskin, V. (2004). *Ontological Semantics*. Cambridge (MA): The MIT Press.

Schubert, L.K. and Hwang, C.H. (2000). Episodic Logic Meets Little Red Riding Hood: A Comprehensive, Natural Representation for Language Understanding. In: *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, L. Iwanska, S. C. Shapiro, Eds., Cambridge (MA) and Menlo Park (CA): MIT/AAAI Press, pp. 111--174.

Shapiro, S.C. (1979). The SNePS Semantic Network Processing System. In: *Associative Networks: Representation and Use of Knowledge by Computers*, N.V. Findler, Ed. New York: Academic Press, pp. 179--203.

Sowa, J.F. 1999. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove (CA): Brooks Cole Publishing Co.

Zarri, G.P. (2005). Integrating the Two Main Inference Modes of NKRL, Transformations and Hypotheses. *Journal on Data Semantics (JoDS)*, IV, pp. 304--340.

Zarri, G.P. (2009). *Representation and Management of Narrative Information, Theoretical Principles and Implementation.* London: Springer.

Zarri, G.P. (2011a). Knowledge Representation and Inference Techniques to Improve the Management of Gas and Oil Facilities. *Knowledge-Based Systems (KNOSYS)*, 24(7), pp. 989--1003.

Zarri, G.P. (2011b). A Structured Metadata Approach for Dealing in an 'Intelligent' Way with Complex 'Narrative' Information. *International Journal of Metadata, Semantics and Ontologies (IJMSO)*, 6(1), pp. 10--22.

Zarri, G.P. (2011c). Differentiating Between "Functional" and "Semantic" Roles in a High-Level Conceptual Data Modeling Language. In: *Proceedings of the 24th International Florida AI Research Society Conference, FLAIRS-24*. Menlo Park (CA): AAAI Press, pp. 75--80.

Objectivity and Reproducibility of Proppian Narrative Annotations

Rens Bod[1], **Bernhard Fisseni**[1,2,3,4], **Aadil Kurji**[1], **Benedikt Löwe**[1,3,4,5]

[1] Institute for Logic, Language and Computation, Universiteit van Amsterdam
Postbus 94242, 1090 GE Amsterdam, The Netherlands
{l.w.m.bod,b.loewe}@uva.nl,aadil.kurji@student.uva.nl
[2] Fakultät für Geisteswissenschaften, Universität Duisburg-Essen
Universitätsstraße 12, 45117 Essen, Germany
bernhard.fisseni@uni-due.de
[3] Isaac Newton Institute for Mathematical Sciences
20 Clarkson Road, Cambridge CB3 0EH, United Kingdom
[4] Fachbereich Mathematik, Universität Hamburg
Bundesstraße 55, 20146 Hamburg, Germany
[5] Corpus Christi College, University of Cambridge
Cambridge CB2 1RH, United Kingdom

Abstract

A formal narrative representation is a procedure assigning a formal description to a natural language narrative. One of the goals of the *computational models of narrative* community is to understand this procedure better in order to automatize it. A formal framework fit for automatization should allow for objective and reproducible representations. In this paper, we present empirical work focussing on objectivity and reproducibility of the formal framework by Vladimir Propp (1928). The experiments consider Propp's formalization of Russian fairy tales and formalizations done by test subjects in the same formal framework; the data show that some features of Propp's system such as the assignment of the characters to the *dramatis personae* and some of the functions are not easy to reproduce.

1. Introduction & Motivation

The formal study of narratives goes back to the Russian structuralist school, paradigmatically represented by Vladimir Propp's 1928 study *Morphology of the Folktale* (Propp, 1958). Researchers in the field of *computational models of narrative* have developed the general Proppian methodology into formal and computational frameworks for the analysis, automated understanding and generation of narratives.[1]

In recent years, there has been an increased interest in the methodological and conceptual issues involved. The enterprise of representing a narrative by a formal structure that can then be used in computational application rests on a number of assumptions:

Assumption E. (*Existence of a structural core*) There is a structural core of narratives; or several, depending on which part of the structure we are interested in.

Assumption O. (*Objectivity of the structural core*) Given a narrative, there is an interpersonal agreement what its structural core is; possibly after some agreement of what part of the structure should be represented.

A formal framework Λ for representing narratives consists of a formal language $\mathscr{L}_\Lambda$, a class of mathematical structures $\mathscr{M}_\Lambda$, and a description of a procedure (called *formalization* in (Löwe, 2011)) of assigning to each natural language narrative N a structure $\Sigma_\Lambda(N) \in \mathscr{M}_\Lambda$. Note that this procedure is not a function in the mathematical sense, but an activity by expert formalizers who follow given guidelines.

In this paper, we explore the validity of Assumption **O**: in particular, we are investigating the following property of formal frameworks Λ:

Property Obj(Λ). Sufficiently trained human formalizers, given the same narrative N will produce the same structure $\Sigma_\Lambda(N)$.

Property **Obj** is an important (and arguably necessary) feature of a formal framework Λ if it is supposed to be the basis of an automatized system. The existence or non-existence of formal frameworks Λ with property **Obj**(Λ) is closely related to Assumption **O**. In (Bod et al., 2011), we described the investigation of **Obj**(Λ) as a natural analogue of the study of annotator agreement in corpus linguistics and computational linguistics: whereas typical annotation tasks involve annotation of sentences or discourses (e.g., (Marcus et al., 1993; Brants, 2000; Passonneau et al., 2006)), the formalization or annotation of a narrative is at the next level of complexity. At the sentence or discourse level, inter-annotator agreement has been studied (e.g. (Carletta et al., 1997; Marcu et al., 1999)), but no such analysis has been done for the formalization of narratives, not even for the oldest and best-known formal approach, Propp's *Morphology of the folktale*, first published in 1928.

We focus on this formal framework, not because it is a particularly good candidate for a framework close to the *stable structural core*, but due to its prominent place in the history of formal representations of narratives. In § 2., we describe the Proppian formal framework and discuss two empirical studies pertaining to it, referred to as **Propp I** and **Propp II**, performed at the *Universiteit van Amsterdam*; in § 3., we discuss the results and future work.

[1] Lehnert's Plot Units, Rumelhart's Story Grammars, Schank's Thematic Organization Points (TOPs), Dyer's Thematic Abstraction Units (TAUs), or Turner's Planning Advice Themes (PATs) are some examples; cf. (Lehnert, 1981; Rumelhart, 1980; Schank, 1982; Dyer, 1983; Turner, 1994).

Ivanko Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
1	Ivanko	Devils			Peasant			
2	Bearlet	Bear/Devil	Bearlet/Wife	Peasant	Peasant			
3	Ivanko	Thieves/Dogs/Devil	Wife	Peasant	Peasant	Grandfather	Horse	
4	Ivanko/Mother	Devil/Peasant			Peasant			
5	Ivanko	Father	Father's Satisfaction	Grandfather	Father	Little Devil	Horse	
6	Ivanko	Devil			Peasant			
7	Ivanko	Devil	Bear	Ivanko/Wife		Peasant	Devil	
8	Ivanko	Devil			Peasant		Horse	
9	Bearlet	Father	Father, Money		Father	Devil	Hare	

Semyons Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
1	Semyons		Elena		Tsar			
2	Semyons	Tsar	Elena		7			
3	Semyons	7th Semyon	Elena	Tsar	Tsar		Kitten/Stone	7th Semyon
4	7th Semyon	Elena's father	Elena	Elena's father	Tsar		Semyon Bros	
5	Semyons	Tsar	Elena	Elena's Father	Tsar	Tsar	Cat	
6	7th Semyon	Tsar	Elena	Elena's father	Tsar		Semyons	
7	Semyons		Elena	Tsar	Tsar		Ship	
8	7th Semyon		Elena		Tsar		6 Semyons	
9	7th Semyon	Tsar	Elena		Tsar		Semyons	Tsar

Shabarsha Test subject	**H**	**V**	**P**	**PF**	**Di**	**Do**	**MH**	**FH**
1	Shabarsha		Gold	Little Devil/Grandad	Master			
2	Shabarsha/Little Devil	Shabarsha			Grandad			
3	Shabarsha	Little Devil	Gold	Grandad	Grandad	Master	Cap	
4	Shabarsha	Little Devil	Gold	Grandad	Master	Master		
5	Shabarsha	Shabarsha	Gold	Grandad	Master		Bear/Hare	Shabarsha
6	Shabarsha	Little Devil	Gold	Grandad	Master		Bear/Hare	
7	Shabarsha	Little Devil/Grandad	Gold		Master	Master	Twine	
8	Shabarsha	Little Boy	Gold	Grandad	Master		Bear/Hare	
9	Little Devil	Shabarsha	Peace	Grandad	Grandad			

Table 1: The assignment of the *dramatis personae* for the three folktales in **Propp I**.

2. Propp's formal system

2.1. Overview of Propp

Working with a corpus of 100 Russian folktales from Afanas'ev's collection *Narodnye Russkie Skazki*, Vladimir Propp developed a formal system to identify each folktale by short annotation strings consisting of symbols representing Proppian *functions* or *narratemes*. In the following, we give a description of the components of the Proppian system relevant for the experiments discussed in this paper. For more details, we refer the reader to (Propp, 1958).

Propp identified seven[2] *dramatis personae* representing roles the characters may play within the tales. They are: the hero (**H**), the villain (**V**), the princess (**P**), the princess's father (**PF**), the dispatcher (**Di**), the donor (**Do**), the (magical) helper (**MH**) and the false hero (**FH**) (Propp, 1958, § 3).

The actions of the *dramatis personae* are described by a set of thirty-one functions described in (Propp, 1958, § 3) by means of examples and more specified subfunctions. These functions are marked by symbols in the order of their occurrence in the folktale; the first seven functions, marked with lowercase Greek letters, are called *preliminary functions*: β Absentation; γ Interdiction; δ Violation, ε Reconnaissance, ξ Delivery, η Trickery, θ Complicity. The *preliminary functions* are not fully developed in (Propp, 1958) and are not included in Propp's own annotation strings. The main functions are: **A** Villainy, **a** Lack, **B** Mediation, **C** Beginning counteraction, $\uparrow$ Departure, **D** First function of the Donor, **E** Hero's reaction, **F** Provision or receipt of magical agent, **G** Spatial transference between two kingdoms, **H** Struggle, **J** Branding, **I** Victory, **K** Liquidation, $\downarrow$ Return, **Pr** Pursuit, **Rs** Rescue, **o** Unrecognized Arrival, **L** Unfounded Claims, **M** Difficult Task, **N** Solution, **Q** Recognition, **Ex** Exposure, **T** Transfiguration, **U** Punishment, **W** Wedding. These functions, instantiated by *subfunctions* marked by superscripts, occur in strict sequential order, i.e., functions have to occur in the folktale in the order they are given in the list above. In the full Proppian system, there are a few specific ways to break strict sequentiality (Propp, 1958, § IX.A): The most important one is that some folktales contain a series of individual tale units, called *moves*. Examples are *trebling*, the triple repetition of moves within the tale, and moves in which a magical agent is obtained in the first move but only used in the second move of the tale. None of the tales we used had *moves* (according to Propp's own annotations), so we did not include this option in our experiment.

2.2. Description of Propp I

Test subjects were trained in the Proppian framework and then asked to annotate four of the folktales formalized in (Propp, 1958). We used the folktales *The Seven Semyons, 147*, *Shabarsha, 151*, and *Ivan the Bear's Son, 152*; in the following, we refer to these folktales as *Semyons*, *Shabarsha*, and *Ivanko*.[3] We chose tales that were available in English translation, and in Propp's annotation had no *moves*

[2]One of these, the *Princess/Princess's Father*, can be split into two with a slightly difficult delineation. In our experiment, we presented the resulting list of eight *dramatis personae*.

[3]In **Propp I**, we also used the folktale *The Enchanted Princess*, but it was too long and omitted in **Propp II**. Due to an

(i.e., retained strict sequential ordering) and used few functions (*Ivanko* uses eight functions, *Shabarsha* six). An annotation of in **Propp I** consisted of (1) the assignment of story characters to the *dramatis personae*, and (2) a list of the functions (group 1) or the functions with corresponding subfunctions (group 2) occurring in the folktale.

Procedure. We had nine test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English. We split them into two groups: Test subjects 1 to 5 were group 1 (no subfunction marking) and test subjects 6 to 9 were group 2 (subfunction marking). Test subjects were instructed that the experiment would last three hours and received a moderate financial compensation for participation.

The experiment started with a 45-minute introduction to Propp's system given by a native speaker of English supported by a projector presentation explaining the relevant fragment of Propp's system. Only a selection of the subfunctions was included (labelled "examples" for group 1 and "subfunctions" for group 2). We analyzed a simple example story, of our own design, as an illustration. A condensed version of the *dramatis personae* and functions was distributed as a leaflet for use during the annotation.

Results. Propp's annotation for *Shabarsha* was $\mathbf{A}^8\,\mathbf{B}^4\,\mathbf{C}\uparrow\mathbf{H}^2\,\mathbf{I}^2\,\mathbf{K}^1\downarrow$; his annotation for *Ivanko* was $\mathbf{A}^9\uparrow\mathbf{H}^2\,\mathbf{I}^2\,\mathbf{K}^1\downarrow$. These consist of the function strings alone and do not include the *preliminary functions*.[4]

We give the results of the assignments of *dramatis personae* in Table 1. The results indicate that the test subjects did not fully understand the Proppian scheme; note in particular the variation in the three main *dramatis personae*, **H**, **V**, and **P** (see below for a methodological remark).

The annotation strings vary widely and are given in Table 2 (subfunctions are marked by superscripts, with a missing subfunction marked by $^{\varnothing}$). Since no two strings are the same, comparison would have to be per function; calculations of statistics per function are not useful because of the variation in the assignment of *dramatis personae* and the small amount of data. The strings are longer than Propp's strings (compare an average of 14.2, 13.2, and 12.8 functions with the Propp's of 6 and 8 for *Ivanko* and *Shabarsha*, respectively).

Methodological Conclusion. Four out of nine test subjects reported that the example story from the presentation was considerably simpler than the folktales.

The variation in the assignment of characters to *dramatis personae* suggests that the description of the *dramatis personae* was not precise enough. For instance, our description of the hero used the words "who is good". Arguably, Shabarsha's behaviour in *Shabarsha* cannot be described as "good", which caused some of the variation in the assignment of the hero.[5]

A number of functions are consistently annotated which do not show up in Propp's own annotations. On the other hand, we see that some of Propp's functions show up in all or almost all annotations strings: e.g., $\uparrow$, **H**, **I**, **K** and $\downarrow$ are reliably reproduced in the *Ivanko* annotation strings. However, since we do not know which events in the tale the annotators marked with these functions, we cannot be sure whether these are actual reproductions of Propp's assignments.

2.3. Description of Propp II

The experiment **Propp II** was a modified version of **Propp I**, taking the problems discussed in § 2.1. into account. We used the same folktales as in **Propp I**. An annotation of a folktale in **Propp II** consisted of (1) a list of the functions occurring in the folktale, and (2) marked text passages for each of the functions that occurred.

The main changes to **Propp I** were: the test subjects were given the assignment of *dramatis personae*; subfunctions were not discussed at all; the example story was from Propp's own corpus. It should be noted that Propp only recorded the annotation strings, so that his choice of *dramatis personae* was extrapolated from (Propp, 1958).[6]

Procedure. We had six test subjects, all students of the *Universiteit van Amsterdam*, and all with native or near-native competence of English. Test subjects were instructed that the experiment would last three hours and received a moderate financial compensation for participation.

The experiment started with a 45-minute introduction to Propp's system given by a native speaker of English supported by a projector presentation explaining the relevant fragment of Propp's system. We gave short descriptions of the *dramatis personae* roughly based on Propp's original text and the the descriptions of the functions from Propp's

oversight, we worked with version 147 of *Semyons* while Propp annotated version 145. This makes it impossible to compare our results to Propp's original annotation, but it does not invalidate the discussion of inter-annotator agreement of our test subjects. We used the translations of Gutermann (Afanas'ev, 1973) for *Semyons* and *Ivanko*, and the translation of Cook (Afanas'ev, 1985) for *Shabarsha*.

In *Semyons*, seven orphans meet the Tsar and pledge to work hard in their professions. The seventh becomes a thief and, with the help of his brothers and their respective talents, journeys to capture Elena the fair as a bride for the Tsar. In *Ivanko*, Ivanko is born of a peasant woman and her kidnapper, a bear. After returning to human society, he causes some damage and is sent to a lake in which devils dwell. Through a series of tricks, Ivanko gains all of the devils' gold and the services of a little devil for a year. In *Shabarsha*, the protagonist Shabarsha takes a day off to earn some money for himself and his boss. He goes to a lake to catch fish, meets a little devil and threatens to evict all of the devils from the lake if they don't pay rent. Through a series of tricks he acquires all of their wealth.

[4]Therefore, we do not take the *preliminary functions* into account for comparison between Propp's original strings and the strings produced by the test subjects. For the sake of completeness, we also list the Propp string for *Semyons* (version 145, cf. fn. 3): $\mathbf{a}^1\,\mathbf{B}^2\,\mathbf{C}\uparrow\mathbf{F}^2\,\mathbf{G}^1\,\mathbf{K}^2\downarrow$.

[5]It is conceivable that the designator "devil" created a connotation in the original audience of the folktale producing a very different reading of Shabarsha's behaviour that cannot be reproduced in contemporary test subjects due to a lack of cultural context and contemporary sympathy for harmless "little devils".

[6]In *Ivanko*, we assigned Ivanko to **H** and the Little Devil and the Grandfather jointly to **V**; in *Semyons*, we assigned the seventh Semyon to **H**, Elena the Fair to **P**, and the Tsar to **Di**; finally, in *Shabarsha*, we assigned Shabarsha to **H** and the Little Devil and the Grandfather jointly to **V**.

Subject		Proppian functions for **Ivanko**
Propp		**A^{9} ↑ H^{2} I^{2}K^{1}↓**
1	β	**↑ G H I ↓**
2	$\beta\ \gamma\ \delta\ \zeta\eta\ \theta$	**A a B C↑ G H I K ↓ U**
3	$\beta\ \gamma\ \delta\ \zeta\eta\ \theta$	**A a B C↑D E H I K ↓ N Ex**
4	$\beta\ \gamma\ \delta\varepsilon\ \zeta\eta\ \theta$	**a B C↑D G H I K ↓ZNQ U**
5	$\beta\ \gamma\ \delta\ \zeta\eta\ \theta$	**a B C↑D E F JI K**
6	$\beta^{\varnothing}\gamma^{2}\delta\varepsilon^{2}\ \zeta\eta^{1}\theta^{1}$	**B^{5} C↑ G^{3} H^{2} I^{2}K^{1}↓**
7	$\beta^{1}\ \delta\varepsilon^{\varnothing}$	**a^{5}B^{2} C↑D^{1}E^{1} H$^{\varnothing}$ I^{2}K^{2}↓ U**
8	$\beta^{1}\ \gamma^{2}$	**↑ G^{3} H^{2} I^{2} ↓ W^{6}**
9	$\beta^{1}\ \zeta\ \theta^{1}$	**a^{5}B$^{\varnothing}$ ↑D^{1}E^{9}F^{9}G$^{\varnothing}$ K^{1}↓**

Subject		Proppian functions for **Shabarsha**
Propp		**A^{8} B^{4}C↑ H^{2}I^{2}K^{1}↓**
1		**a B C↑ G H K**
2	$\beta\ \varepsilon\zeta\eta\ \theta$	**A a B C↑ G H I K ↓ W**
3	$\beta\gamma\ \varepsilon\zeta\eta\ \theta$	**A B C D EF G H I K**
4	$\beta\gamma\ \varepsilon\zeta$	**a B C↑ E I K Pr QU**
5	$\beta\gamma\ \eta$	**a B C D F K ↓ QU**
6	$\gamma^{2}\delta\ \eta^{1}\ \theta^{1}$	**a^{2} H^{2}I^{2}K^{1}↓ N UW^{6}**
7	γ^{2}	**a^{2}B^{2}C↑D^{1} F^{1}G^{3}H^{2}I^{2}K^{1} o W^{6}**
8		**a^{2}B^{1}C↑ G^{3}H^{2}I^{2}K^{1}**
9	$\eta^{\varnothing}\theta^{1}$	**a^{2}B^{2} H^{2} W$^{\varnothing}$**

Subject		Proppian functions for **Semyons**
1	β	**a B C↑ G K ↓**
2	$\beta\ \gamma\ \delta\varepsilon\zeta\eta\theta$	**Aa B C↑ G H K ↓PrRs W**
3	$\beta\ \gamma\ \delta\varepsilon\zeta\eta\theta$	**Aa DEF G K ↓Pr NQTU**
4	$\beta\ \gamma\ \delta$	**a B C↑D G HIK ↓PrRs Q**
5	$\beta\ \zeta\eta\theta$	**a B C↑ F G K ↓**
6	$\beta^{\varnothing}\gamma^{1}\ \delta\ \zeta$	**a^{1}B^{2} ↑ G$^{\varnothing}$ K^{1} Pr N W^{6}**
7		**a^{5}B^{1}C↑ F^{3}G^{3} K^{1}↓ W^{6}**
8	$\beta^{2}\ \gamma^{1}$	**a^{1}B^{1}C↑ G^{3} K^{2}↓ W^{6}**
9	$\beta^{2}\ \gamma^{\varnothing}\delta\ \zeta$	**a^{5}B^{2} ↑ G$^{\varnothing}$ K^{2}↓Pr N W^{6}**

Table 2: The annotation strings for the three folktales in **Propp I** (cf. fn. 3).

Ivanko Test subject		Proppian Functions	**Semyons** Test subject		Proppian Functions	**Shabarsha** Test subject	Proppian Functions
Propp		**A ↑ HIK↓**				Propp	**A BC↑HIK↓**
1	β	**a ↑GHIK↓**	1		**aB ↑GK W**	1	**a HIK N**
2	$\beta\gamma$	**↑ MN W**	2	β	**aB ↑ K↓ W**	2	**aB ↑ MN W**
3	β	**B↑ HI U**	3	β	**aB G oNW**	3	**a C↑HI M UW**
4	β	**↑ HI ↓ U**	4	η	**a ↑G Pr W**	4	**a ↑HI MN**
5	β	**aB↑ HI ↓**	5		**aB ↑ K↓PrRs W**	5	**a ↑H K**
6	β	**aB↑ HIK↓ W**	6	β	**aBC↑GK↓PrRs W**	6	**aBC HIK W**

Table 3: The annotation strings for the three folktales in **Propp II** (cf. fn. 3).

text. We analyzed the folktale (*Ivan Popyalov, 135*) from the Propp corpus. Again the condensed version of the *dramatis personae* and functions was distributed as a leaflet to for use during the annotation. Test subjects were given an assignment of characters to the *dramatis personae* together with each folktale.

Results. We give the results of the function annotation in Table 3. The annotation strings are noticeably shorter than in **Propp I** (on average 6.8 functions per annotator, compared with 13.4 functions in **Propp I** and 6 and 8 functions in the original Propp strings for *Ivanko* and *Shabarsha*, respectively),[7] and more similar to Propp's original strings, but we still do not have matching strings among the test subjects.

It is again not possible to do a serious statistical analysis on the basis of six annotations; we therefore do a qualitative analysis instead. We say that a function occurs *stably* in **Propp II** if it is in at least four of the six annotations. We further distinguish *strong stability* when the marked text of the annotators overlaps, and *weak* otherwise. In *Ivanko*, β, **I** and ↓ were strongly stable and ↑ and **H** were weakly stable (of which ↑, **H**, **I** and ↓ are annotated by Propp); in *Shabarsha*, **a** and ↑ were strong stable and **H** and **I** were weakly stable (of which ↑, **H**, and **I** were annotated by Propp); in *Semyons*, **a**, **B**, **G**, and **W** were strongly stable and ↑ and **K** were weakly stable. Note that in both *Ivanko* and *Shabarsha*, there is a strongly stable function not annotated by Propp (**B** and **a**, respectively).

[7]Most likely, a reason for the much longer strings in **Propp I** was the assignment of superfluous *dramatis personae* by the test subjects in that experiment.

3. Discussion & Future Work

The difference between **Propp I** and **Propp II** show that the assignment of the characters to the *dramatis personae* has an important effect on the assignment of the functions. Even with pre-assigned *dramatis personae*, there are marked differences between Propp's and the test subjects' annotations, and among the test subjects. Some of this effect can be explained by the vagueness of the description of Propp's functions: as an illustration, we mention that subfunction 6 of **W** is listed as "Other form of compensation like a monetary reward". This vague description fits in much more general situations than Propp apparently intended. Making these vague descriptions understandable for the test subjects may require considerably more time and training than we gave the test subjects in our experiments.

The detailed study of human annotations of Propp's framework highlights weaknesses such as vague descriptions of *dramatis personae* and functions, and in general, points to some important obstacles for an automatization of the process of formalization in a computational setting.

In (Bod et al., 2011), we suggested to follow up the studies **Propp I** and **Propp II** with a large-scale inter-annotator study: the results of our experiments suggest that this is not worthwhile. Instead, we should distill the lessons learned from this Proppian case study into studies dealing with other formal representation systems, possibly designed and documented on the basis of the results of this study.

Acknowledgements

The research in this paper was funded by the *John Templeton Foundation* (JTF) via the project *What makes stories similar?* (grant id 20565) and the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO) via the projects *Integrating Cognition* in the VICI programme (DN 277-70-006) and *Dialogical Foundations of Semantics* in the ESF EuroCoRes programme LogICCC (LogICCC-FP004; DN 231-80-002; CN 2008/08314/GW). The second, and fourth author acknowledge the financial support and the kind hospitality of the *Isaac Newton Institute for Mathematical Sciences* (programme *Semantics & Syntax*). All authors would like to thank Ekaterina Abramova and Sanchit Saraf (Amsterdam) for their work in the early set-up phase of the research.

4. References

Aleksandr Nikolayevich Afanas'ev. 1973. *Russian Fairy Tales*. Pantheon. Translation by Norbert Guterman from the collections of Aleksandr Afanas'ev. Folkloristic commentary by Roman Jakobson.

Aleksandr Nikolayevich Afanas'ev. 1985. Shabarsha, translated by Kathleen Cook. In *The Three Kingdoms. Russian Fairy Tales From Alexander Afanasiev's Collection, illustrated by Alexander Kurkin*. Raduga Publisher, Moscow.

Rens Bod, Benedikt Löwe, and Sanchit Saraf. 2011. How much do formal narrative annotations differ? A Proppian case study. In Charles Ess and Ruth Hagengruber, editors, *The computational turn: Past, presents, futures?, Aarhus University, July 4–6, 2011*, pages 242–245, Münster. MV-Wissenschaft.

Thorsten Brants. 2000. Inter-annotator agreement for a German newspaper corpus. In *Proceedings Second International Conference on Language Resources and Evaluation LREC-2000*.

Jean C. Carletta, Amy Isard, Steven Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Michael G. Dyer. 1983. *In-depth understanding: A computer model of integrated processing for narrative comprehension*. Artificial Intelligence Series. MIT Press, Cambridge MA.

Wendy G. Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science*, 4:293–331.

Benedikt Löwe. 2011. Methodological remarks about comparing formal frameworks for narratives. In Patrick Allo and Giuseppe Primiero, editors, *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*, pages 10–28, Brussel. KVAB.

Daniel Marcu, Magdalena Romera, and Estibaliz Amorrortu. 1999. Experiments in constructing a corpus of discourse trees: Problems, annotation choices, issues. In Marilyn A. Walker, editor, *Towards Standards and Tools for Discourse Tagging. Proceedings of the Workshop*, pages 71–78.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19:302–330.

Rebecca Passonneau, Nizar Habash, and Owen Rambow. 2006. Inter-annotator agreement on a multilingual semantic annotation task. In *Proceedings LREC-2006*, pages 1951–1956.

Vladimir Yakovlevich Propp. 1958. *Morphology of the folktale*. University of Texas Press, Austin, Texas, 2nd edition.

David E. Rumelhart. 1980. On evaluating story grammars. *Cognitive Science*, 4:313–316.

Roger C. Schank. 1982. *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge University Press.

Scott Turner. 1994. *The creative process. A computer model of storytelling*. Lawrence Erlbaum Associates, Hillsdale, NJ.

An Experiment to Determine Whether Clustering Will Reveal Mythemes

R. Raymond Lang, John G. Mersch
Computer Science (RL) | Mathematics (JM), Xavier University of Louisiana
1 Drexel Drive, New Orleans, LA 70125
E-mail: rlang@xula.edu, jmersch@xula.edu

Abstract

Claude Lévi-Strauss proposes a universal structure for narrative myths. The structure is expressed as the canonical formula, $f_x(a) : f_y(b) \approx f_x(b) : f_{a-1}(y)$, where the four terms of the formula denote bundles of gross constituent units, his term for predicate relations. The bundles are referred to as mythemes. The deep meaning of a myth is given by associating its semantic content with the terms of the formula. The analytic approach to myths is hindered by (1) circularity between the bundles and their components, and (2) heavy reliance on expert knowledge. This project is to develop a system for the algorithmic identification of these bundles. The investigation is starting with clustering of only word senses (semantemes) and will proceed to clustering of predicate relations. The number of desired clusters is known, and the clustered objects are non-numeric, so an appropriate algorithm is k-mediods, using distance metrics computed with the WordNet::Similarity Perl module. Status of the experiment and planned directions for the work are described.

Keywords: Lévi-Strauss, canonical formula, k-mediods clustering, analysis of myth, mythemes

1. Introduction

Anthropology, psychology, and narrative studies intersect in the study of the myth. Armstrong elaborates on the origins of myth (2005), Lévi-Strauss on their universal structure (1955). He proposes a canonical formula for myth that has guided and informed subsequent scholarship (Maranda, 2001) since he first set it forth.

The claim made of Lévi-Strauss's canonical formula (CF) is that it can make explicit the deep meaning of a myth and so reveal the existential dilemma it addresses. The CF captures the dialectic characteristic of myth. According to the CF, a myth comprises four bundles. Each bundle is made up of relations found in the myth.

Lévi-Strauss begins his analysis by defining gross constituent units (GCUs) as a predicate relation. GCUs are organized into groups, because, as he notes, "the true constituent units of a myth are not the isolated relations but *bundles of such relations*" [emphasis in the original] (Lévi-Strauss, 1955, p. 431). A mytheme corresponds to a bundle of these relations.

A critical step in the application of Lévi-Strauss's method to a specific myth is, given a relation, determine which bundle to associate the relation with. However, this process is circular, since the bundles are not known until the component relations have been properly assigned to bundles. The criteria for a successful analysis are, "the principles which serve as a basis for any kind of structural analysis: economy of explanation; unity of solution; and ability to reconstruct the whole from a fragment, as well as further stages from previous ones" (Lévi-Strauss, 1955, p. 431). This amounts to expert intuition.

The goal of the analysis is to reveal a myth's deep meaning, understood as the existential dilemma, conundrum, or conflict that it is the role of a myth to address (Maranda, 1974). The current work seeks to determine whether the identification of bundles can be performed algorithmically.

2. Meaning and Application of the Canonical Formula

A myth comprises two pairs of bundles. Each pair signifies a binary opposition. The two oppositions are conceptually related. Lévi-Strauss uses the Oedipus myth as an illustration. The two opposed pairs are: (1) overrating versus underrating of family relations, and (2) the rejection of versus the acceptance of humanity's autochthonous origins. The value of family relations is conceptually related to the question of origins, since the former are determined by the latter. The CF

$$f_x(a) : f_y(b) \approx f_x(b) : f_{a-1}(y)$$

captures the abstract relationship among mythemes (Lévi-Strauss, 1955, p. 442).

The CF is a chiasmus rather than a simple analogy (Racine, 2001). It expresses (a) the dialectic exhibited by myth and (b) the contradictory attributes mythic elements embody. Dialectic may manifest along either a temporal dimensional (e.g. summer/winter) or a spatial dimension (e.g. earth/sky). Lévi-Strauss claims "that every myth (considered as the collection of all its variants) corresponds to a formula of [this] type" (1955, p. 442). The interaction of opposites establishes the agonistic tone characteristic of myth.

The terms of the CF denote two contrasting objects and two contrasting functions. The functions and objects combine in a specific fashion to form the four terms of the formula. The surface form of a given myth depends on the interpretation of the terms. The formula does not consider the arrangement of the story's events. Instead, it defines a partition of the events unrelated to their temporal order or arrangement in the story.

3. Clustering

The bundles are aggregates of GCUs (relations). Lévi-Strauss asserts that GCUs "cannot be found among phonemes, morphemes, or semantemes, but only on a higher level.” Because (1) the assertion is unsupported, and (2) an objective of this work is to identify the minimal content required to perform the analysis algorithmically, the initial stage of this work applies a clustering algorithm to word senses (semantemes).

The four bundles are the high-level classification of a story’s GCUs. In this case, since the number of desired clusters is known, an appropriate clustering algorithm is one such as k-means, k-median, or k-medoids. Since the objects under consideration are non-numeric, the k-medoids algorithm (Park, Lee, & Jun, 2006) is an applicable clustering algorithm. Various distance metrics from the WordNet::Similarity software package (Pedersen, Patwardhan, & Michelizzi, 2004) are being evaluated for their utility in producing clusters.

To evaluate the clusters, the two pairs of bundles must exhibit some broad measure of opposition, for example, hunting as opposed to agriculture. If clustering of word senses yields meaningful results, that would constitute a counter-example to Lévi-Strauss’s assertion. There is reason to doubt the assertion (i.e. to regard it as possible that the organization of the semantemes of a myth is related to the myth’s higher-level organization). An application of the CF is a high-level classification supported by a conceptual system that is, in turn, dependent on a sub-conceptual structure (Langacker, 1991). Neural theory of metaphor (Lakoff, 2008) also leads us to expect that sub-level component organization is related to organization of higher levels.

4. Status of the Project

Grimm’s version of “Little Red Riding Hood” is being used as a test case for development purposes. Other versions of the story occur across cultures, it is associated with rituals in some cases (Saintyves, 1989), and it exhibits Campbell’s monomyth (1972).

Story Workbench (Finlayson, 2008) is being used to annotate the story. The marked-up story is preprocessed to produce an array of word senses and matrix of distances, as computed by the WordNet::Similarity package, between the word senses. The k-medoids algorithm as described by Park, Lee, and Jun has been implemented and applied to these word senses and distances.

The system has been applied to an annotation of a single version of Little Red Riding Hood. The algorithm was set to arrange the word senses into four clusters. Almost all the word senses were placed into one large cluster, with the few remaining senses arranged into three much smaller clusters. It is premature to assess the value of clustering on word senses from only a single story. A more complete experiment will use multiple versions of the story. Further experimentation will apply the technique to different sets of myths.

5. Experimental Plan

The plan is to develop techniques for filtering out noise words senses, for example, by clustering only nouns and verbs. Another potential technique for filtering out noise is to utilize index word extraction, which would enable clustering only the word senses that are the most relevant to the story (Wacholder, Klavans, & Evans, 2000).

Also under development is a way to cluster predicate relations, which are the components Lévi-Strauss claims are necessary for finding mythemes. Toward this end, relations in our trial story are being annotated using the semantic role representation in the Story Workbench.

Clustering on predicate relations requires a distance metric. Vector distance is an obvious choice, but this metric requires vectors to be of the same length; however, predicate relations can vary in the number of arguments. Lévi-Strauss deals with complex predicate relations by simplifying a story. He writes, “The technique […] consists in […] breaking down [a] story into the shortest possible sentences, and writing each such sentence on an index card bearing a number corresponding to the unfolding of the story” (Lévi-Strauss, 1955, p. 431). In keeping with this approach, clustering will be limited to predicates over agents and patients. This provides a 3-dimensional vector (predicate head, agent, patient) to which vector distance metrics apply.

WordNet::Similarity gives the distance between two given predicate heads, since each is a particular sense of a specific verb. To compute distances between agents and between patients, these will be assigned to coreference groups, and a suitable WordNet sense will be selected to represent each coreference group. These senses will be used to compute the distance between two agents or two patients. The objective is to treat as the same the word “wolf” and the phrase “you old sinner” used by the hunter to refer to the wolf.

As with word senses in isolation, one difficulty in applying clustering techniques to predicate relations is dealing with noise. If confining attention to two-place predicates, done to reduce noise, is too restrictive, distance metrics for predicates of varying arity will be necessary. Further experimentation will reveal other obstacles as well as possible remediations.

The goal of this project is an algorithmic implementation of Lévi-Strauss’s method; this level of computational narrative interpretation will constitute a meaningful contribution to the repertoire of automated reasoning techniques.

6. Acknowledgements

Grateful acknowledgement for assistance in this project is extended to Theresa Beauboeuf of the Southeastern Louisiana University Department of Computer Science.

In Search of an Appropriate Abstraction Level for Motif Annotations

Folgert Karsdorp, Peter van Kranenburg, Theo Meder, Dolf Trieschnigg, Antal van den Bosch*

Meertens Institute, *Radboud University Nijmegen
Postbus 94264, 1090 GG Amsterdam, The Netherlands, *Postbus 9102, 6500 HC Nijmegen, The Netherlands
{folgert.karsdorp,peter.van.kranenburg,theo.meder,dolf.trieschnigg}@meertens.knaw.nl, a.vandenbosch@let.ru.nl

Abstract

We present ongoing research on the role of motifs in oral transmission of stories. We assume that motifs constitute the primary building blocks of stories. On the basis of a quantitative analysis we show that the level of motif annotation utilized in the Aarne-Thompson-Uther folktale type catalogue is well suited to analyze two genres of folktales in terms of motif sequences. However, for the other five genres in the catalogue the annotation level is not apt, because it is unable to bring to front the commonalities between stories.

Keywords: folktale, motif, annotation, oral transmission, ATU catalogue

1. Introduction

In oral culture artifacts such as stories are propagated through the community and passed on to successive generations. Stories contain all kinds of cultural ideas that are replicated via the process of storytelling. During replication, most elements of a story remain stable producing recognizable variants (lineages) of a cultural artifact. A story like *Little Red Riding Hood* can be told in various ways, i.e. can have many textual forms, but at a more abstract level, the essence of the story remains virtually untouched. How can this be?

Our ultimate goal is to create a model of oral transmission of folktales. We hypothesize that oral transmission of folktales happens through the replication of sequences of motifs. In this view, motifs constitute the primary vehicles of cultural heritage in oral transmission of stories. A prerequisite for building such a motif-based model of oral transmission of stories is to formalize tales as sequences of motifs. Because the manual annotation of motifs is a time-consuming and error prone job with respect to consistency, we wish to create a system for the automatic recognition of motifs. This motif detection system will enable us to analyze large amounts of available data.

The term *motif* immediately gives rise to the question of what exactly is a motif. Without wanting to settle the debate, we propose, as a working hypothesis, that motifs are the simplest meaning-bearing units contributing to the overall plot of a story (Jason, 2007; Van Boven and Dorleijn, 2003). Following Thompson (1946, 415) we add to this that motifs should have “a power to persist in tradition”. Motifs are thus recurring elements found in different stories (variants or types). In this paper we investigate whether motifs form the primary building blocks for stories. With a more or less fixed set of motifs we can analyze a large number of stories. We hope to find evidence for the idea that the way in which motifs can be recombined to produce new stories can best be described with a motif-based story grammar.

In this paper, we investigate a small part of our motif definition, namely what we should conceive as the *simplest* units for the task at hand. That is, what level of description of motifs is appropriate for (1) modeling oral transmission of stories and (2) conceiving stories as sequences of motifs? We will do so by means of a quantitative analysis of the authoritative folktale type catalogue *The Types of International Folktales* by Aarne, Thompson and Uther (henceforth: ATU catalogue) (Uther, 2004). To the best of our knowledge, no similar analysis was conducted before.

The outline of the paper is as follows. We will start with a brief theoretical background about the term motif and the materials used in this study. We then continue with the analysis of the ATU catalogue in which we examine whether the description level of motifs in the catalogue is appropriate for modeling stories as sequences of recurring motifs. The last section offers our conclusions and directions for further research.

2. Different levels of abstraction

We assume that motifs are the simplest meaning-bearing units of a story that have a power to persist tradition. Now, what do we mean by ‘simplest’? Are motifs the elaborate and abstract functions that Propp (1968) distinguishes, or the many thousands of small and hierarchically ordered content units in Stith Thompson’s (1955 1958) *Motif-Index* (henceforth: TMI)?

Propp (1968) recognizes 31 plot units which he calls *functions*, common to a small subgroup of fairy tales. An example of a function is given under (1):

(1) ABSENTION: A member of a family leaves the security of the home environment.

One important aspect of Propp’s theory is that the functions abstract away from specific characters (*dramatis personae*). So, ‘a member’ may be any kind of hero in the story or a member of the family that the hero will later need to rescue. In the TMI we find over 45.000 motifs hierarchically ordered in a tree structure. Many motifs are bound to particular folktale types. Under (2) we list some examples:

(2) Q426 Wolf cut open and filled with stones as punishment;

F911.3 Animal swallows man (not fatally);

F823.2 Glass shoes;

J346 Better be content with what you have, than try to get more and lose everything.

In modeling cultural evolution it is important to realize that the more abstract the level at which we identify motifs, the harder it is to trace lineages with confidence (Dennett, 1995, 357). If the level of comparison is too abstract, we can only identify very general commonalities that are not distinctive enough. Therefore we must take into account the particular forms of expression with which motifs are realized. With this in mind, the rather abstract functions of Propp seem less appropriate for modeling oral transmission of folktales than the more concrete motifs in the TMI, at least if we would use Propp's functions exclusively. Another reason why Propp's functions seem less suitable is that they are only defined for one group of fairy tales, whereas we would like a system that can cope with all kinds of folktales, including genres such as traditional and contemporary legends and jokes.

These objections do not necessarily imply a complete rejection of the usefulness of Propp's functions for modeling oral transmission of stories, yet the more concrete motifs from the TMI seem more appropriate as a starting point in our investigation. In the ATU catalogue, the motifs from the TMI play a key role in the classification of tales into a certain type. Every folktale type contains a short summary of the plot. In this summary we find a sequence of specific motifs that constitute the primary descriptive units of a tale type without an overarching level. An example of a story summary in the ATU catalogue:

> ATU 327A "**Hansel and Gretel**. A (poor) father (persuaded by the stepmother) abandons his children (a boy and a girl) in the forest [S321, S143]. Twice the children find their way back home, following scattered pebbles [R135]. On the third night, birds eat the scattered peas (breadcrumbs) [R135.1]. The children come upon a gingerbread house which belongs to a witch (ogress) [G401, F771.1.10, G412.1]. She takes them into her house. The boy is fattened [G82], while the girl must do housework. The witch asks the boy to show his finger in order to test how fat he is [G82.1], but he shows her a bone (stick) [G82.1.1]. When the witch wants to cook the boy, the sister deceives her by feigning ignorance and pushes her into the oven [G526, G512.3.2]. [...] The children escape, carrying the witch's treasure with them. Birds and beasts (angels) help them across water. They return home."

This folktale type is defined by the motif sequence:

(3)
```
([S321, S143] [R135] [R135.1]
[G401, F771.1.10, G412.1] [G82]
[G82.1] [G82.1.1] [G526, G512.3.2])
```

We take the collection of folktale summaries in the ATU catalogue to be a corpus of stories with motif annotations. We will use this corpus to investigate the question whether we can use the concrete level of motif description utilized in this corpus for extracting a grammar of folktales consisting of sequences of recurring motifs.

As a first step, we aim to establish experimentally that motifs indeed represent the primary recurring building blocks of stories. Different combinations of motifs (some new and some old) give rise to new stories, some of which are variations of already existing types, others give rise to new types. If motifs represent the building blocks of a story – just as words form the basic elements of a sentence – they should recur in different stories. We can consider motifs to be recurring in two ways. First, motifs are 'recurring' if they are found in multiple *variants* of a particular folktale type. The version most widely known today of *Little Red Riding Hood* is based on the Brothers Grimm version. Charles Pernault gives a variant of the story in which little red riding hood is not rescued from the belly of the wolf. Still, both stories share a sufficient number of motifs to conceive them as variants of the same type. Second, motifs can also be said to be recurring if they occur in other story *types*. In both *Little Red Riding Hood* (ATU 333) and *The Wolf and the Kids* (ATU 123), the wolf is "cut open and filled with stones as punishment" which is motif Q426 in the TMI. The ATU catalogue only provides information about folktale types and not about variants. Therefore, 'recurring' in this paper only means recurring in different folktale types.

Genre	# tale types	# motifs
Animal tales	298	478 (1.6)
Tales of magic	223	1573 (7.1)
Realistic tales	200	666 (3.3)
Tales of the stupid ogre	124	184 (1.5)
Religious tales	140	397 (2.8)
Anecdotes and jokes	675	1069 (1.6)
Formula tales	47	80 (1.7)
Total	1707	4447 (2.6)

Table 1: Basic statistics about the contents of the ATU. For each genre the table shows the number of folktale types, the number of motifs and the average number of motifs per tale type.

3. Quantitative analysis of the ATU catalogue

3.1. Statistics

The ATU catalogue lists 2247 unique folktale types divided into seven genres. In our analysis we will only use the 1707 types that explicitly mention the motifs that belong to that type. These 1707 types contain 4447 motif instances and 3698 unique motifs. The fact that motifs are infrequently reused indicates a high *specificity* of the index language; the low average motif sequence length (2.6) indicates a low *exhaustivity* of indexing (Van Rijsbergen, 1979, 13). Table 1 presents some basic statistics per genre. We see that 'Anecdotes and Jokes' are overly well represented. Furthermore, we see that the average motif sequence length is much longer in 'Tales of Magic' than in any other genre.

S_k	k
338	1
6	1
5	1
4	4
3	14
2	65
1	1170

Table 2: Frequency spectrum of subgraph size.

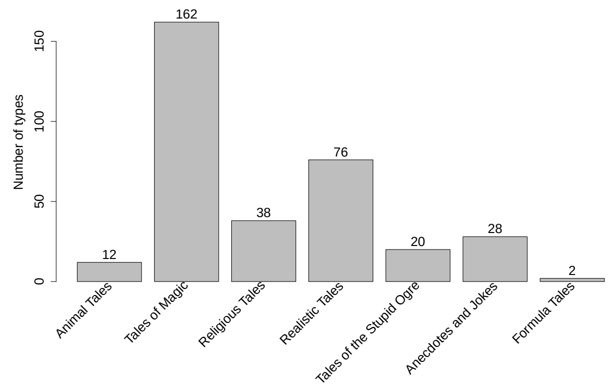


Figure 1: Frequency of genre types in main subgraph.

We are interested in the way folktale types are related in terms of their motif sequences. If certain motifs recur in different folktale types, this provides evidence for the idea that motifs can be used as building blocks for (new) stories. Each folktale type T can be considered as an n-dimensional vector of attribute values:

$$T = (m_1, m_2, m_3, \ldots, m_n) \qquad (4)$$

where m represents a motif and n the number of motifs of T in the ATU catalogue. We construct an undirected graph in which all 1707 folktale types from the ATU catalogue represent the nodes. The nodes are connected to each other if they have any motifs in common. The weight of the edges is defined by the number of overlapping motifs between two types, i.e. the size of their intersection or matching coefficient. Note that in this graph motif sequences are considered sets or 'bags of motifs'. In this paper we do not take into account the particular order of motifs.

The graph contains 1256 subgraphs, but not all subgraphs are equally large. Table 2 shows the frequency spectrum for the size of the different subgraphs. It summarizes the frequency distribution in terms of the number of subgraphs with size S_k per frequency class k. There is one subgraph that consists of 338 folktale types, one subgraph that contains 6 types, and there are 65 subgraphs with a node size of 2.

3.2. Motifs as building blocks

31 percent of the types in the ATU catalogue are connected to other types. The connected folktales form subgraphs of different sizes. One large subgraph containing 338 nodes stands out. In this subgraph, many folktales share motifs which supports the view of motifs as building blocks of stories.

It would be interesting to have insight into which folktale types and genres are represented in this subgraph. Figure 1 clearly shows that most of the types ($N = 162$) belong to the genre of 'Tales of Magic'. This indicates that folktale types in the category 'Tales of Magic' have the largest amount of homogeneity between them in terms of their motif material. In second place come the 'Realistic Tales'. This is interesting as it might reflect the idea that Tales of Magic and Realistic Tales are alike: they both describe adventures and heroes and differ only in the use of magic. Moreover, we see that all distinguished genres in the ATU catalogue are represented in the subgraph. This shows that there are no clear boundaries between the genres in terms of the motifs used. Overall we can state that the existence of overlapping motifs supports the idea that motifs can be recombined to produce new story types.

The subgraph can provide us with information about the relative importance of folktale types based on their positions in the structure of the complete subgraph. This is useful information for a model of oral transmission of folktales, because it adds insight into where many motifs are exchanged. Moreover, because many motifs are shared among many folktale types, these 'exchange centers' provide the strongest case for the view of stories as sequences of recurring motifs.

There are a few ATU types that can be expected to be central nodes, because, at least in narrative practice, they seem to supply other types with one or more of their own motifs. These 'central stations' in the land of Tales of Magic include: (1) *The Dragon-Slayer* (ATU 300), (2) *The Magic Flight* (ATU 313), (3) *Bird, Horse and Princess* (ATU 550) and (4) *Water of Life* (ATU 551). These types play an important role in Propp (1968) and should contain the prototypical motif sequences that Tales of Magic have. We will use the graph to test whether the hypothesized centrality of these types is justified.

One way to approach a node's centrality is by looking at its degree. We are interested in those cases where a folktale type is connected to many other folktales and where many unique motifs are shared. Therefore, we define the degree of a folktale as the number of unique motifs shared with its connected nodes. More formally, we take the size of the union of the intersections between a type T and the types $(T_1, T_2 \ldots T_n)$ connected to T:

$$degree(T_j) = |\bigcup_{i=1}^{n}(T_j \cap T_i)| \qquad (5)$$

Table 3 lists the top 10 nodes in the network. Two out of four tale types that were expected to be central nodes are present in this list: *Bird, Horse and Princess* and *Water of Life*. Compared to the average degree of 2.9, the other two

ATU type		degree
425A	*The Animal as Bridegroom*	18
403	*The Black and the White Bride*	14
425B	*Son of the Witch*	13
875	*The Clever Farmgirl*	12
560	*The Magic Ring*	12
550	*Bird, Horse and Princess*	12
531	*The Clever Horse*	12
400	*The Man on a Quest for his Lost Wife*	11
551	*Water of Life*	10
480	*The Kind and the Unkind Girls*	10

Table 3: Top ten folktale types in terms of their degree.

types also score relatively high with a degree of 7 for the *Dragon-Slayer* and 9 for the *Magic Flight*. *The Animal as Bridegroom* has by far the highest degree. This type list motifs common to many other types, such as B620.1 ("Daughter promised to animal suitor") and D2006.1.1 ("Forgotten fiancee reawakens husband") which is also present in the *Magic Flight*. It remains to be seen whether this type fulfilled an exemplary role in oral transmission or whether it is more likely to be the result of extensive borrowing of motifs from other stories. In any event, these types strongly suggest that, at least in the case of 'Tales of Magic' and to a lesser extent 'Realistic Tales', stories can be created by intertwining motifs from other stores.

3.3. Motifs are not building blocks

We should not, however, jump to any conclusions. Darányi and Forró (2011), for instance, have shown on the basis of cluster analyses that for a small part of the ATU catalogue (the genre of 'Tales of Magic') we can find partially overlapping types. However, for the ATU catalogue as a whole, many motif sets are mutually exclusive and the overlap between folktale types in terms of their motif material is rather sparse. The frequency spectrum in Table 2 shows that 1170 out of 1707 folktale types share no motifs with other types. For these types, the occurrence of a single motif is enough evidence to unambiguously distinguish a certain type. Within the group of tales that are connected to each other, there are 36 types that have completely equal sets of motifs. Again, this brings into question whether we can distinguish multiple folktale types on the basis of their motif material. Finally, there are many types ($N = 983$) that consist of a single motif. Here it becomes unclear what the difference is between a motif and a tale type (Dundes, 1997, 197). For all these cases, it has no added value to define folktale types in terms of sequences or sets of motifs.

The numerous folktale types in the ATU catalogue that consist of only unique motifs constitute a problem for a motif-based story grammar. We want to conceive motifs as the basic elements with which new stories (variants and types) can be produced. However, most motifs in the ATU catalogue are exclusively associated with single tale types. The ATU catalogue does not provide a positive clue that these motifs can recur. This does not exclude that they could recur, for instance in variants of the tale type.

Still, the amount of motifs unique to single tale types will hinder the generalization capabilities of any system induced from this data. To underscore this point more forcefully, let us explain this prediction in more detail.

On the basis of the frequency spectrum in Table 2, we can estimate the probability of finding a folktale type with solely non-overlapping motifs. This can be accomplished by dividing the number of tale types with unique motif material by the total number of tale types. We have $N = 1707$ folktale types and $n_1 = 1170$ types that have completely unique motif material. The probability P of finding a story with only non-overlapping and thus new motifs is $P = 0.69$, which is rather high.

From the viewpoint that the ATU catalogue is a motif-based classification system, all this indicates an important shortcoming of the system. The long tail distribution of types in the system shows that we are dealing with a collection of unrelated exemplars with little predictive power. The system falls short, because the description level of motifs is too specific to describe tales in terms of sequences of recurring motifs.

4. Concluding remarks

We investigated whether the description level of the ATU catalogue types is appropriate for a model of stories as sequences of recurring motifs. We have shown that many ATU types in the genres of 'Tales of Magic' and 'Realistic Tales' share motifs which makes it possible to describe motifs as building blocks to create (new) stories. However, for a model that aims to describe all kinds of folktales in terms of motif sequences, the description level of the ATU catalogue is inappropriate. The degree of specificity and consequently the lack of co-occurring motifs makes it hard to generalize over different stories in terms of their motif sequences. In the ATU catalogue, the majority of motifs cannot be conceived as building blocks for (new) stories. Therefore, we should aim to discover a way of formalization that is more appropriate; one that on the one hand enables us to discover enough commonalities between stories, while on the other ensures that enough distinctive features remain.

Future research should be directed towards a system where motifs exist at different levels of abstraction. In this multi-layered system, low-level motifs are compatible with those found in the TMI and will consist of particular phrases, often as concrete as strings of contiguous words ("the big bad wolf"). At a more abstract level, we will look for non-contiguous co-occurrences of higher-level linguistic elements, such as subject–verb–object triplets and sequences of triplets. This more semantic level of description strives to be compatible with the functions developed by Propp.

5. Acknowledgements

This work has been carried out within the Tunes and Tales project, which is funded by the Royal Netherlands Academy of Arts and Sciences through the Computational Humanities program.

6. References

Sándor Darányi and László Forró. 2011. Detecting multiple motif co-occurrences in the Aarne-Thompson-Uther tale type catalog: A preliminary survey. Presented at the second AMICUS workshop.

Daniel Dennett. 1995. *Darwin's dangerous idea: evolution and the meanings of life*. Penguin Books. Penguin Group, England, London.

Alan Dundes. 1997. The motif-index and the tale type index: a critique. *Journal of Folklore Research*, 34(3):195–202.

Heda Jason. 2007. About 'motifs', 'motives', 'motuses', '-etic/s', '-emic/s' and 'allo/s-', and how they fit together. *Fabula*, (48):85–99.

Vladimir Propp. 1968. *Morphology of the folktale*. Publication of the Indiana University Research Center in Anthropology, Folklore, and Linguistics. University of Texas Press, Austin.

Stith Thompson. 1946. *The Folktale*. Dryden Press, New York.

Stith Thompson. 1955–1958. *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jestbooks, and local legends*. Indiana University Press.

Hans-Jörg Uther. 2004. *The Types of International Folktales: a Classification and Bibliography Based on the System of Antti Aarne and Stith Thompson*, volume 1–3 of *FF Communications*. Academia Scientarium Fennica, Helsinki.

Erica Van Boven and Gillis Dorleijn. 2003. *Literair Mechaniek, Inleiding tot de analyse van verhalen en gedichten*. Uitgeverij Coutinho.

Cornelis Joost Van Rijsbergen. 1979. *Information Retrieval*. Butterworths.

Understanding Objects in Online Museum Collections by Means of Narratives

Chiel van den Akker, [1] Marieke van Erp,[2] Lora Aroyo,[2] Roxane Segers,[2] Lourens van der Meij,[2] Susan Lêgene,[1] and Guus Schreiber[2]

[1] Department of History VU University Amsterdam

[2] The Network Institute, Department of Computer Science VU University Amsterdam

Boelelaan 1105

[1] 1081 HV Amsterdam

[2] 1091 HV Amsterdam

[1] c.m.vanden.akker@vu.nl, [2] marieke@cs.vu.nl

Abstract

In this contribution, we present the narrative model used in Agora, an interdisciplinary project of the history and computer science departments at VU University Amsterdam and two cultural heritage institutions, the Rijksmuseum in Amsterdam and Sound & Vision in Hilversum. In the Agora project, we develop methods and techniques to support the narrative understanding of objects in online museum collections. A first demonstrator is now being tested. Here, our focus is on the specificity of modeling narratives in the heritage and history domain and the solutions Agora offers to specific problems of that domain.

In Agora, we believe that the interpretation of objects in online museum collections is supported by enriching the museum collection metadata with a structured notion of historical events and the (semi-)automatically generation of proto-narratives from those events. Starting from historical theory, three proto-narratives are distinguished: a biographical, a conceptual, and a topological proto-narrative. These proto-narratives are organizations of events based on the theory of narrative and historical theory. Proto-narratives not only take basic characteristics of the narrative into account, but also historical periods and complex historical events.

Keywords: narrative understanding, historical theory, events, online museums

1. Introduction

The Agora-project[1] aims to support the access to and interpretation of objects in cultural heritage collections. Its goal is to help the interpretation process of cultural heritage collections. The notion of narrative is central to this.

In Agora, we believe that the interpretation of objects in online museum collections is supported by enriching the museum collection meta-data with a structured notion of historical events and the (semi-)automatically generation of proto-narratives from those events. These enrichment and modeling efforts provide interesting challenges for information science as the historical domain is still far from charted in terms of formal representation. In this contribution, our focus is on the specificity of modeling narratives in the heritage and history domain and the solutions Agora offers to specific problems of that domain.

This paper is structured as follows. In Section 2, we explain our motivation. In Section 3, we identify issues in narrative modeling. In Section 4, we explain how Agora's modeling of narratives is related to the theory of narrative. In the Section 5, we explain how Agora deals with domain-specific issues of narrative modeling by relating it to historical theory on narrative.

2. Motivation

It is widely acknowledged by scholars and information scientists alike that narratives structure our understanding of the world and each other (Lakoff & Narayanan, 2010; Tuffield et al., 2006; Ricoeur, 1984) Agora is about narratives as a means to interpret objects in online heritage collections. Whereas other work on modeling narratives has been to understand the underlying models that curators use when they for example design an exhibition (e.g., Mulholland et al, 2011), we encourage users to make their own narrative based on (semi-)automatically generated proto-narratives that are built up from events and start from historical theory on narrative.

Although the notion of narrative in history is much debated in historical theory, there is a strong consensus that historians tell stories (Roberts, 2001). The focus of historical theory on narrative has always been the historical monograph. However, information technology provides us with a different medium of interpreting and representing history, providing new perspectives on the past. Therefore the historical monograph should no longer be the prime focus of historical theory on narrative (Rigney, 2010).

In Agora, we model narratives to support the user's interpretation of objects in cultural heritage. As a platform, Agora is an example of a different medium to tell stories. A first demonstrator[2] is now being tested. Agora does not provide its users with ready-made stories; instead users are encouraged to make their own narratives. As such Agora takes the participatory nature

[1] http://agora.cs.vu.nl/

[2] http://agora.cs.vu.nl/agoratouch/

of narratives in online environments seriously. Agora is not only based on narrative theory and historical theory on narrative; it also provides a means to reflect on historical narratives created in online environments.

3. Defining the narrative and its specificity in the heritage and history domain

Historical narratives share with other types narratives (e.g., novels, myths, short stories) its defining characteristics:

- narratives are organizations of events making up a comprehensive whole;
- this comprehensive whole can be referred to as the theme or moral of the story;
- since events are what persons do or goes through, narratives always involve characters;
- events in a narrative make up a chronicle, a sequence of events, one after the other (*fabula*);
- these events are, however, usually not presented in their chronological order; this is so because the *understanding* of a sequence of events does not coincide with their chronological order: the same event may for example initiate a story in one narrative and bring closure to a story in another; the order of presented events is called the *story* or *plot;* and
- *narratives* may be presented by means of different types of media (monograph, film, play, interface, etc.).

These defining characteristics of narratives are based on narrative theory (Bal, 1985) and in accordance with historical theory on narrative (Ricoeur, 1984). There are also differences between historical narratives and other types of narratives (we only discuss differences on the level of the narrative). These differences concern:

- historical *periods* and its relation to events and the narrative;
- *complex* events in relation to periods and the narrative; and
- the *criteria* used to incorporate an event in a narrative.

We will first explain how the defining characteristics of narratives are modeled in Agora. Then we will explain how Agora deals with the issues typical of historical narratives.

4. Objects, events and proto-narratives

Narratives are organizations of events (Aristotle, 2006; Ricoeur, 1984). In Agora, we use the Simple Event Model (Van Hage et al., 2011) to model events with which we enrich the metadata of objects in the museum collections. An event is something that happens at a certain *time* and *place* involving an *actor* as either agent or patient. An event is thus something that an agent *does* (e.g., attacking) or something that an agent *undergoes* (e.g., an earthquake).

Although each event is a concrete particular (Davidson, 2001), it can be described in general terms, hence as a *type*. Following this, four event properties are distinguished: time, place, actor, and type. In Agora, each event is unique because of the "name" it bears. So two events may share all their event properties and still be distinct entities because they have different names (e.g., "Operation Fall Gelb" and "German invasion of Low Countries"). Each event may involve more than one actor and each event may belong to different event-types.

Since museum collections consist of objects, we must relate the notions of events and objects. Two object-event-relations are distinguished:

- an object represents an events (e.g., a painting depicting the battle of Shimonoseki)
- an object is used in or functions in an event (e.g., a canon used in the battle of Shimonoseki)

This enrichment of objects using event-information enables event-centered browsing of collections. In Agora, objects are interpreted by means of events related to those objects.

To provide the user with possible meaningful relations between events, supporting the narrative understanding of objects, we distinguish between three *proto-narratives*: a biographical, conceptual and a topological proto-narrative (Van den Akker et. al., 2011). Two or more events are related because:

- they involve the *same actor* (e.g., Captain François de Casembroot or the vessel "The Medusa"): biographical proto-narrative;
- they belong to the *same type* (e.g., Battle, Imperialism): conceptual proto-narrative; and
- they happen at the *same place* (e.g., Shimonoseki): topological proto-narrative

The choice of a topic is a first step in the narrative understanding of an object. If the user is interested in the vessel "The Medusa", she may want to know more about its role in the "Battle of Shimonoseki" and possible other events "The Medusa" was involved in. If the user is interested in the concept of "Imperialism", she may want to know about other acts of imperialism. If she is interested in the place "Shimonoseki", she may want to know about other events related to this location. The choice of a topic thus determines the interpretation process and the story supporting that interpretation.

The biographical proto-narrative may be about a person (François de Casembroot), a country (Japan), or even an object (The Medusa) if the object happens to be an agent, as might be the case with certain ships. The difference between a country as a location and a country as an "actor" is their relation to an event. If a country does something (e.g., attacking) or undergoes something (e.g., being attacked), then the country is modeled as the "actor"' of the event. If an event happens in a certain country, then that country is modeled as the "location" of the event. It should be noted that an event may have both the country as an "actor"' and as a "place" (e.g., the attack on Shimonoseki has the event properties "Shimonoseki" and "Japan" as "place", and the event properties "Japan" and "The Dutch" as "actors"). It is up to the user to decide whether he or she prefers a biographical, conceptual, or topological proto-narrative, focusing on agents, concepts, or locations.

The three proto-narratives enable users to interpret the history of a certain place, concept or actor. As such these proto-narratives follow the logic of the event-model. We do not include the temporal dimension as a possible type of relationship to form a proto-narrative, as each event-event relation designates a temporal structure.

Proto-narratives are derived from the navigation path of the user browsing the online collection in which objects are enriched with event-information. Providing a choice of automatic generated proto-narratives from the navigation path, the user is allowed to integrate the objects from the museum collection into an overarching narrative, establishing a personal interpretation of those objects. An example of a set of proto-narratives that is derived from a user's navigation path is shown in Figure 1.

The narratives are ranked by the length in number of different objects and events included in the narratives. The interface also indicates the type of each proto-narrative that is generated. Users can click on each narrative to see further details about the narrative, such as of which objects and events it consists.

The proto-narrative is the first dimension along which we can order events; the event properties of the events belonging to a proto-narrative provides a second dimension ordering. One can, for example, group all events by "type" in a narrative on Shimonoseki, showing that this place is more often associated with "battle" than with "trade" or *vice versa*. The events belonging to a topological narrative can be ordered by time, actor, and type. Conceptual narratives can be ordered by time, actor, and place and biographical narratives can be ordered by time, type, and place.

The *first dimension* thus consists of relating events to a topic by selecting those events which have that topic as an event-property. Here the order of events is determined by the order as they appear in the user's navigation path. The *second dimension* consists of ordering the events *in* the proto-narrative on the basis of the event-properties of those events other than the one which was used in the first dimension. These two dimensions allow users to be as specific as they would like to be in their narrative.

The defining characteristics of narratives (Section 3) are all accounted for. If the events in a proto-narrative are ordered by time, then the *fabula* (i.e., the chronological order of events) is established. Since all events happen at a certain time, they can be ordered chronologically. In Agora, events can be part other events, they may overlap in time or occur at the same time in different locations. At this moment, the chronological presentation of events starts with the chronologically first event. If two events start at the same time, the event with the longest duration is

Topic	Narrative Type	objects	events	objects plus events
Nederland	topological	11	4	15
Decolonialization	conceptual	11	4	15
militaire geschiedenis	conceptual	10	3	13
overzeese geschiedenis	conceptual	7	2	9
Indonesië	topological	7	2	9
Ons Vrije Nederland	biographical	7	2	9
politieke geschiedenis	conceptual	6	2	8
tweede kwart 20e eeuw	conceptual	5	3	8

Figure 1: Example view of generated narratives in the Agora Demonstrator

presented first. The reason is that the event with the lesser duration is likely to be part of the event with the longer duration.

Other orderings using the event properties lead to different *stories* on the same topic. The same *fabula* may thus lead to different stories. If for example the events belonging to a conceptual proto-narrative are ordered on the basis of the event-property “actor”, a different sequence of events (story) emerges. The events belonging to the conceptual narrative on “Imperialism” ordered by actor may tell a story of the imperialism of the UK, France, The Netherlands, or the US.

Finally, narratives involve characters, they make up a “comprehensive wholes” and are expressed in a particular medium. Since in Agora one of the event properties is the actor involved, all narratives have characters in them. The two dimensions of ordering events together establish the “comprehensive whole” of the narrative. The narrative medium is the interface.

The Agora demonstrator consists on the one hand of event and facet browsing, and on the other hand of narrative ordering by means of proto-narratives (the first dimension of ordering events) and by reordering of the event-properties (the second dimension of ordering events). This allows for a great variety of narrative perspectives on the content users select while browsing the collection.

5. Periods and Complex Events

In this section, we explain how Agora deals with specific characteristics of the cultural heritage and history domain in modeling narratives.

What criteria are used to incorporate event in the narrative? Agora provides an answer to this question by having users choose a topic after possible proto-narrative are generated from their navigation-path, so that only events having as one of its event-properties the *topic* of the narrative are included.

Typical of the history domain are period names (e.g., Renaissance, Thirty Years War, French Revolution, Decolonisation). The question is how these period names are related to the concept of narrative. In Agora, we propose to view period names as a topic of a conceptual narrative. Period names such as “Renaissance” are not simply names of time lapses (as for example centuries and decades are); they are rather names of narratives. This is in accordance with historical theory on narrative (Ankersmit, 2001).

One could argue that the Thirty Years War and The French Revolution are not period names, but complex events, or a series of events. This is indeed true. It does not, however, contradict the view to model periods as conceptual proto-narratives. Moreover, the conceptual proto-narrative makes clear that entities such as The French Revolution are both period names and complex events. In Agora, a period name *is* a complex event and *vice versa*. A narrative on the French Revolution is a narrative having as a conceptual topic “The French Revolution” and it consist of events having as a type “French Revolution”.

The latter too is in accordance with historical theory on narrative. Period names are not only names of narratives; but these narratives or organizations of events must also can be considered *projects* (Danto, 1985). All events belonging to the French Revolution - events having as a type “French Revolution” - contribute to the project we call “The French Revolution”. Similarly, all objects and events belonging to the Renaissance - events having as event property the type “Renaissance” - contribute to the project we call the “Renaissance”.

We may furthermore add that the identification of period names with conceptual narratives is in accordance with Davidson’s theory of events.

Davidson’s (2001) central claim is that events are concrete particulars, that is, unrepeatable entities with a location in space and time. This definition of events is in agreement with the historian’s conception of the event (White, 2008). So now the question is how this is related to event-types.

Historians not only believe that events are concrete particulars; they are at the same time accustomed to discussing event types (e.g., revolutions, (acts of) colonialism, the rise of cities, etc.). They usually call these event types “structures”, that is, sequences of occurrences changing the course of history (Sewell Jr., 1996). It is crucial to note that the idea of events as concrete particulars does not oppose the idea of events as being of a certain type. For several particular historical events can be instances of the same type (structure). Thus for example, all acts of colonialism are, as particular historic events, non-repeatable, and, at the same time instances of the event-type “colonialism”.

This still allows the distinction between for example Dutch Imperialism in the nineteenth century and Roman Imperialism in the first century A.D.. This is so because the events belonging to a proto-narrative on “Imperialism” can be reordered according to location (e.g. Java and Gaul), time, and actor (the Dutch, the Romans). Obviously, “Dutch Imperialism” and “Roman Imperialism” may also simply be two different topics of two different proto-narratives.

All of this is in line with Davidson’s solution to the problem how an event as a concrete particular can account for recurrence. Davidson (2001) argues that there is just one, non-repeatable, particular event (e.g., colonialism) and all acts of colonialism are subparts of that event. So the *sum* of all acts of colonialism is a particular event, in this case the event “colonialism”. This is how an event as a particular occurrence can account for recurrence. Treating period names as the topic of a conceptual narrative, an organization of events which all have as event property the topic as type, is completely in line with this. It is furthermore in line with the contention that historic events are elements of narratives, as is emphasized by historical theorists (Danto, 1985; Ricoeur, 1984; White, 2008).

6. Conclusion

In this contribution, we presented three proto-narratives that support the narrative understanding of objects in online heritage collections. First we have shown that these proto-narratives share the general characteristics of narratives. Then it was argued that these proto-narratives also take domain specific issues into account.

We are now in the midst of testing the demonstrator in which we have implemented these proto-narratives. The first results show that users benefit from the enrichment of objects by means of a structured notion of event and the proto-narratives derived from the navigation path they have taken. The results of the pilot will be used to further improve the demo. We intend to publish on how to evaluate narrative understanding and the design and results of the first pilot in the near future.

7. Acknowledgements

The Agora project is funded by NWO in the CATCH programme, grant 640.004.801
We would like to thank the anonymous reviewers for their helpful comments on an earlier draft of this paper.

8. References

Akker, C. van den, S. Legêne, M. van Erp, L. Aroyo, R. Segers, L. van der Meij, J. Van Ossenbruggen, G. Schreiber, B. Wielinga, J. Oomen and G. Jacobs (2011). Agora Digital Hermeneutics: Online Understanding of Cultural Heritage. *Proceedings of the 3rd International Conference on Web Science (WebSci'11),* 14 - 17 June, Koblenz, Germany.

Ankersmit, F. (2001). *Historical Representation*, Stanford: Stanford University Press.

Aristotle, (2006). *Poetics*, Newburyport: Focus Publishing.

Bal, M. (1985). *Narratology: Introduction to the theory of Narrative*, Toronto: University of Toronto Press.

Danto, A. (1985). *Narration and Knowledge*, New York: Columbia University Press.

Davidson, D. (2001). *Essays on Actions and Events*, Oxford: Oxford University Press.

Hage, W.R. van, V. Malaisé, R. Segers, L. Hollink and G. Schreiber, (2011). Design and use of the Simple Event Model (SEM), *Web Semantics: Science, Services and Agents on the World Wide Web* 9, pp. 128–136.

Lakoff, G. and S. Narayanan (2010). Toward a Computational Model of Narrative, *Computational Models of Narrative: Papers from the AAAI Fall Symposium,* 11 - 13 November Arlington, VA, USA.

Mulholland, P., A. Wolff, T. Collins, and Z. Zhrahal. (2011). An event-based approach to describing and understanding museum narratives, *Proceedings of the Workshop on Detection, Representation and Exploitation of Events in Semantic Web (DeRiVE 2011), in conjunction with the Tenth International Semantic Web Conference (ISWC 2011).* 23 October 2011, Bonn, Germany.

Ricoeur, R. (1984). *Time and Narrative* Vol. 1, Chicago: University of Chicago Press.

Rigney, A. (2010). When the Monograph is no Longer the Medium: Historical Narrative in the Online Age, *History and Theory* 49, pp. 100-117.

Roberts, G. ed. (2001). *The History and Narrative Reader*, London and New York: Routledge.

Sewell, W. Jr. (1996) Historical Events as Transformation of Structures: Inventing Revolution at the Bastille *Theory and Society* 25, no. 6, pp. 841-881.

Tuffield, M., D. Millard, and N. Shadbolt. (2006) Ontological Approaches to Modelling Narrative. In: *2nd AKT DTA Symposium*, January 2006, AKT, Aberdeen, Scotland.

White, H. (2008). The Historical Event, *Differences: A Journal of Feminist Studies*, vol. 19, no. 2, pp. 9-34.

Indexter: A Computational Model of the Event-Indexing Situation Model for Characterizing Narratives

Rogelio E. Cardona-Rivera, Bradley A. Cassell, Stephen G. Ware, R. Michael Young

Liquid Narrative Research Group
North Carolina State University
{recardon, bacassel, sgware}@ncsu.edu, young@csc.ncsu.edu

Abstract

Previous approaches to computational models of narrative have successfully considered the internal coherence of the narrative's structure. However, narratives are also externally focused and authors often design their stories to affect users in specific ways. In order to better characterize the audience in the process of modeling narrative, we introduce *Indexter*: a computational model of the Event-Indexing Situation Model, a cognitive framework which predicts the salience of previously experienced events in memory based on the current event the audience is experiencing. We approach computational models of narrative from a foundational perspective, and feel that salience is at the core of comprehension. If a particular narrative phenomenon can be expressed in terms of salience in a person's memory, the phenomenon, in principle, is representable in our model. This paper provides the fundamental bases of our approach as a springboard for future work which will use this model to reason about the audience's mental state, and to generate narrative fabula and discourse intended to achieve a specific narrative effect.

Keywords: Narrative understanding and generation, representations, retrieval and indexing, artificial intelligence, cognitive psychology

1. Introduction

Historically, computational models of narrative have focused on representation of the diverse structural properties of narratives (Lebowitz, 1985; Cavazza et al., 2001; Riedl et al., 2003; Szilas, 2003). These models consider only the internal properties of the narrative. Authors, however, intentionally design stories to affect their audience in specific ways (Bordwell, 1989; Holland, 1989). As Szilas (2010) has suggested, a computational model of narrative must go beyond simple story structure and account for how the experiencer receives the narrative.

In this paper, we provide initial steps toward a computational model that accounts for a user's comprehension process during the experience of a narrative. This model, which we call *Indexter*, explicitly reasons about the salience of narrative events in a person's memory as they experience an unfolding story. The salience of a narrative event indicates how recallable the event is in a person's mind. An author's manipulation of the salience of events during a narrative experience is a key means used to affect a reader's comprehension of the story's structure. Salience enables the drawing of connections between new material and earlier parts of the story. Salience prompts expectations about upcoming action. Lack of salience obscures predictions and facilitates surprise or misdirection. A model of narrative that accounts for salience could be linked to existing models that build off of salience to account for a reader's inference-making process (Niehaus and Young, 2010), her feelings of suspense (Cheong and Young, 2006), and her level of surprise (Bae and Young, 2009), along with many other narrative phenomena.

Though our current model focuses on the manipulation of salience in narrative, salience alone is not sufficient for the modeling or creation of effective stories. A story's internal structure clearly plays a role in how a reader understands it (Graesser et al., 2002). Thus, the computational model that we present extends an existing planning-based approach to narrative (Young, 2007), which models coherent story structure (Riedl and Young, 2010). We augment this plan-based approach with information that allows us to model the updates being made to a reader's mental model of the story during *online comprehension*, that is, during the process of experiencing the narrative. To do this, we incorporate elements into the planning model drawn from an empirically verified cognitive model of online comprehension called the *Event-Indexing Situation Model* (Zwaan et al., 1995a; Zwaan and Radvansky, 1998). While we are basing our work on a planning-based knowledge representation previously developed to generate stories, our discussion here does not describe a system that uses this representation in a generative fashion. The work we describe here is preliminary. It is the first step of a four-part research agenda involving:

1. Development of a plan-based knowledge representation for narratives and an algorithm that characterizes the reader's construction of event-indexing situation models.

2. Validation of the predictive power of the algorithm and representation.

3. Integration of the computational model into a generative system.

4. Validation of the generative system in an online comprehension scenario.

A generative system which uses a computational model that characterizes both the internal structure of a narrative and its effects on a reader during online comprehension will lead to the creation of more engaging, effective and understandable stories.

2. Theoretical Bases of our Implementation

Our work has two fundamental bases. The first basis is a cognitive model of online story comprehension, the Event-Indexing Situation Model (Zwaan et al., 1995a; Zwaan and Radvansky, 1998). The second basis is an AI plan-based model of narrative, which follows directly from IPOCL (intentional, partial order, causal link) plans (Riedl and Young, 2010).

2.1. The Event-Indexing Situation Model

The Event-Indexing Situation Model (EISM) is a cognitive model of online narrative comprehension. Cognitive psychologists studying narrative comprehension define a *situation model* as an integrated mental representation of a particular situation in the story world. Situation models are formed by a reader from an amalgamation of information explicitly stated in a narrative and inferred by the reader (see McNamara and Magliano (2009) for a review of several variants of situation models). In particular, the EISM posits that, as we perceive a narrative, we discretize the narrative into *events*, or chunks of narratively important action (Zwaan et al., 1995a). This event segmentation centers around verb phrases in text and character actions in film (Zacks et al., 2009). Each event is indexed by the reader relative to a number of key factors or dimensions including:

- *time index* - the time frame in which the event occurs
- *space index* - the space in which the event takes place
- *protagonist index* - whether or not the event involves the protagonist
- *causal index* - the event's causal status with regards to previous events
- *intention index* - the event's relatedness to the intentions of a character

The EISM makes predictions about the salience of events based on these indices.

2.1.1. The EISM and Memory

Zwaan and Radvansky (1998) discuss the interplay between the EISM and memory in the context of Ericsson and Kintsch's (1995) conceptualization of Short-Term Working Memory (STWM) and Long-Term Working Memory (LTWM). Zwaan and Radvansky point out that,

> It is possible in highly practiced and skilled activities, such as language comprehension, to extend the fixed capacity of the general short-term working memory (STWM) system by efficiently storing information in long-term memory and keeping this information accessible for further processing. This expansion of STWM is called long-term working memory (LTWM) and corresponds to the accessible parts of a previously constructed mental representation in long-term memory.
>
> (Zwaan and Radvansky, 1998)

The STWM is represented in the EISM by a structure known as the *current situation model*.

Definition 1 (Current Situation Model) *The* current situation model *refers to the model of the event that is currently being perceived. This is the model at time t_n, for a given event e_n.*

When an event is perceived, a situation model of that event is built to identify what it's situation model indices are with respect to all previously perceived events. All previously perceived events represent Ericsson and Kintsch's idea of LTWM, which is represented in the EISM by a structure known as the *integrated situation model*.

Definition 2 (Integrated Situation Model) *The* integrated situation model *refers to the model of the events that have been perceived up until right before the event currently being perceived. This is the model for times t_1 through t_{n-1}, for events e_1 through e_{n-1}.*

The STWM maintains retrieval cues to information in LTWM to help with information storage and retrieval. The metaphor of a hash map is useful here: The STWM can be thought of as a set of keys to the values that are held in LTWM. For the EISM, the keys are all the unique situation model indices that exist in the development of a story. Each value in this memory hash map is a list of events that share a particular situation model index of a story. Online comprehension in the EISM is modeled as follows: each incoming event is analyzed (by the audience[1]) to determine which situation model indices it contains. The audience tries to match the incoming event to the most recently *foregrounded* events, or the events that are currently most salient.

The matching between events is done by verifying if there is any overlap between the incoming event's situation model indices and the most recently foregrounded events' indices. If the incoming event does not share any indices with the most recently foregrounded events, then a lookup is done to the memory hash map. If the lookup is successful (meaning that the situation model indices have been encountered before), the corresponding values (the previous events) become foregrounded. The incoming event is then inserted in the memory hash map and associated to the events that have been foregrounded. If the lookup is unsuccessful (meaning that we have encountered a completely novel situation), a new key is created with the new indices, and the key is mapped to the current event in the memory hash map.

2.1.2. Example Interaction Between the EISM and Memory

Consider a story which is perceived by the audience as a sequence of events $e = \langle e_1, e_2, ..., e_{10} \rangle$. In this story, only events e_1 and e_{10} have the same causal index (i.e.,

[1] In this paper, we refer to an individual experiencing a narrative as *the audience*. This term is intended to make no commitment to the medium through which the narrative is experienced, in contrast to terms like *reader*, *viewer* or *player*, which might imply restriction to a specific storytelling context.

they form part of the same causal chain). Recall that, as each individual event is perceived, a current situation model is created for it and it is subsequently integrated with the integrated situation model before the next event is perceived. According to the EISM, when event e_{10} is perceived, it acts like a retrieval cue to event e_1 due to their common causal index. Thus, the EISM will predict that, after having perceived event e_{10}, event e_1 will be more salient in memory than events e_2 through e_9.

The EISM does not make a commitment to determining which indices prove to be stronger predictors of recall. The strength of recall is operationalized in the cognitive psychology literature through various means including word association tasks (Zwaan et al., 1995a), question-answering tasks (Graesser and Franklin, 1990), timed reading tasks (Zwaan and Radvansky, 1998), and narrative summarization tasks (Graesser and Clark, 1985).

2.1.3. Building the indices in the EISM

The EISM establishes a set a criteria for assigning a situation model index to an event, with one criterion for each situational dimension: time, space, protagonist, causation and intention. Since the EISM makes predictions on salience relative to how many indices are shared between events, the criteria for indices is best expressed in terms of when events share an index.[2] The criteria for assigning situation model indices is succinctly described by Zwaan et al. (1995a) and we paraphrase and expand upon it here:

- Two events share a *time* index if they occur in the same time frame. This time frame is identifiable using the criteria employed by Zwaan (1996): two events are assumed to share a time index if they are perceived by the audience in sequential order and neither event contains an explicit discontinuity in time.

- Two events share a *space* index if they occur in the same spatial region.

- Two events share a *protagonist* index if they both involve the story's protagonist. The protagonist index is special in that it contributes to an event's salience, regardless of whether the event has been foregrounded or not. The authors of the EISM distinguish a single character as the protagonist of a story, and the model predicts that any event that deals with the protagonist is more likely to be salient than events that do not deal with the protagonist.

- Two events share a *causation* index if they are related causally. A *direct causal relation* is directed, from one event to another. A direct causal relation from event e_1 to e_2 exists, as specified by Trabasso and Sperry (1985), if it meets the logical criteria of necessity and if the events pass a counterfactual test of the form: if event e_1 had not occurred, then in the context of the story, event e_2 would not have occurred. An *indirect causal relation* between two events e_i to e_n exists if there is a path in the transitive closure of the causal relation from e_i to e_n. Trabasso and Sperry (1985) reference four types of causal relations that can exist between events:

  – *Enablement* is a causal relation that involves events which are necessary but not sufficient to cause other events.
  – *Motivation* and *Psychological Causation* are causal relations that are similar in that they both purposefully effect a change in the world, with the difference that *motivation* is goal-directed whereas *psychological causation* is not.
  – *Physical Causation* involves a naive interpretation of the physical world or of mechanical causality between objects and/or people.

- Two events share an *intention* index if they are part of the same plan to achieve a goal. Goal structures are derived from General Knowledge Structures as identified by Graesser and Clark (1985).

The EISM situational indices are coded dichotomously; that is, two events can either share an index, or not. Zwaan notes that the model may be extended in future work.

2.2. The IPOCL Planning Model

Intentional Partial Order Causal Link (or IPOCL) plans are a data structure for representing stories that explicitly model the events of a story along with the casual, temporal, and intentional relationships between them (Young, 1999; Riedl and Young, 2010). Here we introduce the IPOCL Planning Model, and give a brief formal description of what an IPOCL plan looks like.

A plan is a sequence of steps that describes how a world transitions from its beginning, or initial state, to its end, or goal state (Newell and Simon, 1961). In narrative terms, it describes how the plot of a story causes the story world to transition from beginning to end.

Definition 3 (State) *A* state *is a single function-free ground predicate literal or a conjunction of literals describing what is true and false in a story world. The* initial state *completely describes the world before the start of a plan. The* goal state *is a conjunction of literals which must be true at the end.*

Definition 4 (Planning Problem) *The initial and goal states together make up the* planning problem *to which a particular IPOCL plan is the solution.*

Characters, items, and places in the story are represented as logical constants. The actions which materialize between the initial and goal states make up the plan. Actions are created from templates.

Definition 5 (Operator) *An* operator *is a template for an action which can occur in the world. It is a three-tuple $\langle P, E, A \rangle$ where P is a set of* preconditions*, literals which must be true before the action can be executed, E is a set of* effects*, literals which are made true by the execution*

[2]A situation model index is a property of the event, independent of other events. In other words, each event has an individual time, space, protagonist, causation and intention index.

of the action (Fikes and Nilsson, 1971)*, and A is a set of characters which must consent to the execution of that action* (Riedl and Young, 2010). *For generality, P, E, and A can have variable terms to convey ideas such as "creature x steals item y." An operator for which $A = \emptyset$ is called a* happening*; these actions represent accidents or the forces of nature which are not intended by anyone.*

Definition 6 (Planning Domain) *The set of all available operators is called the* planning domain. *A domain describes all the possible kinds of actions that can occur.*

An instance of an operator, called a step, represents an actual action that will take place in a story.

Definition 7 (Step) *A* step *is a three-tuple $\langle P, E, A\rangle$, where P, E, and A are the preconditions, effects, and consenting characters from the step's operator. If any literals in P or E contain variables, or if any symbols in A are variables, those variables must each be bound to a single constant. The step "the dragon steals the treasure" is an instance of the operator "creature x steals item y."*

Plan steps are partially ordered with respect to time (Sacerdoti, 1975).

Definition 8 (Ordering) *An* ordering *over two steps is denoted $s < u$, where s and u are steps in the plan and s must be executed before u.*

A plan must guarantee that, for each step, all of the step's preconditions are true before it is executed (McAllester and Rosenblitt, 1991). A precondition can be true in the initial state or made true by the effect of an earlier step.

Definition 9 (Causal Link) *A* causal link *is denoted $s \xrightarrow{p} u$, where s is a step with some effect p and u is a step with some precondition p. A causal link $s \xrightarrow{p} u$ implies the ordering $s < u$. A casual link explains how a precondition of a step is met. In other words, p is true for u because s made it so. Step u's* causal parents *are all steps s such that there exists a causal link $s \xrightarrow{p} u$. A step's* causal ancestors *are its causal parents in the transitive closure of the parent relation.*

IPOCL plans contain structures called frames of commitment to explain a character's actions in terms of individual goals (Riedl and Young, 2010).

Definition 10 (Intention) *An* intention *is a modal predicate of the form $intends(a, g_a)$ where a is an actor and g_a is a literal that actor a wishes to be true. A* motivating step *is a step which causes an actor to adopt a goal. It has as one of its effects an intention—a modal predicate of the form $intends(a, g_a)$. A* final step *is a step which achieves some actor goal. It must have g_a as one of its effects.*

The steps which materialize between a motivating and final step make up a frame of commitment.

Definition 11 (Frame of Commitment) *A* frame of commitment *is a five-tuple $\langle S', P, a, g_a, s_f\rangle$ where S' is a subset of steps in some plan P, a is a character, g_a is some goal of character a, and s_f is a final step which has effect g_a. The steps in S' are all the steps which character a takes in order to achieve goal g_a. All steps in S' must be causal ancestors of s_f, and all steps in S' must be ordered before s_f.*

Simply put, a frame of commitment describes the steps an actor takes to achieve some goal, and the step which finally achieves the goal.

The artifact produced by a planner is a plan:

Definition 12 (Plan) *A* plan *is a five-tuple $\langle S, B, O, L, I\rangle$ where S is a set of steps, B a set of variable bindings, O a set of orderings, L a set of causal links, and I a set of frames of commitment. A* complete plan *is guaranteed to achieve the goal from the initial state. A plan is complete if and only if:*

- *For every precondition p of every step $u \in S$, there exists a causal link $s \xrightarrow{p} u \in L$. This means that every precondition of every step is satisfied.*
- *For every step $s = \langle P, E, A\rangle \in S$, and for every character $c \in A$, there exists a frame of commitment $i = \langle S', P, a, g_a, s_f\rangle$ such that $s \in S'$ and $c = a$. This means that every step which is not a happening is a member of some frame of commitment that explains why the characters who carry out that step choose to carry it out. In short, every action is taken for a reason.*
- *For every causal link $s \xrightarrow{p} u \in L$, there is no step $t \in S$ which has effect $\neg p$ such that $s < t < u$ is a valid ordering according to the constraints in O. In other words, it is not possible that a causal link gets undone before it is needed.*

IPOCL plans are formal data structures which can be manipulated by planning algorithms (Riedl and Young, 2010). They model important information about stories which, we claim in this paper, can be modified to operationalize the EISM to predict how well humans remember certain steps.

3. Indexter: a model that characterizes Situation Models using Plan Structures

Indexter is realized by expanding the IPOCL plan representation with information regarding EISM relevant data. As defined, the IPOCL plan representation already captures many of the features needed to represent EISM structures, and the enhancements we outline are straightforward to introduce. By extending an existing knowledge representation used to characterize the structural properties of a narrative, Indexter can characterize both proper narrative structure and the online mental state of the audience which experiences the narrative. This characterization is a foundational approach to a computational model of narrative; our model does not currently characterize specific narrative phenomena such as tension, suspense, expectations, humor, character development, etc. Instead, our model is intended to provide a foundation for prompting the experience of these narrative phenomena thorough the facilitation of the manipulation of salience.

Events in the EISM framework map to steps in an IPOCL plan, assuming that the steps center around verbs (in text) or actions (in film). In the remaining discussion, we use the EISM term *event* and the IPOCL term *step* interchangeably. Recall that the IPOCL model represents elements of the *fabula*, the set of events, characters, locations and the other entities within the story world. The EISM deals with online narrative comprehension which occurs when the audience perceives the narrative's *discourse*, or the telling of the events in a story.

To extend IPOCL with the EISM knowledge representation, we leverage IPOCL for use as a *discourse plan*, a structure which contains the elements from the story that will be included in the story's telling to the audience (Young, 2007). In general, discourse plans do not have to preserve the ordering of events as they occur in the *fabula* (for instance, as in cases of foreshadowing or flashback (Bae and Young, 2009)), nor do they have to contain all the events that occur in the story (e.g., as in the case of temporal ellipsis). However, in the discussion here, we use a fairly straightforward set of extensions to the IPOCL plan representation already used to model *fabula* in order to characterize just those elements of the narrative discourse relevant to the current work. The way we represent all the EISM indices is explained in the following sections.

3.1. Time

Time is implicitly represented in the current IPOCL model. Steps are modeled as executing instantaneously and IPOCL's temporal representation provides a partial ordering over all of a plan's steps' times of occurrence. Rather than extending this base representation with more complex models of time as has been done with temporal planning approaches (e.g., that of Penberthy and Weld (1994)), we approximate this by requiring that each operator in an IPOCL planning domain contain a distinguished variable called the *time frame*. In any IPOCL plan, each step's time frame variable must be bound to one of a list of constants that refer to time frames in the given narrative. These constants, enumerated as part of a planning problem's initial state description, can be defined either by the domain creator or automatically, for instance, by a temporal cluster analysis of the steps in the plan.

3.2. Space

In the current IPOCL model, spatial properties of steps are represented only to the extent that the writer of an operator includes spatial relations in its preconditions and effects. To model where an event occurs, we require that each operator in an IPOCL plan domain contain a distinguished variable called the *location*. In any IPOCL plan, each step's location variable must be bound to one of a list of constants that refer to locations in the given narrative. These constants, enumerated in a planning problem's initial state description, can be defined by the domain creator or automatically, for instance, by inferring a step's location from the bindings of variables that appear in the step's preconditions or effects.

3.3. Protagonist

In the EISM model, the protagonist is single designated character that fulfills the role for an entire story. To model the protagonist, we require that an IPOCL initial state description contain an entry designating a single character as the protagonist for the given story plan. For any given step in a plan, the protagonist index captures whether or not events involve a story's protagonist. The IPOCL plan representation already contains elements that can be used to characterize the protagonist index of each step. Each step designates a set A of consenting characters. A contains zero or more variables that are bound to the agents that consent to the execution of that step. By comparing the designated protagonist for a story with the members of A, we can easily check whether or not the protagonist is involved in the execution of the step.

3.4. Causation

The causation index captures whether or not events have a causal relation. The IPOCL plan representation contains several elements that can be used to represent the causation index. Recall that Trabasso and Sperry (1985) used four types of causal relations in their analysis of causal structure. One of these causal relations (physical causation) is currently not represented in our initial work. The other three causal relation types are represented as follows:

- *Enablement*

  An IPOCL causal link represents an enablement causal relation. In an IPOCL plan, a causal link's originating step s_i is a necessary but not sufficient condition for the subsequent step s_j through $s_i \xrightarrow{p} s_j$, because it is possible for any other step s_k to establish p for s_j, creating a causal link $s_k \xrightarrow{p} s_j$.

- *Motivation* and *Psychological Causation*

  A causal link $s_i \xrightarrow{p} s_j$, where s_i is an IPOCL motivating step represents a motivation causal relation. Recall that a motivating step is a step that causes an actor to adopt a goal. For all the steps taken by that actor in service to that goal, two types of psychological causal relations could occur:

  1. All effects of steps taken in service of the goal that do not establish a causal link to any other step that is executed by the actor are said to be psychologically caused by the action that establishes the effects.
  2. All effects that are made true by the final step (that achieves the effects that define the goal) which are not part of the goal condition are said to be psychologically caused by the final step.

  Psychological causal relations are contextualized by a specific goal state and an actor that intends to achieve it. Informally, they can be thought of as unwanted/unplanned side-effects of actions taken in service of a goal.

3.5. Intention

The intention index characterizes the role that an event plays in a character's plan to achieve a single goal. The IPOCL plan representation currently contains elements that can represent the intention index through the IPOCL frames of commitment. A frame of commitment describes the steps an actor takes to achieve a specific, designated goal condition. We say that two steps share an intention index just when they are part of the same IPOCL frame of commitment.

4. Indexter in action: using the model

Indexter allows for calculating the saliency of any previously experienced event with respect to the current event being perceived. Being able to estimate salience at any point by using our model will allow an AI planner to generate a narrative which can directly operate on the salience of events in the audience's mind. This generative system would be able to actively track and manipulate the audience's mental model to achieve specific narrative phenomena that arise, in part, from the dynamics of salience.

4.1. Calculating Salience

Once an IPOCL plan has been augmented to keep track of EISM information, one way to compute salience is by taking a majority-vote of the indices that are referenced by the event that is currently being perceived. Recall that indices are dichotomously tracked for events. If two events share an index, the value for that index is 1; otherwise, it is 0. Assume that salience can be represented with a real-numbered value between 0 and 1, where 0 represents no salience whatsoever and 1 represents maximum salience. Salience is calculated from the parameter event e_i with respect to the current event being perceived e_n. Each EISM index is assigned a weight coefficient such that the total salience will be between 0 and 1. Under these constraints, an equation to calculate the salience of any event e_i is:

$$\begin{aligned} salience(e_i, e_n) &= w_1 t_{e_n} + w_2 s_{e_n} + w_3 p_{e_n} \\ &\quad + w_4 c_{e_n} + w_5 i_{e_n} \end{aligned} \tag{1}$$

Where t_{e_n} is the time index, s_{e_n} is the space index, p_{e_n} is the protagonist index, c_{e_n} is the causality index, i_{e_n} is the intentionality index for the *event that is currently being perceived* e_n. Each index represents the overlap on that index between any event e_i and the current event e_n. For any situation model index of the current event x_{e_n}, $x_{e_n} = 1$ just when event e_i shares the x index with the current event e_n and $x_{e_n} = 0$ otherwise.

The coefficient w_j represents the contribution (weight) of its respective index to the saliency of the parameter event e_i. The coefficients are restricted to sum to 1, that is, $\sum_{j=1}^{n=5} w_j = 1$.

Clearly, assigning specific weights to the various indices will affect the salience of events in a significant way. Unfortunately, the authors of the EISM do not specify which indices are stronger predictors of recall. In future work, we will seek to determine the values for these indices empirically, through experimental evaluation. For the current discussion, however, a straightforward way to weigh the indices is to assign each an equal value. For five indices, this implies $w_j = 0.2$. Equation 1 then becomes:

$$\begin{aligned} salience(e_i, e_n) &= 0.2 t_{e_n} + 0.2 s_{e_n} + 0.2 p_{e_n} \\ &\quad + 0.2 c_{e_n} + 0.2 i_{e_n} \end{aligned} \tag{2}$$

We use equation 2 to calculate the salience of events in the following example.

Example salience calculation: The Knight's Quest

Consider the following story, in which each event is tagged with an event marker that illustrates the order in which the audience (in this case, the reader) perceives the events:

> A dragon flies to a castle (e_1), steals the treasure in the castle (e_2), and flies off to a cave (e_3). A couple of hours later the knight smiths a sword at the castle (e_4) to prepare for his quest. The following day, the knight travels to the cave (e_5), slays the dragon (e_6), reclaims the treasure (e_7), and returns to the castle (e_8).

In this story, we designate the `Knight` as the protagonist. This story is a totally ordered, text-based realization of a plan that IPOCL can produce (Riedl and Young, 2010). This plan is illustrated in Figure 1, augmented with our extended knowledge representation elements. Time, space, and protagonist indices are indicated in the steps. The steps are grouped into their specific frames of commitment, and are connected by causal links shown as arrows.

To calculate the salience of a given step, we use Equation 2. For example, to calculate the salience of step e_2 = `(steal Dragon Treasure)` at the step e_6 = `(slay Knight Dragon)`, we determine how many indices overlap between event e_2 and event e_6. Event e_2 is not connected in space, time, or protagonist to event e_6. The events are also not in the same frame of commitment. They are, however, connected causally. Using Equation 2:

$$\begin{aligned} salience(e_2, e_6) &= (0.2 \cdot t_{e_6}) + (0.2 \cdot s_{e_6}) + (0.2 \cdot p_{e_6}) \\ &\quad + (0.2 \cdot c_{e_6}) + (0.2 \cdot i_{e_6}) \\ salience(e_2, e_6) &= (0.2 \cdot 0) + (0.2 \cdot 0) + (0.2 \cdot 0) \\ &\quad + (0.2 \cdot 1) + (0.2 \cdot 0) \\ salience(e_2, e_6) &= 0.2 \end{aligned}$$

The salience of all steps relative to e_6 can be calculated in the same manner. For comparison, consider the step e_5 = `(walk Knight Cave)`. Event e_5 shares the time, space, protagonist, causation and intention indices with event e_6, such that the salience of event e_5 at event e_6 is:

$$\begin{aligned} salience(e_5, e_6) &= (0.2 \cdot 1) + (0.2 \cdot 1) + (0.2 \cdot 1) \\ &\quad + (0.2 \cdot 1) + (0.2 \cdot 1) \\ salience(e_5, e_6) &= 1 \end{aligned}$$

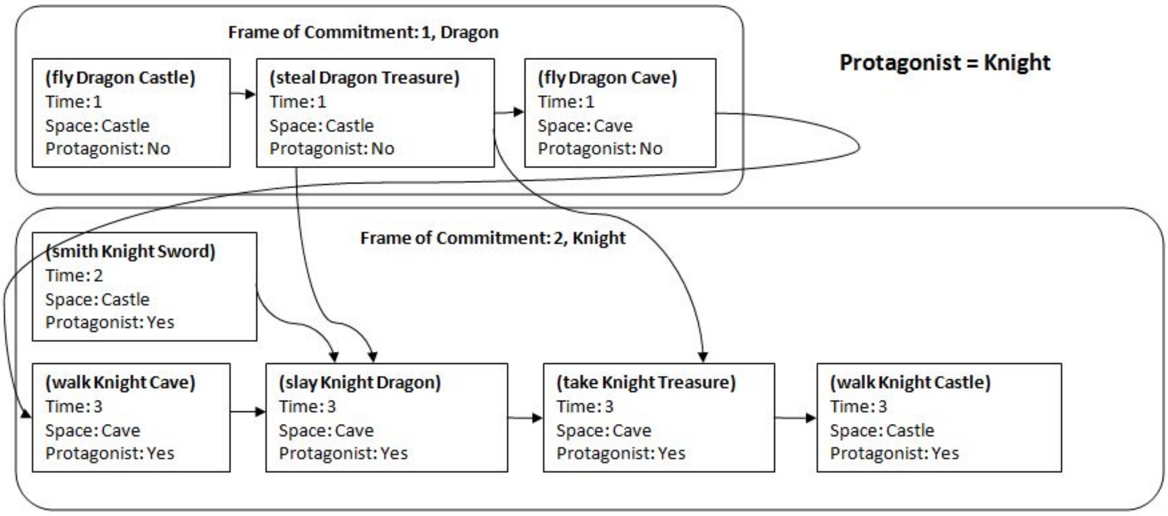


Figure 1: Example IPOCL Plan for *The Knight's Quest*, with extensions indicating aspects of our knowledge representation.

Thus, our computational model would predict that the `(walk Knight Cave)` step is more salient in the reader's mind than the step `(steal Dragon Treasure)` when the reader is reading the narrative at step `(slay Knight Dragon)`.

5. Potential Applications

The computational model we have presented here could be useful as a generative model for both *fabula* and discourse planning. We outline below some of the applications that a generative model would enable. Generally, if a particular narrative phenomenon can be expressed in terms of salience in a person's memory, the phenomenon, in principle, is representable in our model.

5.1. Fabula Planning

Using EISM information, an AI planner could construct the story in a way that it produces a plan according to the expected salience of events, which (if executed) could optimize a person's feeling of suspense (Cheong and Young, 2006) or surprise (Bae and Young, 2009). Alternatively, we could use an AI planner to dynamically construct a story that manipulates salience online, in order to actively affect a reader's inference-making process (Niehaus and Young, 2010). The latter manipulation could be used in educational contexts, with intelligent tutoring systems (Thomas and Young, 2011), as well as entertainment contexts, to affect which future events are narratively afforded (Young and Cardona-Rivera, 2011) by the events perceived in the story at a given point.

5.2. Cinematic Discourse Planning

Most work in cinematic generation reasons about low-level frame-by-frame placement of a camera in a virtual scene (Bares et al., 2000; Christianson et al., 1996). However, in film, cinematographers either explicitly or implicitly frame shot sequences to manipulate the mental state of the viewer (Branigan, 1992). Initial work on a system which reasons about high level narrative goals has been done by Jhala and Young (2010). However, this work does not make extensive use of a knowledge representation to characterize a cinematic's effect on the mental state of a viewer. EISM has been shown to accurately model cognition for film (Magliano et al., 2001). One way for cinematic generators to achieve a higher level of communicative capability is for them to reason about the cognitive effects of shots on viewers. Our model could provide this information. Given a plan step's saliency prediction, an planning system could construct a visual discourse specifically to bring certain information into focus.

Manipulation of focus would prove useful for many application that relate to visual discourse. For example, a character may deliberate on a course of action and decide to change his or her plan. In these cases a cinematographer may want to make events salient which help the viewer infer what a character is currently thinking before the character makes a drastic change in their intentions. Using our model, a discourse planner would be able to reason directly about the effects of shots on the viewer's mental state, allowing for the design of cinematic action from a narrative standpoint.

6. Limitations and Future Work

Our computational model was designed to follow the cognitive EISM very closely. Our intent is to increase the likelihood that our model will demonstrate a similar effectiveness at predicting salience. While the EISM is an empirically verified and very useful framework for characterizing the mental state of an audience during online comprehension, it does not track certain information which would be useful in a generative computational model of narrative. Also, there are details of our implementation which are subject to refinement. We identify some of the limitations from a computational perspective of the EISM and Indexter in the subsections that follow.

6.1. Limitations of the EISM

Previous work on situation models have focused on a single protagonist, and our model is restricted to one protagonist as well. However, stories often include multiple important characters beyond the protagonist (e.g. the antagonist) which may prove to be important indicators of salience. In these cases, it may be useful to extend our model to more than one character. Future work will involve determining the need for extending the protagonist index to account for multiple characters beyond the protagonist.

Space is a very complicated phenomenon. The EISM model of space is a simplification, in that it does not require or provide representations of spatial hierarchies (e.g., rooms within buildings), spaces within spaces, movable spaces (e.g., shipping containers), adjacent spaces, etc. We are interested in improving the representational capacity of our computational model to capture these types of potentially complex spatial relations in a narrative.

As we noted in Section 2.1.3., the current EISM treats situation model indices as binary (Zwaan et al., 1995a): two events are either connected by an index or not. In an effort to accurately depict the cognitive psychology research, our model treats indices in the same binary fashion. We are interested in relaxing this constraint to be able to represent events that are moderately (as opposed to directly) linked via situation models. We hypothesize that moderate situational relations will have a significant, but more gradual effect on saliency. We can think of two potential ways to approach this limitation:

1. *Introduce a distance penalty on the salience score for distant events*. This would be useful when considering events that happened (in the discourse) relatively early in comparison to the event that is currently being perceived. For example, an event further back on a causal chain should be slightly more difficult to recall than one which is closer to the step that is currently being perceived in the causal chain.

2. *Allow for a non-dichotomous indexing of events*. This would be useful when considering events that have a close relationship along one index, but that would be ignored because there is no strict overlap. For the space dimension, Zwaan and Radvansky (1998) have shown that the spatial representation in terms of distance between objects in an environment does affect response time of readers when probed with questions regarding the environment. For example, consider a spatial relationship of two adjacent rooms A and B. If the mental representation of the audience captures this adjacency, mentioning room A in the discourse will elicit transient memory saliency for room B.

6.2. Limitations of Indexter

Indexter calculates the salience between two events. Future work would extend the capabilities of our model to calculate saliency between an event and an object. Experiments have shown the abilities of people to recall objects or rooms, not necessarily a specific event alone (Zwaan et al., 1995b). However, it is not clear how the EISM handles this. Future work will determine how to expand Indexter to handle salience for elements other than events.

Indexter also depends on weights for each index to determine saliency. The weights we used are arbitrary and we allow these to be set manually. Knowing what coefficients accurately represent the predictive power of each index would increase the accuracy of salience calculations. Preliminary research has been done to determine the index weights (Zwaan et al., 1995a), and results suggest that these indices could be narrative or genre-dependent.

In Indexter, salience is calculated by computing a real-numbered value between 0 and 1 and is always calculated with respect to another event. Future work would determine at what point an event becomes salient and if it is dependent on any specific aspect of the story.

7. Conclusion

Narratives are an important part of the human experience, and they are used in diverse contexts well beyond entertainment. Psychologists (e.g., Bruner (1991)) suggest that narratives are key for explaining the ways that humans understand and reason about the world around them; these narrative psychologists posit that people perceive and interpret activities and behaviors by structuring them into a narrative. While many approaches to the development of a computational model of narrative have focused on the models' uses when generating stories, the foundational role that stories play in our cognition suggests that these models are significant also for the insight they provide to us about our own intelligence.

Previous approaches to computational models of narrative have been successful in capturing the diverse structural aspects of narrative. We propose that reasoning about the effects of narratives on their audience is the next step on the path of developing an artificial intelligence system capable of communicating narratives to humans. To this end, we have presented Indexter: a computational operationalization of the Event-Indexing Situation Model, drawn from the field of cognitive psychology. Our model extends previous work in computational models of narrative which uses AI planning constructs. Specifically, we have modified IPOCL plan structures to be capable of tracking situation model indices as a narrative is experienced. This paper presented the foundation of future work, which will leverage our model to predict the salience in memory of previously experienced events in a narrative, use that information to reason about the audience's mental state, and generate narrative *fabula* and discourse to achieve a specific narrative phenomenon.

8. Acknowledgements

This material is based upon work supported in part by U.S. DOE Computational Science Graduate Fellowship No. DE-FG02-97ER25308 and NSF Grants Nos. IIS-0915598 and IIS-0941421. The opinions, findings, and conclusions expressed here are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

9. References

Byung-Chull Bae and R. Michael Young. 2009. Suspense? Surprise! or How to Generate Stories with Surprise Endings by Exploiting the Disparity of Knowledge between a Story's Reader and Its Characters. *Interactive Storytelling*, pages 304–307.

William Bares, Scott McDermott, Christina Boudreaux, and Somying Thainimit. 2000. Virtual 3D Camera Composition from Frame Constraints. *Proceedings of the Eighth ACM International Conference on Multimedia*, pages 177–186.

David Bordwell. 1989. *Making Meaning: Inference and Rhetoric in the Interpretation of Cinema*. Cambridge: Harvard University Press.

Edward Branigan. 1992. *Narrative Comprehension and Film*. Routledge.

Jerome Bruner. 1991. The Narrative Construction of Reality. *Critical Inquiry*, 18(1):1–21.

Marc Cavazza, Fred Charles, and Steven J. Mead. 2001. *Characters in search of an author: AI-based virtual storytelling*. Springer.

Yun-Gyung Cheong and R. Michael Young. 2006. A Computational Model of Narrative Generation for Suspense. *Working Notes of the AAAI Workshop on Computational Aesthetics*, pages 8–14.

D.B. Christianson, S.E. Anderson, L.W. He, D.H. Salesin, D.S. Weld, and M.F. Cohen. 1996. Declarative Camera Control for Automatic Cinematography. In *Proceedings of the National Conference on Artificial Intelligence*, pages 148–155.

David E. Copeland, Joseph P. Magliano, and Gabriel A. Radvansky. 2006. Situation Models in Comprehension, Memory, and Augmented Cognition. In Chris Forsythe, Michael L. Bernard, and Timothy E. Goldstein, editors, *Cognitive Systems: Human Cognitive Models in Systems Design*, chapter 2, pages 37–66. Lawrence Erlbaum Associates, Mahwah, New Jersey.

K. Anders Ericsson and Walter Kintsch. 1995. Long-term working memory. *Psychological Review*, 102:211–245.

Richard E. Fikes and Nils J. Nilsson. 1971. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4):189–208.

Arthur C. Graesser and Leslie F. Clark. 1985. *Structures and Procedures of Implicit Knowledge: Advances in Discourse Processes*. Ablex Publishing.

Arthur C. Graesser and Stanley P. Franklin. 1990. QUEST: A cognitive model of question answering. *Discourse Processes*, 13(3):279–303, July.

Arthur C. Graesser, Brent Olde, and Bianca Klettke. 2002. How Does the Mind Construct and Represent Stories? *Narrative Impact: Social and Cognitive Foundations*, pages 231–263.

David Herman. 2000. Narratology as a cognitive science. *Image & Narrative (Online Magazine of the Visual Narrative)*, 1(1).

Norman N. Holland. 1989. *The Dynamics of Literary Response*. New Work, NY: Columbia University Press.

Arnav Jhala and R Michael Young. 2010. Cinematic Visual Discourse: Representation, Generation, and Evaluation. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(2):69–81, June.

Michael Lebowitz. 1985. Story-telling as planning and learning. *Poetics*, 14(6):483–502.

Joseph P. Magliano, Jason Miller, and Rolf A. Zwaan. 2001. Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15(5):533–545.

David McAllester and David Rosenblitt. 1991. Systematic Nonlinear Planning. In *Proc. of AAAI*, pages 634–639. Massachusetts Institute of Technology Press.

Danielle S. McNamara and Joseph P. Magliano. 2009. Toward a Comprehensive Model of Comprehension. *Psychology of Learning and Motivation*, 51:297–384.

A. Newell and H. A. Simon. 1961. GPS, a program that simulates human thought. *Computers and thought*, pages 279–293.

James Niehaus and R. Michael Young. 2010. A method for generating narrative discourse to prompt inferences. In *Proceedings of the 3rd Workshop on Intelligent Narrative Technologies*, pages 1–8.

J. Scott Penberthy and Daniel S. Weld. 1994. Temporal Planning with Continuous Change. In *Proceedings of the National Conference on Artificial Intelligence*, pages 1010–1015.

Mark O. Riedl and R. Michael Young. 2010. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, pages 164–7.

Mark O. Riedl, C.J. Saretto, and R. Michael Young. 2003. Managing Interaction Between Users and Agents in a Multi-agent Storytelling Environment. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 741–748.

Earl D. Sacerdoti. 1975. The nonlinear nature of plans. In *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, pages 206–214.

Nicolas Szilas. 2003. IDtension: a narrative engine for Interactive Drama. *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment Conference*, pages 187–203.

Nicolas Szilas. 2010. Requirements for Computational Models of Interactive Narrative. *AAAI Fall Symposium on Computational Models of Narrative*.

James M. Thomas and R. Michael Young. 2011. Dynamic Guidance for Task-based Exploratory Learning. In *Artificial Intelligence in Education*, pages 369–376. Springer.

Tom Trabasso and Linda L. Sperry. 1985. Causal relatedness and importance of story events. *Journal of Memory and Language*, 24:595–611.

R. Michael Young and Rogelio E. Cardona-Rivera. 2011. Approaching a Player Model of Game Story Comprehension Through Affordance in Interactive Narrative. In *Proceedings of the 4th Workshop on Intelligent Narrative Technologies*, pages 123–130.

R. Michael Young. 1999. Notes on the use of Plan Structures in the Creation of Interactive Plot. In

Proceedings of the AAAI Fall Symposium on Narrative Intelligence, pages 164–167.

R. Michael Young. 2007. Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 8:177–208.

Jeffrey M. Zacks and Khena M. Swallow. 2007. Event Segmentation. *Current Directions in Psychological Science*, 16(2):80–84, October.

Jeffrey M. Zacks, Nicole K. Speer, and Jeremy R. Reynolds. 2009. Segmentation in Reading and Film Comprehension. *Journal of Experimental Psychology: General*, 138(2):307–27, May.

Rolf A. Zwaan and Gabriel A. Radvansky. 1998. Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–85, March.

Rolf A. Zwaan, Mark C. Langston, and Arthur C. Graesser. 1995a. The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science*.

Rolf A. Zwaan, Joseph P. Magliano, and Arthur C. Graesser. 1995b. Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2):386.

Rolf A. Zwaan. 1996. Processing Narrative Time Shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5):1196–1207.

Towards Finding the Fundamental Unit of Narrative: A Proposal for the Narreme

Alok Baikadi, Rogelio E. Cardona-Rivera

North Carolina State University
{abaikad, recardon}@ncsu.edu

Abstract

Verb- and action-based event representations have been the cornerstone of narrative representation. However, these suffer from a lack of specificity as to the level of abstraction being discussed. For example, a single verb-based event can be elaborated *ad infinitum*, generating arbitrarily many new verb-based events. In this position paper, we present a proposal for the fundamental unit of narrative, which we call the *narreme*. Our contribution is two-fold. First, we present the structure of the narreme, which encodes the state of the narrative, not the state of the world. Second, we present the ways narremes can be combined, which gives rise to the structure of the narrative itself. These combinations have special properties which account for the causal, temporal and intentional relationships between the events that make up a narrative. Lastly, we present an interpretation of common narrative tasks within the context of the narreme.

1. Introduction

Many approaches to computational models of narrative discretize the narrative into *events* that are typically defined in terms of verbs, for the case of text, and actions, for the case of films (Riedl et al., 2003; Szilas, 2003; Chambers and Jurafsky, 2009; Elson and McKeown, 2009; Jhala and Young, 2010). While this level of abstraction is useful as an initial step toward a computational model of narrative, the distinction of what constitutes an event is arbitrary. Moreover, the flexibility of these units for incorporation into hierarchical structure presents a problem when trying to identify a suitable level of abstraction for action in a narrative. This in turn makes it difficult to compare approaches to computational models of narrative that differ in the level of abstraction used. For example, the action of walking to the store to buy milk could be decomposed *ad infinitum* into subsequences of actions. Consider one such decomposition: getting up, exiting the house, driving the car into the parking lot, entering the store, buying the milk. To help disambiguate what is meant by an event, this position paper presents a proposal for the fundamental unit of narrative, which we call the *narreme*.

2. Related Work

2.1. The History of the Narreme

The term narreme is borrowed from Dorfman (1969), who also used the term to refer to the fundamental unit narrative, similar to the phoneme in phonology or the morpheme in morphology. However, Dorfman is unclear as to how these narremes could be combined to form a narrative. Dorfman's narremes also suffer from the same ambiguity of abstraction as events.

2.2. Barthes' Narrative Units

In essence, narremes are similar to Barthes' (1966) characterization of narrative units. However, Barthes characterizes several types of narrative units, with varying degrees of importance:

- *functions* are narrative units that provide the basis of the narrative. They can be informally described as action-reaction sequences. For example, a telephone ringing is a function which associates the telephone ring to someone picking the phone up.
- *indices* are narrative units that expand upon the functions by providing detailed descriptions of the actions that take place. If a telephone was ringing softly, then the adjective "softly" is an index on the function of the telephone ringing.

Barthes indicates that these narrative units are combined hierarchically and sequentially, but makes no commitment as to how this combination would work. Barthes' theory has lead to successful efforts to computationally model certain types of narratives (Cavazza et al., 2001). Despite this success, Barthes' approach conflates the distinction narratoligsts (e.g. (Bal, 1997)) make between a narrative's *fabula*, or the story behind the telling, and the narrative's *discourse*, or the telling itself. This distinction is important for decoupling the modeling of aspects that relate to the story (e.g. the actual interactions of the characters (Szilas, 2003) or the narrative's conflict (Ware and Young, 2010)) and the modeling of aspects that relate to the telling (e.g. the communicative intent of the story's author (Young, 2007)). Our definition of narremes operates at the level of fabula.

2.3. Narrative Change

The narratologist Rimmon-Kenan (2002) defines a useful notion of events that we build off to define the narreme:

> To make this a bit more useful for the purpose of the present study, one might add that when something happens, the situation usually changes. An event, then, may be said to be a change from one state of affairs to another.

Our definition uses a similar notion of change as a criterion for distinguishing narremes from each other.

3. The Narreme

One of the fundamental properties of narrative is the concept of change. An individual narreme encodes the state of the narrative, along one or several dimensions in narrative space. This dimension is known as a *narrative axis*.

Definition 1 (Narrative Axis) *A* narrative axis *is a dimension which captures changes between world states. The dimension can be any measure that allows for quantization in categorical or numeric units other than world time.*

The world time represents the true total ordering of events relative to the story world. It is the "clock time" related as the story moves forward. This is contrasted with *narrative time*.

Definition 2 (Narrative Time) Narrative time *is the relative time to the Point of View character(s) in the narrative. Narrative time is monotonically increasing through the development of the fabula.*

While both narrative time and world time are often aligned, it is possible for one to depart from the other. Consider as an example a time travel narrative. Narrative time progresses forward from the point of view of those characters, while they experience different segments of world time. Given the previous definitions, we define the narreme as follows:

Definition 3 (Narreme) *A* narreme *is the basic unit of narrative structure. It encodes the state of the narrative, rather than the state of world in which the narrative takes place. A narreme is atomic along one or more narrative axes over narrative time.*

Narremes do not necessarily exclude the notion of verb-based event representations; it is possible for a verb-based representation to encode a unit of change along a narrative axis. Narremes make a commitment to a level of abstraction insofar as a particular narrative axis defines one. A narrative axis is, in essence, a criterion for determining a level of abstraction. For example, the narratologist Hogan (2011) claims that a narrative is composed of minimal units of emotional temporality. These minimal affective units could be one of several dimensions that narremes describe. The combination of narremes gives rise to the narrative's structure.

4. The Narrative Structure

The narrative structure is made up of connections between narremes. These connections form a graph structure with the narremes as nodes. An edge exists between two nodes, exactly when there is a change along at least one narrative axis. These edges have several properties which are important to consider:

- *There are no self loops.* Since a pair of narremes are connected when there is a change along a narrative axis, there cannot exist a link between a narreme and itself.
- *The edges are directed.* Two narremes are connected when there is a change along at least one narrative axis. A narrative axis is defined by changes over narrative time. Since narrative time is monotonic, these connections imply an ordering, which means the edge must be directed.
- *The graph is acyclic.* Because edges exist over narrative time, and narrative time is monotonic, there cannot be a loop in the graph.

These properties reveal that the edges induce a directed acyclic graph structure over narremes. These properties are necessary, but not sufficient in our definition of narrative. Narratologists (e.g. Bal (1997)) consider that the key ingredients in a fabula are the causal, temporal, and intentional relationships between the events that make up the narrative. Therefore, we must be able to reconstruct these relationships from our graph structure:

- *Temporal relationships* follow from the definition of narrative time.
- *Causal relationships* occur between sets of edges between narremes. A narreme causally relates subsets of incoming edges to subsets of outgoing edges.
- *Intentional relationships* occur between an incoming edge along a narrative axis and a subset of the causally related outgoing edges. An empty intentional relationship denotes unintentionality between this narreme and the preceding one.

Finally, multiple narremes may be connected to the same narreme, along different axes. Every narrative axis is independent of the others when forming edges between narremes. Put simply, a single narreme can affect several future ones, though not all in the same way.

5. Final Thoughts

Our definition of narreme is not inconsistent with current computational models of narrative. Rather it simply allows to specify the level of abstraction that these models should operate at. This representation allows a basis of comparison for different approaches to common narrative tasks, including comprehension, generation and inclusion in an interactive system.

Comprehension can be modeled as the reconstruction of the sequence of narremes. Gernsbacher (1990) described a narrative as a set of instructions which allow you to reconstruct a situation. Comprehension is then the mental process of creating a graph between the various narremes described in the discourse.

Generation can operate over the narrative space by simply searching the space of narremes until a suitable narrative is found. Given their atomicity, narremes can be exchanged indiscriminately, allowing evolutionary approaches to narrative generation.

Interactivity can accommodate narratives by allowing users to act freely within the scope of a single narreme. An interactive narrative system would then concern itself with transitioning the user from one narreme to the next, focusing on maintaining the story structure, while allowing the user a space of interactions within a narreme.

Although we have defined a formal approach the identifying the fundamental unit of narrative, future work is necessary. For instance, identifying the dimensionality of the narrative space (i.e. number of narrative axes available for the narremes) is paramount. However, we hope that future models will capitalize on the definitions that we have presented here and that our work will help focus the search for a common encoding of computational models of narrative.

6. References

M. Bal. 1997. *Narratology: Introduction to the theory of narrative*. Univ. of Toronto Press.

Roland Barthes. 1966. Introduction á l'analyse structurale des récits. *Communications*, 8(1):1–27.

Marc Cavazza, Fred Charles, and Steven J. Mead. 2001. Characters-based Interactive Storytelling. *IEEE Intelligent System, special issue on AI in Interactive Entertainment*, pages 17–24.

Nathaneal Chambers and Dan Jurafsky. 2009. Unsupervised Learning of Narratie 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*, 2:602–610.

Eugene Dorfman. 1969. *The narreme in the medieval Romance epic: An introduction to narrative structure*. Univ. of Toronto Press.

David K. Elson and Kathleen R. McKeown. 2009. Extending and Evaluating a Platform for Story Understanding. In *AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II*, pages 32–35.

Morton A. Gernsbacher. 1990. *Language comprehension as structure buillding*. Erlbaum, Hillsdale, NJ.

Patrick C. Hogan. 2011. *Affective Narratology: The Emotional Structure of Stories*. Univ. of Nebraska Press.

Arnav Jhala and R. Michael Young. 2010. Cinematic Visual Discourse: Representation, Generation, and Evaluation. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(2):69–81, June.

Mark O. Riedl, C.J. Saretto, and R. Michael Young. 2003. Managing Interaction Between Users and Agents in a Multi-agent Storytelling Environment. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, volume 34, pages 741–748.

Shlomith Rimmon-Kenan. 2002. *Narrative Fiction: Contemporary Poetics*. Routledge.

Nicolas Szilas. 2003. IDtension: a narrative engine for Interactive Drama. *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment Conference*, pages 187–203.

Stephen G. Ware and R. Michael Young. 2010. Modeling Narrative Conflict to Generate Interesting Stories. In *Proceedings of the Sixth Artificial Intelligence and Interactive Digital Entertainment Conference*, pages 210–215.

R. Michael Young. 2007. Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 8:177–208.

People, Places and Emotions:
Visually Representing Historical Context in Oral Testimonies

Annie T. Chen, Ayoung Yoon, Ryan Shaw

University of North Carolina at Chapel Hill
School of Information and Library Science
E-mail: atchen, ayyoon@email.unc.edu, ryanshaw@unc.edu

Abstract

This paper presents visualizations to facilitate users' ability to understand personal narratives in the historical and sociolinguistic context that they occurred. The visualizations focus on several elements of narrative – time, space, and emotion – to explore oral testimonies of Korean "comfort women," women who were forced into sexual slavery by Japanese military during World War II. The visualizations were designed to enable viewers to easily spot similarities and differences in life paths among individuals and also form an integrated view of spatial, temporal and emotional aspects of narrative. By exploring the narratives through the interactive interfaces, these visualizations facilitate users' understandings of the unique identities and experiences of the comfort women, in addition to their collective and shared story. Visualizations of this kind could be integrated into a toolkit for humanities scholars to facilitate exploration and analysis of other historical narratives, and thus serve as windows to intimate aspects of the past.

Keywords: historical narratives, personal narratives, visualizations, emotion identification, comfort women

1. Introduction

As individuals travel through life, the experiences they partake in, their inner states, and their interactions with the world become unique life journeys. However, their life paths also cross, and they themselves are set in particular social ecological contexts. Thus, personal narratives can be invaluable sources for understanding individuals in the past and present, as well as acquiring a broader sense of their common experiences in different social and cultural settings through the passage of time.

Theories of narrative construction and coherence have often examined the importance and yet, difficulty, of making sense of time in narrative. Narration is distinguished by ordering and sequence; narrators create plots from disordered chronological experience (Cronon, 1992, p.1349). Personal narratives link temporal properties to spatial ones; "they look back on and recount lives that are located in particular times and places… the narratives themselves are produced in particular times and places" (Laslett, 1999, p. 392). Space and time are also key dimensions along which we construct our understanding of narratives (Zwaan, 1999).

While space and time structure our personal narratives, emotions give them meaning. As Jones (2005) writes, "Life is inherently spatial, and inherently emotional… Each spatialized, felt, moment or sequence of the now-being-laid-down is … mapped into our bodies and minds to become a vast store of past geographies which shape who we are and the ongoing process of life" (p. 205-206). Without these emotions there is perhaps a question as to whether the traces of our experience would continue to linger with us: "There is also inevitably in each memory the expression of emotion; it is almost as if these memories could not exist if there had not been strong emotion felt and then expressed in the face, body, or gut" (Singer & Salovey, 1993. p. ix).

These emotional attachments are what distinguish personal narratives from simple chronologies of life events. The narrator plays a critical role in the shaping of narrative: "With narratives, people strive to configure space and time, deploy cohesive devices, reveal identity of actors and relatedness of actions across scenes. They create themes, plots, and drama. In doing so, narrators make sense of themselves, social situation, and history" (Bamberg & McCabe, 1998: □). Personal narratives and the emotions that are reflected within them "serve as a window to identity" (Horrocks & Callan, 2006). Thus, through narrative, readers can come to see how individuals make sense of themselves and their world.

However, in perusing narrative, it is not always easy for readers to make meaning of what they read. This difficulty may arise from a variety of factors such as space and time discontinuities and the inherent diversity of stories and life experiences. It is with regard to these difficulties that text mining and visualization methods may be of assistance.

This paper proposes a number of visualizations to facilitate users' ability to understand personal narratives in the historical and sociolinguistic context that events unfolded. The visualizations focus on several elements of narrative – time, space, and emotion – to explore a particular corpus: oral testimonies of Korean "comfort women," women who were forced into sexual slavery by Japanese military during World War II. The methods also leverage shared resources that are often used in text mining, emotion detection and sentiment analysis.

1.1 Personal Narratives in Historical Context: Oral Testimonies of Korean "Comfort Women"

The Japanese military sexual slavery system, or the "comfort women" system, was in operation from 1932 to 1945, during the period of the Manchurian and Pacific wars (for more detail, see Yoshiaki, 2000; Stetz & Oh, 2001; Chung, 1997). The exact number of women who were drafted into the sexual slavery system is still controversial, but it is generally estimated at 200,000 or more. The wide mobilization of military sex slaves was in the context of mobilizing human resources from occupied territories as part of the war effort. The majority of them were Korean, aged 14-19, from the rural lower classes (Chung, 1997), but women from China, Taiwan, and the Philippines were also forced to serve as "comfort women."

Due to various complexities including the power relationship in East Asia, diplomatic relations between Korea and Japan, and efforts by the Japanese government to keep the military sexual slavery system secret, the existence of "comfort women" was not revealed until 50 years after the war ended. In addition, as the experience was a "shameful" part of an individual's personal past, the victims were reluctant to identify themselves or to be formally identified as "comfort women" (Chung, 1997). While feminists, human rights activists, and historians have worked to raise public awareness of this chapter of history, the individual stories of "comfort women" are neither part of the official national histories of the countries involved, nor exist as part of their collective memories.

As the testimonies of the comfort women include experiences, perceptions, and emotions, testimonies can be seen as one form of personal narrative, trauma narratives, in a historical context. However, they are also different from personal narratives, as they are known to often contain more political tendencies, engagements of readers' sympathy, and more possibilities of intentional narrator intervention (Stephen, 1994; Beverley, 1991; Kaplan, 1991; Sommer, 1988).

The "comfort women" narratives are similar to other personal narratives in that they may be fragmented, and that they may also have factual errors, omissions, and contradictions. At the same time, their narratives are surprisingly detailed, including the names of ships that transported them from place to place, the names of their companions, and the names of the small towns by which they passed – their memories particularly vivid, persistent, and somatic, as has often been observed with trauma narratives (Misztal, 2003). In addition, extreme events connected their experiences to certain emotions, which reflect how they see and understand those experiences and actions.

1.2 Techniques for Mining People, Places and Emotion

Text mining techniques have previously been used for extracting information about historical events and displaying them using maps and timelines (e.g. HiTiME, Yamamoto et al., 2011). The Historical Timeline Mining and Extraction (HiTime) Project has developed a text analysis system for the recognition and extraction of historical events and facts from primary and secondary historical sources such as biographies, brochures, letters and old newspaper articles (http://ilk.uvt.nl/hitime/). ThemeRiver employs a river metaphor to depict changes in thematic variations over time in a large document collection (Havre, Hetzler, & Nowell, 2000).

There has also been substantial research on automated methods for identifying emotional expression in narrative. Many studies employed the Linguistic Inquiry and Word Count software, which provides statistics on the presence of words representing emotional and cognitive processes, as well as various linguistic patterns (Pennebaker & Francis, 1996) (e.g. Bantum & Owen, 2009; Liess et al., 2008). SentiProfiler incorporates Wordnet-Affect to support the visual examination of sentiment in Gothic literature (Kakkonen & Kakkonen, 2011). Plaisant et al. (2006) demonstrated how text mining and visualization could be used to explore erotics in a corpus of letters between Emily Dickinson and her sister-in-law. Pennebaker and Gonzales (2009) illustrated how linguistic patterns in blog posts might comprise historical memories and reflect the social dynamics of traumatic events.

Considering the literature, it becomes apparent that though there has been previous work in highlighting temporal, spatial and emotional aspects of narrative, extant systems do not readily support the visual integration of these three elements of narrative. However, it is also evident that these elements are inextricably intertwined, both in experience and memory. Tools that facilitate visual synthesis of these aspects of historical narrative could be of invaluable assistance to scholars. Thus, the aim of this paper is to propose methods for textual analysis in the spirit of casting light upon the historical and sociolinguistic context of the narratives, as well as the life course of the individuals whose stories are being told.

2. Methodology

The narratives employed in this analysis were compiled from two anthologies of translated interview content, compiled by the Korean Council for the Women Drafted for Military Sexual Slavery by Japan and Korea Chongshindea's Institute (Howard, 1995; Schellstede & Yu, 2000). To render the content to digital format, the content was scanned and visually inspected to correct any errors. Person and place names were identified using the Stanford Named Entity Recognizer (Finkel, Grenager, & Manning, 2005).

Emotional content was identified using a lexicon that was constructed based on Wordnet-Affect (Strapparava & Valitutti, 2004). In order to understand more about what the women in the testimonies thought and how they viewed themselves, two categories were added: Cognition and Self-reflexivity. Sentences were identified as involving cognitive processes if at least one of the following words appeared in any tense: feel, think, believe, and wonder. Sentences were labeled self-reflexive if there was reference to "myself" within the sentence. The selection of these words was partially guided by Raskovsky, Slezak, Wasser, and Cecchi's (2010) study of instropection in texts, and by the authors' own reading of the narratives. Following the extractions, visualizations were generated using PHP scripts.

3. Visualizing Personal Narratives

3.1 Juxtaposition of Life Paths

The purpose of the first visualization is to assist the viewer to examine the life courses of individuals as compared to others (Fig. 1). Each row represents the life course of one woman, and the constituent elements are places that she mentions in her testimony. The places appear in order of appearance in the text. The paths are aligned based on the places selected by the viewer. In Figure 1, the focal point of "Shinuiju" is selected. This visualization enables the user to identify and peruse testimonies that share commonalities.

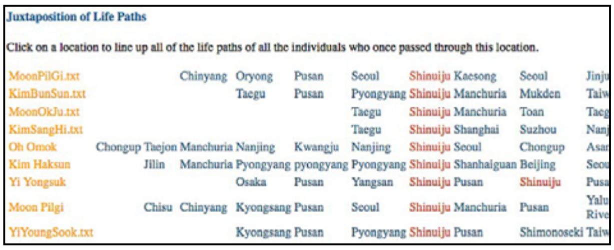


Figure 1: Juxtaposition of Life Paths

Some of the challenges in modeling these paths were differing levels of granularity in place names, as well as the lack of place names. In some cases, the names of the places that women were located were never mentioned, perhaps because they were unclear about where they were taken.

Given that many of the places that the comfort women stayed over the course of their lives may be unfamiliar to the reader, and that they traveled from place to place so often that it would be difficult to grasp even for those who are familiar, a map representation was generated using Google Maps API V.3. This representation allows the reader to see the paths taken by the women throughout the narrative (Fig. 2).

Figure 2: Spatial Depiction of a Life Path

3.2 People, Places and Emotion

As the literature review demonstrated, the experience of being is inherently spatial, temporal and emotional. Thus, this visualization was conceived to facilitate user exploration of this multidimensional landscape. Scanning the interface below from left to right, one can quickly acquire an overview of the affective content of the text, as well as significant people and places (Fig. 3). As the user hovers over the circles, the sentences that contain affective content are displayed in an info-bubble.

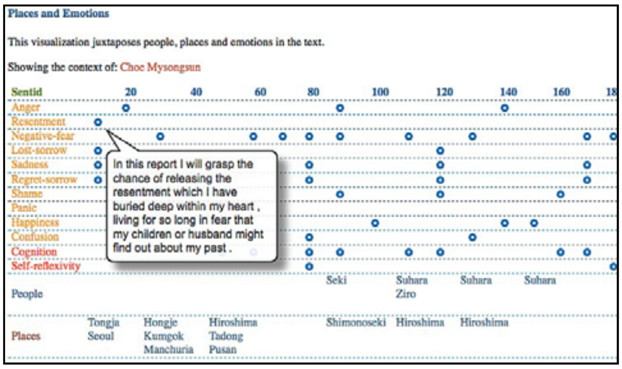


Figure 3: People, Places and Emotions

4. Discussion

4.1 Contextualizing Experiences Lived

As Cronon (1991) writes, "to recover narratives people tell themselves about the meanings of their lives is to learn a great deal about their past actions and about the way they understand those actions" (p. 1369). The visualizations presented in this paper can facilitate users' exploration of narratives; assist them to make connections between temporal, spatial, emotional and cognitive elements; and help them to understand the comfort women as individuals, as well as in terms of their collective experience.

It goes without saying that there are similarities in their experiences. As the visualizations show, many of those who later became comfort women were taken to Shimonoseki, where the women were dropped off before being assigned to other locations. It is also possible to quickly see that fear is by far the most common emotion felt by the women, not surprising given their experience. However, the consistent presence of other emotions such as sadness, regret, anger and hopelessness also contextualize their experiences.

Aside from discovering similarities, the visualizations can assist the reader in other ways as well. For instance, if one flips through the People, Places and Emotions visualization (Fig. 3) for several women, one might notice that sadness and regret are present at the end of many narratives, particularly those generated from Howard (1995). Through this, we can perhaps see the work of the editor to end each testimony with the parting thought from the focus individual regarding their past, their desires for reparations or apologies, and so on.

Examining the sentences that appear as one hovers over Sadness and Regret at the ends of the narratives, the reader may realize that though there are similarities in the women's attitudes, their different attitudes also shine through: "It was bad enough that I had to suffer what I did. I am bitter when I think of this, but I am not going to blame others any longer"[1] (Yi Yongsuk), "Of course Japan is to blame, but I resent the Koreans who were their instruments even more than the Japanese they worked for" (Kim Tokchin), and "Who would be able to guess what inner agony I suffer in my heart?" (Choe Myongsun). In the words of Kim Haksun, "Once I am dead and gone, I wonder whether the Korean and Japanese governments will pay any attention to the miserable life of a woman like me."

By highlighting the women's thoughts and references to themselves, the Cognition and Self-Reflexivity categories provide yet another view of the individual characters of the women. For example, the People, Places and Emotions visualization enables users to follow Kang Tokkyong through the narrative, experiencing her abandonment with her as she finds herself alone in a truck, and then witnessing her defiant spirit with utterances such as these: "If such a thing happened now, I would kill myself by biting my tongue off," and "I tried to throw myself off of the ship as we crossed the sea to Korea, but this woman sensed what was going on and followed me everywhere, making it impossible for me to take my own life."

As the above utterances demonstrate, though the women featured in these testimonies share similarities of experience, there also aspects of their experiences and their reactions to them that are different. In principle, simply by moving one's mouse over the People, Places and Emotions visualization for each woman, the user can acquire a taste of these differences, and then click into the testimonies for a deep perusal.

This visualization is meant to support Wertsch's distributed approach to collective remembering, in which, though collective memory is inherently social, there is not "a single system of uniform knowledge and belief," but rather, a need for "collaboration between those focusing on individual remembering and those concerned with collective phenomena" (Wertsch, 2009, p. 132). It may also stem the tide that Greene (2004) has observed of the focus of memory studies shifting away from individual remembering.

4.2 Implications for Shared Resources and Future Work

The visualizations discussed in this paper potentially contribute to the dialogue on shared resources in various ways. First, this study employed extant resources for the mining and visualization of a particular type of narrative, and therefore serves as an example of the applicability of these tools to this type of narrative. In the case of emotions, there were a significant number of false positives due to words in the lexicon that could take on different meanings. Future work could integrate a mechanism for word sense disambiguation or a machine learning approach to emotion identification.

[1] This excerpt and all following excerpts are from Howard (1995).

The visualizations in this paper primarily facilitated user exploration of the nexus of time, space and emotion. Various other aspects of narrative might be visualized in a similar fashion. For example, topic modeling techniques might be used to extract common themes and motifs from the narratives, and then the motifs could be juxtaposed with the other elements of time, space and emotion. Other aspects of the narrative such as active/passive voice, frequency of pronouns, etc., might also be integrated to provide additional methods of exploring context and mood.

In addition, the testimonies visualized in this paper were obtained from translated interview content. The techniques used in the visualizations might be applied to content in other languages, such as Korean and Chinese. An interface facilitating comparisons of testimonies in multiple languages might enable researchers to explore differences in representation due to translation, editorial style, linguistic structure and culture.

The techniques used in these visualizations could also be applied to other narratives. Historical testimonies serve as a memory of the experiences of particular groups, such as Holocaust survivors[2], Iraqi refugees[3], and survivors from other genocides[4]. As "the notion of testimony expresses urgency, a story that must be told because of the struggles it represents" (Stephen, 1994, p. 224), testimonies have increasingly gained attention from various fields as windows to unknown or little-known "truths," and to promote social justice. Visually representing individuals' life traces could be a way of representing collective experiences involving marginalization, repression, and oppression, thus granting access to intimate aspects of the past.

5. Conclusion

This paper sought to design visualizations that would be helpful for analyzing historical narrative. These visualizations enable viewers to easily see the sequence of places for any one individual, spot similarities and differences in their life paths, and form an integrated view of spatial, temporal and emotional aspects of narrative. These types of visualizations could be integrated into a toolkit for humanities scholars to assist them in exploring and analyzing narratives.

6. Acknowledgements

This work was partially supported by National Science Foundation grant IIS 0812363.

[2] United Holocaust Memorial Museum: http://www.ushmm.org/research/collections/resourcecenter/testimony/; USC Shoah Foundation Institute: http://tc.usc.edu/vhitc/%28S%28r0cji5mclbl3nv551z4w0lj3%29%29/default.aspx

[3] Refugee Council USA: http://www.rcusa.org/index.php?page=congressional-hearings

[4] The Holocaust Memorial Day Trust: http://hmd.org.uk/resources/survivor-stories

7. References

Bamberg, M., & McCabe, A. (1998). Editorial. *Narrative Inquiry, 8*(1), pp. iii--v.

Bantum, E. O., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological Assessment, 21*(1), pp. 79--88.

Beverley, J. (1991). The margin at the center: On testimonio (Testimonial Narratives), In Smith, S. and Watson, J. (Eds.). *De-colonizing The Subject*. Minneapolis: University of Minnesota Press.

Cronon, W. (1992). A place for stories: Nature, history, and narrative. *Journal of American History,* 78(4), pp. 1347--1376.

Chung, C. S. (1997). The Origin and Development of the Military Sexual Slavery Problem in Imperial Japan. *Positions*, *5*(1), pp. 219--255.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs Sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pp. 363--370.

Greene, A. (2004). Individual remembering and 'collective memory': Theoretical presuppositions and contemporary debates. *Oral History, 32*(2), pp. 35--44.

Havre, S., Hetzler, B., & Nowell, L. (2000). ThemeRiver: Visualizing theme changes over time. Paper presented at the *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pp. 115--123.

Horrocks, A. & Callahan, J. (2006). The role of emotion and narrative in the reciprocal construction of identity. *Human Resource Development International*, 9(1), pp. 69--83.

Howard, K. (Ed.) (1995). *True Stories of the Korean Comfort Women*. Y. J. Lee (Trans.). London: Cassell.

Jones, O. (2005). An ecology of emotion, memory, self and landscape. In Davidson, J., Smith, M., & Bondi, L. (Eds.), *Emotional Geographies.* Hampshire, England: Ashgate Publishing Limited, pp. 205--218.

Kakkonen, T., & Kakkonen, G. G. (2011). SentiProfiler: Creating comparable visual profiles of sentimental content in texts. In *Proceedings of Language Technologies for Digital Humanities and Cultural Heritage Workshop*, pp. 62--69.

Kaplan, C. (1991). Resisting autobiography: Out-law genres and the transitional feminist subject. In Smith, S. and Watson, J. (Eds.). *De-colonizing The Subject*. Minneapolis: University of Minnesota Press.

Laslett, B. (1999). Personal narratives as sociology. *Contemporary Sociology, 28*, pp. 391--401.

Liess, A., Simon, W., Yutsis, M., Owen, J. E., Piemme, K. A., Golant, M., & Giese-Davis, J. (2008). Detecting emotional expression in face-to-face and online breast cancer support groups. *Journal of Consulting and Clinical Psychology, 76*(3), pp. 517--523.

Misztal, B. A. (2003). Chapter 6: Studying memory. In *Theories of Social Remembering*. Maidenhead: Open University Press, p. 142.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion, 10*(6), pp. 601--626.

Pennebaker, J. W. & Gonzales, A. (2009). Making history: Social and psychological processes underlying collective memory. In Boyer, P., & Wertsch, J. V., *Memory in Mind and Culture*. Cambridge: Cambridge University Press, pp. 171--193.

Plaisant, C., Rose, J., Yu, B., Auvil, L., Kirschenbaum, M. G., Smith, M. N., Clement, T., & Lord, G. (2006). Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces. *JCDL '06*, pp. 141--150.

Raskovsky, I., Slezak, D. F., Wasser, C. D., & Cecchi, G. (2010). The emergence of the modern concept of introspection: A quantitative linguistic analysis. *NAACL Young Investigators Workshop 2010.*

Schellstede, S. C., & Yu, S. M. (Eds.). (2000). *Comfort Women Speak: Testimony by Sex Slaves of the Japanese Military: Includes New United Nations Human Rights Report.* New York: Holmes & Meier.

Singer, J. A, & Salovey, P. (1993). *The Remembered Self: Emotion and Memory in Personality.* New York: Free Press.

Sommer, D. (1988). 'Not Just a Personal Story': Women's testimonios and the plural self. In Brodzki, B., & Schenck, C. (Eds.), *Life/Lines: Theorizing Women's Autobiography.* Ithaca: Cornell University Press, pp. 170--130.

Stephen, L. (1994). *Hear My Testimony: Maria Teresa Tula, Human Rights Activist of El Salvador.* Boston: South End Press.

Stetz, M. & Oh, B.C. (Eds.). (2001). *Legacies of the Comfort Women of World War II.* London: M. E. Sharp.

Strapparava, C., & Valitutti, A. (2004). Wordnet-affect: An affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC 2004), pp. 1083--1086.

Wertsch, J. V. (2009). Collective memory. In Boyer, P., & Wertsch, J. V., *Memory in Mind and Culture*. Cambridge: Cambridge University Press, pp. 117--137.

Yoshiaki, Y. (2000). *Comfort Women.* New York: Columbia University Press.

Yamamoto, M., Takahashi, Y., Iwasaki, H., Oyama, S., Ohshima, H., & Tanaka, K. (2011). Extraction and geographical navigation of important historical events in the Web. In K. Tanaka, P. Fröhlich, and K. S. Kim (Eds.). *W2GIS 2011, LNCS 6574*, pp. 21--35.

Zwaan, R. A. (1999). Five dimensions of narrative comprehension: The Event-Indexing Model. In Goldman, S. R., Graesser, A. C., Broek, P. V. D, (Eds.). *Narrative Comprehension, Causality and Coherence: Essays in Honor of Tom Trabasso*. Mahwah, NJ: Lawrence Ehrlbaum Associates, Inc., pp. 93--110.

TrollFinder: Geo-Semantic Exploration of a Very Large Corpus of Danish Folklore

Peter M. Broadwell, Timothy R. Tangherlini
U.C.L.A.
Box 951537, Los Angeles, CA 90095-1537
broadwell@humnet.ucla.edu, tango@humnet.ucla.edu

Abstract

We propose an integrated environment for the geo-navigation of a very large folklore corpus (>30,000 stories). Researchers of traditional storytelling are largely limited to existing indices for the discovery of stories. These indices rarely include geo-indexing, despite a fundamental premise of folkloristics that stories are closely related to the physical environment. In our approach, we develop a representation of latent semantic connections between stories and project these into a map-based navigation and discovery environment. Our preliminary work is based on the pre-existing corpus indices and a shared-keyword index, coupled to an index of geo-referenced places mentioned in the stories. Combining these allows us to produce heat maps of the relationship between places and a first level approximation of the story topics. The heat maps reveal concentrations of topics in a specific place. A researcher can use these topic concentrations as a method for building and refining research questions. We also allow for spatial querying, an approach that allows a researcher to discover topics that are particularly related to a specific place. Our corpus representation can be extended to include multimodal network representations of the corpus and LDA topic models to allow for additional visualizations of latent corpus topics.

Keywords: computational folkloristics, GIS, heat maps, text mining, geospatial search

1. Introduction

The study of folklore (folkloristics) has, since its inception, been predicated on the comparison of multiple variants of a culturally expressive form both across time and across space. Within the broader discipline of folkloristics, the study of traditional narrative—fairy tale, legend, epic, ballad, etc.—has been the main focus of this approach. Despite the underlying emphasis on the comparison of story variants in time and in space, the realization of a meaningful geo-temporal representation of a folklore corpus has largely eluded folklorists.

In our work, we present a method for navigating a large folklore collection (>30,000 individual stories) that includes visualizations of the frequency with which topics, keywords or a series of keywords appear in relationship to a place name. These frequencies are visualized as heat maps, and can assist a researcher in developing hypotheses related to (a) the distribution of narrative motifs and allomotifs across a geographic area (Thompson, 1951; Dundes, 1964), (b) the emergence of locally determined "tradition dominants" (Eskeröd, 1947), (c) tradition participants' cognitive maps of an area, be it a country, a region or a smaller locality (Lynch, 1960), and (d) the interaction between concepts in the overall corpus and individual repertoires of individuals (Tangherlini, 2010. Our approach to corpus visualization allows the researcher to move beyond the traditional one story-one classifier indexing schemes that are standard for most folklore collections, and allows her to discover latent patterns based on story semantics. These patterns are visualized on top of historically relevant maps. Finally, our implementation of spatial querying allows a researcher interested in a specific area, or type of area (e.g. the heath), to develop queries that return latent semantic patterns visualized as heat maps as a result of the queries.

2. History of the Problem

The first systematic folklore theory was labeled "the historic geographic method" (Krohn, 1926), and was based on the comparison of a large number of variants of a single story or motif across a broad geographic area (in some studies, global), and over a very wide range of time (in some studies, millennia). The main goal of this approach was to find the geographic and temporal origin for any given type of folkloric expression known as the *urform*, or original form. Consequently, early folklorists were more interested in mapping where a particular folklore variant was collected (Anderson, 1923) than they were in exploring the place name referents that are a common feature of folk expressions. In the maps based on this early folklore method, time was persistent, so that a story attested in the ninth century could appear side-by-side with a story variant collected in the late nineteenth century.

Carl Wilhelm von Sydow called into question the idea that one could use these types of maps to determine anything other than the fact that a motif or story type had been attested in a particular place at some time (1943), a view that was refined by Anna Birgitta Rooth several years later (1951). Their work effectively put an end to the use of maps to "discover" the original forms of folk expressions.

Recent folklore scholarship has recognized the power of geographic representations of aspects of a folklore corpus as part of a more nuanced description of the dynamics of folklore. This recognition is tempered with an understanding that these maps show distributions of topics as related to tradition participants' conceptualizations of where various phenomena (ghosts, trolls, elves, witches, etc.) exist. In short, geographic representations of a folklore corpus are excellent tools for understanding the distribution of topics and motifs across an area, and cannot be used to answer questions about the origins of any particular expression. Understanding the distribution of topics and motifs can help researchers develop sophisticated, geographically aware research questions based on a consideration of latent patterns in the corpus as a whole. Franco Moretti (2000) labels this type of broad approach "distant reading."

Below, we illustrate how heat maps, derived from (a) the indexed representation of a folklore collection and (b) a semantic indexing of that same collection can assist in developing a distant reading approach to a folklore corpus. The maps reveal interesting patterns of distribution for closely related topics that can inform a historically conditioned reading of the tradition. Why were witches so commonly associated with Grinderslev? What may lie behind the close spatial relationship between mentions of cunning folk and ministers? The researcher can subsequently explore these questions on a "close reading" level, by drilling down to the underlying stories, thereby combining the benefits of pattern discovery across the entire corpus with specific interrogations of individual corpus occurrences.

3. Corpus Selection and Preparation

Over the past decade, various folklore archives have begun digitizing their holdings, most notably the Finnish Literary Society and the Dutch Folktale Database (Venbrux & Meder, 2004). As part of this process, many archives have geo-encoded place names associated both with the collection of a specific item as well as the place names mentioned in that item. This geo-encoding of story related place names is a critical step in the realization of a geographically aware "distant reading" approach to folklore study.

Our work is based on the largest collection of folklore produced by a single person. This Danish collection, the Evald Tang Kristensen collection, is housed at the Danish Folklore Archives at the Royal Library in Copenhagen, and comprises approximately 24,000 hand-written manuscript pages. Tang Kristensen, whose active collecting career spanned approximately fifty years from 1876-1925, collected stories and songs from more than 3,500 people. During the past ten years, we have digitized this collection, along with information concerning the informants and the places they lived.

The majority of the collection consists of legends—believable mono-episodic retrospective narratives told as true, often detailing encounters with supernatural phenomena—and descriptions of everyday life in the rural parts of the Danish peninsula, Jutland. For this work, we have concentrated on two of Tang Kristensen's main published collections of these types of narratives, and their supplementary volumes: *Danske sagn* (1892-1901), *Danske sagn, ny række* (1928-1939); *Gamle folks fortællinger om det jyske almueliv* (1891-1894) and *Gamle folks fortællinger om det jyske almueliv, tillægsbind* (1900-1902). These printed collections, based on his field collections, comprise twenty-five volumes, and 31,086 individual stories.

The volumes were scanned and processed using an OCR program that had been trained on a small subset of the printed books. The subsequent output was chunked into stories using a simple regular-expression matching procedure. The accuracy of the chunking was then manually checked against the printed collections and corrected. The indices describing stories told by informants and place names mentioned in stories for each of these collections were also scanned, aligned and integrated. This integration of the indices allowed us to develop a set of metadata for each story, including story source (informant name), story collection place, and places mentioned in the story.

All of the place names related to stories—places of collection and places mentioned—were subsequently geo-referenced, at least to the level of parish (in the late nineteenth century, Denmark was organized into *amt* [county], *herred* [district] and *sogn* [parish]). The geo-referencing of these place names was a three-part process. First, place names that could be unambiguously matched to an existing gazetteer of place names adjusted to align with late nineteenth century orthographic conventions were automatically assigned the corresponding geo-reference. Second, place names that were accompanied with a topographic reference number from an earlier index of Danish parish names (Skjelborg, 1967) were assigned to the corresponding parish's coordinates. Since parishes generally have a radius of approximately five kilometers, this lack of absolute precision does not overly distort the results of visualizations and spatial queries. Finally, place names that could not be assigned coordinates by these two methods were assigned coordinates in a semi-supervised process using DDupe (Bilgic et al., 2006).

As part of our meta-data representation of each of the stories, we included the topic index assignments from the original collections' topic indices. As with many printed folklore collections, these indices are constrained by the limitation that each story can only have one classification. Consequently, stories that span multiple topics can easily be misclassified, making it difficult for researchers to discover these stories. In earlier work, we have described a multi-modal network-based classification scheme that allows for the discovery of stories that span multiple classifications (Abello, Broadwell & Tangherlini, 2012). In this work, we avoid many of the pitfalls of the one story-one classification problem by adding a simple keyword representation of the corpus. This representation can be refined and expanded in future work.

Keywords for the test corpus were derived in RapidMiner (Mierswa et al., 2006). Common Danish stop words were removed from the corpus, and all words appearing in more than 200 and fewer than ten documents were similarly removed. The remaining ~10,000 words were lemmatized to standard Danish dictionary lemmata using the CST online lemmatizer for Danish (Jongejan & Haltrup, 2010) reducing the list by approximately 40%. Future work on the lemmatization of keywords will rely on a dictionary particularly tuned to the vocabulary of late nineteenth century Danish folklore (Feilberg, 1977).

4. Spatial Data Preparation

Many of our research methods follow those applied by Mendoza Smith, Kuznetsova, Smith and Sugihara to the USC Shoah Foundation's Visual History Archive, a very large collection of approximately 30,000 videotaped testimonies provided by witnesses to the Holocaust (Mendoza Smith et al., 2011). To facilitate exploration of the video archive by historians, the interviews were chunked into one-minute segments and human experts then tagged each segment with keywords drawn from a custom thesaurus of topics, geographic places and historical periods. These associations were stored in a database and incorporated into the Visual History Archive's online search interface. Geo-semantic visualization and exploration of the archive could therefore progress by examining the *co-occurrences* within testimony segments of topical (i.e., semantic) keywords and other keywords representing specific geographic locations.

Exploration of latent geo-semantic phenomena in the full corpus of >30,000 Danish folk stories likewise involves the discovery and visualization of topics and place co-occurrences within individual stories. For the purposes of our study, the locus of co-occurrence is the individual folk story; a topic that appears in the same story with a geographic location is considered to have co-occurred once with that place. To store and tally all such co-occurrences in the text corpus, we first ran a full-text search through each of the stories to find full-word matches of the approximately 10,000 keywords in our word list. We subsequently collapsed this keyword set via stemming and the CST lemmatizer to approximately 6,000 lemmatized keywords. We stored the ~343,000 story-to-keyword associations generated by the search in a MySQL database. We also included in this database the contents of the aforementioned collection indices listing the places mentioned in each story, of which there are 25,000 unique associations. These tables, in addition to a table that links Tang Kristensen's original story topic indices to individual stories, enabled us to run database queries to find all co-occurrences of topics and geographic locations in the text corpus.

5. Heat Maps

5.1 Methods

We use ESRI Corporation's ArcMap software to generate geospatial visualizations of the co-occurrences of specific keywords and groups of keywords with geographic locations. Other software tools we evaluated for this analysis included R, Matlab, Google Earth and Google Fusion Tables. To facilitate a geographic "distant reading" of latent geo-semantic relationships in the corpus, we mapped point locations that co-occur with certain keywords or topic indices that are of interest to researchers. Each of these points was also assigned a "z" value indicating the number of stories in which the place co-occurred with the topic in question. We then ran a spatial interpolation analysis on these points and z-values and displayed the results on a historical map of Denmark, thereby highlighting regions in which the topic had an unusually high number of co-occurrences. The heat maps for several topics can also be overlaid on the map to expose potential geospatial interactions between related or oppositional topics (see Figures 1-3). Via a process of comparison, we determined that the Natural Neighbor interpolation method (Sibson, 1981) produced the most useful results. In particular, the regions it generated appeared the most robust to changes in parameters such as the cell size of the output raster. We chose a spatial interpolation method rather than a kernel density estimation algorithm because there seemed to be little utility in assuming that the spatial density of any given story topic would exhibit a normal distribution.

5.2 Preliminary Results

If Danish legends are any indication, witches were a common nuisance in nineteenth century Denmark. Certain areas of Denmark were known for their historical association with witches, although a person simply reading through the corpus would be hard-pressed to pinpoint these locations. The heat map for the keyword *heks* (witch) highlights several areas that have significant "hot spots" for this topic. Of particular note is the area around Grinderslev (see Figure 1). Grinderslev was the

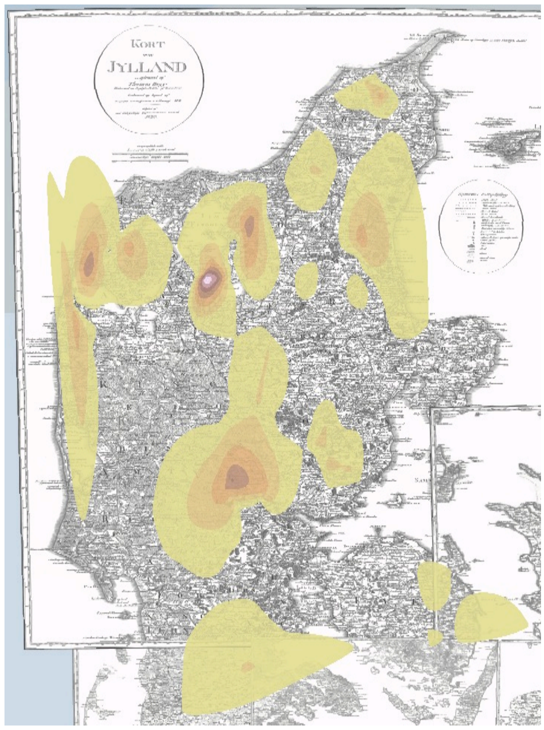


Figure 1: Heat map of the keyword *heks* (witch), with a hot spot surrounding Grinderslev

site of a well-known Augustinian monastery, *Grinderslev kloster*, founded in the twelfth century. The monastery was built near a holy spring, *Breum kilde*, but was abandoned in the aftermath of the Reformation. The spring at Breum was subsequently associated with witchcraft, and in 1686, Anne Madsdatter and her sister were burned at Breum, the last witch burning in Denmark (Bruun, 1920). Although this episode is well known in the study of Danish witchcraft, the persistent relationship between the area surrounding Grinderslev and stories about witchcraft has not been recognized previously, suggesting a topic for further, in-depth inquiry.

By the time Tang Kristensen was collecting folklore at the end of the nineteenth century, the social and political landscape was undergoing considerable change. Prior to the constitutional reforms of 1849 and their subsequent implementation over the next several decades, local power shifted dramatically from the church ministers who had been appointed by Copenhagen to locally appointed authorities. This challenge to the central authority of the Lutheran church was manifest in storytelling in the conflict between ministers on the one hand and cunning folk on the other hand (Tangherlini, 1999). The conflict was not evenly distributed throughout the country, and various parishes became strong supporters of the local minister as the most appropriate defender of local interests. This endorsement of the minister was often expressed in stories where a minister successfully defends against some form of supernatural threat, most often a ghost or witchcraft. Equally common, however, were stories that endorsed a cunning man or woman in this role of local defender. Interestingly, the endorsement of a local person as the protector of the local interests was often found in the stories of the emerging middle class of farm owners, and their farmers' party (*Venstre*) that sought local solutions to local problems. These debates over the relative merits of local control emerge in a tantalizing form in a heat map that plots the hot spots for the topic categories of ministers versus cunning folk (Figure 2).

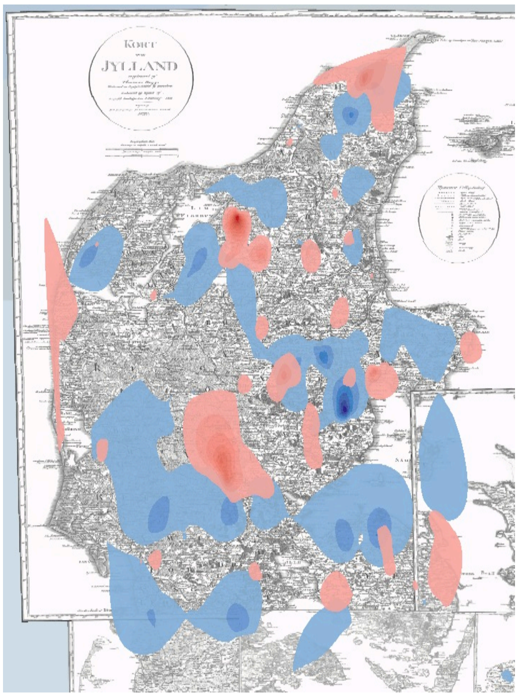


Figure 2: Heat map of the indices *præster* (ministers) in blue, and *kloge mænd og koner og deres bedrifter* (cunning men and women and their activities) in red

The proximity of these hot spots—and their overlap in certain regions—closely tracks the areas where these debates were most prominent in the Danish political landscape, such as the hot spots in the Nykøbing Mors area in north central Jutland. Of particular note is a very significant hot spot around the main Jutlandic city of Århus. Århus was in a constant struggle with Copenhagen over the economic and political control of Jutland.

Since the heat maps do not tell us whether the stories are positive or negative endorsements of the abilities of the minister or cunning folk, the researcher must investigate these stories in greater detail. At the same time, the heat map has revealed a latent pattern in the relative distribution of these topics across the Danish landscape. This pattern is intriguingly congruent with hot spots of contemporaneous political debates concerning local control. It also highlights areas, such as Skive and Nykøbing Mors in Jutland, where the Lutheran church was particularly strong (and where there were numerous Evangelical sects developing), while also highlighting areas where cunning folk were very active. For example, Vindblæs, in north central Jutland, was the center of activity for one of the best-known folk healer dynasties in Danish history (Rørbye, 1976).

Among the most common tasks for ministers and cunning folk in Danish legend tradition was to deal with the surprisingly frequent problem of haunting, manifest as ghosts (*spøgelse*) and revenants (*gengangere*). The main strategy for eliminating the threat of the haunt was to conjure it down (*at mane* or *nedmane*). As with the other heat maps, a heat map that represents the concentration of these three keywords produces some very interesting results (Figure 3).

Figure 3: Heat map of the keywords *spøgelse* (ghost) in purple and *gjenganger* (revenant) in blue and *mane* (to conjure down) in yellow

Although the terms *spøgelse* and *genganger* are roughly equivalent, the former generally describes an ethereal form whereas the latter generally describes a more corporeal form for the haunt. This heat map illustrates Eskeröd's concept of "tradition dominants" well, where one form of haunting dominates the local belief vocabulary, forcing out other possible allomotifs for the motifemic slot of "ghostly threat" (Dundes, 1964). So, for example, along the west coast in Ringkøbing country, and along the east coast in Mols and on the island of Samsø, the tradition dominant is the *spøgelse*. In other parts of Denmark, such as Hjørring county and the area surrounding Vejle both types of haunts appear, suggesting a nuanced distinction between these two types of haunts.

The hot spot for the conjuring of haunts in the area near Skive raises some interesting questions. This area is also one of the hot spots for cunning folk and, to a lesser extent, ministers (Figure 2). This area was also one of the places where the debate over the control of local churches was most intense. Indeed, a researcher aware of the significant political and religious battles that were current in these areas at the end of the nineteenth century would be immediately drawn to these patterns related to ghostly threat and the attempts to battle that threat through the supernatural intervention of conjuring.

5.3 Influence of Population Density

We conducted a study to quantify the degree of correlation between the population density of a region and the frequency with which places in that region appeared in stories from the corpus. If most such appearances were confined to a few concentrated population centers, then the efficacy of topic heat map visualizations would be considerably diminished. First we reconstructed the approximate boundaries of Jutland's 84 *herreder* (a now-defunct administrative unit) at the time of the 1901 Danish census by calculating the convex hull (Barber et al., 1996) of all places registered as belonging to a given *herred*. We then calculated the areas of these regions and used the population figures from the 1901 census to compute the population density of each *herred* (Figure 4).

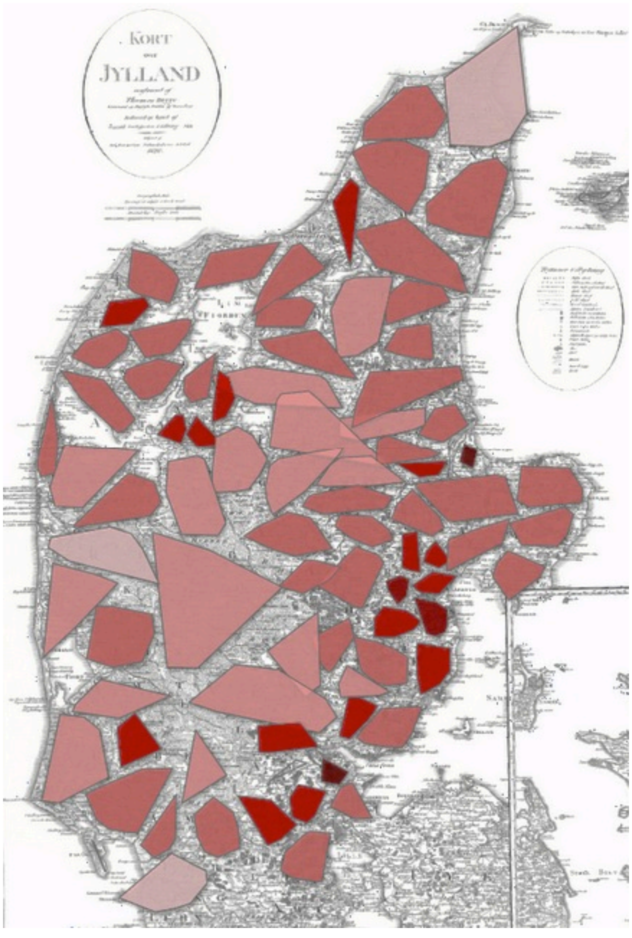


Figure 4: Choropleth map of the population densities of *herreder* (counties) in Jutland according to the Danish census of 1901. Very light red = 10-20 persons/km^2, light red = 20-75 persons/km^2, red = 75-120 persons/km^2, dark red = 120-160 person/km^2.

We also computed the "places mentioned" density of a *herred*, defining the numerator of this ratio as the total number of times each place within a region was mentioned in a story.

We found a highly significant and moderately strong correlation (Pearson coefficient r = 0.58624) between a region's population density and the "places mentioned" density of that area. This result supports the findings in Tangherlini (2010) that people are most likely to tell stories about places that are close to them. Despite the relatively flat population distribution of Jutland in the late nineteenth century and Tang Kristensen's habitual avoidance of large cities when collecting folklore, it is not surprising that the stories he collected still exhibit a general bias towards places where people lived.

Comparisons of specific topic heat maps (Figures 1-3) to Figure 4's population density map, however, indicate that the degree of correlation between a region's population and the geographic loci of individual semantic topics can vary widely. A future software environment for geo-semantic corpus exploration therefore might be most effective if rather than attempting to normalize away the influence of population density on topic hot spots, it presents population density as another dimension for the researcher to consider. For example, the software tool could overlay a topic heat map with the population density map. It could then quantify the correlation between a topic's geographic distribution and regions of high population density as an indication of the topic's affinity for populated areas.

6. Geospatial Queries

6.1 Methods

Visualizing geo-semantic relationships via heat maps can serve to highlight sets of geographic locations at which a particular phenomenon or group of phenomena may be especially prevalent and therefore worthy of further study. It is also the case, however, that researchers may pursue a line of inquiry that is effectively the inverse of the heat map approach: choosing a location of interest and proceeding to search for the topics that might be especially prominent at that location or in its vicinity. The database structures that we built to facilitate the generation of heat maps and "hot spot" regions also enable this type of investigation. In particular, researchers can execute spatial queries on the story-to-keyword and story-to-place relationships in the database to obtain a list of topics associated with places that lie within a given radius from a specified geographic location.

If the spatial region is large, or if there is a high density of story locations in the region, it may not be practical to list all of the keywords that co-occur with the places within the search region. To address the problem of how to rank the query results so that the user sees the most potentially relevant keywords first, we adopted techniques commonly used by online search engines. The first option is to return the keywords that co-occur with the most places in the region, ranked in descending order by raw co-occurrence totals. This approach may produce valid results, but it tends to favor keywords that occur in many stories and thus are more likely to be prevalent within most regions on the map. Therefore, this first search ranking may not return the words most characteristic of the region.

An alternative search ranking strategy is to normalize the co-occurrences of the keywords within the specified region with their global co-occurrence counts across the entire corpus, thus penalizing the most commonly occurring words. This approach does tend to rank highly the keywords that are particularly endemic to the region, but it also strongly favors keywords that appear rarely both throughout the corpus and within the specified

region, to the detriment of keywords that are more common globally but are also highly prevalent within the region.
Our third search ranking technique is on average the most successful at privileging keywords that are characteristic of a region but not exceedingly rare. It is a variation on the well-known “term frequency * inverse document frequency” (TF-IDF) formula, which is commonly used in text search engines to address the problem of finding a document that most closely matches a user’s search query. The TF-IDF algorithm weights the number of times each term in the query matches a given document in the corpus (its term frequency) by the term’s inverse document frequency score, which is the ratio of the total number of documents in the corpus to the number of documents in which the search term appears. Thus, terms that appear in many documents in the corpus are not allowed to have an undue influence on the results of the search query. At the same time, very rare terms are prevented from outranking terms that appear many times within the document. Our algorithm for ranking spatial query results uses the TF-IDF formula, but with the key difference that it treats locations as documents. This variation on the TF-IDF algorithm therefore may be termed RF-IPF, or “region frequency * inverse place frequency,” and has the following formula:

$$\mathbf{RF\text{-}IPF = RF * \log(|P| / |p \in P : t \in p|)}$$

where:
RF = region frequency: the number of places in the region that co-occur in stories with the keyword **t**
$|\mathbf{P}|$ = total number of places mentioned in the corpus
$|\mathbf{p \in P : t \in p}|$ = total number of places that co-occur in stories with **t**

6.2 Preliminary Results

Table 1 lists the first thirteen keywords returned by the ranking algorithms described above when applied to the results of a spatial query for all keywords co-occurring in stories with places that lie within five kilometers of the Grinderslev monastery, located at 56.697 N, 9.06788 E. As noted in the discussion below, researchers may differ on which ranking algorithm produces the most insightful results; therefore, a search interface that combines all three search rankings might be the best approach.
Whereas the heat maps offer researchers a visualization of the regional saturation of a particular topic or keyword, the spatial query returns the most common topics or keywords related to a particular place. This approach can be part of a multi-part research strategy that incorporates the distant reading approach with the more traditional close reading approach.
In the keyword lists below, it is worth noting the prevalence of certain words that occur in at least two of the lists, as these are all closely connected to the regime of witches. These include *sølv*, *vælte*, and *at læse*. Witches are generally considered to be difficult to catch; a well-known strategy is to shoot them with a silver bullet or, as is more common, a silver button (Danish farmers had silver buttons, but did not commonly have access to silver bullets). One of the main threats that witches posed was to the local economy and consequently there are numerous stories of witches tipping over hay-laden wagons through the use of curses. Often, these curses are read (*at læse*) from the most common book of incantations, known as *Cyprianus* and referred to colloquially as “the Black Book.” The keyword list confirms not only the presence of witches in the landscape (without the word for witch ranking high, but rather with words related to witchcraft ranking high), but also suggests a particular threat of the witch to the community (here the disruption of normal farming activity by tipping over wagons). Other words in these lists are equally suggestive of witches, including *at flyde*, to flow, as witches often stole milk from cows, having it flow magically into their own milk buckets.

Raw	Normalized	RF-IPF
bande (to curse)	**kusk** (carriage driver)	**paste** (to take care of animals)
hale (a tail, prob of a snake)	**reste** (remainder)	**flyde** (to flow)
sølv (silver)	**grønning** (village green)	**hale** (tail)
tigge (to beg)	**rådelig** (recommended)	**grønning** (village green)
vælte (to tip)	**boel** (a large farm)	**borggård** (fort)
læse (to read)	**kristenblod** (Christian blood)	**sølv** (silver)
flyde (to flow)	**om kap** (race or competition)	**bande** (to curse)
paste (to take care of animals)	**indhylle** (enshroud)	**søkke** (to sink down)
øre (ear)	**søkke** (to sink down)	**herre** (lord)
herre (lord)	**mæt** (sated)	**læse** (to read)
lindorm (supernatural snake)	**konfirmation** (confirmation)	**mane** (to conjure down)
stille (to place)	**tjørn** (hawthorn)	**lindorm** (supernatural snake)
østen (to the east)	**mane** (to conjure)	**vælte** (to tip over)

Table 1: Output of three keyword ranking algorithms when applied to the results of a spatial query. The first ranking algorithm also returns the following suggestive keywords in positions 14-18: **stol** (chair or stool), **ræd** (scared), **hund** (dog), **skyde** (to shoot), and **Cyprianus** (book of Satanic incantations).

The wordlists, however, also raise another set of related beliefs, particularly those in supernatural creatures, such as the *lindorm*, a serpent that often threatens to topple local churches, and ghosts and revenants, whom either the local minister or a cunning person is called on to conjure down (*at mane*).
The obvious benefit of these spatial queries is the quick connections a researcher can make between a place and its local environment and topics that have a high frequency in that place. The selection of keywords or topics for geo-spatial visualization can proceed quickly, and the researcher can then build hypotheses about the relationship between these topics and the local area.
In the above example, for instance, the connection between witches, curses and Satanic forces such as the *lindorm* and the *Cyprianus* is immediately apparent. An intriguing appearance in these lists are stories that include conjuring (keywords *mane* and *at søkke*), suggesting an area that is conceptually linked not only to witches, but also ghosts (who by the late nineteenth century had been

theologically linked to Satan), and strategies for dealing with ghosts.

7. Conclusions and Future Work

We have demonstrated two computational techniques—heat map visualizations and spatial queries—which enable researchers to discover and explore latent geo-semantic relationships within a very large corpus of text narratives. Heat map visualizations aggregate topic occurrences at individual spatial points to facilitate a "distant reading" analysis of large-scale geo-semantic phenomena within the corpus, and also suggest smaller sub-regions and related stories that may be worthy of detailed "close reading." Spatial queries are useful for finding keywords or topic groups that may be particularly applicable to a given location or region. We also discuss algorithms for ranking the results of a spatial query.

Preparing our folklore corpus for computational analysis entailed a significant amount data cleaning and processing; it is our hope that this process may in the future become more streamlined as sophisticated software tools for large-scale data ingest, automated tagging and indexing become available. Similarly, we would prefer that the investigation of geo-semantic relationships within a large corpus take place via a dedicated software platform such as a faceted browsing system that would generate topic heat maps and bounding regions automatically, rather than requiring manual data entry into ArcMap. This system also could allow researchers to perform spatial queries and to browse the results interactively.

Much of the process of interpreting the significance of data visualizations remains subjective, and thus a software platform that allows for the rapid generation and comparison of multiple geospatial visualizations would facilitate this effort. Such a system could also automatically identify regions of potential interest by constructing spatial contour polygons around areas with high densities of a particular keyword. The system then could inform the researcher whether a specific place lies within one or more high-density topic regions, or identify the areas in which two or more such topic regions intersect.

At present, our automated semantic analysis of the text corpus employs a keyword-based "bag of words" model, supplemented by limited, manually constructed topic indices. The use of more sophisticated techniques for automated text analysis and characterization would improve our ability to identify significant semantic phenomena in the texts, which could then be explored via the geographic visualization and search techniques described above. Such computational analysis techniques include topic modeling using Latent Dirichlet Allocation, which would provide a more sophisticated set of aggregated keywords for spatial visualization and queries. Stories also could be aggregated into groups via multimodal network clustering techniques. Additionally, named-entity recognition, sentiment analysis, and automated narrative decomposition could be used to discover roles, causal relationships, and significant event types in each story, which could then be cross-referenced with co-occurring locations and plotted as geographic heat maps.

8. Acknowledgments

Funding for this work was provided through the American Council of Learned Society's Digital Innovation Fellowship, and NSF Grant # IIS-0970179e "Network Pattern Recognition for the Humanities" (Lewis Lancaster, PI). We would like to thank the anonymous referees for their helpful suggestions.

9. References

Abello, J., P. Broadwell, T. Tangherlini. (2012). The trouble with house elves: Computational folkloristics and classification. Communications of the ACM, *forthcoming* (accepted December 2011).

Anderson, W. (1923). Kaiser und Abt. Die Geschichte eines Schwanks. FF Communications 42. Helsinki: Suomalainen Tiedeakatemia.

Barber, C., D. Dobkin, H. Huhdanpaa. (1996). The Quickhull algorithm for convex hulls. In ACM Transactions on Mathematical Software 22, pp. 469-483.

Bilgic, M., L. Licamele, L. Getoor, B. Schneiderman. (2006). D-Dupe: An interactive tool for entity resolution in social networks. In Proceedings of IEEE Symposium on Visual Analytics Science and Technology 2006 (VAST '06).

Bruun, (1920). Danmark: Land og Folk, volume 3. Copenhagen: Gyldendal.

Dundes, A. (1964). The morphology of North American Indian folktales. FF Communications 195. Helsinki: Suomalainen Tiedeakatemia.

Eskeröd, A. (1947). Årets äring. Etnologiska studier i skördens och julens tro och sed. Nordiska Museets Handlingar 26. Lund: Håkan Ohlsson

Feilberg, H. (1977). Bidrag til en ordbog over jyske almuesmål. 4 volumes. Copenhagen: Dansk Historisk Håndbogsforlag.

Jongejan, B. and D. Haltrup. (2010). The CST Lemmatiser. Center for Sprogteknologi, University of Copenhagen.

Kristensen, E.T. (1891-1894). Gamle folks fortællinger om det jyske almueliv, som det er blevet ført i mands minde, samt enkelte oplysende sidestykker fra øerne. 6 volumes. Kolding: Sjodt & Weiss.

Kristensen, E.T. (1892-1901). Danske sagn, som de har lydt i folkemunde, udelukkende efter utrykte kilder. 7 volumes. Aarhus: Aarhus Folkeblads Bogtrykkeri.

Kristensen, E.T. (1900-1902). Gamle folks fortællinger om det jyske almueliv, som det er blevet ført i mands minde, samt enkelte oplysende sidestykker fra øerne. Tillægsbind. 6 volumes. Århus: Forfatterens forlag.

Kristensen, E.T. (1928-1939). Danske sagn, som de har lydt i folkemunde, udelukkende efter utrykte kilder saml. og tildels optegnede af Evald Tang Kristensen. Ny Række. 7 volumes. Copenhagen: Woels Forlag.

Krohn, K. (1926). Die folkloristische Arbeitsmethode begründet von Julius Krohn und weitergeführt von Nordischen Forschern. Instituttet for sammenlignende kulturforskning. Ser. B 5. Oslo: H. Aschehoug.

Lynch, K. (1960). The Image of the City. Cambridge: The MIT Press.

Mendoza Smith, R., A. Kuznetsova, M. Smith, P. Sugihara. (2011). Geo-semantic data mining and knowledge discovery for a large collection of historical narratives. Final Report: UCLA Institute for Pure and

Applied Mathematics Research in Industrial Projects for Students.
Mierswa, I, M. Wurst, R. Klinkenberg, M. Scholz, T. Euler. (2006). YALE: Rapid prototyping for complex data mining tasks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), ACM Press, 2006.
Moretti, F. (2000). Conjectures on world literature. In New Left Review 1, pp. 54-68.
Rooth, A.B. (1951). The Cinderella cycle. Lund: Gleerup.
Rørbye, B. (1976). Kloge folk og skidtfolk. Kvaksalveriets epoke i Danmark. Copenhagen: Politiken.
Sibson, R. (1981). A brief description of natural neighbor interpolation (Chapter 2). In V. Barnett, Interpreting Multivariate Data. Chichester: John Wiley. pp. 21–36.
Skjelborg, Å. (1967). Topografisk ordning af Danmarks egne med alfabetisk nøgle. Copenhagen: Dansk folkemindesamling.
Sydow, C.W.v. (1943). Finsk metod och modern sagoforskning. In Rig, pp. 1-23.
Tangherlini, T. (2000). How do you know she's a witch?: Witches, cunning folk, and competition in Denmark. In Western Folklore 59, pp. 279-303.
Tangherlini, T. (2010). Legendary performances: Folklore, repertoire and mapping. In Ethnographia Europaea 40, pp. 103-115.
Thompson, S. (1955-58). Motif-index of folk-literature; a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books, and local legends. Bloomington: Indiana University Press.
Venvrux and T. Meder. (2004). Authenticity as an analytic concept in folklore: A case of collecting folktales in Friesland. In Ethnofoor 17, pp. 199-214.

A Hybrid Model and Memory Based Story Classifier

Betul Ceran, Ravi Karad, Steven Corman, Hasan Davulcu

Arizona State University
P.O. Box 87-8809
Tempe AZ 85287-8809
betul@asu.edu, rkarad@asu.edu, steve.corman@asu.edu, hdavulcu@asu.edu

Abstract

A story is defined as "an actor(s) taking action(s) that culminates in a resolution(s)." In this paper, we investigate the utility of standard keyword based features, statistical features based on shallow-parsing (such as density of POS tags and named entities), and a new set of semantic features to develop a story classifier. This classifier is trained to identify a paragraph as a "story," if the paragraph contains mostly story(ies). Training data is a collection of expert-coded story and non-story paragraphs from RSS feeds from a list of extremist web sites. Our proposed semantic features are based on suitable aggregation and generalization of <Subject, Verb, Object> triplets that can be extracted using a parser. Experimental results show that a model of statistical features alongside memory-based semantic linguistic features achieves the best accuracy with a Support Vector Machine (SVM) based classifier.

Keywords: Story, Non-Story, Classification, Feature Extraction

1. Introduction

In this paper, we utilize a corpus of $16,930$ paragraphs where $3,301$ paragraphs coded as stories, and $13,629$ paragraphs coded as non-stories by domain experts to develop a story classifier. Training data is a collection of Islamist extremist texts, speeches, video transcripts, forum posts, etc., collected in open source. A story is defined as "a sequence of events leading to a resolution or projected resolution." We investigate the utility of standard keyword based features, statistical features that can be extracted using shallow-parsing (such as density of POS tags and density of named entities), and a new set of semantic features in development of a story classifier. Our study is motivated by the observation (Halverson and Corman, 2011) that interrelated stories that work together as a system are fundamental building blocks of (meta-) narrative analysis.

Computational models of stories have been studied for many different purposes. R.E. Hoffman et al. (2011) models stories using an artificial neural network. After the learning stage, they compare the story-recall performance of the neural network model with that of schizophrenic patients as well as normal controls. The most common form of classification applied on to the domain of stories tackles the problem of mapping a set of stories to predefined categories. One of the popular applications is the classification of news stories to their topics (Masand et al., 1992; Billsus and Pazzani, 1999).

Gordon investigated a similar problem to detect stories in conversational speech (Gordon and Ganesan, 2005) and weblogs(Gordon and Swanson, 2009). They use a confidence-weighted linear classifier with a variety of lexical features, and obtained the best performance with unigrams with precision = 66%, recall = 48%, F-score = 0.55. They applied this trained classifier (with 5002 blogs) to classify weblog posts in the ICWSM 2009 Spinn3r Dataset.

In this paper, we focus on discriminating between stories, and non-stories. The main contribution of this paper is the introduction of a new set of features based on linguistic subject, verb, object categories that we named as *triplet based verb features* which are motivated by the definition of "story" as "actors taking actions that culminate in resolutions.". Our proposed semantic features are based on suitable aggregation and generalization of <Subject, Verb, Object> triplets that can be extracted using a shallow-parser. Experimental results show that the combination of POS features, with semantic triplet-based features achieves highest accuracy with a Support Vector Machine (SVM) based classifier. We obtain precision of 0.706, recall of 0.559 and and F-measure of 0.634 which shows a 12% boost in precision and 5% boost in recall, an overall 10% boost in F-measure due to the utility of triplet based features.

2. System Architecture

2.1. Data Collection

Our corpus is comprised of $16,930$ paragraphs from extremist texts collected in open source. Stories were drawn from a database of Islamist extremist texts. Texts were selected by subject matter experts who consulted open source materials, including opensource.gov, private collection/dissemination groups, and known Islamist extremist web sites and forums. The texts come from groups including al-Qaeda, its affiliates, and groups known to sympathize with its cause. The subject matter experts selected texts which they believe contained or were likely to contain stories, defined as a sequence of related events, leading to a resolution or projected resolution.

Extremists texts are rarely, if ever, composed of 100% stories, and indeed the purpose of this project is to enable the detection of portions of the texts that are stories. Accordingly, we developed a coding system consisting of eight mutually-exclusive and exhaustive categories story exposition, imperative, question, supplication, verse, annotation, and other along with definitions and examples on which coders could be trained. After training coders achieved reliability of Cohens Kappa = .824 (average across eleven randomly sampled texts). Once reliability of the coders and process was established, single coders coded the remainder of the texts, with spot-check double coding to ensure reliability was maintained.

Figure 1: System Architecture

2.2. Human Annotation: Story vs. Non-Story Coding

This study only considers story vs. non-story codes. Our rationale for building a classifier using this data for training is as follows. A story typically contains several components. First, there must be an actor or actors. This can include politicians, mujahidin, and everyday people, etc. Second, the actors must be performing actions. This can include fighting, preparing for a battle, talking to others, etc. Third, the actor's actions must result in a resolution. Resolutions can include a new state of affairs, a new equilibrium created, a previous equilibrium restored, victory, etc. Stories are differentiated from non-stories as following: Because they describe actions, stories will have a lower proportion of stative verbs than non-stories. Stories will include more named entities, especially person names, than non-stories. Stories will use more personal pronouns than non-stories. Stories may include more past tense verbs ($i.e.$, X resulted in Y, X succeeded in doing Y, etc.) than non-stories. Stories may repeat similar nouns. For example, "mujahedeen" may be mentioned in the beginning of the story and then again at the end of the story. Paragraphs with stories in them have different sentence lengths than paragraphs without stories in them.

2.3. Feature Extraction

In this paper we investigate the utility of standard keyword based features, statistical features based on shallow-parsing, and a new set of semantic features to develop a story classifier.

- **Keywords:** TF/IDF measure (Robertson, 2004) is calculated for each word contained in the whole paragraph set. Then a certain number of terms, in our case $20,000$, with the top TF/IDF values are selected as features. Then term-document frequency matrix is created out these keyword features.

- **Density of POS Tags:** Part of Speech (POS) Tag Ratios (Brill, 1992) for each document is calculated with respect to numbers of tokens.

- **Density of Stative Verbs:** Some other statistical features are also included in all experiments, such as the number of valid tokens and the ratio between observed stative verbs and total number of verbs in a paragraph.

- **Semantic Triplets Extraction:** We present our semantic triplet extraction methods in Section 3. We also discuss how triplets from stories and non-stories are aggregated and generalized to form memory-based features for verbs.

2.4. Support Vector Machine (SVM) Classifier

SVM (Joachims, 2001) is a supervised learning technique which makes use of a hyperplane to separate the data into two categories. SVM is originally proposed as a linear classifier (Boser et al., 1992) but later improved by the use of kernel functions to detect nonlinear patterns underlying the data (Cortes and Vapnik, 1995).There are various types of kernel functions available (Chang and Lin, 2011). In this study, we use RBF kernel defined as $K(x_i, x_j) = e^{\|x_i - x_j\|}$, where $x_{i,j}$ are data points (Keerthi and Lin, 2003).

2.4.1. Training and Testing

The corpus contains 1,256 documents containing both story and non-story paragraphs. There are a total of 16,930 paragraphs, where 13,629 paragraphs classified reliably as non-stories, and 3,301 paragraphs classified as stories by domain experts. In our evaluations, we performed 10 fold cross validation with the document files as follows: we break documents into 10 sets of size n/10, where n is total number of documents (1,256). During the training phase, both story and non-story paragraphs from 9/10 documents are used as the training set, their features are extracted, and a classifier is trained. During the testing phase, the remaining 1/10th of the documents are used; the features for both stories and non-stories are extracted, and matched to the features extracted during the training phase. Doing this evaluation, we are ensuring that training and test data features are in fact coming from different documents. We calculate precision, recall for each iteration of the 10 fold cross validation and we report mean precision, recall for both both stories and non-stories.

3. Semantic Triplet Extraction

We follow a standard verb-based approach to extract the simple clauses within a sentence. A sentence is identified to be complex if it contains more than one verb. A simple sentence is identified to be one with a subject, a verb, with objects and their modifying phrases. A complex sentence involves many verbs. We define a triplet in a sentence as a relationship between a verb, its subject and object(s). Extraction of triplets (Rusu et al., 2007; Jonnalagadda et al., 2009; Hooge Jr, 2007) is the process of finding who (subject), is doing what (verb) with/to whom (direct objects), when and where (indirect objects/and prepositions). Our triplet extraction utilizes the information extraction pipeline shown in Figure (2).

3.1. Pronoun Resolution

Interactions are often specified through pronominal references to entities in the discourse, or through co references where, a number of phrases are used to refer to the same entity. Hence, a complete approach to extracting information from text should also take into account the resolution of these references. Our pronoun resolution module (Lee et al., 2011; Raghunathan et al., 2010) uses a heuristic approach to identify the noun phrases referred by the pronouns in a sentence. The heuristic is based on the num-

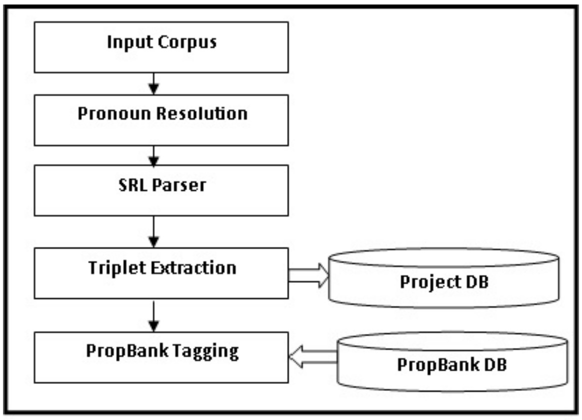


Figure 2: Triplet Extraction Pipeline

ber of the pronoun (singular or plural) and the proximity of the noun phrase. The closest earlier mentioned noun phrase that matches the number of the pronoun is considered as the referred phrase.

3.2. Semantic Role Labeler (SRL) Parser

SRL parser (Punyakanok et al., 2008) is the key component of our triplet extractor. To extract the subject-predicate-object from an input sentence, important step is identifying these elements in a sentence and parse it. SRL parser does exactly the same. SRL is propriety software developed by Illinois research group and its shallow semantic parser. The goal of the semantic role labeling task is to discover the predicate-argument structure of each predicate that fill a semantic role and to determine their role (Agent, Patient, Instrument etc). As shown in the following example, SRL is robust in identifying verbs, and their arguments and argument types accurately in the presence of syntactic variations.

Numbered arguments (A0-A5, AA): Arguments define verb-specific roles. They depend on the verb in a sentence. The most frequent roles are A0 and A1 and, commonly, A0 stands for the agent and A1 corresponds to the patient or theme of the proposition.

Adjuncts (AM-): General arguments that any verb may take optionally. There are 13 types of adjuncts: AM-ADV - general-purpose, AM-MOD - modal verb, AM-CAU - cause, AM-NEG - negation marker, AM-DIR - direction, AM-PNC - purpose, AM-DIS - discourse marker, AM-PRD - predication, AM-EXT - extent, AM-REC - reciprocal, AM-LOC - location, AM-TMP - temporal, AM-MNR - manner.

References (R-): Arguments representing arguments realized in other parts of the sentence. The label is an R- tag prefixed to the label of the referent, e.g. [A1 The pearls] [R-A1 which] [A0 I] [V left] [A2 to my daughter-in-law] are fake.

3.2.1. SRL System Architecture

SRL works in four-stages, starting with pruning of irrelevant arguments, identifying relevant arguments, classifying arguments and inference of global meaning.

Pruning - Used to filter out simple constituents that are very unlikely to be arguments.

Argument Identification - Utilizes binary classification to identify whether a candidate is an argument or not. The classifiers are applied on the output from the pruning stage. A simple heuristic is employed to filter out some candidates that are obviously not arguments.

Argument Classification - This stage assigns labels to the argument candidates identified in the previous stage.

Inference - In the previous stages, decisions were always made for each argument independently, ignoring the global information across arguments in the final output. The purpose of the inference stage is to incorporate such information, including both linguistic and structural knowledge. This knowledge is useful to resolve any inconsistencies of argument classification in order to generate final legitimate predictions.

3.3. Triplet Extraction

Our triplet extraction algorithm processes SRL output. The SRL output has a specific multi-column format. Each column represents one verb (predicate) and its arguments (A0, A1, R-A1, A2, etc) potentially forming many triplets. For a simple sentence, we can read one column and extract a triplet. For complex sentences with many verbs, we developed a bottom-up extraction algorithm for detecting and tagging nested events. We will illustrate our approach using the following example.

Example Paragraph: *"America commissioned Musharraf with the task of taking revenge on the border tribes, especially the valiant and lofty Pashtun tribes, in order to contain this popular support for jihad against its crusader campaign. So he began demolishing homes, making arrests, and killing innocent people. Musharraf, however, pretends to forget that these tribes, which have defended Islam throughout its history, will not bow to US"*

Our algorithm produces the following triplets for the example paragraph above:

Event	Subject	Verb	Object
	America	commission	Musharraf
	America	take	revenge
	Musharraf	demolish	homes
	Musharraf	make	arrests
	Musharraf	kill	innocent people
	Musharraf	pretend	$E1$
$E1$	Musharraf	forget	$E2$
$E2$	tribes	defend	Islam
$E2$	tribes	not bow	to US

Table 1: Extracted Triplets

3.3.1. Bottom-Up Event Tagging Approach

In the example above, consider the triplet <Musharraf, pretend, $E1$>. Here the object column of the verb $pretend$ has an $A1$ argument including three other verbs (forget, defend and bow). That is, argument $A1$ is itself complex, comprising other triplets. So we tag argument $A1$ with a nested event ($E1$), and recursively process $A1$ with our triplet extraction rules. We achieve this nested processing through a bottom-up algorithm that (i) detects simple verb occurrences ($i.e.$ verbs with non-verb arguments) in the SRL parse tree, (ii) extracts triplets for those simple verb occurrences using the following **Triplet Matching Rules**, (iii) replaces simple verb clauses with an event identifier, thus turning all complex verb occurrences into simple verb occurrences with either non-verb or event arguments, and applies the following **Triplet Matching Rules**.

3.3.2. Triplet Matching Rules

We list four matching rules below to turn simple SRL columns into triplets:

1. A0, V, A1: <SUBJECT, VERB, DIRECT OBJECT>
2. A0, V, A2: <SUBJECT, VERB, PREPOSITION>, if direct object A1 not present in column.
3. A0, V, A1, A2-AM-LOC: <SUBJECT, VERB, DIRECT OBJECT, location (PREPOSITION)>
4. A1, V, A2: <DIRECT OBJ, VERB, PREPOSITION>

3.3.3. Triplet Extraction Accuracy

The triplet extraction accuracy is based on SRL accuracy. SRL has precision of 82.28%, recall of 76.78% and f-measure 79.44% (Punyakanok et al., 2008).

3.3.4. Triplet Based Feature Extraction

For each verb (V) mentioned in a story (S), or non-story (NS) we stemmed and aggregated its arguments corresponding to its SUBJECTs, OBJECTs and PREPOSITIONs to generate following set-valued "semantic verb features" by using the training data:

- Argument list for S.V.Subjects, S.V.Objects, S.V.Prepositions for each verb V and story S.
- Argument list for NS.V.Subjects, NS.V.Objects, NS.V.Prepositions for each verb V and Non-Story NS.

For each test paragraph P, for each verb V in P, we extract its typed argument lists P.V.Subjects, P.V.Objects and P.V.Prepositions. Then, we match them to the argument lists of the same verb V. A match succeeds if the overlap between a feature's argument list (e.g. S.V.Subjects, or NS.V.Subjects) covers the majority of the test paragraph's corresponding verb argument list (e.g. P.V.Subjects).

4. Generalized Verb Features

4.1. VerbNet(VN) Main Classes:

Generalization and reduction of features is an important step in classification process. Reduced feature representations not only reduce computing time but they may also yield to better discriminatory behavior. Owing to the generic nature of the curse of dimensionality it has to be assumed that feature reduction techniques are likely to improve classification algorithm.

Our training data had 750 and $1,754$ distinct verbs in stories and non-stories, yielding $750*3 = 2,250$ and, $1,754*3 = 5,262$ verb features for stories and non-stories respectively, and total of $7,512$ features. VerbNet (VN) (Kipper et al., 2008) is the largest on-line verb lexicon currently available for English. It is a hierarchical domain-independent, broad-coverage verb lexicon. VerbNet index has $5,879$ total verbs represented, and these verbs are mapped into 270 total VerbNet main classes. For example, the verbs mingle, meld, blend, combine, decoct, add, connect all share the same meaning (i.e. to bring together or combine), and hence they map to verb class "mix" numbered 22.1. With the help of VerbNet and SRL argument types of the verbs, we mapped all occurrences of our verbs in stories and non-stories to one of these 270 VerbNet main classes. This mapping enabled us to reduce our verb features to $268*6 = 1,608$ verb features. The number 6 is used in the previous equation since each verb class can lead to at most 6 features as V.Subject, V.object and V.preposition for its story and non-story occurrences. We started with $7,512$ verb features, and after mapping these verb features to their verb category features we ended up with $1,608$ features only. In the generalization process, we faced a problem of verb sense disambiguation. There are some verbs which can be mapped to different senses, and each sense belongs to a different verb class. For example, the verb "add" can be used with the sense mix (22.1) or categorize (29.2) or say (25.3). To solve this problem, we used argument types extracted using SRL for the ambiguous verbs. Then, we performed a look-up for each verb in the PropBank database to identify the matching verb sense with same type of arguments, and its verb class. PropBank (Palmer et al., 2005) is a corpus that is annotated with verbal propositions, and their arguments - a "proposition bank". In the look-up process, there is a chance that we may encounter more than one verb sense for the input verb matching the corresponding argument types. In this case, we picked the first matching verb sense listed in PropBank.

5. Experimental Evaluations

In this section, we evaluate the the utility of standard keyword based features, statistical features based on shallow-parsing (such as density of POS tags and named entities), and a new set of semantic features to develop a story classifier. Feature extraction and matching is implemented using JAVA and classification is performed using LIBSVM (Chang and Lin, 2011) in MATLAB.

Feature Set	Precision	Recall	F-measure
POS	0.133	0.066	0.088
POS + Keywords	0.632	0.535	0.579
Triplets	0.548	0.321	0.405
POS + Triplets	**0.706**	**0.559**	**0.634**

Table 2: Classifier Performance for Stories

Feature Set	Precision	Recall	F-measure
POS	0.887	0.944	0.914
POS + Keywords	0.774	0.836	0.804
Triplets	**0.850**	0.936	**0.891**
POS + Triplets	0.805	**0.996**	0.891

Table 3: Classifier Performance for Non-Stories

5.1. Effectiveness of Semantic Features

The baseline performance for a dummy classifier which would assign all instances to the majority class (non-story) would achieve 80.5% precision and 100% recall for the non-story category however, its precision and recall would be null for the stories. Hence, not useful at all for detecting stories.

Our proposed model makes use of triplets to incorporate both semantic and structural information available in stories and non-stories. In Table (2), we report the performance of SVM classification with various feature sets. SVM with POS and generalized triplet based features **outperforms** other combinations of standard categories of features in terms of precision and recall. If we compare the performance of POS features alongside keyword-based (second row) vs. triplet-based (fourth row) features, Table (2) shows 12% boost in precision and 5% boost in recall, resulting in 10% boost in F-measure for the story detection due to the utility of triplet based features.

6. Conclusion

This paper proposes a hybrid model with triplet based features for story classification. The effectiveness of the model is demonstrated against other traditional features used in the literature for text classification tasks. Future work includes more detailed evaluations, and also experiments with appropriate generalizations of nouns, adjectives and other types of keywords found in verb arguments.

7. Acknowledgements

This research was supported by an Office of Naval Research grant (N00014-09-1-0872) to the Center for Strategic Communication at Arizona State University.

8. References

D. Billsus and M.J. Pazzani. 1999. A hybrid user model for news story classification. *Lectures-International Centre for Mechanical Sciences*, pages 99–108.

B.E. Boser, I.M. Guyon, and V.N. Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.

E. Brill. 1992. A simple rule-based part of speech tagger. In *Proceedings of the workshop on Speech and Natural Language*, pages 112–116. Association for Computational Linguistics.

C.C. Chang and C.J. Lin. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

C. Cortes and V. Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Andrew S. Gordon and Kavita Ganesan. 2005. Automated story capture from conversational speech. In *K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture*, page 145–152, Banff, Canada. ACM, ACM.

A. Gordon and R. Swanson. 2009. Identifying personal stories in millions of weblog entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA*.

H. L. Halverson, J. R. Goodall and S. R. Corman. 2011. *Master Narratives of Islamist Extremism.* New York: Palgrave Macmillan.

R.E. Hoffman, U. Grasemann, R. Gueorguieva, D. Quinlan, D. Lane, and R. Miikkulainen. 2011. Using computational patients to evaluate illness mechanisms in schizophrenia. *Biological psychiatry*.

D.C. Hooge Jr. 2007. *Extraction and indexing of triplet-based knowledge using natural language processing*. Ph.D. thesis, University of Georgia.

T. Joachims. 2001. A statistical learning learning model of text classification for support vector machines. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 128–136. ACM.

S. Jonnalagadda, L. Tari, J. Hakenberg, C. Baral, and G. Gonzalez. 2009. Towards effective sentence simplification for automatic processing of biomedical text. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 177–180. Association for Computational Linguistics.

S.S. Keerthi and C.J. Lin. 2003. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural computation*, 15(7):1667–1689.

K. Kipper, A. Korhonen, N. Ryant, and M. Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40.

H. Lee, Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky. 2011. Stanfords multi-pass sieve coreference resolution system at the conll-2011 shared task. *CoNLL 2011*, page 28.

B. Masand, G. Linoff, and D. Waltz. 1992. Classifying news stories using memory based reasoning. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 59–65. ACM.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

V. Punyakanok, D. Roth, and W. Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287.

K. Raghunathan, H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. Manning. 2010. A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 492–501. Association for Computational Linguistics.

S. Robertson. 2004. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of Documentation*, 60(5):503–520.

D. Rusu, L. Dali, B. Fortuna, M. Grobelnik, and D. Mladenić. 2007. Triplet extraction from sentences. *Proceedings of the 10th International Multiconference Information Society-IS*, pages 8–12.

A Crowd-Sourced Collection of Narratives for Studying Conflict

Reid Swanson, Arnav Jhala

UC Santa Cruz
1156 High St, Mail Stop SOE3
Santa Cruz, CA - 95064
{reid,jhala}@soe.ucsc.edu

Abstract

In this paper we have described a corpus that provides many of the details required to understand the context and dynamics that factor into a person's emotional state, the perception of consequences and their likely selection of a conflict resolution strategy. We believe this information is useful for meta-reasoning about narrative through a deeper knowledge about people's thought process. It also provides enough detail for attempting a data driven approach for modeling the mental process of conflict resolution in computational agents that respond in a way that people find believable. Finally, we have provided a methodology to extend the corpus, so that over time we may cover a broader spectrum of conflicts and target specific domains when they are needed for applications.

1. Introduction

A combination of our biological traits and environmental experiences uniquely shape our beliefs, values and goals, which make conflict an unavoidable part of human experience that we must constantly manage. It is so pervasive in our daily lives that conflict is one of the most studied phenomena in behavioral psychology (Wall Jr. and Callister, 1995). It is also the fundamental mechanism that transforms temporally and causally related events from being mere chronologies into rich narrative discourse that forms deep connections to the human psyche (Schank and Abelson, 1995). Narrative and conflict are so intimately linked together that the study of one should lead to significant insights and knowledge about the other.

Understanding and modeling the way people choose their resolution strategies in relation to the cause, context and dynamics of an interaction is important for virtually any computational system that interacts with people. This is especially true for narrative understanding and generation systems where the connection to people's experiences is so tightly wound. So, in order to model conflict we need to have a broad understanding of people, such as the issues that cause the most conflicts, the traits and actions that tend to escalate or deescalate situations and discerning how short term and long term interactions affect the strategies selected by a participant.

One way of modeling such a system is to collect and analyze conflict scenarios from the real world. These questions have long been the topic of research in behavioral psychology and there are numerous studies addressing these issues, for example see the meta-analyses of Laursen et al. (Laursen et al., 2001) and Holt and DeVore (Holt and DeVore, 2005). However, the focus of this research is intended for human interpretation and not for building *computational* models. For example, many studies look at correlation and size effects between various factors, such as personality and demographics. These numbers give us insight into the important factors related to conflict, however it is challenging to incorporate them into a computational model because we require explicit details for all the parameters aside from the small set of controlled features.

There are a number of studies where the details of the conflict scenarios are available (Hall, 1973; Sternberg and Soriano, 1984; Wheeler and Ladd, 1982; Thomas, 1974; Leifer and Roberts, 1972; Ayas et al., 2010; Steele, 2009; Goldsworthy et al., 2007). However, even in these cases there is not sufficient information to enable the type of automated reasoning we would like to accomplish. For example, Wheeler and Ladd (Wheeler and Ladd, 1982) is one of the larger collection of scenarios we have seen containing 22 hypothetical situations. These scenarios tend to be very short one sentence descriptions that provide very little text in terms of the relationships between the people or other relevant contextual factors that may have a significant impact on how someone might behave. In contrast Sternberg and Soriano (Sternberg and Soriano, 1984) provide more detailed descriptions, but only 9 were investigated. While each scenario encapsulates more of the intricacies necessary for a deep representation, the breadth is insufficient to model a wide range of behaviors and conflict types.

To address these challenges we present a corpus, and a methodology for extending it, that can be used for modeling and studying conflict across a variety of parameters. Currently the corpus contains 196 narrative descriptions of conflicts written by 133 unique participants, hypothetical responses to these situations from other participants and several types of annotations. We also collect and include anonymized personality profiles and demographic information for most of our participants. Of the participants we do have this information for, 37 were men and the other 57 were women.

In this paper, we present the process and initial collection of a standard corpus of conflict scenarios and responses that can be used to learn computational models of conflict resolution[1]. The corpus is intended to cover the breadth and depth required to enable reasoning about the static and dynamic contextual factors that influence the way conflicts begin, the way they escalate and de-escalate depending on the conflict resolution strategies chosen by the participants, and

[1]The corpus can be obtained at the following URL https://games.soe.ucsc.edu/project/siren

the short and long-term consequences of their outcomes.

2. Related Work

Many narratologists have argued that conflict is central to storytelling (Brooks and Warren, 1979; Ryan, 1991; Herman et al., 2005). It not only creates interesting and engaging experiences and conversations but also leads readers to actively form expectations of outcomes. Recently, efforts to build computational models of narrative have led to a detailed characterization of the parameters of conflict. Cheong et al (Cheong et al., 2011) have proposed a model of inter-personal conflict that includes several dimensions including *participants, causes, effects,* and *outcomes*. Their model is implemented in several interactive games that explore a variety of different types of conflict, such as conflicts related to *fairness*. It is modeled from conflict scenarios that are collected through interviews of children about conflict in school settings. Evaluation of the model is carried out through self-reported perception of players about the degree of fairness in resource allocation. Ware et al (Ware et al., 2011) have identified 7 dimensions of conflict and implemented them in a plan-based narrative generation algorithm. Of the 7 dimensions (participants, subjects, duration, balance, directness, intensity, and resolution), they have demonstrated through a user study that *balance, directness, intensity,* and *resolution* are recognizable properties of conflict. This work makes available a large-scale corpus of a variety of conflict situations and annotations that can be used to validate and extend existing computational models. All parameters of conflict scenarios that are measured in earlier work are available in the current dataset.

Conflict is also a fundamental part of our daily lives that has been extensively studied by psychologists and sociologists to understand the contextual parameters and dynamics of this type of human interaction. There are two aspects of this research that is particularly relevant to our work. First, we are interested in the contextual variables that cause, escalate, de-escalate and influence people's resolution strategies (Barki and Hartwick, 2004; Jehn et al., 2008; Holt and DeVore, 2005; Koza and Dant, 2007; Steele, 2009; Gelfand et al., 2008; Canary and Cupach, 1988; Kaushal and Kwantes, 2006; Wall Jr. and Callister, 1995; Laursen et al., 2001). These factors are important in designing templates to solicit particular types of scenarios and for characterizing the scenarios we collect. Second, we are interested in the dynamics of conflict as they play out (Koza and Dant, 2007; Rahim, 2000; Wall Jr. and Callister, 1995). Conflicts are not static incidents that happen all at once, but continuous interactions that play out through time. This aspect is often neglected but has been shown to be extremely important in determining the behavior of people (Joshi, 2008; Mannix, 2003; Andrade et al., 2008; Koza and Dant, 2007; Rahim, 2000; Wall Jr. and Callister, 1995). For example, Joshi noticed that people do not often repeat resolution strategies during a conflict, but will often repeat the strategy of their adversary, regardless of any contextual factors. This progression is important to recognize and capture to what ever extent possible.

Our work is an attempt to incorporate and expand the parameters and models of these communities to build a corpus that provides a rich set of data for learning computational models of conflict usable for deep narrative understanding and generation.

3. General Methodology

In this section we describe a four step collection and annotation process that we used to build this corpus. Although our goal is to cover as many scenarios and variables as possible, there will always be some gaps in the data. We do not consider this corpus to be complete, but a starting point at which more data can be added for specific purposes using the methodology described in this paper. The steps in this process are:

Pre-requirement Require non-identifiable demographic and personality information from users (Section 4.)

Step 1 Collect narratives of real-world conflict scenarios and several pieces of semi-structured information about the experience (Section 5.)

Step 2 Convert to hypothetical scenarios with different contextual parameters (Section 6.)

Step 3 Gather responses to the hypothetical scenarios from our participants (Section 7.)

Step 4 Annotate responses according to their resolution strategy (Section 8.)

We first require all of our participants to go through a qualification process. In this phase we gather several pieces of non-identifiable personal information about each contributor. In step 1, we begin by collecting narrative descriptions of real world conflicts as recalled by the participant. The second step is a manual process of converting these actual conflict scenarios into generic hypothetical situations. The third step uses these hypothetical situations to query other participants what they would do in a similar situation. Finally, the fourth step involves annotating the responses along 8 dimensions we have developed for categorizing conflict resolution strategies.

We would like our corpus to be as broad as possible so that it covers as many of the topics and parameters that have been identified in related work. Manual collection of personal experiences (Mangione, 1996; Isay, 2007) can be time consuming and expensive because they usually require physical resources such interview locations and the presence of an expert interviewer. It is also difficult to avoid bias, without great care and expense, in these approaches because the sample population is often drawn from a similar pool of participants.

In this work we utilize crowd-sourcing techniques in an attempt to maintain the advantages of manual collections strategies, while leveraging large Web communities to enable broad coverage that is often difficult to achieve with limited resources. Although we plan to pursue other online communities to expand our corpus, we currently use Amazon's Mechanical Turk to collect our data. Mechanical Turk consists of over 100,000 workers from 100 different countries, a majority of which are found in the United States. At least within the United States, these workers have a wide range of demographic profiles, which is fairly consistent with the population of Internet users as a whole. Although

there are a few biases, such as an over representation of younger users and women (Ipeirotis, 2010).
For each component of the corpus described in the following sections, the data is collected using a simple web-based survey from both open-ended and closed-ended response types from Mechanical Turk workers in the United States. An example of the type of data for each step that will be discussed in subsequent sections is provided in Table 1.

4. Participant Screening

In order to allow a deeper analysis of our corpus and to enable more robust computational models of conflict in the future, we also require all participants to complete two questionnaires prior to working on any of our tasks. First, we required a standard five factor personality instrument (Rammstedt and John, 2007) and second we a set of ten demographic questions described below.
The first two questions were about age and gender, which have been shown to correlate with many types of behaviors. The third question asks for the location of the participant in order to capture cultural differences between different regions of the country or between rural and urban areas. Fourth, we ask about the education level of each participant, which we believe could be a significant factor in the types of strategies a person tends to employ.
In previous research, Bryant (Bryant, 1992) found that the popularity and type of social status a person has can be a strong indicator of the type of strategy one is likely to engage. Questions 5-7 are posed to try to capture an indirect measure of the sociability of the participant through various metrics, such as the number of phone contacts, social networking friends and text messages they receive. We also asked a few other questions (8-10) that provide slightly more information about the daily life of the participant, such as the amount of physical activity they perform, the amount of television they watch and the amount of time they spend playing video games.

5. Real World Conflicts

5.1. Corpus Collection

The first step in building our corpus is to collect narrative descriptions of actual conflicts from our participants. In our survey, we specifically asked for real conflicts in which the participant was one of the party's in a dispute. To solicit these scenarios we provided a web-based form that asked the participants to describe a situation in which they were in a conflict with another person. We did not provide a precise definition, however we advised them that these could be *verbal or physical disagreements, or it could be a situation in which the actions of someone else made you uncomfortable.* As further guidelines, we also asked the participants to provide as much of the following details in their recounting:

- *What happened to initiate the conflict and what was it about?*
- *What steps did you take to try to resolve the conflict?*
- *How did the other person respond to your steps?*
- *What was the outcome?*

In an initial pilot study we learned that providing such an open-ended task is problematic for two reasons. First, it appears that workers on Mechanical Turk are averse to this type of task in general, regardless of monetary reward[2]. Second, without any guidance there is a worry the narratives will only focus on a small range of topics that are common and easy to recall.
We believe the workers are hesitant to work on these types of open-ended tasks for several reasons. First, the user only has the opportunity to perform this task a small number of times, in contrast to most tasks, which can be worked on repetitively for long periods of time. Even when the participant can obtain a much higher reward per unit of time, this may not be enough to offset the value to the user (in terms of stress, time, or other factors) of having to find another suitable task. Similarly, most tasks require very little creative thought, whereas this one engages the participant to compose original content. This necessitates much more focus and attention than the typical task and may dissuade many people from telling us their stories.
From our experience in designing other tasks, such as the one in section 7, providing the user with seed of information from which to start can greatly facilitate the effort and lead to much greater participation. In order to help prompt the user and potentially encourage a broader range of topics, we considered providing templates that explicitly asked for conflicts covering a particular set of parameters and issues that we have identified from other conflict related research. For example, asking for conflicts involving a supervisor whom you disagreed with over a policy at work. However, the number of templates is too large to enumerate and we felt would bias the corpus to heavily towards the parameters we chose to include. Instead, we left the survey completely open-ended but provided several high-level suggestions pertaining to the common causes and types of conflicts that have been identified. For example, *You caught someone cheating or lying*, *You were forced to spend time with someone you do not like*, or *Someone teased you or made an inappropriate remark about your religion, culture or ethnicity*. An example conflict scenario that was written by one of our participants is provided in step 1 of Table 1.
Following these guidelines we used a simple web-form to solicit free-text narrative descriptions of a conflict the participant had been in with another person. Once the participants were finished describing their story they were asked a series of 11 questions related to the conflict. These questions are summarized and reproduced in abbreviated form in Table 2. While the abbreviations capture the spirit of the response, we tried to provide more descriptive text in the survey to help avoid ambiguity. For example, it is made clear that we are asking for the shortest matching duration. We also try to qualify the subjective responses with personally relatable qualifications, such as *minor consequences* are things that would not significantly impact future actions. In addition to the questions above, the participants were asked to select as many issues as they thought

[2]We varied the reward from $1.00 to $0.20 and saw no noticeable difference in the rate at which people performed the task.

Step 1	**Scenario**	*My best friend wrecked my car speeding. We got into an intense argument because she did not show any remorse for what she had done. We did not speak for about 2 years because of this. As a result I missed a few days of work because I had no transportation until my rental car was ready. In the end we realized we were too good friends to let this come between us. She could have lost her life.*		
	Issues	Fairness, Resource, Theft, Asking a favor, Deception, Breaking a promise		
Step 2	**Hypothetical**	*You let a close friend borrow your car and you just found out they got into a severe accident while driving recklessly.*		
Step 3		**Participant 1**	**Participant 2**	**Participant 3**
	1st Preference	I would check on my friend to make sure they were ok.	Make them pay for it.	I'd first make sure they were OK.
	2nd Preference	I would help my friend with whatever they needed.	Never let them borrow my stuff again.	I'd be mad because it was their fault and never let them drive my car again.
	3rd Preference	I would call insurance to let them know.	Curse myself for being dumb enough to let them borrow it in the first place.	I'd call insurance to figure out what to do.
	Personality	Not Available	0.4**E**, 0.3**A**, 0.7**C**, 0.2**N**, 0.6**O**	1.0**E**, 0.6**A**, 0.7**C**, 0.7**N**, 0.6**O**
	Demographics	26-40, Male, College, AR	26-40, Male, College, CA	26-40, Female, College, VT
(Repeat) Step 2	**Shared**	*You borrowed a friend's car for the weekend and got into a wreck that totaled the car while speeding around a corner.*		
	Hypothetical 1	When you tell your friend they express concern for your safety.		
	Hypothetical 2	When they find out, they file a lawsuit against you.		
	Hypothetical 3	When they find out, they tell you they will never let you borrow anything again.		

Table 1: An example of the first three steps in developing the corpus.

	Question	Valid Responses
(1)	How long has it been since you were in the conflict?	Week, 3 Months, Year, Longer
(2)	How severe did you believe the consequences would be if you had not confronted the situation?	None, Minor, Moderate, Severe
(3)	Did you get upset or frustrated during the interaction?	Not at all, Slightly, Moderately, Extremely
(4)	Was the other person the same sex as you?	Yes, No
(5)	How would you categorize your relationship with the other party?	Stranger, Acquaintance, Friend, Close friend, Romantically interested, Romantically involved, Spouse/domestic partner
(6)	How long had you known the other party before the conflict?	Never, Week, Couple Months, Year, Longer
(7)	How long did you know the other party after the conflict?	Never, On going, Week, Couple months, Year, Longer
(8)	Was the person related to you in any of the following ways?	Sibling, Parent, Child, None
(9)	Was this person a coworker?	Supervisor, Employee, Peer, Yes: something else, No
(10)	Was this person in school with you?	Teacher, Student, Classmate, Yes: something else, No
(11)	Were you satisfied with the outcome?	Completely dissatisfied, Slightly dissatisfied, Slightly satisfied, Completely satisfied

Table 2: A summary of the questions accompanying a scenario description.

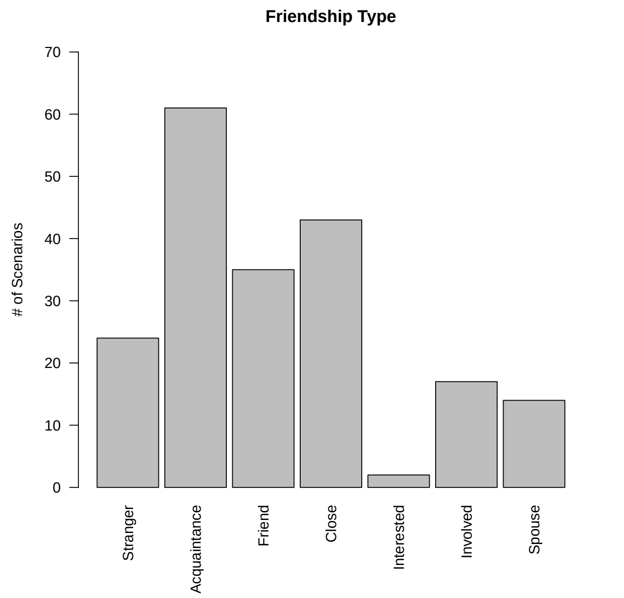
were appropriate to their conflict from a list of 21 common causes of conflict (or type in their own). For example, some of the issues were *fairness*, *social pressure*, *property disputes* or *trust*.

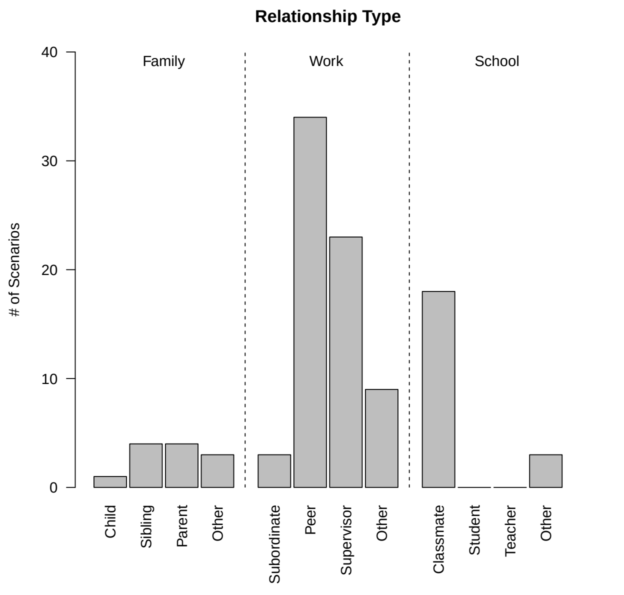
5.2. Corpus Statistics

Most conflicts are between peers at work who know each other, but do not consider themselves friends (Figure 1). Unfortunately, there are relatively few conflicts between family members and school related relationships. To extend the corpus in the future it would be advisable to use targeted surveys or interviews to fill in these domains. Most reported conflicts happened in the distant past (Figure 2). There is a fairly even distribution of satisfaction levels represented in the corpus, with the exception of very recent conflicts. Given that these conflicts are recent and memorable enough to recall, it is unsurprising that there is little ground between the extreme values.

Over all temporal durations, the conflicts primarily occurred between people who had known each other for a long time (Figure 3a). The two parties usually did not remain acquainted after the conflict despite most of the conflict involving long term acquaintances, although the cause is not clear from this analysis.

In general the level of emotion is biased toward the high end of the scale, while the perceived severity of consequences is more moderate (Figure 3b) On a 4-point scale, these values ranged from *Not at all* to *...continued feel worked up for some time after the conflict* and *No consequences what so ever* to *Severe consequences that would cause serious life altering changes* respectively. We were surprised to see that the reported level of emotion was greatest for scenarios that

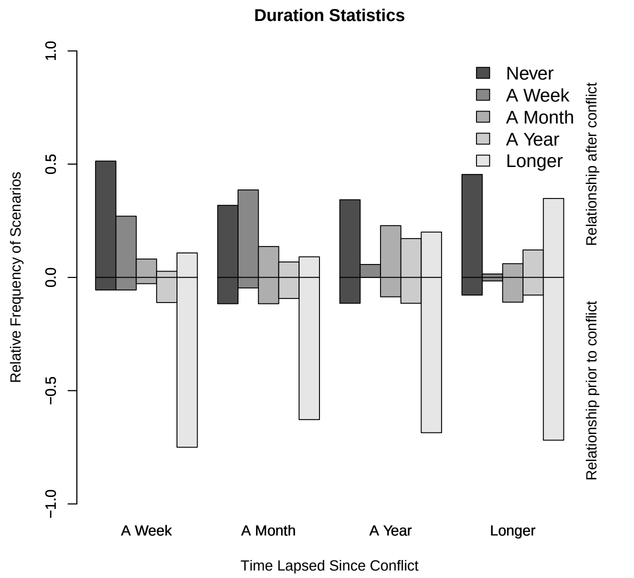


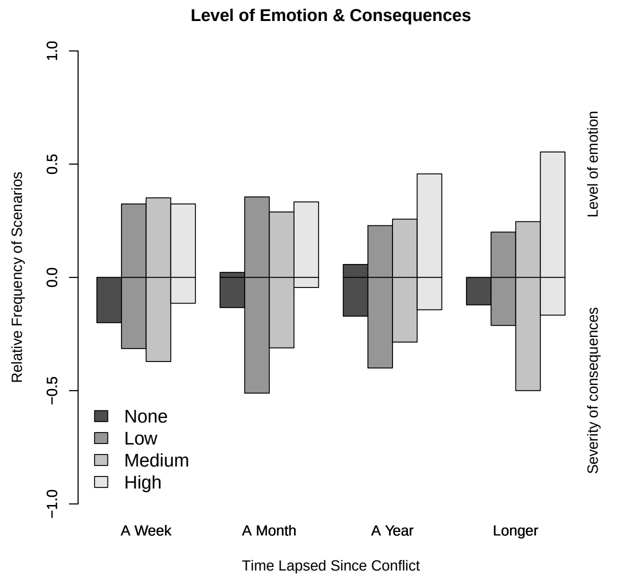


(a) The reported friendship to the other party involved. *Close* indicates a close friend, *Interested* indicates someone they are romantically interested in, but not involved and *Involved* is someone they are romantically involved with, but not married to.

(b) The reported relationship type to the other party. The left section shows the number of scenarios involving various family members, the middle is for work relationships and the right shows school relationships.

Figure 1





(a) The relative frequency of scenarios for how long the parties knew each other before (bottom) and after (top) the conflict as a function of time since the conflict.

(b) The relative frequency of scenarios for the level of emotion felt during the conflict (top) and the perceived severity of consequences (bottom) as a function time since the conflict.

Figure 3

happened in the distant past, especially compared with very recent conflicts. Similarly, participants generally perceived the consequences to be more minor for recent conflicts than for ones that happened much farther in the past. Although it is not entirely clear, we suspect two things may be going on. One, people have a hard time understanding long term consequences and may underestimate their effects. Two, looking back people see all the things that have changed in their life since that point in time. With this perspective they may attribute more weight to those things they remember or find important, whether or not those events were the actual cause.

Figure 4 shows the number of scenarios for the top 20 issues participants said their conflicts were about. The participants were free to choose multiple issues and/or write in their own. The most common thing in our corpus that

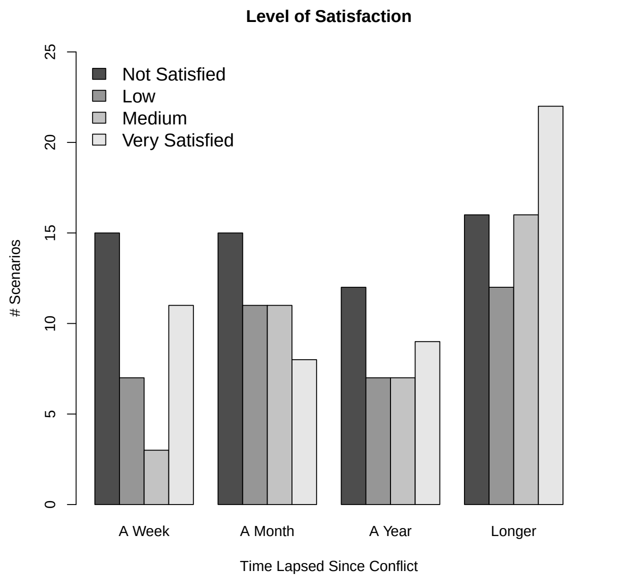


Figure 2: The reported level of satisfaction with the outcome as a function of time since the conflict occurred.

brings people into conflict are issues of fairness. It is unsurprising that *fairness* was the most common issue since this can broadly be interpreted to encompass many of the other issues.

6. Creating Hypothetical Situations

Although we believe the conflicts collected in the previous section provide a valuable resource in itself, it is still difficult to learn computational models entirely from narrative prose. This section describes the next step in our process, which will help break the problem into smaller pieces and allow us to learn more about the behavior of individual

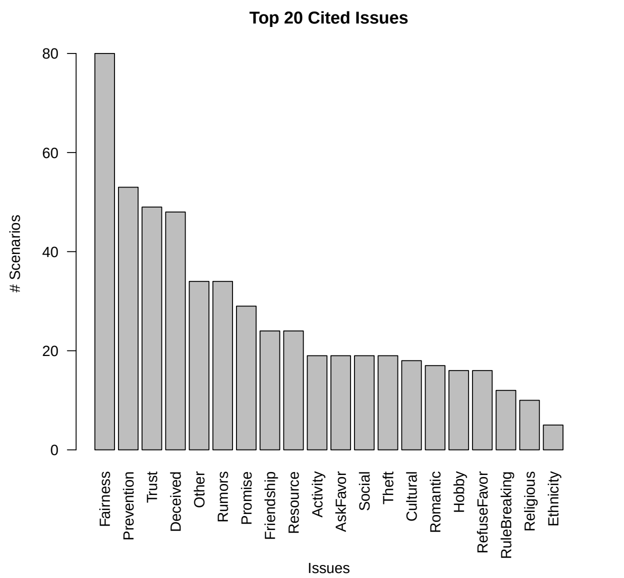


Figure 4: The 20 most common issues participants labeled their conflicts.

participants. In this stage, we manually converted the actual conflict scenario into a generic hypothetical situation, which can then be posed to independent participants to see how they would respond in that situation. This is primarily done by changing the perspective to the second person, such as, *Your best friend wrecked your car...* and focusing on the main issue involved.

At least one envisioned use of the corpus is to analyze and model how slight variations in the context of a situation can affect the perceived consequences, the emotional level involved and ultimately how a person would choose to resolve the issue. There are many factors which could change the way someone would react to the situation in Table 1. For example, changing the relationship from a close friend to an employee would probably dramatically change the dynamics of the response for many people. Similarly, providing details of the type and value of the car might also significantly effect the way many people would deal with the situation. Finally, providing some indication of the history of interaction (and prior conflicts) can also affect the way people assess and choose their strategies for resolving conflicts.

Because of these issues, we considered keeping the scenarios as close as possible to their original form, while only making small changes to the various roles and parameters. We could create a set of new scenarios from the original by altering the relationship type of the adversary, the gender or other analogical mappings, such as the type of car, that might alter perception of the situation, such as the level of consequences. For example, *Your boss wrecked your car while speeding...* or *Your friend who has been unreliable before, wrecked your car while speeding...*

We chose not to pursue this strategy at this time for several reasons, although we believe it is worth exploring in the future. The primary concern with this approach is that it causes extra difficulty in obtaining reliable data in step 3, described in section 7. We expect that participants will perform many tasks repetitively, which could lead habituation and prevent them from reading the text closely enough to make the types of nuanced distinctions we are trying to acquire.

In this initial study, we wanted to try to gather as diverse a collection of responses (described in the next section) to these scenarios as possible, since it was unclear exactly how many participants we would be able to attract. So, we took a different direction and tried to extract the essential facets of the scenario that we believed would resonate with the largest number of participants. As the number of scenarios in the corpus grows we expect to cover an adequate amount diversity simply through the number of unique stories that will enable learning robust models of conflict resolution. These generic situations will also provide a neutral starting point for creating alterations to the context in the future.

Our general methodology for transforming the actual scenarios was to try to capture only the core event by removing any references to subjective matters, such as emotional descriptions, or to the details of how the situation was resolved. When possible, we also tried to make the hypothetical situations gender neutral, so that the same scenario could be performed by an arbitrary worker, since it is difficult to control up front. Additionally, it was sometimes necessary to change a more substantial component of a story in order to appeal to a wider audience, while still maintaining the core issue involved. For example, the event which causes the conflict in Table 1 is the crashing (and wrecking) of the car by the close friend while speeding. Through a subjective analysis of this story we concluded that the author was making a qualitative judgment that the speeding was an inappropriate behavior and was at least partially to blame for the accident. However, in a generic version of the event, such as "*Your friend wrecked your car while speeding*", the word speeding might not carry the same connotation to all the readers as the original conflict scenario intended. For example, some readers might not regard speeding as an issue or make the connection that the speeding is the supposed cause of the accident. In these cases we tried to reword the scenario to maintain the original connotation with as few loaded words as possible. Step 2 in Table 1 gives an example of a complete transformation for the the scenario described in step 1.

7. Responses to Hypothetical Conflicts

The hypothetical situations created by the process in the previous section are then used to build a library of conflict resolution strategy responses from the other participants. For each scenario we ask the participant to read the description and to tell us the most likely thing he or she would say or do if they were in that situation. We also provide them with two other text boxes for the participants to type in other alternatives ranked by the likelihood they might try them. We also ask the participant to tell us (on a 4 point scale) if they have ever been in a similar situation as the scenario they have been presented with. Step 3 in Table 1 presents the responses of three participants to the hypothetical situation in step 2.

As a final step in the corpus collection phase, we manually aggregated all of the similar responses and created a new hypothetical scenario from the perspective of the responder for each unique and applicable response type. For example, the 3rd preference of Participant 1 and Participant 2 are considered the same type and so only one new hypothetical situation would be created for these two responses. (Repeat) step 2 in Table 1 presents three transformations from the original hypothetical situation and a particular response type. For most of the hypothetical situations there is a shared context and only a few sentences at the end are altered to reflect the specific details of the given response. When presented to the user, however, the shared and non-shared text is presented as a single narrative text.

Creating these new hypothetical situations introduces a new set of challenges that often requires subjective judgments to be made in terms of reframing the situation from the other party's perspective. There are an infinite number of ways that we could pose the situation from the other perspective by changing the back story (e.g., why they borrowed the car, why they were speeding, etc.), each of which has the potential to significantly change the resulting responses.

From the perspective of the original conflict scenario the assignment of blame is often very one sided, such that the other party is clearly in the wrong. If we translated the second round of hypothetical situations with this information encoded we would be severely biasing our scenarios by asking the participants to respond to situations in which they are described as being clearly at fault.

Our strategy for dealing with this problem was to try to change the context to be more neutral so that the other party (now the subject of the new hypothetical) is not necessarily completely at fault or had some reasonable explanation for their actions. For example, in the *car* example we changed the wording of *driving recklessly* back to *speeding* and added *around a corner*. This wording is important because although it suggests impropriety on the part of the driver, it allows the possibility that the accident could have been a reasonable mistake that could have happened to anyone (e.g., taking an unexpected turn faster than expected). By targeting this level of ambiguity we hope that the responses given will require each individual to fill in the missing information from their own background, personality and experiences, which is what will help us uncover the correlations and parameters needed to build computational models from our data.

After creating these new hypothetical situations we had Mechanical Turk participants provide 3 responses they might do in this new situation. At this point we now have two turns of responses that cover a wide range of contextual variables. Although we would ideally prefer a corpus of real-time conflict interactions, we believe the two turns from opposite perspectives still allows for a basic analysis of some aspects of the dynamic temporal nature of conflict that is often neglected in other research in this area (Joshi, 2008; Mannix, 2003; Andrade et al., 2008).

8. Conflict Resolution Strategies

The final step in our process is to annotate the responses obtained in the previous section by the most appropriate conflict resolution strategy the utterance suggests. Conflict resolution strategies are most often categorized using the Thomas-Kilmann Conflict Mode Instrument (TKI) (Thomas, 1974). The TKI classifies conflict resolution strategies along the two dimensions: *assertiveness* and *cooperativeness*. Depending on the *combination* of *assertiveness* and cooperativeness they propose five distinct types of conflict resolution strategies: *competitive*, *collaborative*, *compromising*, *accommodating* and *avoiding*.

However, we have found two problems with using these five categories for our corpus. First, we performed an initial inter-rater reliability study to assess the level of agreement that could be achieved. We collected 354 randomly choses responses from 12 hypothetical scenarios. On average about 9 participants provided 3 responses for each of the 12 scenarios for a total of 354 ratings that the authors independently annotated. This resulted in a relatively low unweighted κ of 0.506. Second, similar to other researchers (Sternberg and Soriano, 1984; Joshi, 2008; Rahim, 2000; Hall, 1973), we have found that these two dimensions and five categories do not cover the types of responses in our corpus.

The two dimensions in the TKI can be seen as describing the manner of action (*assertiveness*) and characterizing the the roles of the participants (*cooperativeness*). However, when looked at it in this way, the model is underspecified. There are often cases in our corpus when a response is both assertive and non-cooperative, but is highly misleading to call it *competition* as in the TKI. Similarly, this categorization cannot distinguish between a bully who is physically intimidating another person versus two parties who are competing for a scarce resource (e.g., a promotion at work). In this section we describe a new set of dimensions that offer a more fine-grained classification, which spans a wider range of possible responses. However, since each dimension is much more specific we hope to achieve higher agreement for each of these facets.

To extend the characterization from this viewpoint, we identified six core dimensions that further subcategorize the manner of action and roles of the participants. We also include two additional dimensions that are core to the categorization, but allow for a more detailed classification with little extra effort. *Active* expresses whether the individual's action is an active step or is directly acknowledging the conflict. Most responses will be active, but there are some cases, such as ignoring the other party, hanging up or walking away that might be seen as passive if they are done to avoid the confrontation altogether. *Aggressive* specifies whether the response was hostile towards the other party, for example shouting or intimidation.*Interest* describes which party is intended to benefit from the action taken. For example, making sure your friend is not hurt is done in the interest of your friend (most likely), whereas making your friend pay for the damage is probably in your self interest. *Yielder* describes who is intended to yield or give up their position. Although *interest* and *yielder* appear similar on the surface, we believe they capture an important conceptual difference that allows us to represent complex semantics in resolution strategies. For example, when you apologize it is usually done in the interest of the other party and you are willingly admitting your mistake (i.e., yielding your position). On the other hand, you could also do something in the interest of the other party that also requires them to yield their position, for example taking the car keys away from someone who has been drinking. *Solver* represents which party is responsible for resolving the situation according to the response. For example, demanding that the other party pay for the wrecked car implies the other party is responsible for a satisfactory resolution. Whereas, calling the insurance company could be seen as having a 3rd party mediate the situation. 3rd party resolution is also common in school and work environments where individuals in conflict often appeal to their teacher or boss. *Involvement* specifies the scope of involvement of different parties in the conflict. For example, calling someone a liar in a closed room only involves the conflicting parties. However, calling someone a liar in front of other people in order to lower the esteem of your adversary is involving an external 3rd party into the resolution strategy. *Economic* indicates whether the response uses some form of monetary compensation or pressure as a means to resolving the conflict. Generally, the strategies can be well defined without this extra information, for example as a type of compromise, but it has been used by other researchers (Sternberg and Soriano, 1984) and seemed reasonable for our corpus. *Connotation* does not provide any objective information to the responses, however, given enough annotated data paired with demographic information we could use this label to learn the perceptions of different actions across demographic and cultural groups. We are currently examining the viability of annotating our corpus along these dimensions using non-expert Mechanical Turk raters. In a preliminary study we have seen comparable levels of agreement across all dimensions to our expert annotations on the TKI labels. However, we leave detailed results and further analysis for future work.

9. Discussion

This corpus is important to our research in several ways. We are currently developing a social simulation designed to improve the understanding and literacy of different conflict resolution strategies in a safe environment removed from one's peers. Our corpus is an attempt to provide a central repository that will eventually cover all the significant contextual and dynamic factors surrounding interpersonal conflict. Currently we are exploring using these descriptions as inspiration for authoring scenarios relevant to different demographic groups in our simulation environment. We are also using the semi-structured annotations provided along with the scenarios, responses and conflict resolution labels as a principled method for defining and setting parameters for controlling non-player character behavior.

Our goal is to create in-game scenarios that resonate with the user and to model non-player characters in a way that consistently speak and act according to a profile that is recognizable to the player in the real world. We are not trying to teach a particular conflict resolution strategy or the "right way" to manage conflict. Instead we would like to increase the vocabulary of strategies a player is familiar with by encouraging him or her to experiment with different behaviors. In doing so, the user can can observe various outcomes in different contexts and with different non player characters in the world.

The corpus we have described provides a base set of narratives that cover a wide range of contextual factors surrounding human conflict. This corpus has been useful to us in developing the scenarios and parameters of our social simulation. This corpus is also easily extendable to support many other types research in this area. Following the steps outlined in the paper it is easy to augment the corpus with additional narratives or target specific domains. We have also contributed a number of annotations that help characterize different aspects of the stories and the people who contributed them. This corpus and initial annotations offer a standard library from which more detailed annotations can be layered on top. For example, using tools, such as the *Story Workbench* (Finlayson, 2008), which facilitate the syntactic and semantic analysis of narrative discourse. In the future we hope this corpus will be used and extended for a variety of purposes surrounding narrative and conflict analysis.

10. References

Leticia Andrade, Donde Plowman, and Dennis Duchon. 2008. Getting past conflict resolution: A complexity view of conflict. *Management Department Faculty Publications*, January.

Tuncay Ayas, Metin Deniz, Mücahit Kagan, and M. Fuat Kenç. 2010. An investigation of conflict resolution strategies of adolescents. *Procedia - Social and Behavioral Sciences*, 2(2):3545–3551.

Henri Barki and Jon Hartwick. 2004. Conceptualizing the construct of interpersonal conflict. *International Journal of Conflict Management*, 15(3):216–244, July.

Cleanth Brooks and Robert Penn Warren. 1979. *Understanding fiction*. Prentice-Hall.

Brenda K. Bryant. 1992. Conflict resolution strategies in relation to children's peer relations. *Journal of Applied Developmental Psychology*, 13(1):35–50, March.

Daniel J. Canary and William R. Cupach. 1988. Relational and episodic characteristics associated with conflict tactics. *Journal of Social and Personal Relationships*, 5(3):305 –325.

Yun-Gyung Cheong, Rilla Khaled, Corrado Grappiolo, Joana Campos, Carlos Martinho, Gordon P.D. Ingram, Ana Paiva, and Georgios Yannakakis. 2011. A computational approach towards conflict resolution for serious games. In *Foundations of Digital Games*, Bordeaux, France.

Mark Alan Finlayson. 2008. Collecting semantics in the wild: The story workbench. In *In*, page 4653. AAAI Press.

Michele J. Gelfand, Lisa M. Leslie, and Kirsten M. Keller. 2008. On the etiology of conflict cultures. *Research in Organizational Behavior*, 28:137–166.

Richard Goldsworthy, Nancy Schwartz, Sasha Barab, and Anita Landa. 2007. Evaluation of a collaborative multimedia conflict resolution curriculum. *Educational Technology Research and Development*, 55(6):597–625, December.

Jay Hall. 1973. *Conflict management survey: a survey of one's characteristic reaction to and handling of conflicts between himself and others*. Teleometrics Int'l.

David Herman, Manfred Jahn, and Marie-Laure Ryan. 2005. *Routledge Encyclopedia Of Narrative Theory*. Routledge, February.

Jennifer L. Holt and Cynthia James DeVore. 2005. Culture, gender, organizational role, and styles of conflict resolution: A meta-analysis. *International Journal of Intercultural Relations*, 29(2):165–196, March.

Panagiotis G. Ipeirotis. 2010. Demographics of mechanical turk. Technical report, NYU, New York, March.

Dave Isay. 2007. *Listening Is an Act of Love: A Celebration of American Life from the StoryCorps Project*. Penguin Press HC, The, 1ST edition, November.

Karen A. Jehn, Lindred Greer, Sheen Levine, and Gabriel Szulanski. 2008. The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Decision and Negotiation*, 17:465–495, April.

Anupama Joshi. 2008. Conflict resolution between friends during middle childhood. *The Journal of Genetic Psychology*, 169(2):133–148, June. PMID: 18578297.

Ritu Kaushal and Catherine T. Kwantes. 2006. The role of culture and personality in choice of conflict management strategy. *International Journal of Intercultural Relations*, 30(5):579–603, September.

Karen L. Koza and Rajiv P. Dant. 2007. Effects of relationship climate, control mechanism, and communications on conflict resolution behavior and performance outcomes. *Journal of Retailing*, 83(3):279–296, August.

Brett Laursen, Benjamin D. Finkelstein, and Noel Townsend Betts. 2001. A developmental meta-analysis of peer conflict resolution. *Developmental Review*, 21(4):423–449, December.

Aimee Dorr Leifer and Donald F. Roberts. 1972. Children's responses to television violence. *Television and Social Behavior*, Vol. 2:43–180.

Jerre Mangione. 1996. *The Dream and the Deal: The Federal Writers' Project, 1935-1943*. Syracuse University Press, June.

Elizabeth Mannix. 2003. Editor's comments: Conflict and conflict Resolution-A return to theorizing. *The Academy of Management Review*, 28(4):543–546, October. ArticleType: misc / Full publication date: Oct., 2003 / Copyright 2003 Academy of Management.

M. Afzalur Rahim. 2000. *Managing Conflict in Organizations: Third Edition*. Quorum Books, 3 edition, December.

Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1):203–212, February.

Marie-Laure Ryan. 1991. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press, Bloomington, IN, USA.

Roger C Schank and Robert P Abelson. 1995. Knowledge and memory: The real story. *Knowledge and Memory: The Real Story*, pages 1–85.

Godfrey Steele. 2009. Emotion and conflict management in elementary schools. *SSRN eLibrary*, June.

Robert J Sternberg and Lawrence J Soriano. 1984. Styles of conflict resolution. *Journal of Personality and Social Psychology. Vol 47(1)*, 47(1):115–126, July.

Kenneth W Thomas. 1974. *Thomas-Kilmann conflict mode instrument*. XICOM.

James A. Wall Jr. and Ronda Roberts Callister. 1995. Conflict and its management. *Journal of Management*, 21(3):515–558.

Stephen G. Ware, Brent Harrison, R. Michael Young, and David L. Roberts. 2011. Initial results for measuring four dimensions of narrative conflict. In *The Fourth Workshop on Intelligent Narrative Technologies at the 2011 AI and Interactive Digital Entertainment Conference (AIIDE)*, Stanford, California.

Valerie A. Wheeler and Gary W. Ladd. 1982. Assessment of children's self-efficacy for social interactions with peers. *Developmental Psychology*, 18(6):795–805.

Towards a Culturally-Rich Shared Narrative Corpus: Suggestions for the Inclusion of Culturally Diverse Narrative Genres

Victoria Romero, James Niehaus, Peter Weyhrauch, Jonathan Pfautz, Scott Neal Reilly
Charles River Analytics
625 Mt. Auburn St.
Cambridge, MA 10238 USA
E-mail: vromero@cra.com, jniehaus@cra.com, pweyhrauch@cra.com, jpfautz@cra.com, snreilly@cra.com

Abstract

This paper proposes that the inclusion of culturally diverse narrative genres should be an explicit goal when developing a shared narrative corpus. We argue that narrative genres from under-represented cultures and from cultures relevant to specific applications of computational narrative research should be prioritized. We offer the example of Mexican narcocorridos as a narrative genre that satisfies both these criteria.

Keywords: culture, shared resources, story bank

1. Introduction

At the inaugural meeting of the Computational Models of Narrative Workshop in 2009, one participant noted, "as far I know, every society in the world has stories" (Finlayson, Richards & Winston, 2010, p 97). While it is obvious that narratives are universal, it is difficult to specify exactly what makes them universal. At the workshop, the decision was made to establish a shared resource for narrative researchers. This paper suggests that as narratives are selected for inclusion in shared resources, special attention should be paid to including narratives from a variety of cultures. This will facilitate empirical research on the universality of narrative and enhance our understanding of how narrative operates within broader cultural and socio-political contexts. This paper offers criteria for developing a culturally diverse story bank and suggests Mexican *narcocorridos* (drug ballads) as a narrative genre that may satisfy these criteria and make a valuable contribution to a shared resource.

2. Building a Culturally Rich Story Bank

We suggest that a careful consideration of narrative samples from non-Western cultures is required to create a culturally diverse story bank. For an example of how *not* to approach cross-cultural research, review the field of cognitive psychology. Before the 1970s, virtually all cognitive research was conducted on samples in the US and Europe. Then, in the 1970s, Rosch and others (e.g. Rosch, 1975) found that fundamental psychological phenomena which were assumed to be universal, were actually strongly influenced by culture. Soon, there was broad acknowledgement that cross-cultural work was critical to understanding psychological processes. The field of cross-cultural cognitive psychology grew quickly, but focused heavily on comparisons of Western (US and Western European) samples with East Asian samples, generally drawn from Japan, Taiwan, and South Korea. As a result, we have a wealth of knowledge on the cognitive differences between East Asian and Western populations, but after 30 years of "cross-cultural" research, we still lack a well-rounded understanding of the interaction between culture and cognition. Latin American, Middle Eastern, South Asian, and African cultures are dramatically underrepresented in the current psychological literature, causing significant problems when we try to understand what motivates young men in Sub-Saharan Africa to join bands of pirates or young men in Mexico to risk their lives and the lives of their families to join drug cartels. We are forced to extrapolate psychological theories to these populations with no empirical evidence as to whether the theories are applicable. Our narrow cultural focus has also limited our understanding of the full spectrum of psychological phenomena, and the extent to which our own ways of thinking and feeling are affected by our culture.

As the computational narrative community builds shared resources, we recommend that inclusion (or prioritization) criteria be developed to ensure these resources support inquiry into the cross-cultural and universal aspects of narrative. To begin, we suggest prioritizing narrative samples from cultures not yet represented in shared resources and from cultures relevant to specific applications of narrative research. At the 2009 Computational Models of Narrative Workshop, four applications were specified: 1) filtering and making sense of incoming information, 2) detecting and producing propaganda, 3) understanding and influencing other cultures, and 4) helping others tell their own stories (Richards, Finlayson, & Winston, 2009).

Bearing in mind these criteria, and considering our own research, we offer the example of Mexican narcocorridos as a rich, socially relevant narrative sample that meets the under-representation and applicability criteria. The genre originated in a culture often under-represented in the social and behavioral sciences, and is directly relevant to understanding a socio-political problem with serious ramifications for the security of Mexico and its neighbors. Narcocorridos are a tool of influence wielded by those who sympathize with violent cartels. They bear similarities to the Islamic texts that are the subject of current research (Finlayson, 2011). Analyzing narcocorridos will complement research on Islamic extremist texts by illuminating of how narrative influence operates in two different cultural contexts.

3. Mexican Narcocorridos

Narcocorridos, otherwise known as "drug ballads," are a popular form of lyrical music in Mexico and the US states along Mexico's northern border (Simonette, 2006). Narcocorridos are a contemporary form of corridos, ballads that tell stories, typically of a protagonist fighting against overwhelming odds. For example, corridos from the early 1900s told the stories of Mexican revolutionaries (Paredes, 1963; Simonett, 2001a) [1]. Narcocorridos emerged with the rise of the cocaine trade in the 1970s and 1980s. The genre has exploded in popularity in the last ten years with the escalation of the drug wars in Mexico. Narcocorridos typically glorify the stories of individuals within the cartels, raising them to the status of mythical heroes. The lyrics are not drawn from extant stories, they are written to reflect current events, often referencing contemporary politicians, drug kingpins, and conflicts. In fact, some narcocorridos are so current and accurate that law enforcement agencies regard them as useful intelligence sources. Other narcocorridos are allegorical, but nonetheless reflect current social contexts.

Violence related to drug trafficking in Mexico has increased dramatically since 2006, when newly elected President Felipe Calderón initiated military action against the cartels (Rawlins, 2011). Warring cartels have divided the country into territories, and have more control than the federal government in some regions. Cartels have adopted tactics learned from Islamic extremists, mimicking extremists' use of YouTube to broadcast beheadings and other gruesome acts to punish, frighten, and intimidate anyone considering resistance (Johnson, 2010). In 2008, the US Joint Forces Command (USJFCOM) cited the Mexican government as one of the world's least stable governments, second only to Pakistan. The report noted that the Mexican government was at risk for "rapid and sudden collapse" due to the increasing instability caused by organized crime and drug cartels.

Narcocorridos are a pragmatic candidate for early entry into a shared narrative resource for several reasons. First, their lyrics are easy to obtain; an internet search immediately yields dozens of songs with linguistically accessible Spanish lyrics (although some familiarity with Mexican slang is required). Second, because narcocorrido musicians are covered heavily in the Mexican and Mexican-American media, we can estimate the popularity of different narcocorridos, facilitating research into the narrative characteristics that resonate with audiences. Finally, the methods required to analyze prose and analyze lyrics will have numerous similarities. Like stories, narcocorridos have settings, characters, plots, story trajectories, resolutions, themes, metaphors, and many other characteristics of prose. To the extent that lyric analysis requires different methods than prose analysis, our understanding of both narrative forms will be enhanced. Excluding musical forms of narrative from the corpus may mean excluding genres that are important means of storytelling in some cultures.

Understanding the similarities between narcocorridos and narratives from other strategically important cultures is key in understanding how narratives operate as propaganda. For example, there are similarities between narcocorridos and the Taliban-produced videos promoting Taliban resistance to ISAF forces in Afghanistan. The manner in which the Taliban videos glorify and mythologize violence is similar to both narcocorridos and American gangster rap (in fact, many Taliban videos use gangster rap as a soundtrack). These Taliban narratives are distributed similarly to narcocorridos – as viral videos, over pirate radio stations, and through DVDs in the marketplaces[2]. There is great concern among strategic communications experts that both the Taliban recruitment videos and narcocorridos build support among their audiences and draw new recruits to their respective organizations (Garcia, 2006; Seib, 2011). Both Taliban narratives and narcocorridos effectively cast villains as heroes by simultaneously leveraging and subverting cultural norms and values. We may learn a great deal about narratives as tools of influence by studying how these types of narratives take advantage of cultural norms and memes to glorify violent anti-social behavior that is highly detrimental to the very people embracing the narratives.

4. Conclusion

Narcocorridos use narrative to effectively leverage traditional Mexican values and sentiments in a way that is both broadly appealing and highly influential. We believe that analyzing narcocorridos will advance our understanding of how narratives operate within cultural contexts and how they interact bi-directionally with the socio-political context. By comparing these narratives to narratives with similar functions from other cultures (such as the Taliban recruitment narratives or Islamic extremist narratives) we may learn which aspects of these narratives cross cultural boundaries, and which are culture-specific, leading to a deeper understanding of how narratives affect politics and security.

[1] The *banda* music that accompanies the narrative is rooted in the fusion between Mexican and German music in the 1890s; banda bears many similarities to polka (Simonette, 2001b).

[2] Narcocorridos are banned from the airwaves in some Mexican states. To avoid regulation and cartel reprisals, much of the production is done across the border, in Los Angeles, California. Narcocorridos are broadcast from stations in the US Southwest, distributed by CDs and DVDs in Mexican markets, and virally through YouTube and other social media (Schoichet, 2010).

5. References

Finlayson, M. A. (2011). Corpus annotation in service of intelligent narrative technologies. *AAAI Technical Report*, WS-11-18, 17 – 20.

Finlayson, M. A., Richards, W., & Winston, P. H. (2010). Computational Models of Narrative: Review of the Workshop. *AI Magazine*, pp. 97-100.

García, M, M. (2006). 'Narcoballads' : The psychology and recruitment process of the 'Narco'. *Global Crime*, 7 (2).

Johnson, T. (2010). Why are beheadings so popular with Mexico's drug gangs? *McClatchy Report*, posted April 1,www.mcclatchydc.com/2010/04/01/91481/beheadings-become-signature-of.html

Paredes, A. (1963). The ancestry of Mexico's corridos: A matter of definition, *The Journal of American Folklore,* 76 (301), pp. 231- 235.

Rawlins, A. (2011). Mexico's Drug War. *Council on Foreign Relations*.

Richards, W., Finlayson, M. A., & Winston, P. H. (2009). *Computer Science and Artificial Intelligence Laboratory Technical Report : Advancing Computational Models of Narrative*. Cambridge, MA :Massachusetts Institue of Technology.

Rosch, R. (1975). Universals and cultural specifics in human categorization. In R. Brislin, S. Bochner, & L. Lonner (Eds.), *Cross-cultural Perspectives on Learning*. New York: Halstead Press.

Schoichet, C. (2010). Narcocorridos: Mexican Ruling Party Proposes Ban on Songs that Glorify Drug Trafficking. *El Paso Times*. January 22, 2010.

Seib, P. (2011). *Public Diplomacy, New Media, and Counterterrorism*. Los Angeles: Figueroa Press.

Simonette, H. (2001a). Narcocorridos: An emerging micromusic of Nuevo L.A. *Ethnomusicology*, 45(2), pp. 315-337.

Simonette, H. (2001b). *Banda: Mexican Musical Life Across Borders*. Middletown, CT: Wesleyan Press.

Simonette, H. (2006). Los gallos valiente: Examining violence in Mexican popular music. *Transcultural Music Review*, 10, pp. 1-19.

USJFCOM. (2008). Joint Operating Environment : Challenges and Implications for the Future Joint Force. Norfolk VA.

Towards a Digital Resource for African Folktales

Deborah O. Ninan, Odetunji A. Ọdẹ́jọbí

Computer Science and Engineering Department, Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀, Ọ̀ṣun State, Nigeria
jninan@oauife.edu.ng, oodejobi@oauife.edu.ng

Abstract

This paper explores the development of a digital resource that is amenable to the formal specification of African folktales. The ultimate aim of this project is to develop computational structure and models for the narrative underlying African Folktale. We collected a number of Yorùbá folktales with corresponding English translations. We then analysed their components and the structure of the narrative that they embodied. The requirements of a markup language to capture the content and structure of the narratives was proposed. Ongoing work is aimed at the development of a framework for the computational model and the automatic generation of folktales based on this data.

1. Introduction

African folktales (AFT) have being studied, explored and admired for their aesthetics and moral teachings. Their interesting and engaging narratives have been a subject of many intellectual discourse. Here, we examine a formal treatment of the narratives underlying these folktales with the aim to present useful computational model for the development of digital artefacts such as software. Our approach becomes necessary in the background of the benefits of modern computer software used in entertainment and education. Subjecting African folktales to computational analysis is necessitated by the need to expand their applications as well as share their intellectual resource with a broader audience. In this paper therefore, we examine the features of African folktales and the development of a digital resource that is amendable to formal specification of the narratives underlying them. In indigenous African communities, and amongst the Yorùbá people in particular, folktales are essential tool for educating the youth on the culture and moral values of the society. We shall be focusing specifically on folktales of Yoruba origin. We think that the richness of this class of folktale provides interesting features and characteristics that can be found in most folktales of African origin.

In this work, we collected and analysed a number of folktales from Yorùbá origin. We observed, generally, that they share a similar structure, content, form and configuration in terms of the temporal and spatial properties of their narratives. It is safe to speculate, therefore, that a formal representation mechanism can be devised to model their narratives. Such a mechanism can then be developed into computational artefacts amenable to mathematical treatment and a tool for automatic generation of narratives.

The rest of this paper is organised as follows. In section 2. we provide a discussion of folktales in African context. Section 3. contains a brief description of features of African folktales. In section 4. we discuss our efforts at developing the digital resource for AFT and briefly states the markup system used in this work and section 5. concludes this paper.

2. Folktales in the African Context

Education in the African cosmology involves the use of different and diverse medium of instruction including myths, legends, proverbs, tongue twisters, riddles, jokes and folktales. Of these, the folktales continue to be a potent instrument of education and entertainment. A unique aspect of African folktales is that they exhibit, in a vivid manner, the richness and fertile legacy of the culture that produced them. Folktale narratives facilitates the creation of abstract world, away from reality, through the ingenuous use of actors and props. Hence, a folktale creates a fictional world for the exploration of and explanation of abstract entities and ideas. This presents a simplified world view that is easily comprehensible, especially by children and youth who are usually the target of folktales. Intuitively, such folktales serve as a mechanism for learning about simple concepts which can be used in gaining understanding and insight into more complex and real-life situations.

In the African cosmology, folktales serve as a means of handing down traditions and customs from one generation to the next. The storytelling tradition has thrived for generations in the absence of formal documentation, e.g. in the form of printed material. Folktales are used to mould children behaviour and personality, as they provide a means for impacting many lessons about life and living. Folktales are used to teach children about their heritage, culture and codes of behaviour. In most African settings, the audience is usually children, age twelve and below, and the narrator is almost always one of the aged women in the neighbourhood. African folktales are found in other continents, such as in North America, South America, and the West Indies.

A peculiar feature of Yoruba folktales is that the audience is encouraged to interpret the narrative. In that sense, knowledge is attained not just by receiving information, but also by interpreting the information and relating it to the learner's experience. Yoruba folktales seems to impact on the learners the ability to organize, structure, and use information which are essential to complex problem solving.The Yorùbá folktales are numerous, among these are stories, riddles or fables, histories, myths, songs, proverbs etc.

Although many theories of narrative have been proposed, several processes pertaining to narrative remain inadequately formalized and, hence, beyond full mechanization (Michael, 2010). Certain premises that are central and necessary, in any attempt to understand narratives have being highlighted (Michael, 2010). According to him, the first premise is that narratives cannot be understood out of context. This is because what counts as a narrative in one environment may not count as such in another. There is therefore the need to encode and reason with the domain knowledge about environments in folktale modeling.

Building storytelling systems has been one of the biggest challenge in the history of Artificial Intelligence (AI) research (Peinado and Gervás, 2005b). Some work have addressed the intelligence underlying narrative. For example, (Brooks, 1997) described a system for graphically representing and manipulating non-linear cinematic narrative such that the narrative material can be treated as a programmatic expression of computation. The work also developed a model called agent stories for generating non-linear cinematic stories for new digital media and concluded that a graphical method of seeing and manipulating the structure of multi-linear story material would be a valuable tool for trying new structural arrangements quickly as well as for helping to write new story material. Lang (1999) has proposed a declarative model for simple narratives that points to a concrete implementation and directly address the question of what constitutes a story. The implementation assumed that stories will have only a single protagonist, enforcing a fixed point of view and so cannot represent some kinds of tales.

It was observed by (Peinado and Gervás, 2005a) that a majority of storytelling projects reuse a couple of main plots and only change elements of the story world, like characters, objects, places, etc. Such story may sometimes seem different to the reader, but from the point of view of narratology this approach has no guarantee of success in terms of creativity. Gervás et al. (2005) showed that the usual ontological tools are unable to offer complete and reliable solutions for representing and exploiting narrative information. They pointed out that there is a need to take care of those connectivity phenomena like causality, goal, indirect speech, co-ordination and subordination, etc., that link together the basic elementary events. To address this problem, Zarri (2005) proposed the Narrative Knowledge Representation Language (NKRL), expressly specified and implemented for dealing with non-fictional narratives and temporal information. He also observed that W3C proposals such as the RDF(S), OWL or OWL2 are, in their standard format, unable to supply a basis for representing elementary events on a computer. Efforts directed at addressing this problem (Peinado and Gervás, 2005a; Tuffield et al., 2006) have identified three key approaches towards narrative generation: content modeling, story modeling and user modeling. A three-layer architecture for narratives ontological model which considered narratives as important form of knowledge representation have been suggested. This approach has become a key task to machine accessible knowledge in both expressing and understanding narratives and related concepts. Szilas (2010) has opined that existing computational models of narrative need to be improved in two directions: (i) a broad model of narrative is needed, (ii) model must also take into account the fact that a story lived by the user, is different from a story observed by an audience. A fundamental requirement of interactivity need to be added to this modeling approach for building computational models for interactive narrative, which is central to AFT modeling.

In summary, as suggested by Michael (2010) and Ontanon and Jichen (2011), several processes pertaining to narratives remain inadequately formalized limiting the prospect for full mechanization. This is particularly so for African folktales. It has been proposed that a general formal framework that attempts to make precise such processes and related notions is important. In addition the importance of certain premises that narratives are expected to adhere to, and the formal implications that these have in terms of the computability of the various relevant notions have been suggested. Ontanon and Jichen (2011)'s recent work suggested the importance of domain knowledge in story generation and particularly in analogy-based story generation (ASG) based on the construct of knowledge container in case-based reasoning, presented a theoretical framework for incorporating domain knowledge in ASG. They concluded that proper vocabulary; a better similarity measure; a large variety of source stories and incorporating domain knowledge into the mapping algorithm is fundamental for high quality narrative generation. Our work derives its design philosophy from the premise that African folktales are composed of narratives which share similar characteristics with those presented in the reviewed work. Hence, it should be possible to use many of the models proposed in the literature as given while extending other feature that are limited in expressing African folktales.

3. Features of African Folktales

In this section we describe the features of African folktales(AFT) that we consider to be interesting in the context of computational modeling. As stated earlier, the discussion here is based on Yorùbá folktales but the features discussed are shared by folktales of African origin. First we discuss the attributes of the contents of AFT and provide a formal description of its structure.

3.1. Components of African Folktales

The various settings of African folktales include farmlands, towns, marketplaces etc. The characters could be animals, humans or spirits who all take the form of humans and the story usually has one plot which actually starts with introduction graduates into climax and ends with the conclusion. We illustrate our discussions here using the story documented in Table 3 with the English translation in Table 4 in the Appendix.

3.1.1. The actors

African folktales comprise narratives that reflect a culture where animals abound and form an integral part of the community. Consequently, humans exist in the same world as the monkey, elephant, giraffe, lion, zebra, crocodile and rhinoceros along with a wide variety of birds such as the ostrich, parrots, vulture and the eagle. Animals can also occupy positions such as a king, priest, wife, husband, chief, etc, as well as engage in occupations such as farmers, herbalist, blacksmith, and bricklayer usually reserved for humans in the real world. The animals and birds in this world can also take on human characteristics of greed, jealousy, honesty, loneliness, etc. Character is said to be an aspect of narrative deeply intertwined with plot. According to (R. M, 2000), character directly influences an agent's choice for action, which in turn contributes directly to an unfolding plot. The stories in the folktale illustrate the consequence of each of these behaviour to provide valuable moral lessons. The actors sometimes include visible spirits such as *Yemọja* (Mermaids) and invisible spirits (e.g. *iwin*), trees (*ìrókò*) and other inanimate objects such as rocks or rivers. These objects are empowered with the ability to speak, work, laugh, sing, etc. in order to convey the message of the story. It is instructive to note that animals such as the tortoise is a popular actor associated with being very cunning and clever. Its actions usually result in disgrace and regret at the end of the story (see story in Table 4). There are exceptions to this though, as the same tortoise can be cast as being wise and reasonable in other tales. Babalola compiled a collection of Yoruba folktales which reveal the importance of animals and their interrelationship in the Yoruba culture(Babalola, 1979).

3.1.2. The props

The prop of a story is the surroundings in which the tales take place. The majority of the narratives are situated in the village setting where the individual objects become props. The farmland, rivers and hills, the blacksmith workshop, Herbalist place (see story in Table 4), the market, etc. are examples of settings within which a story can be told. In African folktales the environment reveals the vastness of the land and educate the reader about the climate, such as the dry (Harmattan) and raining seasons. It also include periods of draught when the rain has stopped for several years. In the rainy season the hills are described as being slick with mud. The *acacia* trees swaying in a gentle breeze, muddy streams that are home to fish, hippos and crocodiles, moss covered rocks, and giant ant hills that serve as a "back scratcher" for huge elephants. All these give the listener a sense of the variety of life in a typical African jungle world. The king's palace, festivals such as the *Egúngún* festival, dressing and fashion, etc. are usually conceptualised and described in the context of African world view. There are also spirit underworld (*Àjà-ilẹ̀*) and under-rivers inhabited by beings of various shapes and disposition.

3.1.3. The songs

A unique attribute of folktale of Yorùbá origin is that they almost always contain songs that are sung at particular stages of the narrative. Many of the folktales have musical participation by the audience that adds much to making the tale more interesting and enjoyable. In the Yorùbá folktales it is common for the audience to answer questions aloud, to clap their hands in rhythm to word repetition (chorus), and to join in the chorus. In the story in Tables 3 and 4 the song is:

babalawo mo wa bebe *alugbinrin*
ogun to se fun mi lekan *alugbinrin*
o ni ki n mama mowo b' enu *alugbinrin*
o ni ki n mama mese b' enu *alugbinrin*
gbongbo lo yo mi tere *alugbinrin*
mo fowo kan 'be mo mu b' enu *alugbinrin*
babalawo mo wa bebe *alugbinrin*

Most of the songs are repetitive as the same chorus are used repeatedly. There are terms in the song that do not serve any semantic function but only to serve the purpose of creating a rhythm. The song is the main mechanism for interactiveness in the folktale narratives as it invokes the audience's active participation and it serves to sustain attention, reinforce the story and improve audience experience of the narratives. It has been discussed that the act of singing communicates the basic emotional state such as, fear, anger, joy, sadness, surprise and disgust and conveys information about group membership such as age, gender culture and social groups (Welch, 2005). Singing was pointed out as a cultural transformational activity.

3.1.4. The themes

Themes in Yorùbá folktales are usually stated, by the storyteller, at the beginning of the story. It is normally a pointer to the lesson that is expected to be learned in the story. The concept embodied in the theme may be conveyed by a set of characters with stereotypical traits belonging to the human, animal, and metaphysical realms. AFT teach specific moral lessons. Most African Folktales explain why an animal looks or acts in the particular way they do, or how an animal came to have specific character trait. The tales provide causal explanations or reasons for common and uncommon things and usually end in proverbs. The origin and meaning of most Yorùbá proverbs can be traced to folktales (Babalola, 1979).

3.1.5. The plot

The plot in an African story suggests inter-related sequence of events that seem to follow a unique cause and effect pattern that emphasise, and are conceptualised around, the community. The community in this case include the physical world of the living and the meta-physical world of the ancestors (the living dead). Underlying the plot are three important stages: (i) the beginning where the community is at peace; (ii) the middle when a member of group of individual acts in a manner to violate the peace; and (iii) The end when normalcy is restored, or there is a return to the peaceful state of the community. The general aim is to show the consequence of unacceptable behaviour and the price required for bringing back the community into its peaceful state. When certain member of the community act in a manner to compromise the order in the community, he or she has not only offended the living, he has also offended the ancestors as well. Persistent violations are considered unacceptable behaviour. The communal structure therefore imposes some obligation on the individual member of the community and the need for the entire society to ensure compliance. The plot is generally about how well this pattern of events accomplishes some social order and/or disorder. In the case of an imbroglio, the plot underlying the narratives may include multiple inferences to real and imaginary actors.

4. AFT Markup system

Our preliminary study on the subject suggests that a key digital resource for the computational study of narrative, that is, quality corpora, are not available. After a preliminary theoretical exploration of the language for representing the narrative underlying AFT we consider the development of the requisite language resource very crucial. In this section, therefore, we first address the general description of the specific characteristics as well as explore the formal representation of structure and contents of AFT narratives. Our approach at this stage of the research is to extend the eXtensible Mark-up Language (XML) (Zeng, 2010) for marking up the collected folktales. In order to capture the description of the folktale accurately, we employ a hierarchical meta-language approach. In this approach, a narrative is described starting from the most abstract. The initial abstract representation is further described at a more concrete level until all the items have being completely described. The meta-data schemas proposed in this work aim to provide a profile for narrative for extending the context of a folktale by providing a description that emphasised specific attribute of the task. The technique we adopted allow us to use the same tag for similar entities in the Yorùbá version and the translated English version of the narrative. This way we can generate a cross index between the two versions of the story. For example, Table 1 documents the marked-up section of the first paragraph of the story in 3. As can be seen in the English translation in Table 2, the tag is similar and in the lexical items for the tag, comes from the English language. For example, the tag <actor> is used to annotate the main actor as <actor id=1> Ìjàpá </actor> in the Yoruba version and <actor id=1>tortoise</actor> in the English version. We have redefined and extended a number of XML tags in this manner and so far we have collected about twenty (60) story items. We hope to continue to expand the digital resource as our research progresses.

Table 1: Ìjàpá àti oníṣègùn

<actor id=1> Ìjàpá </actor> àti <actor id=2> Yọ́nrínbo</actor> ìyàwó re ti jẹ́ tọkọ taya fún ọ̀pọ̀lọ̀pọ̀ ọjọ́ ṣùgbọ́n tí wọn kò rí ọmọ bí. Èyí ba <actor id=2> Yọ́nrínbo</actor> inú jẹ́ púpọ̀, ó sì pinnu láti wá ìrànlówó lọ sí ọ̀dọ̀ <actor id=3>bàbá oníṣègùn</actor>. <actor id=3>Bàbá oníṣègùn </actor> se ọbè àsèje kan tí ó lágbára láti mú ni lóyún fún <actor id=2>Yọ́nrínbo</actor>. Ó sèé pọ̀ mọ́ ọbè ẹran aládídùn. O gbe fún <actor id=1> ìjàpà</actor> kí ó gbe lọ sí ilé fún ìyàwó rẹ yọ́nrínbo. <actor id=1>Ìjàpá</actor> gbe ìkòkò ọbẹ̀ aládídùn yii on forí lé ònà <prop id=3> ilé re </prop>.

Table 2: Tortoise and the medicine soup

<actor id=1>Ijapa</actor> the <actor id=1>tortoise</actor> and his wife <actor id=2>Yorinbo</actor> have been married for a long time but do not have a child. This made <actor id=2>Yorinbo</actor> very sad. <actor id=1>He</actor> decided to seek help from a <actor id=3>medicine man</actor>. The <actor id=3>medicine man</actor> prepared a powerful potion that would make <actor id=2>Yonribo</actor> pregnant. <actor id=3>He</actor> mixed it into a delicious smelling beef stew and handed it to <actor id=1>Ijapa</actor> the <actor id=1>tortoise</actor> to give to his wife. The <actor id=1>tortoise</actor> carried the < prop id = 2 pot > and set off for <prop id=3> home </prop>.

5. Conclusion

In this presentation, we have provided an exposition of African folktales in the context of computational analysis and synthesis. We have discussed the feature of the folktales that are interesting from the point of view of computational rendering. We identify the need to develop a domain specific digital resource for use in this work and we have discussed a mark-up scheme based on the XML for annotating and cross-indexing our AFT digital resource collection. Our ongoing work is focused on the computational modeling and the automatic generation of folktales based on the digital resource that we have proposed in this paper.

6. References

A. Babalola. 1979. *Akojopo alo ijapa*. Number ISBN 0 19 575227. university Press Limited, Three Crowns Building, Eleyele road, Jericho, Ibadan.

K. Brooks. 1997. Programming narrative. In *Proceedings. 1997 IEEE Symposium on Visual Languages (Cat. No.97TB100180)*, pages 380–386.

P. Gervás, B. Dı'Agudo, F. Peinado, and R. Hervás. 2005. Story plot generation based on cbr. *Journal of Knowledge-Based Systems*, 18(4-5):235–242, month.

R. Lang. 1999. A declarative model for simple narratives. In *Proceedings of the AAAI fall symposium on Narrative Intelligence*, pages 134–141.

L. Michael. 2010. Computability of narrative. In *Proceedings of the Computational Models of Narrative: Papers from the AAAI Fall Symposium*, pages 47–54.

S. Ontanon and Z. Jichen. 2011. On the role of domain knowledge in analogy-based story generation. In *Twenty-Second International Joint Conference on Artificial Intelligence*, pages 1717–1722.

F. Peinado and P. Gervás. 2005a. Creativity issues in plot generation. In *IJCAI-05 Workshop on Computational Creativity)*, pages 45–52.

F. Peinado and P. Gervás. 2005b. A generative and case-based implementation of proppian morphology. In *The 17th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (ACH/ALLC)*, pages 129–131.

Young R. M. 2000. Creating interactive narrative structures: The potential for ai approaches. In *AAAI Spring*

Symposium in Artificial Intelligence and Interactive Entertainment, AAAI.

N. Szilas. 2010. Requirements for computational models of interactive narrative. *Artificial Intelligence*, pages 62–68.

M. M. Tuffield, D. E. Millard, and N. R. Shadbolt. 2006. Ontological approaches to modelling narrative. In *Proceedings of 2^{nd} AKT Doctoral Symposium*.

G. F. Welch. 2005. *Singing as communication*. Oxford University Press., New York.

G. P. Zarri. 2005. An n-ary language for representing narrative information on the web. *SWAP*.

M. L. Zeng. 2010. Domain-specific markup languages and descriptive metadata: their functions in scientific resource discovery. *Encontros Bibli: Revista Eletronica de Biblioteconomia e Ciencia da Informacdcao*, pages 164–176. DOI = 10.5007/1518-2924.2010v15nesp2p164.

Appendix

Table 3: Ìjàpá àti oníṣègùn

Ìjàpá àti Yọ́nrínbo ìyàwó re ti jẹ́ tọkọ taya fún ọ̀pọ̀lọpọ̀ ọjọ́ ṣùgbọ́n tí wọn kò rí ọmọ bí. Èyí ba inú Yọ́nrínbo jẹ́ púpọ̀, ó sì pinnu láti wá ìrànlówó lọ sí ọ̀ọ̀ baba oníṣègùn. Bàbá oníṣègùn se ọbè àsèje kan tí ó lágbára láti mú ni lóyún fún Yọ́nrínbo. Ó sèé pọ̀ mọ́ ọbè ẹran aládídùn. O gbe fún ìjàpà kí ó gbe lọ sí ilé fún ìyàwó rẹ yọ́nrínbo. Ìjàpá gbe ìkòkò ọbẹ̀ aládídùn yii on forí lé ònà ilé re.

Láìpẹ́, ọbè aládídùn náà bẹ̀rẹ̀ sí íta sánsán sí ìjàpá ní imú. Títa sánsán náà pọ̀ tí ìjàpá pinnu láti yojú wo ohun tí babaláwo se sínú ìkòkò náà. Ní gbàt ó yọjú wòó ó rí àwọn ẹran ńlá kòndù kòndù nínú ọbẹ̀ àsèjẹ yi. Ìjàpá lérò wípé dájúdájú Yọ́nrínbo kò lè dá nìkan je ọbè náà tán. Yọ́nrínbo kò leè jeun púpọ̀, mo ní láti ràn án lọ́wọ́. Nítorínáà ó mú ẹran ńlá kòndù kan ó sì báa jẹẹ́. Ó tún mú èkejì, ó tún mú èketa, ọbè yi mà ti dùn jù o ìjàpá so sínú ara rẹ̀. Ó tún mú ìkerin, títí kò fi kàá mó. Títí ó fi se àkíyèsí pé ẹran ti ku eyọ kan nínú ìkòkò. Ìgbà yí ni àyà ìjàpá já pé òun ti fi gbogbo ọbè Yọ́nrínbo jẹ tán. Ṣùgbón láìpé láì jìnnà ó gbàgbè, ó lérò wípé bí onísègùn bá jé alágbára lóòtó bí wọ́n se sọ, ó yẹ kí eyọ ẹran kan tí ó sẹ́ ku lè tó láti mú kí Yọ́nrínbo lóyún. Ó dé ìkòkò náà ó sì tèsíwájú ní ònà ilé rè. Bí Ìjàpá tí nlo ní ona ilé rẹ ni ikùn rẹ bẹ̀rẹ̀ síí wú, ìgbà tí ó se ikùn bẹ̀rẹ̀ síí dun. O dùn ún tó béè tí ó fi pinnu láti padà sí òdò bàbá onísègùn láti bẹ̀bẹ̀ fún ìrànlówó. Bí ó ti ńlo ni ó bẹ̀rẹ̀ sí ní kọ orin wípé,

bàbá láwo mo wá bèbè *alugbinrin*
ògùn tó ṣe un mi lerekan *alugbinrin*
ó ní kí n mámà mówó b enu *alugbinrin*
ó ní kí n mámà mèsè b enu *alugbinrin*
gbòngbò lónà yo mí tèrè *alugbinrin*
mo fọwọ́ kan bè mó mu b enu *alugbinrin*
bàbàláwo mo wá bèbè *alugbinrin*

Bí ó ti ńlọ ni ikùn rè ńwú síi. Ìgbà tí ó fi máa dé ilé bàbá onísègùn ikùn rẹ ti le ó sì ńdùn ún púpò tó bẹ́ẹ̀ tí ó fẹ́rẹ̀ má leè sòrò mó. Ìgbà tí ó dé ilé bàbá onísègùn ó jẹ́wọ́ ohun tí òun ṣe ó bẹẹ onísègùn náà láti ran lọ́wọ́. Ṣùgbọ́n ó seni laanu pé onísègùn kò ní ẹ̀rọ̀ sí wàhálà ìjàpá. Béè ni ikùn ìjàpá wú 'u títí kò lè sọ̀rọ̀ mọ́ títí ó fi kú léhìn ọjọ́ díè.

Table 4: Tortoise and the medicine soup

Ijapa the tortoise and his wife Yorinbo have been married for a long time but do not have a child. This made Yorinbo very sad. He decided to seek help from a medicine man. The medicine man prepared a powerful potion that would make Yonribo pregnant. He mixed it into a delicious smelling beef stew and handed it to Ijapa the tortoise to give to his wife. The tortoise carried the pot and set off for home.

Very soon, the beef stew aroma became overpowering and the tortoise thought he should take a peek into the pot. Inside were very large juicy chunks of meat and Ijapa thought, "Surely Yorinbo cannot finish these by herself. She has a small appetite, I will have to help her out". So he helped her to eat one big juicy chunk of meat. Then two, and three. The beef stew was really delicious. Ijapa thought surely, the medicine man was in the wrong profession. And then four. Soon, he was no longer counting, until he noticed that there was only one piece left.

The tortoise was shocked at what he had done. He had eaten Yorinbo's potion. But not one to dwell too long on his mistakes, he shrugged it off. After all, if the medicine man was as powerful as they say, the remaining one piece of chunky juicy meat should be enough to make Yorinbo pregnant. So he covered the pot and continued on his way.

As the tortoise continued on his way home, his stomach began to ache. His stomach ached so badly, he decided to return to the medicine man to seek some help. He began to sing:

babalawo mo wa bebe *alugbinrin*
ogun to se un mi lekan *alugbinrin*
o ni ki n mama mowo bo enu *alugbinrin*
o ni ki n mama mese bo enu *alugbinrin*
gbongbo lo yo mi tere *alugbinrin*
mo fowo kan obe mo mu bo enu *alugbinrin*
babalawo mo wa bebe *alugbinrin*

By the time he got to the medicine man, his stomach was huge and hard and ached so much that he could hardly talk. But he managed to confess what he had done and pleaded for help. Unfortunately, the medicine man could not help him and Ijapa the tortoise had to face the consequence of his action. His stomach continued to grow and ache and after several days of agony, Ijapa died.

Formal Models of Western Films for Interactive Narrative Technologies

Brian Magerko and Brian O'Neill
Georgia Institute of Technology
85 5th St. NW, Technology Research Bldg. #319, Atlanta, GA 30308-1030
E-mail: magerko@gatech.edu, boneill@cc.gatech.edu

Abstract

Interactive narrative technologies have typically addressed the authoring bottleneck problem by focusing on authoring a tractable story space (i.e. the space of possible experiences for a user) coupled with an AI technology for mediating the user's journey through this space. This article describes an alternative, potentially more general and expressive approach to interactive narrative that focuses on the procedural representation of story construction between an AI agent and a human interactor. This notion of procedural interaction relies on shared background knowledge between all actors involved; therefore, we have developed a body of background knowledge for improvising Western-style stories that includes the authoring of scripts (i.e. prototypical joint activities in Westerns). This article describes our methodology for the design and development of these scripts, the formal representation used for encoding them in our interactive narrative technology, and the lessons learned from this experience in regards to building a story corpus for interactive narrative research.

Keywords: interactive narrative, story corpus, improvisation, scripts, cognition

1. Introduction

The field of interactive narrative technologies (INTs), where researchers create AI-driven approaches to computer-mediated story experiences for human users, is heavily constrained by the prospect of content authoring. No matter what technical approach a particular researcher is exploring, it is typically difficult to show that a system works in a compelling fashion without a non-trivial effort in writing story content in a machine-readable form. This is due to interactive narratives containing both a formal, computational element (e.g. the programming behind getting the Holodeck to work) and an aesthetic one (e.g. the story content that has to be encoded in the Holodeck so it can involve users in story-based experiences). The work presented in this article discusses how a focus on splitting INT research into *background knowledge* for the formulation of stories and *processes* that operate on that knowledge to enable multiple agents to collaboratively create a story can be used to address this issue of authoring in a novel way.

Interactive narratives normally involve exposing a user to a *story space* (i.e. a bounded experience where multiple possible stories can be experienced) where larger story spaces mean more possible personalized user experiences and, most importantly, more content authoring by the designer. The story space can be thought of as "the space of intended experiences" for the user; in other words, the author / designer's vision (Magerko 2007a). The AI-based technology employed (normally called a *story manager*) typically serves as a guide through that space (see Roberts and Isbell 2007 for a survey of the field). "Guiding the user" could mean helping the user stay within the bounds of the story space and not executing actions that could lead to the story stopping (Young et al. 2004; Magerko 2007a). It could alternatively mean selecting story content that fits the system's perception of what would be most enjoyable to the user (Thue et al. 2007; Yu and Riedl 2012).

The most successful interactive narrative to date, *Façade,* reportedly took over five man years to author content for a relatively short work (30 minutes max for a successful story) compared to other media forms (e.g. 30-60 minute television shows, 90-150 minute films, or 40 hour digital games). *Facade* represents story content as *beats* (i.e. atomic moments of interesting narrative content), which are dynamically selected by the system as the user interacts with the characters in the story world (Mateas and Stern 2002). Other representations include planning operators, story graphs, and Proppian functions (Roberts and Isbell 2007). Each of these representations are used by the designers of systems to create the space of possible experiences for the user through the intentional authoring of story events / beats / etc. for a user to potentially experience.

The story spaces of the systems, such as *Façade*, referred to above are bounded by the content authored for the experience; in other words, the only scenes experienced by a user of such a system are the ones hand authored by the designer. While this may not be a problem itself – many systems have been built with this authoring constraint - it is a limitation of INTs that human storytellers do not have. People have the ability to draw on their personal experiences, on other stories they have heard or told, etc. and create something wholly new. Even if that new thing is an amalgam of older stories and experiences, the generative process of combining these narratives into a new one is a creative process in and of itself; very few stories told are wholly new and unique.

There is a subset of human storytelling that deals with the real-time generation of story content as a key part of the story experience for those involved. These domains (e.g improvisational theatre, Live Action RolePlay, tabletop roleplaying, etc.) often provide a similar kind of experience to those INTs attempt to provide (Flowers, Magerko, and Mishra 2006; Magerko et al. 2009). Our empirical observation of improvisational actors (Magerko et al. 2009; Baumer and Magerko 2009; Baumer and Magerko 2010; Fuller and Magerko 2011)

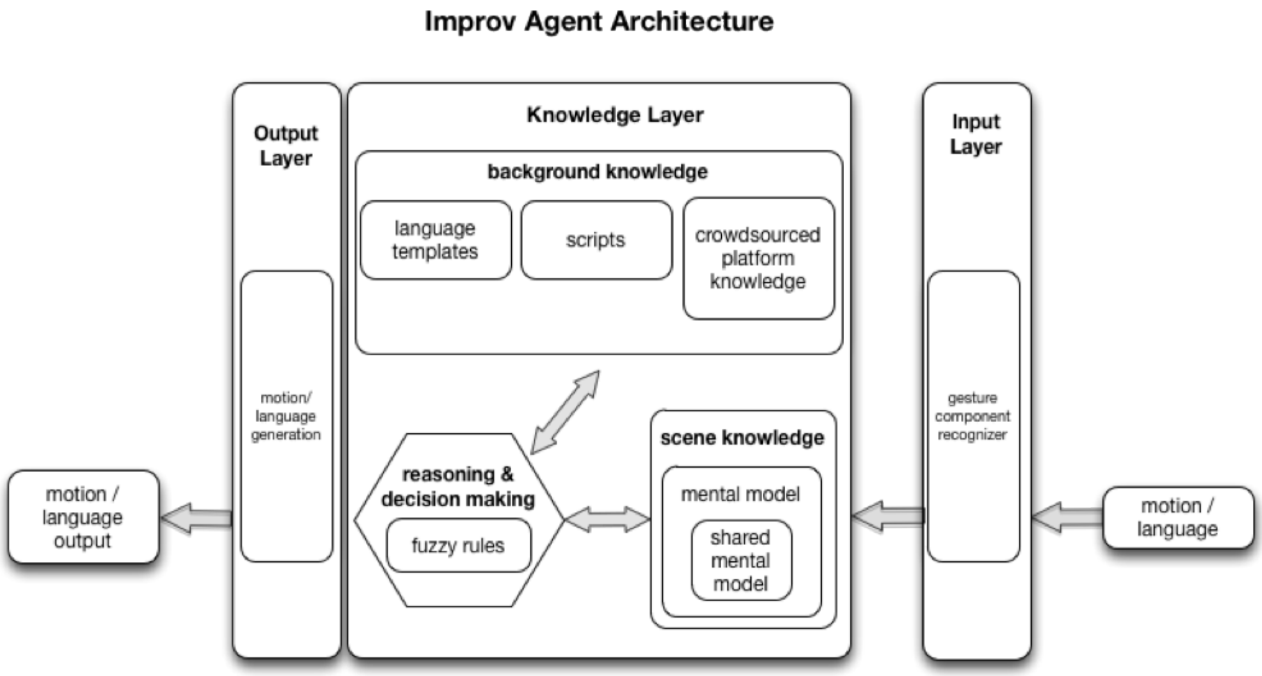


Figure 1. An architectural view of the *Digital Improv Project*. The significant components are a) the gestural / language input layer using a Microsoft Kinect, b) the background knowledge, c) the knowledge that is collaboratively constructed about the scene, d) the reasoning mechanism, and e) the gestural and language output layer.

has lead to the unsurprising conclusion that improvised stories by professionally trained actors are rarely anything close to the verbatim retelling of stories or personal experiences of the past. Rather, improvisers weave knowledge from a myriad of personal and cultural sources, in real-time, to co-create a new story with their fellow improvisers on stage. This approach to story generation is wholly different from anything seen in traditional interactive narrative technologies that rely heavily on pre-authored story spaces

As we have argued elsewhere (O'Neill et al. 2011), the notion of *story co-creation* is a particularly powerful one for INT research. Story co-creation refers to a story space that is bounded by the basic knowledge the agents involved know plus the functions they have for presenting, combining, and altering that content (much like improvisational actors do on stage) as opposed to having a centralized intelligent agent (i.e. a drama manager) that has privileged information about what the story can and cannot contain. We contend that story co-creation is an understudied, but potentially powerful stance on how to build INT systems. By taking a stance of making co-creative systems, we are placing the user in a position that has the same privileges as the computer; the user is no longer limited by only the vision of the designer and what has been encoded in the story space. This is not necessarily the only kind of interactive narrative people want (e.g. different entertainment media exist with varying amounts of user agency / control over the experience), but it is a direction for the field that is both underexplored and potentially fruitful given the plentiful examples of co-creative experiences that humans enjoy.

Based on our empirical study of improvisers (Magerko et al. 2009; Fuller and Magerko 2011; O'Neill et al. 2011), we have concluded that in order to build a co-creative interactive narrative experience, we must develop a system that has: a) similar *background knowledge* (see Figure 1) to the other agents (human or AI) in the scene, b) a model of the *scene knowledge* related to the story that is being negotiated / communicated as part of the performance, and c) the processes that operate on both knowledge sets to maintain the scene knowledge base correctly and to collaboratively construct a scene with the other actor(s). This article focuses on the first construct, background knowledge. We have focused on two main knowledge representations for background knowledge, as described in detail in this article: fuzzy concepts (based on Lakoff's *prototype theory* and fuzzy logic (1989)) and scripts (based on Schank and Abelson's seminal work on this formal psychological construct of temporal events in our daily lives (1977)).

This article briefly covers our previous research and the related work in script-based representations and improvisation in interactive theatre. It then discusses our background knowledge representation and the use of scripts as a key element in representing joint activities (i.e. what agents are doing together) in a scene. It closes with a discussion on the process we have employed in creating our story corpus and the lessons learned for building a shared repository of corpora for the interactive narrative community.

2. Related Work

Previous Research

Our work on the *Digital Improv Project* has led us to build interactive narrative technologies that are based on a formal study of the socio-cognitive processes employed by improvisational actors (Magerko, Fiesler, and Baumer 2010; Magerko, Dohogne, and DeLeon 2011). We have built formal models of how actors negotiate the details of a scene as part of a real-time performance without any agent necessarily having any privileged knowledge about the story (though this is possible in certain improv games) in a system called *Party Quirks* (Magerko, Dohogne, and DeLeon 2011) and are currently modeling how improvisers establish the *platform* (i.e. the introductory details about what characters are in the scene, where they are, and what they are doing together) in a game called *Three Line Scene* (O'Neill et al. 2011). *Party Quirks,* based on the real-life improv game of the same name, involves a party host who has three guests with previously assigned "quirks" (e.g. is a robot or is a pirate who is afraid of treasure) that the host has to guess during the scene. This game rarely involves story and is more focused on the representation and communication of character, which is why we focused on it as our first major system.

In terms of building a complete interactive narrative work, we have reasoned that building the platform should be our first major task in narrative co-construction before moving to the middle and conclusion sections of an improvised scene. Our most recent INT effort, *Three Line Scene* (also inspired by a real-life improv game), builds on our work in *Party Quirks* to enable an AI and a human to establish the platform of a scene based in the Old West (i.e. involving cowboys, bandits, gunfights, etc.). The nature of *Three Line Scene* is to establish the details of a scene within three lines to quickly and solidly get the platform agreed on so the scene can progress. Our future work will address other processes related to the co-creation of novel improvised stories, such as how the *tilt* (i.e. the main conflict) in a scene is negotiated and resolved (Brisson, Magerko, and Paiva 2011) and how conceptual blending is employed during performances to create new knowledge structures in the scene.

Scripts

Schank and Abelson argue that people use scripts to represent and navigate well-known situations (1977). Specifically, these scripts are a predetermined series of actions that define those situations, typically built from a person's experiences in those situations. While these scripts are malleable to the specifics of a situation (such as what food is ordered in a restaurant), the overall sequence and content of a script (how to behave in a restaurant) is rarely altered. In Schank's model, a script represents a causal chain. Actions early in the script explicitly enable the latter elements, and the scripts themselves may have preconditions before they can be retrieved and applied to a situation. The inference capabilities of humans allow us to recognize and apply scripts based on a small number of observed events. People may need to be able to recognize a ten step script based solely on the first and last events encoded in that script. In addition to telling us how to act in a given situation, scripts allow us to understand stories that others tell us. We use these scripts to fill in the gaps of a story when details are omitted. When such information is left out, we can assume that the omitted details happened according to the script. Finally, in Schank's model of scripts, two or more people in the same situation may operate from different scripts. For example, at a restaurant, a customer, a waiter, and a cook would all operate from separate scripts. Thus these scripts are tailored to a typical experience from a single perspective.

Orkin developed plan networks as a means of displaying collections of pathways through a given scenario (2007). Orkin sought to learn the common interactions between a customer and a waiter in a typical restaurant scenario. He observed thousands of interactions between players in a virtual environment known as *The Restaurant Game*. Players were tasked with acting as either a customer or a waiter, and each interaction between players represented a new plan in the network. Orkin visualized these plan networks as directed graphs, where each node was a discrete event and a directed edge indicates that one event immediately followed another in one or more of the observed interactions. With a large enough number of observations, any individual path through the plan network graph can be seen as a valid interaction.

While plan networks and Schankian scripts both aim to describe the typical behavior in a common interaction, plan networks model the behavior of all parties involved, as opposed to the Schankian approach which takes a single perspective to the interaction. Additionally, plan networks focus on the temporal sequence of events and ignore the issue of causality. However, plan networks do allow for an understanding of multiple pathways through a scenario, unlike Schank's model of scripts. This feature of having multiple paths has mapped well onto our formulation of genre-specific scripts for improv theatre, as described in the next section, and heavily contributes to our formulation of background knowledge of improv actors.

Co-creation in Interactive Narrative

Co-creation has been sparsely applied in interactive narrative systems. Co-creation is closely related to the concept of *agency*, which has been described as the impression a user has of how much control they have in a story (Thue et al. 2010). Co-creation refers to the actual generation of content in a story; in other words, the amount of co-creation in an experience is related to how much of a scene is built on elements that were introduced in the scene as opposed to being pre-authored. Co-creation, therefore, depends heavily on procedural definitions of story creation.

Procedural representations of story creation in interactive narrative (as opposed to drama management techniques) are not commonplace in the field. One particular system of note, Fairclough's *OPIATE*, attempted to procedurally represent Propp's functions from Propp's formal analysis of Russian folktales (Fairclough and Cunningham 2004). This allowed the system to recognize when a situation matched the conditions for a function and allow it to dynamically assign roles to characters, plot elements to be instantiated, etc. While this system represented a procedural set of rules that was both heavily restricted to a particular domain (Russian folktales) and was not necessarily conducive to modern expectations of interactive narrative experiences (Tomaszewski and

Binsted 2007), it was a significant work in the exploration of procedural definitions for INT systems. OPIATE could only essentially involve the user in Russian folktales, but the story space was defined by the knowledge in the world plus the definitions for how to apply that knowledge; in other words, story elements were not concretely pre-authored beforehand. While this work may have suffered from an over-constraining story domain, it did create a precedent for authoring story knowledge in a procedural form that removed the computer from the privileged role it normally assumed in INTs and attempted to create a more open-ended, less specifically defined story space for the user to explore and contribute to. Other notable systems include Swartjes' improv theatre-inspired investigation of object creation (Swartjes 2010) and Zhu's representation of status in the domain of real-world interactive theatre (Zhu, Ingraham, and Moshell 2011).

We contend that this focus on procedurality in interactive narrative systems is one that has significant potential for the future of the field. In order to build such a system, however, we need to arrive at a clear understanding of how to represent the knowledge an agent will employ and the processes that will operate using that knowledge. We refer to the knowledge that is used by processes in a co-creative experience as *background knowledge*, which is described in the next section.

3. Background Knowledge for Improvisation

Fuzzy Concepts

If agents are going to improvise together, they need to be able to refer to similar story constructs during improvisation – just like in any kind of collaboration or conversation. This requires agents to have significant enough overlap in their knowledge base – before the scene begins – to have anything sensible to say to each other. As mentioned earlier, we have worked on a system that constructs the platform (i.e. initial details) of a scene. Our initial work focused on how to formally represent the *character prototypes* (Magerko, Dohogne, and DeLeon 2011) that improvisers employ and how those prototypes are physically communicated on stage (e.g. a pirate taking a swig from a bottle of rum). This work has subsequently been extended to cover the major elements of scene platforms: character, location, and joint activity (Sawyer 2003).

Our main formalism for representing knowledge in this framework has been inspired by Lakoff's *prototype theory* (1989) and the corresponding subfield in logic known as *fuzzy logic*. Prototype theory suggests that we have shared cultural constructs that describe elements of our world (e.g. tables, superheroes, puppies, etc.). These constructs (prototypes) are not easily expressed in Boolean logic; tables are not *always* made of wood and superheroes do not *always* wear capes. Rather, prototypes are described as having degrees of memberships in different categories (e.g. superheroes have a strong, but not 100% membership in the category *wears capes* because not all superheroes wear capes, though many do). We refer to *degrees of association* as a bidirectional degree of membership (e.g. *pirates* are associated with *peglegs* strongly and vice versa). We have found that this epistemological theory fits very well with our data collected on human improvisers (Magerko, Dohogne, and DeLeon 2011). Fuzzy logic is a representation that affords exactly this kind of relation between knowledge and categories. Elements are described as having degrees of membership (DOM) to each set in the world. For example, *superhero* would have a degree of membership in *wears cape*, *made of wood*, *eats spinach*, and any other set that is included in our world state.

We use the above formalism for describing prototypes in the platform for an improvised scene. As shown in Figure 2, we have relationships between the gestural **M**otions performed by an agent or human improviser (via a Microsoft Kinect interface) and the semantic **A**ctions that those motions could represent. For example, waving your hand in the air could be strongly associated with the *saying hello* set, medium with the *dancing* set, and close to 0 for the *bandaging a wound* set. Actions have associations with **C**haracters (e.g. *bandaging a wound* would be highly a member of the *doctor* set and perhaps medium for *pirate*) and **J**oint **A**ctivities (e.g. bandits are highly associated with the *robbing bank* activity). These different sets of DOM values, as shown in Figure 2, can be used to infer new knowledge from a gestural input, to scene elements, to finally an output entailed by the new scene knowledge that has been inferred (e.g. seeing the other actor point their hand → they are pointing a gun → cowboys point guns, so perhaps they are a cowboy → bandits are in

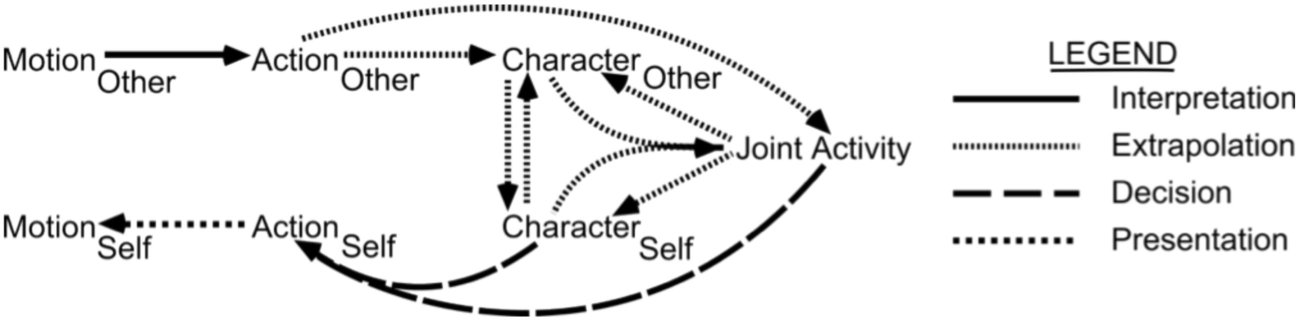


Figure 2. A depiction of the knowledge involved in reasoning about platform. Each arrow represents a table of degree of associations between the two sets. For example, all actions have degrees of associations with all characters. Therefore, if the Other actor executes an action, the Self agent (the one going through this thought process) can entail possible characters the Other may be portraying based on the degree of association values.

scenes with cowboys → I am a bandit → bandits also point guns → I should make the *gun pointing* motion and say "Reach for the sky!"). This process is described in more detail in (O'Neill et al. 2011).

While our fuzzy representation has worked well in the improv systems we have built so far, we quickly found during our design of *Three Line Scene* that there is a major issue with using fuzzy logic to represent one particular aspect of a scene's platform: joint activities. Joint activities (JAs) do have DOM associations with other scene elements (e.g. *bandits* are highly associated with *robbing a bank*), so having a fuzzy representation as part of the way we describe JAs makes sense. However, JAs also have a decomposition that needs to be observed; in other words, a joint activity like *robbing a bank* can actually be decomposed into multiple actions. Furthermore, these multiple actions are temporal in nature. In the *robbing a bank* joint activity, the bandit should not leave with the money before he says "Stick 'em up!" to the banker. While decomposition could be encapsulated by DOM values (i.e. the actions in a joint activity are highly associated with that activity), temporality cannot. Therefore, we have introduced a second formalism into our architecture, as shown earlier in Figure 1: scripts.

Scripts for Improvisation

For the purposes of our *Three Line Scene* system, we have focused on an Old West domain, which has definable genre characteristics (e.g. has cowboys, gunfights, saloons, etc.) and was deemed a large enough story space to allow for interesting improvised scenes without being too large (i.e. untractable) or too small. We identified typical joint activities by watching canonical Old West. Once we had selected our set of joint activities, we re-watched relevant scenes from the films and listed the pertinent actions in the joint activity. We asked multiple people to watch the same scene and then cooperatively authored scripts from these lists of events to build a corpus of script information about Old West stories.

We considered crowdsourcing approaches for identifying Old West joint activities and building the scripts. Crowdsourcing can give a good set of responses, but setting the problem up for such an approach can be cumbersome. Orkin was able to crowdsource typical restaurant interactions, but doing so required thousands of interactions in a virtual environment (2007). We considered asking people to build scripts using Amazon Mechanical Turk, but we only would have been able for the next event at any given time, rather than the whole script. We quickly became dissatisfied with the time and complexity requirements for collecting scripts from the crowd and opted for the lower cost option of mining genre examples for script information instead.

As suggested earlier, we represent scripts using a modified form of Orkin's plan networks. These plan networks allow scripts to be represented as a collection of possible sequences of discrete events. Each network is represented as a directed graph, where nodes represent individual events and arcs connect events to other events that could potentially occur next. Each node in the script structure contains information about the specific action, what character the improviser is portraying, and what other objects or characters are involved in the action. Additionally, nodes that could potentially be the first or last event in a script are tagged as such. Arcs only represent possible successors -- no assumptions can be made about causal relationships between events whether or not they are connected by an arc.

Cycles are permissible in plan networks. For example, in a *bar* script, one could imagine returning to an earlier point in a script after finishing a drink in order to order a new one. However, it is often possible for cycles to exist in a network that stop making sense if repeatedly traversed. In a *Western shootout* script, a cycle may exist between two characters drawing their guns. This cycle exists so that either character may draw first. The

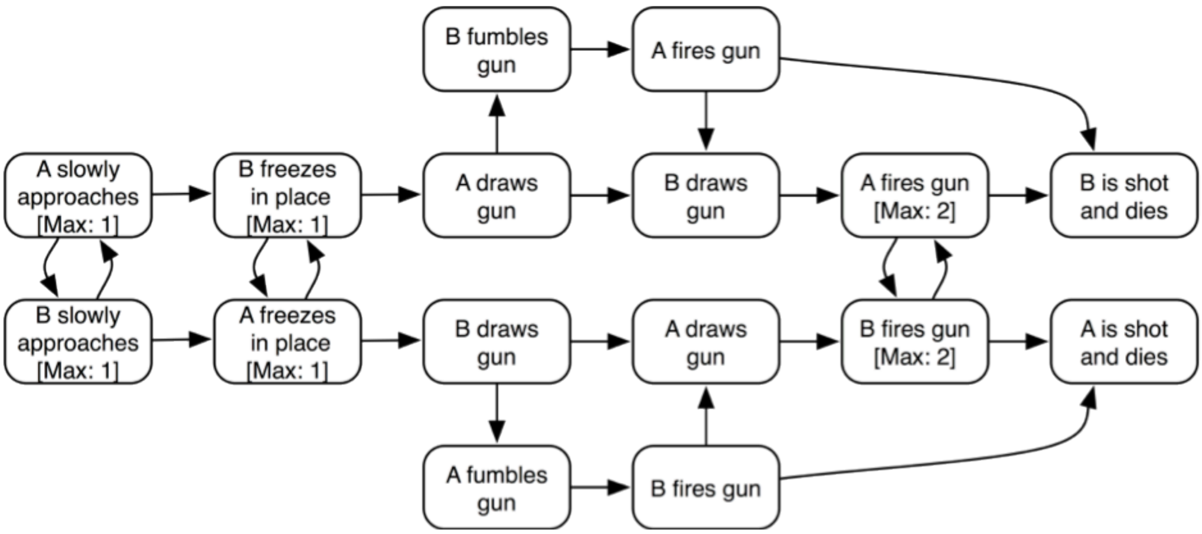


Figure 3. A script representation of an Old West shootout. The two characters described here, denoted as A and B, approach the shootout, and possibly draw and fire. Note the loop counters in the nodes representing each character entering the shootout, freezing, and firing their weapons. The two left-most nodes are tagged as acceptable starting points for the script, while the nodes representing a character getting shot are tagged as possible endpoints for the script.

consequence of such a cycle, unfortunately, is that both characters may repeatedly draw their guns. We have two approaches to managing and preventing these possibly infinite cycles. The first approach is the incorporation of "loop counters" in our representation of plan networks. Each node in such a cycle is tagged with the maximum number of times that it can be traversed. Our second approach is the separation of paths through the plan network so that there is no cycle. Separating the cycle into multiple paths allows us to enforce consequences to a specific order of events. For example, in the *Western shootout* script shown in Figure 3, each character drawing a gun has been separated into multiple nodes. In addition to preventing a character from drawing his gun multiple times, this approach further allows us to restrict who shoots first based on who drew first.

In the *Three Line Scene* system, we have a script for each joint activity in our knowledge base. There are two circumstances in a scene that require script retrieval. In the first, one of the improvisers believes he knows the joint activity in the scene. Each joint activity has only one script, so the script can easily be retrieved. In the other case, an improviser is trying to determine the joint activity (and the relevant script) based on the actions he has observed from the other improviser. Our *Three Line Scene* knowledge structure relates joint activities to individual actions that an improviser might take (as shown in Figure 2). Therefore, based on a particular action, an improviser can find one or more relevant scripts that may apply. The improviser can then apply further information from the scene so far to narrow the field (e.g., the characters that the improvisers are playing), or if no other information is available, choose a script to follow for the time being until confirming or conflicting information is presented.

Our background knowledge for joint activities is a hand-authored corpus that is comprised of a set of plan networks like the gunfight one illustrated in Figure 3. While this corpus is currently not especially large, it is sufficiently large enough for us to conduct the *Three Line Scene* project on platform establishment and informs us about knowledge authoring and representation issues for the future. This corpus, which is available from the authors to the academic community by request, is one that we intend to a) use as a knowledge base for our current *Three Line Scene* project; b) retain as a corpus of background knowledge for future, more complex improv agents; and c) continue to refine and augment to build a corpus that includes more genre-specific scripts as well as scripts from other genres that could be blended with the Western scripts as part of the improvisational process of creating new stories.

4. Discussion

This article has described how we as a research group have generated our own corpus of scripts to serve as background knowledge for an interactive narrative technology based on improvising scenes. This process has involved the hand authoring and peer reviewing of scripts compiled from genre examples in Western films. It would be an incredible boon to our work – and other projects in the INT field, undoubtedly – to have a pre-existing story bank that already had multitudes of genre scripts (and other platform elements) for us to rely on (Finlayson 2011). Such a story bank would allow us to move our efforts away from knowledge representations / background knowledge and focus more heavily on the procedural knowledge involved in improv (i.e. how to negotiate scene elements, how to computationally blend background knowledge with scene to create new story elements, etc.). However, our experience in authoring for *Three Line Scene* has encouraged some reflection on the concept of a story bank and how the idiosyncrasies of INT research projects may not fully benefit from such an effort without careful deliberation and awareness of the field.

The particulars of our knowledge representation (i.e. scripts and the fuzzy mappings illustrated in Figures 2 & 3) are not commonly used in other interactive narrative projects. As mentioned earlier, other systems rely on planning operators, beats, complete plans, story graphs, or Proppian functions to logically encode story elements (Roberts and Isbell 2007). As we have observed earlier, story representation in an INT is directly related to the affordances of the system for the AI involved (Magerko 2007b). In other words, what story representation is used in a system influences what the AI in that system can and cannot do. This has a direct relevance on the potential use of a story bank for INT research. If a particular AI-based approach does not map well to the affordances of the representation used in a story bank, then, best case, that approach necessitates its own story knowledge and cannot make use of a story bank, and worst case, that approach falls out of favor because it is inconvenient given the particular representation used in a story bank. We call this issue the *AI affordance problem*.

The affordance issue suggests that a general corpus or story bank should have malleable guidelines for representations. Rather than having a pre-defined logical representation, some initial core representation should be decided on which, in turn, can be added to in time with the needs of new projects coming to light. This could be designed with a decentralized (i.e. conventions agreed upon by the group of users) or centralized (i.e. a governing body reviews proposals for alterations / additions and formally agrees on new modifications) organizational mindset. Regardless of the approach, some intentional design in the governance of the representation used in a story bank needs to be taken to incorporate the myriad different logical approaches used in INT research and ones that are yet developed. A corpus needs to be as nimble and adaptive as the field using it, else it may either fall into disuse or bias the field towards the representational status quo.

A second issue with a general corpus for interactive narrative research is the nature of subjectivity in even the simplest story elements. When developing our earlier improv system, *Party Quirks*, that focused on representing and communicating character prototypes, we found very quickly during user testing that our hand authored prototype information rarely directly matched with our users' background knowledge. In other words, the *subjectivity problem* is reflected in how our conceptualization of character prototypes did not reflect everyone's view of the same prototypes.

This observation forced us to reconsider how we were authoring data about prototypes. Rather than rely on hand authoring, we opted for crowdsourcing our character prototype information. We created tasks on Amazon Mechanical Turk that allowed us to get a large amount of data fairly quickly about how strongly / weakly associated each of our character prototypes were for each of the possible actions in the world. We have employed this process again in the development of *Three Line Scene*, creating a separate crowdsourcing task for each connection (except **M**otion → **A**ction) seen in Figure 2. This process has allowed us to build a sizeable dataset that represents the views of a much larger population than our research lab in terms of the degrees of association between prototypes in an Old West story world. That dataset is then probabilistically sampled by our intelligent agents at the beginning of a scene to create a unique actor with background knowledge that is drawn from the particular views of our crowdsourcing subject pool.

Our crowdsourcing solution is one potential way to address the subjectivity problem. Whether it is with crowdsourcing or some other approach, the subjective nature of stories (e.g. the affect of a scene, what the definition of a prototypical character is, what the theme of a story was, etc.) need be captured to fully represent story elements as viewed by potential observers. If a corpus only intends to capture the typically non-subjective aspects of stories (e.g. the occurrence and ordering of events) then this issue may be avoided, though it will be decidedly sparse and, in some cases, non-subjectivity of content may be difficult to agree upon without getting data from an outside population anyway.

Our work on the *Digital Improv Project* is intended to serve as an exemplar of the kinds of research in interactive narrative technologies at present that are directly related to the collection of logical representations of stories. Our particular representation is decidedly different from those in other systems, but none are proven to be the absolutely correct formalism. There is potential for interactive narrative systems to be bootstrapped enormously with access to a large body of knowledge about stories (Gordon and Swanson 2009; Yu and Riedl 2012). However, if we as a community intend to collaborate on such an effort, we need to keep in mind issues like the affordance and subjectivity problems to develop a tool that helps our work in the future and reflect the variety of approaches used at present.

5. Acknowledgements

We would like to thank all of the members of the ADAM Lab at Georgia Tech who have put their hard work into the *Digital Improv Project* effort, including Chris DeLeon, Peter Dohogne, Joshua Faubel, Daniel Fuller, Rania Hodhod, Jonathan Pelc, Andreya Piplica, Matthew Postema, Mark Riedl, Dannielle del Rosario, and Emmy Zhang. This work was funded by NSF grants: IIS 0757567, 1036457, 0929326, 1037711, 1127359, 0840122, and 1129840.

6. References

Baumer, A., and B. Magerko. 2009. Narrative Development in Improvisational Theatre. In *Proceedings of 2nd International Conference on Interactive Digital Storytelling*, 140–151. Guimarães, Portugal.

———. 2010. An Analysis of Narrative Moves in Improvisational Theatre. In *Proceedings of 3rd International Conference on Interactive Digital Storytelling*. Edinburgh, Scotland.

Brisson, A., B. Magerko, and A. Paiva. 2011. Tilt Riders: Improvisational Agents Who Know What the Scene is About. In *Proceedings of 11th International Conference on Intelligent Virtual Agents*, 35-41. Reykjavik, Iceland.

Cavazza, M., F. Charles, and S. J Mead. 2002. "Character-based interactive storytelling." *IEEE Intelligent Systems*, 17 (4): 17–24.

Fairclough, C., and P. Cunningham. 2004. "A Multiplayer OPIATE." *International Journal of Intelligent Games and Simulation*, 3 (2): 54-61.

Finlayson, M.A. 2011. Corpus Annotation in Service of Intelligent Narrative Technologies. In *Proceedings of the 4th Interactive Narrative Technology Workshop at the Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*, Palo Alto, CA.

Flowers, A., B. Magerko, and P. Mishra. 2006. "Gamemasters and interactive story: A categorization of storytelling techniques in live roleplaying." *Futureplay*, London, Ontario.

Fuller, D., and B. Magerko. 2011. Shared Mental Models in Improviational Theatre. In *Proceedings of 8th ACM Conference on Creativity and Cognition*. Atlanta, GA.

Gordon, A., and R. Swanson. 2009. Identifying personal stories in millions of weblog entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA*.

Lakoff. 1989. Cognitive models and prototype theory. In *Concepts and conceptual development*, ed. Ulric Neisser, 63-100. CUP Archive.

Magerko, B. 2007a. "Evaluating Preemptive Story Direction in the Interactive Drama Architecture." *Journal of Game Development* 2 (3): 25-52.

———. 2007b. "A comparative analysis of story representations for interactive narrative systems." *Proceedings of Third Annual International Artificial Intelligence and Interactive Digital Entertainment Conference*, Palo Alto, CA.

Magerko, B., P. Dohogne, and C. DeLeon. 2011. Employing Fuzzy Concepts for Digital Improvisational Theatre. In *Proceedings of the Seventh Annual International Artificial Intelligence and Interactive Digital Entertainment Conference*. Palo Alto, CA: AAAI Press.

Magerko, B., C. Fiesler, and A. Baumer. 2010. "Fuzzy Micro-Agents for Interactive Narrative." *Proceedings of the Sixth Annual International Artificial Intelligence and Interactive Digital Entertainment Conference*. Palo Alto, CA.

Magerko, B., W. Manzoul, M. Riedl, A. Baumer, D. Fuller, K. Luther, and C. Pearce. 2009. An Empirical

Study of Cognition and Theatrical Improvisation. In *Proceeding of the Seventh ACM Conference on Creativity and Cognition*, 117–126. Berkeley, CA.

Mateas, M., and A. Stern. 2002. “A Behavior Language for Story-Based Believable Agents.” *IEEE Intelligent Systems* 17 (4): 39-47.

O’Neill, B., A. Piplica, D. Fuller, and B. Magerko. 2011. A Knowledge-Based Framework for the Collaborative Improvisation of Scene Introductions. In *Proceedings of the 4th International Conference on Interactive Digital Storytelling*. Vancouver, CA.

Orkin, J.D. 2007. Learning plan networks in conversational video games. MsC Thesis in Media Arts and Sciences Thesis, MIT Media Laboratory.

Roberts, D. L, and C. L Isbell. 2007. Desiderata for managers of interactive experiences: A survey of recent advances in drama management. In *Proceedings of the First Workshop on Agent-Based Systems for Human Learning and Entertainment (ABSHLE 07)*, Honolulu, Hawaii.

Sawyer, R. K. 2003. *Improvised dialogues: Emergence and creativity in conversation*. Ablex Publishing Corporation.

Schank, R.C., and R.P. Abelson. 1977. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Vol. 2. Lawrence Erlbaum Associates Hillsdale, NJ.

Swartjes, I. 2010. Whose Story is it Anyway? How Improv Informs Agency and Authorship of Emergent Narrative. Ph.D. thesis, University of Twente, Enschede, The Netherlands.

Thue, D., V. Bulitko, M. Spetch, and T. Romanuik. 2010. “Player agency and the relevance of decisions.” In *Proceedings of Third International Conference on Interactive Digital Storytelling*: 210–215, Edinburgh, Scotland.

Thue, D., V. Bulitko, M. Spetch, and E. Wasylishen. 2007. Interactive storytelling: A player modelling approach. In *Proceedings of Third Annual International Artificial Intelligence and Interactive Digital Entertainment Conference,* Stanford*, CA*.

Tomaszewski, Z., and K. Binsted. 2007. The limitations of a Propp-based approach to interactive drama. In *Intelligent Narrative Technologies: Papers from the AAAI Fall Symposium*. Arlington, VA.

Young, R. M, M. O Riedl, M. Branly, A. Jhala, R. J. Martin, and C. J. Saretto. 2004. “An architecture for integrating plan-based behavior generation with interactive game environments.” *Journal of Game Development* 1 (1): 51–70.

Yu, H., and M. O. Riedl. 2012. A Sequential Recommendation Approach for Interactive Personalized Story Generation. In *Proceedings of the 11th International Conference on Autonomous Agents and Multi Agents Systems*. Valencia, Spain.

Zhu, J., K. Ingraham, and J. Moshell. 2011. “Back-Leading through Character Status in Interactive Storytelling.” *Proceedings of Fourth International Conference on Interactive Digital Storytelling*: 31–36. Vancouver, Canada.

Detecting Story Analogies from Annotations of Time, Action and Agency

David K. Elson

Columbia University
New York City

Abstract

We describe the Story Intention Graph (SIG) as a model of narrative meaning that is amenable to both corpus annotation and computational inference. The relations, focusing on time, action and agency, can express a range of thematic scenarios and lend themselves to the automatic detection of story similarity and analogy. An evaluation finds that such detection outperforms a propositional similarity metric in predicting human judgments of story similarity in the Aesop domain.

1. Introduction

Narrative constructs are key to way we perceive, understand and reflect upon information (Bartlett, 1932). We understand stories and events in the context of previous stories we have heard and previous events we have experienced. This can be seen in the many allegories and metaphors that have a narrative basis—some government's austerity measures threaten to "kill the goose that laid the golden eggs," an overly zealous individual may "cry wolf" too many times, some particularly dangerous turn took us "out of the frying pan and into the fire."

An algorithm capable of finding structural similarities between stories can greatly assist us in our need to filter, search, and otherwise organize the many stories to which we are exposed on a daily basis, from news articles to fiction and personal communication. Much like a trained language model allows us to recognize n-grams as being more than the sum of their parts, a data bank of encoded stories would let us identify "narrative idioms" that recur and are likely to appear in future stories. To accomplish this, we need a symbolic model for representing narratives that is sufficiently formal to allow us to algorithmically detect meaningful analogies, yet general enough so that manual tagging of existing stories is feasible (for building the data bank) and automatic tagging is plausible. In a sense, we aim to accomplish automatically the type of structuralist analysis of similarities and trends that Propp performed on Russian folk-tales (Propp, 1969) and Bremond on French folk-tales (Bremond, 1970), using manual annotation as a bootstrap.

We describe the **Story Intention Graph** (or SIG) as a set of discourse relations and coreferent entities designed to meet these dual goals. The relations and entities can be manifested as node and arc types in a semantic network, and a particular instance of the SIG model, a "SIG encoding," represents a narrative as a connected graph. To use Formalist terms (Bal, 1997), the SIG captures the underlying sequence of story-world events (the *fabula*) as well as their selection and ordering in the surface rendering of the story (the *sjužet*). Unlike most prior models of narrative discourse that have been proposed for discourse annotation, this SIG has an emphasis on agency, encoding the links between an action and the intention of its agent (Bundgaard, 2007), between a goal-driven action and its outcome (van den Broek, 1988), between a goal and its subgoal or superordinate goal (Stein and Albro, 1996), between an event and an affectually impacted agent (Graesser et al., 1994), and more. We have used a custom software tool to collect a corpus of 70 SIG encodings, collectively called **DramaBank**.

This paper summarizes the SIG model (Section 3) and its utility in modeling a range of narrative tropes, then presents an approach that leverages it to find similarities between story encodings (Section 4). We compare and evaluate several approaches to finding narrative analogies before concluding in Section 5.

2. Related Work

The notion of diagramming narrative as a semantic network is sometimes seen in cognitive psychology (Graesser et al., 1991; Trabasso and van den Broek, 1985). Artificial intelligence originally saw narrative as emerging from scripts, plans, agent interactions or models of common sense (Cullingford, 1981; Wilensky, 1983). Story grammars were also in vogue for a brief period (Prince, 1973; Rumelhart, 1975; Mandler and Johnson, 1977). More recent work in semantic story understanding tends to employ first-order logic (Mueller, 2004; Mueller, 2006; Zarri, 2010) and other formal representations for plans and strategies (Hobbs and Gordon, 2005). Story generation presents its own unique challenges (Gervás et al., 2006) but can also use a planning framework (Riedl and Young, 2005).

Some recent studies have striven to find discourse patterns among stories statistically (Chambers and Jurafsky, 2008; Gordon and Swanson, 2009) or build classifiers that adopt Lehnert's (1981) notion of recombinable *plot units* as a discourse model (Appling and Riedl, 2009; Goyal et al., 2010; Nackoul, 2010). Our SIG model bears some similarities to Lehnert's plot units, but has a greater expressive range by adopting a "theory of mind" approach to literature

(Palmer, 2007). This emphasis on the internal states of discrete, intentional agents is also featured in Wiebe's (2005) model of private state frames, as well as Grosz and Sidner's (1986) model of speaker intention and a recent computational treatment by Chen and Fahlman (2008). Other recent work has adopted the theory-of-mind approach to reading a text, with its emphasis on epistemic differences between agents, in order to model real-life narratives (Löwe et al., 2009; Nissan, 2008). Our attempt to create a "DramaBank" of annotated narratives runs parallel to the "StoryBank" approach of Finlayson (2008); the latter focuses more broadly on many aspects of sentential-level discourse coherence.

The problem of detecting and generating analogies has a long history as well (see (French, 2002) for a review), but not traditionally in the narrative sense. When attempted (e.g., (Winston, 1980; Finlayson, 2009)), a method of narrative analogy detection is sensitive to the choice of representation used (Löwe, 2010), so the design of the representation is an integral aspect of any approach to analogy detection. As the current inquiry is no exception, we emphasize the SIG as a means for describing meaningful temporal and agentive relationships among stories.

3. Story Intention Graphs

The SIG is a constructionist model, in that it brings out coherence at both local and global levels: what events happen, when, why, and to whom. In each encoding, a discourse is connected to a representation of its meaning in a single, integrated graph.

In the encoding the discourse is divided up into fragments, typically of clause or sentence length. Each fragment is represented by a **Text (TE)** node. Text nodes are chained together by **followed by (f)** arcs so that the order of nodes in the chain reflects the order in which the fragments appeared in the original discourse (the "telling time").

Events that occur in the *fabula* of the story-world, as opposed to fragments of the story's telling, are represented as separate coreferent entities called **Proposition (P)** nodes. Text nodes connect to equivalent Proposition nodes with **interpreted as** and their order in the story *fabula* is also determined by **followed by** arcs.[1] Because of this dichotomy, the SIG can represent disfluencies in narration such as flashbacks when "story-world time" and "telling time" diverge.

The remaining "interpretative" nodes and arcs describe a reader's cumulative cognitive situation model (Zwaan and Radvansky, 1998) over the course of comprehending the entire narrative, including both content that is directly stated in the discourse and content that the reader infers. In this context, a proposition is represented by an **Interpretative Proposition (I)** node. A **Belief (B)** node acts as a frame, inside of which the content of other nodes is understood to be a state of the story-world in the mind of some particular, discrete **agent**. A **Goal (G)** node is similar to a Belief, except that the nodes and arcs inside a Goal frame are understood to be the state of the story-world as <u>desired</u> by the discrete agent. Agency frames—goals and beliefs—can be nested indefinitely to model theory-of-mind interpretations of narrative meaning (for instance, that Alice wants Bob to believe that Alice believes that Bob has some property).

P nodes connect to the interpretative frames and nodes through six arcs: **interpreted as (ia)**, **implies (i)** and **actualizes (ac)** are "actualizing," that is, indicating a positive functional relationship; **prevents/ceases (pc)** indicates a negative functional relationship; **attempt to cause (ac)** and **attempt to prevent (ap)** indicate agent <u>intention</u> to either trigger (actualize) or prevent/cease. The first four differ in their directness: *Interpreted as* indicates direct equivalence, *implies* indicates obvious entailment; *actualizes* indicates an positive but indirect causal relationship; *prevents/ceases* a negative, indirect causal relationship.

An example SIG encoding for part of the Aesop fable "The Wily Lion" (Table 1) is shown in Figure 1. Three TE

[1]This section summarizes the SIG schemata but, for brevity, omits and simplifies certain details such as a state-interval model of time. See (Elson, in review) for further details.

A Crow was sitting on a branch of a tree with a piece of cheese in her beak when a Fox observed her and set his wits to work to discover some way of getting the cheese. Coming and standing under the tree he looked up and said, "What a noble bird I see above me! Her beauty is without equal, the hue of her plumage exquisite. If only her voice is as sweet as her looks are fair, she ought without doubt to be Queen of the Birds." The Crow was hugely flattered by this, and just to show the Fox that she could sing she gave a loud caw. Down came the cheese, of course, and the Fox, snatching it up, said, "You have a voice, madam, I see: what you want is wits."
A Lion watched a fat Bull feeding in a meadow, and his mouth watered when he thought of the royal feast he would make, but he did not dare to attack him, for he was afraid of his sharp horns. Hunger, however, presently compelled him to do something: and as the use of force did not promise success, he determined to resort to artifice. Going up to the Bull in friendly fashion, he said to him, "I cannot help saying how much I admire your magnificent figure. What a fine head! What powerful shoulders and thighs! But, my dear friend, what in the world makes you wear those ugly horns? You must find them as awkward as they are unsightly. Believe me, you would do much better without them." The Bull was foolish enough to be persuaded by this flattery to have his horns cut off; and, having now lost his only means of defense, fell an easy prey to the Lion.

Table 1: "The Fox and the Crow" (top) and "The Wily Lion", from Jones (1912).

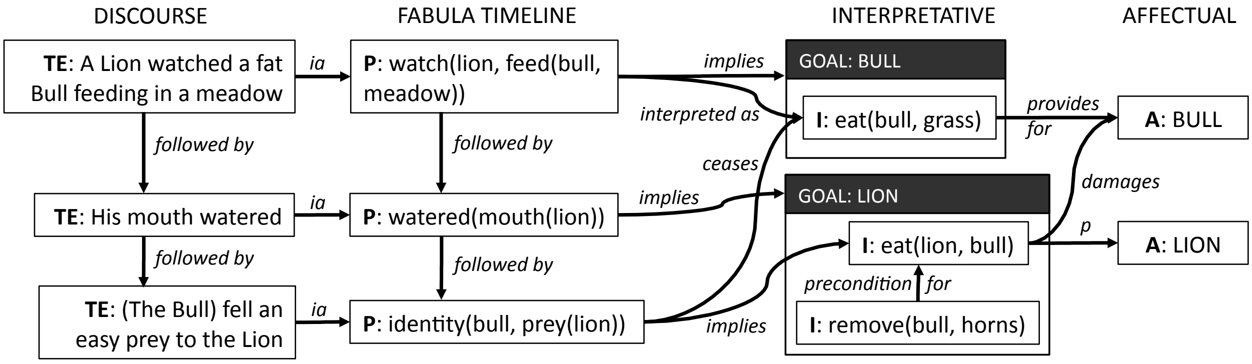


Figure 1: Example SIG encoding for a non-contiguous fragment of "The Wily Lion".

nodes contain text spans and are connected to three P nodes with equivalent propositions. There are two interpreted goals in this encoding: The bull has a goal to eat grass (his will to live is implied), and the lion has a goal to eat the bull. Note that frames themselves refer to mental states: The first sentence implies that the bull has a desire to eat grass, and directly asserts that it is, in fact, eating grass, such that he begins the story with a satisfied goal; later, the "goal content" of eating grass is ceased when the lion eats the bull. This event coreference—the same action is desired, achieved and then lost—forms the basis of the SIG approach to modeling narrative cohesion.

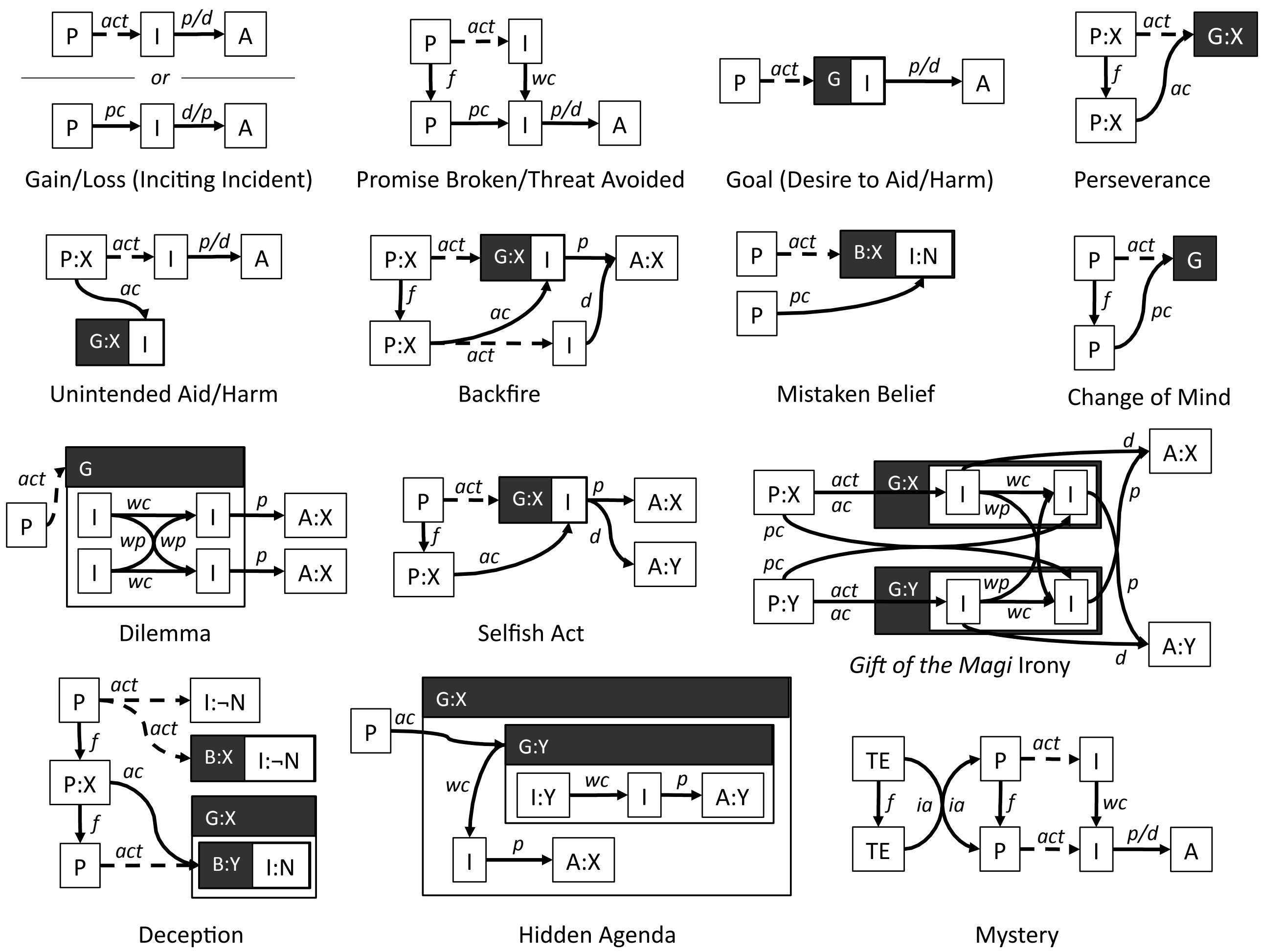
A plan is modeled as a chain of connected nodes inside a Goal frame. Each node is a "subgoal" that leads to the ultimate goal at the end of the chain. The connections are directed arcs that indicate causality, as expected by the agent associated with the frame. Specifically, a **would cause (wc)** relation traverses from one interpretative frame or proposition to another interpretative frame or proposition. It signifies that in the belief context of the originating node, an actualization of the originating node would causally lead to (is both necessary and sufficient for) an actualization of the destination node. **Would prevent (wp)** is its complement, signifying a belief that the actualization of the originating node would cause the destination node to be prevented/ceased. Two other relations, **precondition for (pf)** and **precondition against (pa)**, signify a belief that actualization of the originating node is necessary, but not sufficient, for the actualization or prevention/cessation (respectively) of the destination node. This schematic bears resemblance to a partial-order plan, the key difference being that a SIG plan is an annotator's interpretation of a narrated agent's intentions (Suh and Trabasso, 1993), rather than a solvable system (Riedl and Young, 2004).

Finally, the affectual impact of a P node or actualized I node can be indicated through the combination of an **Affect (A)** node (which indicates a particular agent) and a **provides for (p)** or a **damages (d)** arc (which indicates positive or negative impact, respectively). For instance, in Figure 1 the bull's eating of grass is intrinsically good for the bull, while the lion's eating of the bull is good for the lion but bad for the bull. In a properly formed SIG encoding, every goal is annotated with its affectual impact either through a direct arc or through a path through a plan. The bull's removing its horns, for instance, is indirectly good for the lion because it satisfies part of the lion's plan.

This set of nodes and arcs forms a basic vocabulary and syntax from which complex narrative structures can be constructed. This can be seen through the enumeration of "SIG patterns"—compound relations serve as fragments of abstract narratives. We can define a set of *a priori* SIG patterns to represent a range of narrative scenarios, in a manner similar to Lehnert's enumeration of plot units but with a greater emphasis on temporal and agentive (theory-of-mind) relationships. Notably, these patterns are defined only in terms of node and arc permutations, without any notion of particular propositional content within P and I nodes (except to identify the agent, such as P:X for agent X). We have identified 80 patterns, fourteen of which are shown in Figure 2, in several categories: affectual status transitions (e.g., gain, loss, mixed blessing), single-agent goals (problem, obstacle), outcomes (backfire, lost opportunity, recovery, peripeteia), beliefs (surprise, anagroisis, false dawn), dilemmas, two-agent interactions (selfless act, conflict, coercion, betrayal), persuasion and deception (mutual deception), time (flashback, suspense), mystery, and contradictory points of view. The dotted arc labeled *act* represents any of the arcs that actualize (*ia*, *i* or *ac*). Though an incomplete list of possible, thematically relevant patterns, they outline the range of what can be expressed by permuting these relations.

The DramaBank collection project, underway and publicly available,[2] elicits SIG encodings for stories in various genres from trained annotators. Each machine-readable record includes a reproduction of the source text as well as the nodes and relations of the annotator's encoding (serialized as first-order predicates). The collection includes

[2] http://www.cs.columbia.edu/~delson



Pattern	Example	Pattern	Example
Gain	John made a sale.	Change of Mind	Oscar briefly took up the violin.
Promise Broken	The train arrived, but skipped the station.	Dilemma	Betty wanted to be both a full-time chef and a full-time mom.
Goal	Mary dreamed of being published.	Selfish Act	Zach refused to give the old lady his seat on the bus.
Perseverance	Phil courted Megan for years.	*Gift of the Magi* irony	Della sold her hair to buy a chain for Jim's watch, but in the meantime, Jim sold his watch to buy Della a set of beautiful combs.
Unintended Harm	Lou's party, while fun, helped to spread a nasty flu.	Deception	Paul gave a check to the jeweler that he knew would bounce.
Backfire	Francis argued for a better grade, but annoyed his teacher into a deduction.	Hidden Agenda	The fox challenged the crow to demonstrate her singing ability, so that she would drop a piece cheese that the fox desired.
Mistaken Belief	It was clear out, but Yaël thought it was raining.	Mystery	Hillary jumped out of the burning building. She was performing a stunt for an action movie.

Figure 2: Fourteen examples of SIG patterns, compounded relations that represent common narrative scenarios.

60 encodings covering 25 of Aesop's fables (Jones, 1912), as well as 10 encodings covering 8 samples of longer and more varied narrative discourse: a news article (Wall Street Journal), literary short fiction ("An Alcoholic Case" by F. Scott Fitzgerald, "The Gift of the Magi" by O. Henri, and "The Lady with the Dog" by Anton Chekhov), contemporary nonfiction (an excerpt from *Sled Driver*, by Brian Shul), and epic poetry (*Beowulf* and *The Battle of Maldon*). For the 60 Aesop encodings, annotators supplied precise propositional content in P and I nodes according to a controlled vocabulary of nouns, verb frames, and modifiers (Elson and McKeown, 2009). For the longer and more complex texts, annotators constructed SIG encodings that only indicated an agent for each P and I node in order to accelerate the process. Further details on the collection process appear in Elson and McKeown (2010; 2012).

4. Analogy Detection

Given the SIG model as a representation, we define a narrative analogy between two narratives as a SIG encoding that is *covered* by part or all of the SIG encodings of two or more constituent stories. An encoding that covers a second encoding has a graph structure that isomorphic to a subgraph of the second encoding. For instance, if two encodings both feature an Proposition which is *interpreted as* a Goal frame containing an Interpretative Proposition, which *provides for* an Affect node, both encodings cover the "Desire to Aid" pattern in Figure 2, and thus the stories are analogous in that they both involve an abstract character with a desire to aid itself or another agent.

When two I or P nodes are found to be counterparts (analogous) within the isomorphism, we can also compare the propositions themselves using hypernym trees that correspond to each predicate and argument in the controlled vocabulary. As we describe in Elson and McKeown (2010), the analogous proposition would feature the least general predicates and arguments that are hypernyms to both of the constituent propositions. "A Lion watched a fat Bull," for example, would match "A Fox observed a Crow" with the generic "An animal perceives a second animal," and a scoring heuristic would judge this to be a fairly close match (a strong analogy). Our prior work used this technique alone, without any graph isomorphisms except for temporal sequencing, to find story analogies; here, we use this **propositional similarity** algorithm as a baseline approach.

This section ignores propositional similarities in exploring two methods for detecting story analogies based on isomorphisms alone: **static pattern matching**, a top-down approach, and **dynamic analogy discovery**, which is bottom-up.

4.1. Static pattern matching

In the previous section, we described a subset of 80 graph fragments we have identified that express common narrative scenarios in terms of SIG relations. As a top-down approach to finding analogies between two encodings, we can take these as *a priori* features for measuring analogical strength—more similar stories will have more SIG patterns in common.

The first step is to define and apply a set of closure operations that define transitive arcs that can be derived from the arcs made explicit by the annotator. For instance, an *attempt to cause* an event which would have a *positive* affectual impact on some agent should be equivalent to an *attempt to prevent* an event which would have a *negative* impact, as both are essentially attempts to effect a net positive change for the agent in question. The closures we have identified allow analogies to be detected despite minor variations in graph structure. Once the transitive arcs are in place, we use a theorem prover (Prolog) to determine whether either of the stories in question covers each pattern at least once, compile two vectors from these 80 features and calculate the cosine similarity between the vectors.

As a baseline check for the validity of this approach, we leverage the fact that DramaBank contains 60 encodings of 26 unique fables, including 40 homogeneous pairs of encodings (same source story, different annotators) and 1,015 heterogeneous pairs (different stories). We would naturally expect that the static similarity scores for homogeneous pairs be significantly higher than those for heterogeneous pairs—while we expect differences between parallel encodings of the same stories, given the subjective nature of story understanding and the flexibility of the SIG model, these difference should not exceed those between opposing stories. We do, in fact, find this to be the case: By the two-tailed Student's *t*-test, homogeneous pairs are more similar to $p<.001$. If either the metric had an unacceptable precision or recall for detecting analogies, or homogeneous encodings did not have measurable inter-annotator agreement, we would not see such significant results.

4.2. Dynamic analogy discovery

A third approach to analogy detection finds the largest isomorphic subgraph between two encodings in such a way that observes the semantic constraints of the model; in effect, this finds the most complete and detailed continuous chunk of overlap between two stories (limited, of course, to those overlaps which can be expressed by the model's relations).

We model our algorithm after the ACME model (Holyoak and Thagard, 1989) for finding analogies in connectionist networks. After applying the same transitive closure rules to each encoding, our approach first seeds a set of small "globs" that represent potential isomorphisms between two encodings, then grows each glob by following outgoing corresponding arcs to identify and add new analogous node pairs. In other words, if each node in a certain node pair connects to an unseen node via the same relation, the new nodes are paired. Each glob contains a *binding* which lists not only the discovered node pairs, but pairs of analogous agents as well—as the glob grows, the agent bindings must remain consistent for the analogy to be valid. If agent X in one story is bound to agent Y in a second story, the glob cannot expand to include a node pair in which X would bind to an agent other than Y.

The seeding process begins by considering all possible analogical node pairs among the interpretative nodes (goals, plans and beliefs). If there are multiple node pairs to which a singe glob can expand, the glob forks into two with each descendant taking a "route." To avoid intractable growth, aggressive memoization is used to avoid considering the same glob twice.

Once a glob has expanded to the point where no additional node pairs can be added, it determines which pairs of P nodes in the *fabula* timeline would be consistent with its binding, then adds as many P-node pairs as possible by using the Needleman-Wunsch (1970) alignment algorithm to find the longest path of node pairs that is internally consistent and compatible with the glob binding.

The result at this point is a set globs that relate to different parts of the agentive content (multiple disjoint isomorphisms). We combine as many as possible into a final analogy by examining each glob in descending order of size, and adding it to the largest glob with which it has a compatible binding. Thus our final result is a set of mutually incompatible analogical bindings that align not only timeline propositions, but agentive content found to be isomorphic between the two encodings. We give each glob a score by counting the relations, nodes and agents found to be analogous in its binding. The top-scoring (largest) glob becomes the top-line result—a dynamically generated isomorphic subgraph joining together two encodings.

Results

We have found this algorithm to return substantive analogies, as measured by the sizes of the isomorphic subgraphs that are found: 8.8 bound node pairs, 1.5 agentive bindings and 14.1 analogous relations on average (including inferred, transitive relations) across 1,015 heterogeneous encoding pairs in DramaBank. We also find again that homogeneous pairs yield significantly larger analogies than heterogeneous pairs ($p<.001$), more than 50% larger on average.

The largest analogies found in the corpus, by the number of bound node pairs, were between two particular encodings of "The Wily Lion" and "The Fox and the Crow". This is an initial check on our approach, as while we did not develop the dynamic analogy algorithm using this pair of encodings, we did select these two fables for inclusion in the collection in part due to their strong analogical connection. By drawing each bound node pair into a single compound node, we visualize this analogy as a single hybrid encoding in Figure 3. In this case, there are 11 aligned timeline propositions, two goal frames (one nested within the other as part of a four-stage plan), and two Affect nodes. The overall result is that "the fox is like the lion" and "the crow is like the bull"—in both stories, one is an inciting agent who devises a plan to have a victim devise and execute its own plan that would benefit the inciter. After some persuasion, the inciter's plan succeeds.

4.3. Evaluation Against Gold Standard

In order to evaluate whether we are finding meaningful analogies with each approach, we conducted an evaluation to determine the extent to which we can approximate human ratings of story similarity.

Using Mechanical Turk, we presented raters with each pair of Aesop fables among the 26 we collected, and asked them to rate the degree of similarity on a three-point Likert scale. Our prompt asked for "similarities about story structure and content, such as similarities in plots (what happens) and characters (desires and personality traits)." We presented each story pair to three annotators. The unanimous agreement on the Likert question was 46.3%, with another 50.4% of cases showing a two-to-one majority. To control for nonsense input (as is always a concern with Mechanical Turk), we identified and discounted those individuals whose rate of participation in unanimous agreement was less than 20%; this affected 3.9% of the total vote count. We took the arithmetic mean of ratings for each pair as its canonical similarity.

We then trained a linear regression model on 100 predictor variables separated into three sets, one for each of our three similarity metrics. Variables regarding propositional similarity included the number of overlapping propositions between the two encodings and the closeness of the overlaps. Each of the 80 static SIG patterns was included as a variable. For the dynamic analogy metric, we included various features relating to the largest detected analogy: number of node pairs, number of agent bindings, types of relations found, and so on. These distributions were normalized and fit against the similarity ratings using M5 attribute selection, and evaluated using cross-validation. We ran the evaluation for all combinations of variable sets to gauge the relative impact of each.

The results are shown in Table 2. Propositional overlap variables by themselves were weak predictors of story similarity ratings, as compared to the other two sets, with a Pearson correlation coefficient of only 0.06. The variables regarding static SIG patterns and dynamic analogies were highly influential by comparison, with correlations exceeding .20; the combination of all variables yields a model which correlates with similarity ratings at .33. This model makes progress toward the prediction of story similarity, with an F-statistic of $p<.0001$. The root-mean-square error is .19, compared to .20 for the model with only propositional predictors. In fact, we note that the model including

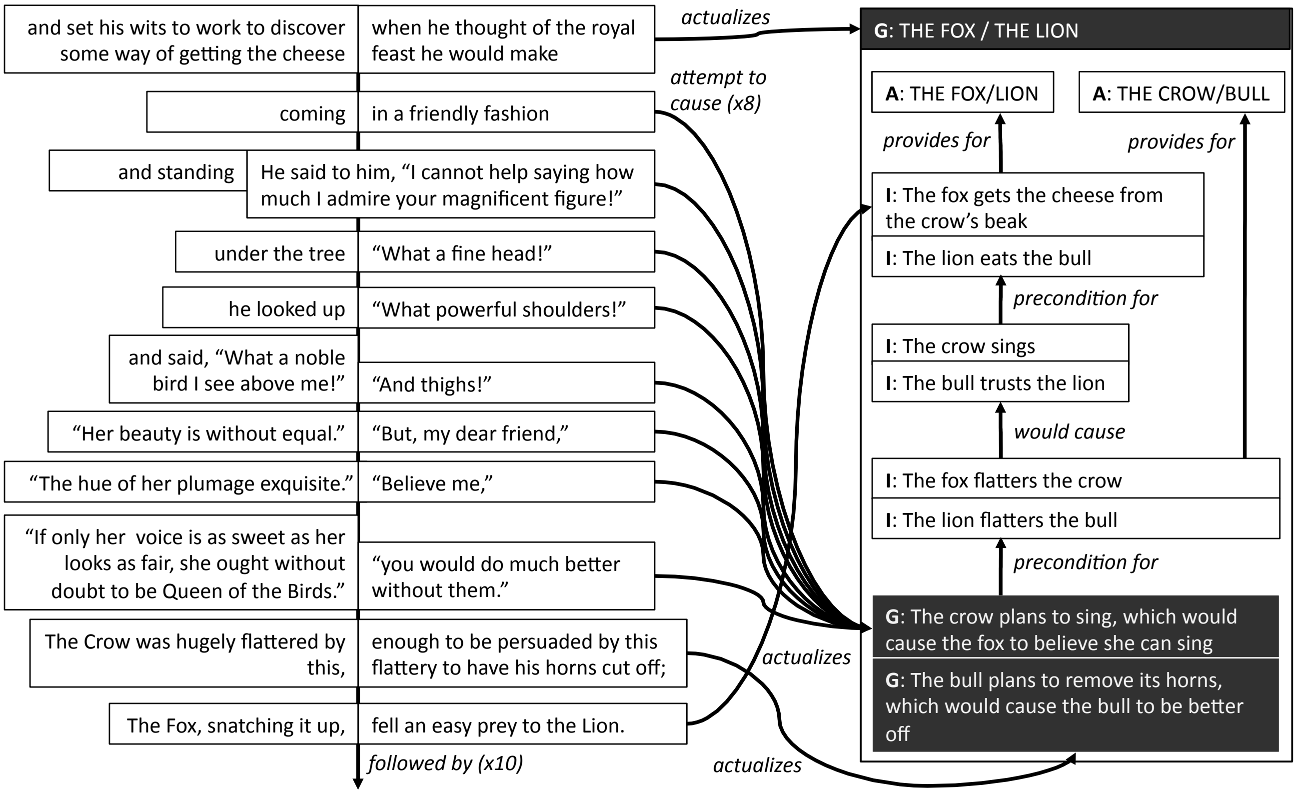


Figure 3: Analogy procedurally discovered between encodings of "The Fox and the Crow" and "Wily Lion".

Predictor Variable Sets	R-Square	RMSE	F-Statistic
Propositional (P)	0.0551	0.1986	p<.0191
Static (S)	0.2729	0.1923	p<.0001
Dynamic (D)	0.2117	0.1948	p<.0001
P+S	0.2724	0.1924	p<.0001
P+D	0.2174	0.1947	p<.0001
S+D	0.3257	0.1893	p<.0001
P+S+D	0.3299	0.1891	p<.0001

Table 2: Cross-validated performance of various linear regression models against story similarity ratings.

all but propositional predictors performed virtually as well as the all-inclusive model, as measured by both correlation coefficient and RMSE. Propositional modeling, while labor-intensive, did not provide helpful returns on the story similarity task.

The largest caveat of these results is the particularly lopsided distribution of similarity ratings—to most raters, nearly all story pairs had few to no appreciable similarities. Only 99 of these encoding pairs, less than 10%, were rated above 0.5. An increase in the amount of training data, or an expansion of the raters' notion of story similarity, would create a smoother distribution for training our models.

5. Conclusion

We have described a novel set of discourse relations intended to model narrative in a manner suitable for both corpus annotation and algorithmic treatment, for purposes of detecting tropes, similarities and analogies across multiple encodings. The SIG model, featured in a collection of 70 encodings of narratives in various genres, features not only narrated events, with their temporal and modal relationships, but coreferent entities: agents, goals, plans, beliefs, attempts, outcomes and affectual impacts, whether stated or implied. These pieces can be permuted to abstractly describe a range of common narrative scenarios. We also described three approaches to detecting analogies, and found that the top-down and bottom-up approaches that leveraged the model's structure outperformed a baseline of propositional similarity against human ratings of story similarity, suggesting that the SIG relations correspond meaningfully to the analogy retrieval task. In future work, they may also lend themselves to a generative model, trained on DramaBank encodings, of story *fabula* and its telling in discourse.

6. References

D. Scott Appling and Mark O. Riedl. 2009. Representations for learning to summarize plots. In *Proceedings of the AAAI Spring Symposium on Intelligent Narrative Technologies II*, Palo Alto, California.

Mieke Bal. 1997. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Toronto, second edition.

Frederic C. Bartlett. 1932. *Remembering: a study in experimental and social psychology*. Cambridge University Press, Cambridge.

Claude Bremond. 1970. Morphology of the french folktale. *Semiotica*, 2(3):247–276.

Peer F. Bundgaard. 2007. The cognitive import of the narrative schema. *Semiotica*, 165(1–4):247–261.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics (ACL-08)*, pages 789–797, Columbus, Ohio.

Wei Chen and Scott E. Fahlman. 2008. Modeling mental contexts and their interactions. In *Proceedings of the AAAI 2008 Fall Symposium on Biologically Inspired Cognitive Architectures*, Arlington, Virginia.

M. Cullingford. 1981. Sam. In R. Schank and C. Riesbeck, editors, *Inside Computer Understanding: Five Programs Plus Miniatures*, pages 75–135. Erlbaum, Hillsdale, New Jersey.

David K. Elson and Kathleen R. McKeown. 2009. A tool for deep semantic encoding of narrative texts. In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*, pages 9–12, Suntec, Singapore.

David K. Elson and Kathleen R. McKeown. 2010. Building a bank of semantically encoded narratives. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Malta.

David K. Elson. 2012. Dramabank: Annotating agency in narrative discourse. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey.

Mark Alan Finlayson. 2008. Collecting semantics in the wild: The story workbench. In *Proceedings of the AAAI 2008 Fall Symposium on Naturally-Inspired Artificial Intelligence*, Arlington, Virginia.

Mark Alan Finlayson. 2009. Deriving narrative morphologies via analogical story merging. In B. Kokinov, K. Holyoak, and D. Gentner, editors, *New Frontiers in Analogy Research*. NBU Press, Sofia.

Robert M. French. 2002. The computational modeling of analogy-making. *Trends in Cognitive Sciences*, 6(5):200–205.

Pablo Gervás, Birte Lönneker-Rodman, Jan Christoph Meister, and Federico Peinado. 2006. Narrative models: Narratology meets artificial intelligence. In *Proceedings of Satellite Workshop: Toward Computational Models of Literary Analysis, 5th International Conference on Language Resources and Evaluation (LREC 06)*, pages 44–51, Genoa, Italy.

Andrew S. Gordon and Reid Swanson. 2009. Identifying personal stories in millions of weblog entries. In *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, San Jose, California.

Amit Goyal, Ellen Riloff, and Hal Daumé III. 2010. Automatically producing plot unit representations for narrative text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*, Cambridge, Massachusetts.

Arthur C. Graesser, Kathy Lang, and Richard Roberts. 1991. Question answering in the context of stories. *Journal of Experimental Psychology: General*, 120:254–277.

Arthur C. Graesser, Murray Singer, and Tom Trabasso. 1994. Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3):371–395.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Linguistics*, 12(3):175–204.

Jerry R. Hobbs and Andrew S. Gordon. 2005. Encoding knowledge of commonsense psychology. In *Proceedings of the 7th International Symposium on Logical Formalizations of Commonsense Reasoning*, pages 107–114, Corfu, Greece.

Keith J. Holyoak and Paul Thagard. 1989. Analogical mapping by constraint satisfaction. *Cognitive Science*, 13:295–355.

V. S. Vernon Jones. 1912. *Aesop's Fables: A New Translation*. Avenel Books, New York.

Wendy G Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science: A Multidisciplinary Journal*, 5(4):293–331.

Benedikt Löwe, Eric Pacuit, and Sanchit Saraf. 2009. Identifying the structure of a narrative via an agent-based logic of preferences and beliefs: Formalizations of episodes from csi: Crime scene investigation. In Michael Duvigneau and Daniel Moldt, editors, *Proceedings of the Fifth International Workshop on Modelling of Objects, Components, and Agents*, pages 45–63.

Benedikt Löwe. 2010. Comparing formal frameworks of narrative structures. In *Computational Models of Narrative: Papers from the 2010 AAAI Fall Symposium*, Menlo Park, California.

Jean M. Mandler and Nancy S. Johnson. 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9(1):111–151.

Erik T. Mueller. 2004. Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research*, 5(4):307–340.

Erik T. Mueller. 2006. Modelling space and time in narratives about restaurants. *Literary and Linguistic Computing*, 4.

David Douglas Nackoul. 2010. Text to text: Plot unit searches generated from english. Master's thesis, Massachusetts Institute of Technology.

Saul B. Needleman and Christian D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.

Ephraim Nissan. 2008. Nested beliefs, goals, duties, and agents reasoning about their own or each other's body in the timur model: A formalism for the narrative of tamerlane and the three painters. *Journal of Intelligent and Robotic Systems*, 52:515–582.

Alan Palmer. 2007. Universal minds. *Semiotica*, 165(1–4):202–225.

Gerald Prince. 1973. *A Grammar of Stories: An Introduction*. Mouton, The Hague.

Vladimir Propp. 1969. *Morphology of the Folk Tale*. University of Texas Press, second edition. Trans. Laurence Scott. Originally published 1928.

Mark Riedl and R. Michael Young. 2004. An intent-driven planner for multi-agent story generation. In *Proc. of the 3rd Int. Joint Conf. on Autonomous Agents and Multi Agent Systems (AAMAS 04)*.

Mark Riedl and R. Michael Young. 2005. Story planning as exploratory creativity: Techniques for expanding the narrative search space. In *Proceedings of the 2005 IJCAI Workshop on Computational Creativity*, Edinburgh.

David Rumelhart. 1975. Notes on a schema for stories. In D.G. Bobrow and A. Collins, editors, *Representation and Understanding: Studies in Cognitive Science*, pages 231–236. New York Academic Press, Inc.

Nancy L. Stein and Elizabeth R. Albro. 1996. Building complexity and coherence: Children's use of goal-structured knowledge in telling stories. In Michael G. W. Bamberg, editor, *Narrative Development: Six Approaches*. Erlbaum, Mahwah, New Jersey.

Soyoung Suh and Tom Trabasso. 1993. Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, 32:279–300.

Tom Trabasso and Paul van den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24:612–630.

Paul van den Broek. 1988. The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, 27(1).

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evalution (formerly Computers and the Humanities)*, 39(2-3):165–210.

Robert Wilensky. 1983. Story grammars versus story points. *Behavioral and Brain Sciences*, 6:529–623.

Patrick H. Winston. 1980. Learning and reasoning by analogy. *Communications of the ACM*, 23:689–703.

Gian Piero Zarri. 2010. Representing and managing narratives in a computer-suitable form. In *Representing and Managing Narratives in a Computer-Suitable Form*, Arlington, Virginia.

Rolf A. Zwaan and Gabriel A. Radvansky. 1998. Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–185.

Story Comparison via Simultaneous Matching and Alignment

Matthew P. Fay
mpfay@mit.edu

MIT Computer Science and Artificial Intelligence Laboratory,
32 Vassar Street
Cambridge MA, 02139

Abstract

Story understanding is an essential piece of human intelligence. If we are to develop artificial intelligence with the cognitive capacities of humans, our systems must not only be able to understand stories but also to incorporate them into the thought process as humans do. The techniques I present enable efficient gap filling through story alignment. The approach demonstrated leverages the solid foundation of bio-informatics alignment techniques to create the simultaneous matching and alignment algorithm for story comparison. The algorithm provides a large improvement in efficiency in solving the matching problem, reducing the search space from 10^{30} nodes to 535 nodes in an example narrative comparison. The technique enables effective story comparison as an important step towards enabling higher level narrative intelligence.

Keywords: Story comparison, story understanding, plot prediction, alignment, matching, stories, narrative

1. Vision

Story understanding is an essential piece of human intelligence. Stories exist in countless forms and varieties, all seamlessly integrated into every facet of our lives. Stories fuel human understanding of the world. Narrative acts as a cognitive Swiss army knife, simultaneously facilitating the transfer of knowledge, culture, and beliefs while also powering our high level mental faculties. If we are to develop artificial intelligence with the cognitive capacities of humans, our systems must not only be able to understand stories but also to incorporate them into the thought process as humans do.

To take story understanding to the next level, I focus on the problem of story comparison. People intuitively use story comparison to draw on old experiences to construct and understand new ideas. As an example, consider a student studying *Hamlet* for the first time. The student may discover acts of revenge and draw the connection to a previous viewing of Disney's *The Lion King*. Innately the student might predict the plot by using the events of *The Lion King* as a template. The human ability to perform analogy fuels gap filling enables generation new ideas from old ones. (Schank, 1990).

I present a novel computational story comparison method which allows for efficient and effective story comparisons with broad applicability. The simultaneous matching and alignment algorithm I demonstrate is capable of providing a polynomial time solution to an otherwise exponential time problem. For example, when working with a version of the *Macbeth* story, my algorithms give a decrease in processing required by reducing a search space from over 10^{30} nodes to only 546 nodes.

2. Foundations

For my work, I leverage the Genesis story understanding system, in development at MIT CSAIL. (Winston, 2011). The high level objective for Genesis is to provide a computational framework for story understanding with a focus on modeling how the human mind reasons.

Genesis contains a variety of knowledge representations. These include, but are not limited to, classes (Vaina and Greenblatt, 1979), transitions (Borchardt, 1994), trajectories (Jackendoff, 1985), goals, persuasion, social relations, and object properties. Each can be expressed easily both in English and in Genesis's representations (Winston, 2011). These representational knowledge types enable Genesis to understand a wide array of story level information.

The representational knowledge is encoded in semantic role frames in Genesis. These frames form the underlying structure for the story analysis that Genesis performs and will be referred to as *story events* for the purposes of this paper. Figure 1 shows visual examples of story events encoded in Genesis.

Genesis also supports a variety of higher level knowledge representations. Two of the most prominent are *commonsense* and *reflective knowledge*. Commonsense knowledge describes cause and effect connections between story events. Examples statements in English include "If XX is killed then XX is dead" and "YY is happy because YY won the lottery." Reflective knowledge encompasses common trends and themes that may occur in stories. The notions of "Revenge" and "Pyrrhic Victory" are good examples.

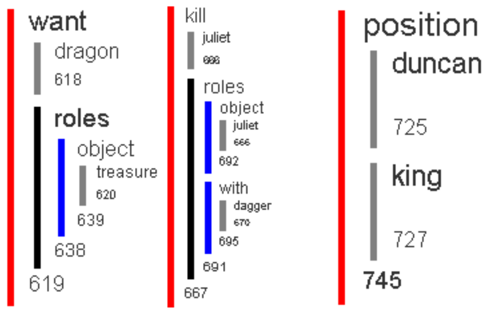


Figure 1: Representational knowledge about stories is stored in hierarchical structures in Genesis. Genesis generates these representations from English sentences. The three sentences shown are from left to right “The dragon wants treasure.”, “Juliet killed herself with the dagger.”, “Duncan is the king.”

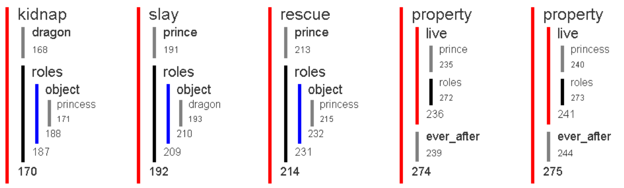


Figure 2: An example of a simple story understood by Genesis is, “The dragon kidnapped the princess. The prince slew the dragon. The prince rescued the princess. The prince and the princess lived happily ever after.” Genesis represents the story as the sequence of story events shown.

3. Story Comparison via Alignment

3.1. Stories as Sequences

The tools currently available in Genesis create a few potential choices for how to best represent a story. Using a library of common sense knowledge, Genesis creates a graph of how the story events relate to each other using causal rules. Another representation of the story is the set of higher level plot units that occur within the story, such as acts of revenge, that characterize the sets of interesting themes in the story. I focus on the event level elements for representing stories and enabling comparisons. In particular, I define a story to be a sequence of time ordered story events. Figure 2 shows an example of a story as a sequence.

A more complex graph representation of this story could offer increased information on how events in a story are connected and fit into the larger puzzle. For example, Genesis might know that kidnapping the princess gives the dragon control over her, and that killing someone allows the killer to take his victim’s things, which is why the prince acquires the princess. However, the additional information comes at a high cost. For example, one potential approach for doing story comparison would be to compare a causal graphs of all the events that occur in the stories. Unfortunately, graph comparison methods are NP-Hard (Cook, 1971). Because a goal for of my work is to enable rapid story comparison, this cost is prohibitive.

3.2. Sequence Alignment

Alignment algorithms have been researched heavily and have been proven to be powerful tools in the study of bio-informatics due to the enormous lengths and quantities of DNA and protein sequences that must be analyzed. By choosing to represent stories as sequences, I am able to leverage previous work from that field to have a solid baseline for my story comparison techniques.

The canonical algorithm used for the global alignment of pairs of DNA sequences is the Needleman-Wunsch algorithm (Morgenstern et al., 1998) (Needleman and Wunsch, 1970). The algorithm provides the best alignment between two sequences of maximum length n in $O(n^2)$ time while allowing for both insertions and deletions. Needleman-Wunsch is only able to do pairwise comparisons between sequences, which is somewhat limiting. However, it serves as a powerful baseline for moving sequence alignment into the story domain. Additionally, my work is transferable to other comparison methods. For example, later work could use the Smith-Waterman algorithm, which can do partial sequence alignment. (Smith et al., 1981) Other techniques which also allow for out of order element alignment and multiple sequence alignment will also be explored and build on this work.

3.2.1. Technical Review of Needleman-Wunsch

The inputs to the Needleman-Wunsch algorithm are two sequences, A and B, and a similarity matrix, S, that compares all element types that exist in the domain. The goal is to find an alignment between the sequences with a maximal alignment score where the score is calculated by summing over the similarity of each pair of elements in the alignment. Gaps are allowed anywhere in the alignment and are given a constant gap penalty d.

To find the optimal alignment, the algorithm first constructs a matrix F with each row representing ordered elements from the first sequence and each column representing ordered elements from the second sequence. Letting i be the row index and j be the column index, the values $F_{i,0}$ and $F_{0,j}$ are initialized to be $F_{i,0} = i * d$ and $F_{0,j} = j * d$ respectively. The rest of the elements of F are generated via the recursion:

$$F_{i,j} = max(F_{i-1,j-1} + S_{Ai,Bj}, F_{i,j-1} + d, F_{i-1,j} + d) \tag{1}$$

Once all elements of F have been generated, $F_{n,m}$, where n is the length of A and m is the length of B, represents the maximal alignment score. To obtain the alignment that gives this score, the algorithm walks through F starting at $F_{n,m}$ by moving either $(0, -1)$, $(-1, 0)$, or $(-1, -1)$. The walk ends once the algorithm arrives at $F_{0,0}$. Each horizontal or vertical step represents a gap in a sequence for the alignment. Each diagonal step represents a pair of matched elements in the alignment.

$X:$	A	G	A	C	T	A	G	T	Total
$Y:$	C	G	A	-	T	-	-	T	Score
Score:	5	10	10	0	10	0	0	10	45

Table 1: The alignment of the DNA sequences "AGACTAGT" and "TCGATT" is shown above. The first two rows show the sequences being aligned with dashes inserted corresponding to gaps in the alignment. The bottom row is the similarity between each pair of elements between sequences in the alignment.

3.3. Story Alignment

The Needleman-Wunsch algorithm works extraordinarily well for DNA where the number of types of elements is small and static. Moving into the story domain causes a great increase in complexity. Unlike traditional Needleman-Wunsch in the bio-informatics domain, story alignment requires a scoring metric that is able to compute the similarity of events on the fly due to the unbounded nature of the domain of story events. An additional issue that must be resolved is that of object and actor continuity, which will be described in in the *Matching Problem* section.

3.3.1. Scoring

A fast and effective method for calculating story event similarity is necessary to align stories via the Needleman-Wunsch algorithm. For the purpose of this work the similarity between two story events depends on the structure of a story event and the similarity of the objects in that structure.

Events in Genesis exist in a hierarchical structure. In Genesis, the sentence "The dragon kidnapped the princess" shares a common structure with "The prince rescued the princess," but not with "They lived happily ever after." As a general rule, the structure of an event depends on the presence of indirect objects, direct objects, and subjects. Parts of speech such as adverbs and prepositional phrases typically become properties of the objects within the event structure. The similarity of two events is calculated using the recursion in Equation 2. A_i represents a sub-event or an object of event A. If at any time, A and B do not have the same number of objects or sub-events, or the structure of A or B is different, the similarity is 0.

When two objects are compared, such as prince and dragon, the properties of those objects are examined. In Genesis, every object has a number of threads containing definitions and other properties of the object. (Vaina and Greenblatt, 1979). These threads contain definitional knowledge acquired from the story and from WordNet about the objects (Stark and Riesenfeld, 1998). A number of a possible threads can exist for any particular word, but Genesis has disambiguation capabilities that allow it to select the most pertinent definition. (Winston, 2011) The similarity functions compares the percent of matching qualities in the thread of one object with that of the other object. *TS* is a function that returns the number of shared qualities of the objects' threads. In Equation 2, the shared count is divided by $max(A_i, B_i)$ which returns the length of the longer thread of properties.

$$S(A,B) = \begin{cases} TS(A_i,B_i)/max(A_i,B_i) & A_i, B_i \in \text{obj} \\ \prod Sim(A_i,B_i) & otherwise \end{cases} \tag{2}$$

4. Matching Problem

The event similarity scoring metric alone is not enough to allow effective story alignment due to the matching problem. Consistency in the relations between objects in the stories is required so that alignments of stories to make sense. Table 2 provides an example of story alignment with poor matching. In this story the roles the objects are playing in the first story do not remain consistent throughout the alignment with the second story.

Story A	Story B
Mary has the ball.	Sally has the ball.
—	John has the gift
Mary gives the ball to Sally.	John gives the gift to Tim.
Sally has the ball.	Tim has the gift.

Table 2: Story alignment requires object consistency in order for the alignments produced to mirror human expectations. In this example, there is no matching of objects between stories. While the events align well at an individual level, the overall story is lost.

To correct this problem, I augment the scoring metric with a list of pairings between objects in the two stories. These relations act as a strong constraint. An object that has been paired to another object will score a 0 in similarity to any other object. This constraint forces the continuity of object relations between stories. However, the matching problem is that there are exponentially many different sets of pairings between the objects of two stories, so brute force search is infeasible.

4.1. Simultaneous Matching and Alignment Algorithm

The simultaneous matching and alignment algorithm is a powerful technique used to solve the matching problem. The algorithm uses a *match tree* data structure that is used to construct the search space of object pairings and search through them. While a full match tree would contain exponentially many nodes, the simultaneous matching and alignment algorithm significantly constrains the search space.

A match tree consists of a number of match nodes arranged in a tree. Each match node has 3 primary components: a list of objects, L_A, from a story, A; a list of objects, L_B, from a story, B; and a list of pairs of objects from A and B, L_{AB}. The tree initially starts with a single root node. The root node's the list of pairs, L_{AB} is empty and the two lists of objects, L_A and L_B, are the complete sets of objects from A and B.

The rest of the match tree is generated in phases. In each phase, a level of the tree is constructed. Each node that has no child nodes and has at least one object in either L_A or L_B will have a set of child nodes constructed. If both L_A and L_B have at least one object in them, then the first object from L_A is chosen as the current O_A. Then, in turn, each object from L_B is chosen as O_B. A child node is created with the parent's L_A, L_B, and L_{AB}. The pair (O_A, O_B) is added to the child's L_{AB}. The child node also has O_A and O_B removed from their respective lists L_A and L_B. This process of child node creation repeats until O_A is paired with each object in LB for the given parent node. O_A is also paired with a $null$ object. This pairing represents O_A not corresponding to any object from story B. In the case that one of L_A or L_B is empty only a single child node is created which pairs the the $null$ object with an object from whichever list is not empty. A node for which L_A and L_B are both empty is a leaf node and will have no children constructed. This process repeats through any number of phases until the entire tree has been constructed, and no more child nodes can be created.

Once the process is complete, each child node contains a unique set, L_{AB}, of relationships between the objects of stories A and B. Each leaf node is a unique and complete set of pairing between the objects between A and B. To determine which sets are the best, the alignment algorithm can be run using the match set from each leaf node. The sets of pairings that yield the highest alignment score would be the best match set for comparing the two stories. While this search would be over an exponential number of nodes, the simultaneous matching and alignment algorithm provides a method for directing the search and greatly improving efficiency.

The simultaneous matching and alignment algorithm constrains the construction of the match tree by directing the search towards nodes which exhibit the highest potential for good alignments. Whenever a new node is created, the two stories are aligned based on the current set of object pairings represented in that node. The score from the alignment algorithm provides a measure of how good of a match set is contained in the node. The key feature of the match tree is that a child node can only further constrain the alignment as compared to its parent. This guarantees that the alignment score of a parent acts as an upper bound on the score for its child nodes. Therefore, the simultaneous matching and alignment algorithm proceeds by only generating the child nodes of the current highest scoring node that has no children. If a leaf node is found with maximal score, that node is guaranteed to have the best possible match set for aligning the two stories.

The simultaneous matching and alignment algorithm is a fast technique for finding the best possible match set from the set of exponentially many possible match pairs. In many cases, the algorithm yields the exact solution in polynomial time. While in the worst case can take exponential time, this tends to occur for only stories which have no good alignments regardless of match set. In cases where a fast solution is required, an approximate solution can always be guaranteed to be found in polynomial time by discarding all but a set number of the highest scoring nodes after each round of node generation and scoring.

5. Results

5.1. Matching Efficiency

Stories can be aligned and compared rapidly and accurately using Simultaneous Matching and Alignment. The simple *give* example shown in Table 2 previously required a search of 353 nodes to find the proper match set. The simultaneous matching and alignment technique reduces the nodes needed to only 17. The efficiency of the algorithm is even more pronounced on more complex stories. One of the most illustrative examples is a comparison of two different versions of *Macbeth* from the Genesis story corpus. These renditions of *Macbeth* have approximately 35 entities taking part in 62 story events. Previously, the story comparison would require constructing over 10^{30} nodes to find the best match set. However, the same result can now be achieved with only a 546 node search.

5.2. Gap Filling

Story C	Story D
Mary has the ball.	John has the gift.
—	John gives the gift to Tim.
Sally has the ball.	Tim has the gift.

Table 3: Stories can be aligned even when gaps exist in the event sequences, because of the capabilities of the underlying alignment algorithm.

Story alignment can be used to make intelligent predictions about unknown events. New events can be imagined by comparing stories and inspecting their differences. Consider *Story C* and *Story D* shown in Table 3. In this case the stories align well despite the gap of missing information in *Story C*. The missing event can be generated from the story event from *Story D* that aligns with the gap. The event's objects can be replaced with the corresponding objects from the match set used in alignment to yield the event, “Mary gives the ball to Sally.” which is an appropriate way to fill in the gap in *Story C*.

6. Contributions

First, I designed an algorithm for story comparison through sequence alignment. The algorithm successfully leverages important research from across domains to achieve effective story alignment. Second, I developed the simultaneous matching and alignment technique to solve the matching problem, reducing an otherwise intractable search to polynomial running time. Next, I implemented the algorithms into the Genesis story understanding system to test the work on a variety of story understanding problems. Finally, a salient demonstration of the algorithm is that it reduces the search space of a problem with over 10^{30} nodes to only 546.

7. Acknowledgments

Research on *Genesis* has been supported, in part, by the National Science Foundation (IIS-0413206), the Office of Naval Research (N00014-09-1-0597), the Air Force Office of Scientific Research (A9550-05-1-0321), and the Defense Advanced Research Projects Agency (FA8750-10-1-0076).

8. References

G.C. Borchardt. 1994. *Thinking between the lines: computers and the comprehension of causal descriptions*. The MIT Press.

S.A. Cook. 1971. The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158. ACM.

R.S. Jackendoff. 1985. *Semantics and cognition*, volume 8. The MIT Press.

B. Morgenstern, K. Frech, A. Dress, and T. Werner. 1998. Dialign: finding local similarities by multiple sequence alignment. *Bioinformatics*, 14(3):290.

S.B. Needleman and C.D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453.

R.C. Schank. 1990. *Tell me a story: A new look at real and artificial memory.* Charles Scribner's Sons.

T.F. Smith, M.S. Waterman, and W.M. Fitch. 1981. Comparative biosequence metrics. *Journal of Molecular Evolution*, 18(1):38–46.

M.M. Stark and R.F. Riesenfeld. 1998. Wordnet: An electronic lexical database. In *Proceedings of 11th Eurographics Workshop on Rendering*. Citeseer.

L.M. Vaina and R.D. Greenblatt. 1979. The use of thread memory in amnesic aphasia and concept learning.(note 0).

P.H. Winston. 2011. The strong story hypothesis and the directed perception hypothesis. In *2011 AAAI Fall Symposium Series*.

Similarity of Narratives

Loizos Michael

Open University of Cyprus
loizos@ouc.ac.cy

Abstract

The task of *recognizing narrative similarity* is put forward as a concrete metric of success for machine narrative understanding. For this task, one seeks to determine which of two narratives is more similar to a third target narrative. As a first step towards building machines that achieve this goal, we investigate herein the notion of narrative similarity through a computational lens. We approach similarity as a balancing act between a listener's search for commonalities between stories, and an author's quest to guard a story's intended inferences.

Keywords: story, abstraction, intended purpose, authorial intentions, tug-of-war, elaboration tolerance, recognizing narrative similarity.

1. Prologue and Sneak Preview

Story telling lies at the heart of human communication, as evidenced by the widespread presence of stories across cultures (e.g., creation and flood myths), and the prominence of story reading and story understanding as educational objectives across the educational ladder. It is, by extension, a natural and desirable goal to investigate how to build artificial agents to understand stories, and how to pragmatically measure the extent to which they succeed in doing so.

A concrete yardstick for measuring progress and comparing different approaches is arguably beneficial for pushing research on Computational Modelling of Narrative (ComMoN) forward. We suggest the task of Recognizing Narrative Similarity (RNS) as such a natural metric. In the RNS task, a corpus of triples of stories is provided, with each triple designating a target story, and labeling one of the remaining two stories as being more similar than the other one to the target story. The success of a machine in the RNS task is determined by the degree of accuracy with which it is able to predict the obscured labels of new triples, after an initial training phase and access to a set of labeled triples.

As a first step in understanding how humans may label stories in the RNS task, or how machines may be designed to succeed in this same RNS task, we focus in this work on investigating what constitutes narrative similarity in a formal sense. We identify two main aspects of narrative that, we posit, play a critical role in determining narrative similarity:

(IP) The *intended purpose* of a narrative aims to capture its author's view, and determines constraints on what inferences are expected to be drawn from the narrative.

(AS) The *abstract structure* of a narrative aims to capture a listener's view, and determines inferences that are actually drawn at various levels of narrative abstraction.

A narrative's author may be viewed in a broad and generic way, as a placeholder for whomever or whatever determines the purpose of the given narrative: the actual author, typical humans, some narrator, the developer of an RNS task, etc.

Narrative similarity results, then, from an interplay between an author's intentions and a listener's interpretation of narratives, in what can be aptly paralleled to a balancing act or a tug-of-war game: the listener "pulls" to find a common abstraction between narratives; the author "pulls" to prevent over-abstraction from ignoring the narrative intentions. Our precise formulation reduces similarity to searching (for abstractions) and checking for satisfiability (of intentions), both of which have a long history in Artificial Intelligence.

In attempting to formalize in this work the RNS task and those aspects of similarity that are both computable and, we believe, appropriate for machines to employ when tackling the RNS task, we make certain working assumptions:

(W1) Stories transcend modalities (e.g., natural language, comic strips, video). Issues related to particular modalities are investigated by other communities, and it is our contention that the ComMoN community should not delve into those territories. Instead, here *we assume that every story is given in a formal representation that is, itself, the direct object of investigation.* Whenever a story is, nonetheless, given in natural language hereafter, this serves only as an aid for this work's readers, and should be treated as if non-present as far as the formal framework is concerned.

(W2) Encoding facts, actions, static and temporal laws, and commonsense knowledge has been at the center of the formal foundations of Artificial Intelligence since its beginning (McCarthy, 1959; Mueller, 2006). *These same types of knowledge and formal representations are arguably rather pertinent and sufficiently rich* for encoding narratives, background knowledge, beliefs, and ascribed intentions, as needed for interpreting stories and recognizing similarity.

(W3) Listeners use their background knowledge and beliefs when interpreting a story, and may ascribe intentions to the story's author on what inferences they are supposed to draw from the story. Building machines with access to appropriate background knowledge and beliefs, and with the ability to ascribe appropriate intentions is, undoubtedly, a big research problem. But even *assuming a black-box solution to this problem — as it is done herein — leaves a lot to be said* about how a machine may recognize narrative similarity. It is this latter problem that we attempt to tackle in this work.

(W4) Labeling triples for the RNS task is biased by the annotator's own background knowledge and beliefs, and way of ascribing intentions. *This bias does not nullify the RNS task's utility for the ComMoN community.* Similar issues are present in the task of Recognizing Textual Entailment

(RTE), yet that task has served well its purpose as a benchmark for NLP research, and as a pitch for attracting interest and funding (Dagan et al., 2005). As was the case for the RTE task, we expect that future research may show the RNS task to be amenable to a learning-theoretic semantics that resolves the need for human annotators (Michael, 2009).

2. When are Narratives Similar?

When comparing two narratives, one has more choices than classifying them as being either identically the same or distinct. Indeed, looking into the wealth of narratives that are available, it is not uncommon to find narratives that although superficial different, they are similar in certain respects. For instance, are these two narratives similar?

> *Alice called in sick that morning, hoping to finish some of the chores that had been piling on for some time now. Instead, she ended up breaking her wrist while saving an old lady from a speeding car, and spent the day at the hospital. Imagine her surprise when her boss called her that evening to see how she was doing... and to thank her for saving his mother!*

> *Bob and Charlie had not studied for the history quiz, and decided to skip class that day. Little did they know that the teacher had decided to give the quiz on the next day, and take the class for a surprise field trip, instead.*

On the one hand, the number of actors in the narratives is the same. Both narratives talk about people deciding to skip doing something they were supposed, or expected, to do on a daily basis. In both cases, their plan backfires, since the intended goal behind their decision is not fulfilled. And in both cases, the story ends with an element of surprise for the listener. In these respects, the two narratives are similar. On the other hand, the number and gender of the protagonists in the two narratives differs. In the first narrative Alice gets caught lying for calling in sick, whereas the second narrative does not seem to suggest the same for Bob and Charlie; presumably they were able to produce a seemingly valid excuse for missing school that day. Whereas Alice ends up knowing about her plan backfiring, Bob and Charlie do not. Furthermore, the first narrative ends up suggesting that Alice might even be somehow rewarded for her deed, despite her lie, whereas this does not seem to be the case for the second narrative. More subtle, and perhaps debatable, is the difference that Alice skipped work with the purpose of doing something presumably unrelated to work, while Bob and Charlie skipped school to avoid going to school. In these respects, the two narratives are dissimilar.

It should be clear that narrative similarity cannot be judged on a single universally-determined dimension or scale. Depending on the aspects of a narrative that one considers, two narratives may be said to be similar to some extent. In particular, similarity cannot be judged based solely on shallow features of narratives such as their length, the number of actors or events they contain, etc. Changing one of these shallow features might not affect the essence of the narrative, but it may also affect it critically in certain cases.

Two are the main aspects that we shall argue determine narrative similarity: intended purpose and abstract structure.

First, similarity should be judged with respect to some *intended purpose* of the narratives (cf. *(IP)*). If the main purpose of a narrative is, for instance, to create suspense by prolonging the resolution of a situation, then a second narrative with a conflicting purpose cannot be said to be similar to the first one. As in our example earlier, the purpose is not necessarily communicated through the narrative itself. We shall assume, therefore, that it is given externally for the purposes of investigating narrative similarity (cf. *(W3)*).

We shall be using the term “intended purpose” in the broadest possible sense, to capture the intentions of a narrative’s author of what is to be communicated. Thus, if a narrative includes a German actor, and its author intends to implicitly ascribe to the actor the hard-working nature often attributed to Germans, but not the generally taller stature they may have compared to Cypriots, then only the former of the statements should be included in the narrative’s purpose.

A narrative will generally have many purposes, differing in degree of importance (cf. Definition 10). In our last example, for instance, it might be more important to communicate that the actor is an adult human, than to communicate the actor’s exact working habits. If one were to give up, for whatever reason, one of these purposes, then the less important one would be given up. It would then be conceivable to talk about two narratives being similar to a less-than-perfect extent if they shared the most important, but not all, of their intentions (cf. credulous similarity in Definition 14).

Second, similarity should not be judged on the specifics (be that syntax, type of prose, length, etc.) of the narratives — unless, of course, those relate to the intended purpose of the narratives — but rather on an *abstraction* of the narratives that suffices to carry their meaning (cf. *(AS)*). In a typical cautionary tale, for instance, many details are not critical, and can be replaced by others. So, when narrating Aesop’s fable of the lion and the mouse, the two characters are, to some extent, placeholders for a mighty and feared one and a seemingly insignificant one. The characters could have easily been replaced with a shark and a sea bass, without essentially affecting the meaning of the story. And in that case, we would like to say that the two versions of the fable are similar. By the same reasoning, many of the details that the fable recounts are not critical. Whether the lion lies in a cave when first meeting the mouse, and whether it is captured by a net or some other type of trap, is immaterial.

We shall be using the term “abstraction” when a narrative is less specific than another along axes such as the following: First, the plot structure of the first narrative shall be a generalization of that of the second one. A story that spans three days, for instance, could be generalized into a story that spans a single period, without distinguishing what happened when within that period. Second, the events or facts of the first narrative shall be a subset of those of the second one. So, some aspects of a story can be omitted altogether (cf. syntactic abstraction in Definition 13). Third, the events or facts of the first narrative shall be generalizations of those of the second one. In natural language, for instance, the term “hypernym” is used for a word that describes a broader notion of which a second word is a type.

This is not to suggest that we shall be dealing explicitly with natural language and hypernyms in the sequel (cf. *(W1)*). Our aim here is to present a formal framework that accepts as input the choice of what constitutes a generalization of an event or fact (cf. semantic abstraction in Definition 13 and its dependence on a domain). Thus, it will be possible to apply the framework to specific settings, where, for instance, events are given in terms of images or sounds, and their generalization is determined by knowledge analogous to hypernyms but appropriate for images or sounds.

The two aspects that we have discussed, intended purpose and abstract structure, capture the two opposing forces acting on a narrative: The first force comes from the listener, who can be seen as seeking to view the structure of two narratives sufficiently abstractly so as to coincide. The second force comes from the author, who can be seen as seeking to limit the extent of abstraction by insisting that the intended purpose of each narrative is respected (cf. Definition 14).

3. A Computational Framework

We shall formally define in this section the necessary notions for identifying narrative similarity. In line with *(W1)* and *(W2)*, the framework follows a formal treatment of what constitutes a narrative, and how knowledge and intentions necessary to interpret a narrative are represented. Precisely in the spirit of *(W2)* was the development of earlier work (Michael, 2010a), which we shall employ and extend herein. We shall first start by recounting some basic definitions pertaining to narrative from that earlier work.

3.1. Background Definitions

We assume throughout that representations are based on a propositional language $\langle \mathcal{F}, \mathcal{A} \rangle$, comprising a finite set $\mathcal{F}$ of fluents, a finite set $\mathcal{A}$ of actions, and an implicit set of logical connectives and the entailment operator $\models$ of Propositional Calculus. A time-line $\langle \mathcal{T}, \preceq^t \rangle$ comprises a countable set $\mathcal{T}$ of time-points, and a total ordering $\preceq^t$ over it; henceforth, the non-negative integers with their natural ordering. To define a narrative, we start with the definition of a discourse: a partially-ordered collection of events and facts.

Definition 1 (Discourse)**.** *A* ***discourse*** *is a triple $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ comprising a finite set $\mathcal{C}$ of clauses of the form $\texttt{occurs}(A, S)$ and $\texttt{holds}(L, S)$, a finite set $\mathcal{S}$ of states, and an acyclic partial ordering $\preceq^s$ over $\mathcal{S}$, where $A \in \mathcal{A}$, L is a literal over $\mathcal{F}$, and $S \in \mathcal{S}$.*

Embedding a discourse into a time-line amounts to choosing specific time-points for the states in the discourse, so that the partial order of events and facts is respected.

Definition 2 (Embedding)**.** *An* ***embedding of*** *a discourse $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$* ***in*** *a time-line $\langle \mathcal{T}, \preceq^t \rangle$ is a set of clauses of the form $\texttt{occurs}(A, T)$ and $\texttt{holds}(L, T)$ that results by substituting a time-point in $\mathcal{T}$ for each state in $\mathcal{S}$, so that if the time-points $T_1, T_2 \in \mathcal{T}$ are substituted, respectively, for states $S_1, S_2 \in \mathcal{S}$ such that $S_1 \preceq^s S_2$, then $T_1 \preceq^t T_2$.*

A domain represents the constraints (cf. *(W2)* and *(W3)*) a discourse is expected to satisfy, according to some listener.

Definition 3 (Domain)**.** *A* ***domain over*** *a time-line $\langle \mathcal{T}, \preceq^t \rangle$ is a finite set $\mathcal{D}$ of clauses of the form $\texttt{occurs}(A, T)$, $\texttt{holds}(L, T)$, $\texttt{static}(\varphi)$, $\texttt{causes}(\varphi, L)$, where $A \in \mathcal{A}$, L is a literal over $\mathcal{F}$, φ is a formula over $\mathcal{A} \cup \mathcal{F}$, and $T \in \mathcal{T}$.*

A semantics to the clauses in a domain is given in model-theoretic terms, by insisting that actions occurrences and fact observations are respected, static constraints are satisfied, and causal change is brought about (Mueller, 2006).

Definition 4 (Model)**.** *Consider a domain $\mathcal{D}$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$. An* ***assignment to*** *$\langle \mathcal{T}, \preceq^t \rangle$ is a mapping M of each pair of $X \in \mathcal{A} \cup \mathcal{F}$ and $T \in \mathcal{T}$ to a truth-value $M(X, T)$. The truth-assignment over $\mathcal{A} \cup \mathcal{F}$ that is induced by projecting / restricting M to a given time-point $T \in \mathcal{T}$ is denoted by $M(T)$. A* ***model of*** *$\mathcal{D}$ is an assignment M to $\langle \mathcal{T}, \preceq^t \rangle$ such that for each $A \in \mathcal{A}$, each literal L over $\mathcal{F}$, each formula φ over $\mathcal{A} \cup \mathcal{F}$, and each $T \in \mathcal{T}$: (i) $M(T) \models A$ if and only if $\texttt{occurs}(A, T) \in \mathcal{D}$; (ii) $M(T) \models L$ if $\texttt{holds}(L, T) \in \mathcal{D}$, and $M(T) \models \varphi$ if $\texttt{static}(\varphi) \in \mathcal{D}$; (iii) $M(T+1) \models L$ if $M(T) \models \varphi$ for some φ such that $\texttt{causes}(\varphi, L) \in \mathcal{D}$; and (iv) $M(T+1) \models L$ if $M(T) \models L$ and $M(T) \not\models \varphi$ for every φ such that $\texttt{causes}(\varphi, \neg L) \in \mathcal{D}$. A domain $\mathcal{D}$ is* ***consistent*** *if there exists a model of $\mathcal{D}$.*

A narrative is a discourse compatible with a given domain.

Definition 5 (Narrative)**.** *Consider a domain $\mathcal{D}$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$. A* ***narrative w.r.t.*** *$\mathcal{D}$ is a discourse $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ that the union of some embedding of which in $\langle \mathcal{T}, \preceq^t \rangle$ with $\mathcal{D}$ gives rise to a consistent domain.*

To accommodate discourses not compatible with all of a domain, a domain may rank subsets of its clauses according to how plausible a narrative is considered when satisfying those clauses. Strict clauses, whose violation nullifies the plausibility of narrativeness, are include in all subsets.

Definition 6 (Default Domain)**.** *A* ***default domain over*** *a time-line $\langle \mathcal{T}, \preceq^t \rangle$ is a triple $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ comprising a domain $\mathcal{D}$ over $\langle \mathcal{T}, \preceq^t \rangle$, a subset $\Delta \subseteq 2^{\mathcal{D}}$ of domains over $\langle \mathcal{T}, \preceq^t \rangle$, and a transitive preference relation $\preceq^d$ over Δ.*

3.2. Interpreting a Narrative

What inferences follow given a narrative w.r.t. a domain? By Definition 5, a narrative has an embedding whose union with the domain is consistent; i.e., the union has at least one model. In general, there may be multiple embeddings that give rise to consistent unions, and each consistent union may have multiple models. Each model resulting from this process is a possible way to interpret the given narrative.

In the case of a default domain, there is an additional aspect of the process that may give rise to narrative interpretations: the sub-domain of the default domain w.r.t. which the narrative is to be understood. Not all sub-domains are eligible, however. First, some sub-domains will not give rise to consistent unions, no matter what embedding of the narrative is considered. Second, among the remaining sub-domains, some are less preferred than others, and the interpretations they give rise to should be excluded from consideration.

Definition 7 (Admissibility)**.** *Consider a default domain $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$. A domain $\mathcal{D}_1$ is* ***admissible for*** *a discourse $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$* ***w.r.t.*** *$\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ if: (i)*

$\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *is a narrative w.r.t. a domain* $\mathcal{D}_1 \in \Delta$*; and (ii) for every domain* $\mathcal{D}_2 \in \Delta$ *w.r.t. which* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *is a narrative, it holds that if* $\mathcal{D}_1 \preceq^d \mathcal{D}_2$ *then* $\mathcal{D}_2 \preceq^d \mathcal{D}_1$.

Each of the admissible sub-domains of a default domain is used to give rise to interpretations: the listener's inferences.

Definition 8 (Interpretation)**.** *Consider a default domain* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ *over a time-line* $\langle \mathcal{T}, \preceq^t \rangle$*. An* ***interpretation of*** *a discourse* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ ***w.r.t.*** $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ *is a model of the domain resulting by the union of an embedding of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *in* $\langle \mathcal{T}, \preceq^t \rangle$ *with a domain that is admissible for* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *w.r.t.* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$.

Each interpretation of a discourse assigns a truth-value to every fluent and action in the language $\langle \mathcal{F}, \mathcal{A} \rangle$ for every time-point in the time line $\langle \mathcal{T}, \preceq^t \rangle$. This assignment is consistent, in the sense of Definition 4, with both the discourse and some subset of the constraints that account for the listener's background knowledge and beliefs. Thus, interpretations can be used to draw inferences from a discourse according to the listener's own knowledge and beliefs.
A discourse is a plausible narrative if it can be interpreted, possibly with some of the listener's weakest beliefs being retracted (cf. Definition 5 where all beliefs are respected).

Definition 9 (Plausible Narrative)**.** *Consider a default domain* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ *over a time-line* $\langle \mathcal{T}, \preceq^t \rangle$*. A discourse* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *is a* ***plausible narrative w.r.t.*** $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ *if there exists an interpretation of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *w.r.t.* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$.

3.3. Intended Purpose

The intended purpose of a discourse is encoded in modal clauses of the form $\mathsf{intend}(\kappa)$, with κ a monotone propositional formula over clauses of the form $\mathtt{occurs}(A, S)$ or $\mathtt{holds}(L, S)$. Roughly, then, the intended purpose of a discourse is interpreted as encoding its author's intention that the listener infers some implicit part of the discourse, even if not explicitly stated somewhere (cf. *(IP)* and *(W3)*).

Definition 10 (Intended Purpose)**.** *An* ***intended purpose of*** *a discourse* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *is a triple* $\langle \mathcal{P}, \Pi, \preceq^p \rangle$ *comprising a set* $\mathcal{P}$ *of* $\mathsf{intend}(\kappa)$ *clauses, a subset* $\Pi \subseteq 2^{\mathcal{P}}$*, and a transitive preference relation* $\preceq^p$ *over* Π*, where each* κ *is a monotone propositional formula over clauses* $\mathtt{occurs}(A, S)$ *and* $\mathtt{holds}(L, S)$*,* $A \in \mathcal{A}$*,* L *is a literal over* $\mathcal{F}$*, and* $S \in \mathcal{S}$.

Much like a default domain, the intended purpose of a discourse determines a relative degree of importance of subsets of intentions, as manifested through the ordering $\preceq^p$.
The semantics of a set $\mathcal{P}_0 = \{\mathsf{intend}(\kappa_i) \mid i \in [n]\}$ of intentions is defined by first interpreting the set $\mathcal{P}_0$ as the single intention $\mathsf{intend}(\kappa)$ with $\kappa = \bigwedge_{i \in [n]} \kappa_i$, and assuming that κ is represented in a disjunctive normal form.[1] Each term of κ — corresponding to a set of $\mathtt{occurs}(A, S)$ and $\mathtt{holds}(L, S)$ clauses — shall be called an ***extension of*** $\mathcal{P}_0$. Each extension shall be viewed as an implicit part of a discourse. Since, in general, there may be many extensions, this may give rise to many implicitly-extended discourses.

[1]It is assumed that a unique DNF representation is implicitly prescribed by an intended purpose $\langle \mathcal{P}, \Pi, \preceq^p \rangle$ for each $\mathcal{P}_0 \in \Pi$.

Definition 11 (Discourse Extension)**.** *Consider a discourse* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *and an intended purpose* $\langle \mathcal{P}, \Pi, \preceq^p \rangle$ *of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$*. A discourse* $\langle \mathcal{C}_i, \mathcal{S}_i, \preceq^s_i \rangle$ *is an* ***extension of*** $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ ***under*** $\mathcal{P}_i \in \Pi$ *if* $\mathcal{S}_i = \mathcal{S}$*,* $\preceq^s_i = \preceq^s$*, and* $\mathcal{C}_i$ *is the union of* $\mathcal{C}$ *and an extension of* $\mathcal{P}_i$.

Observe that $\mathcal{S}$ and $\preceq^s$ already account for the states that may appear in the extensions of $\mathcal{P}_i$, since $\langle \mathcal{P}, \Pi, \preceq^p \rangle$ is assumed to be an intended purpose of the discourse.

Definition 12 (Intention Satisfaction)**.** *Consider a default domain* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ *over a time-line* $\langle \mathcal{T}, \preceq^t \rangle$*, a plausible narrative* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *w.r.t.* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$*, an intended purpose* $\langle \mathcal{P}, \Pi, \preceq^p \rangle$ *of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$*, and a set* $\mathcal{P}_i \in \Pi$*. Let* $\mathbb{M}$ *be the (non-empty) set of interpretations of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *w.r.t.* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$*. Let* $\mathbb{M}_j$ *be the set of interpretations of* $\left\langle \mathcal{C}_j, \mathcal{S}_j, \preceq^s_j \right\rangle$ *w.r.t.* $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$*, where* $\left\langle \mathcal{C}_j, \mathcal{S}_j, \preceq^s_j \right\rangle$ *is an extension of* $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *under* $\mathcal{P}_i$*.* $\mathcal{P}_i$ *is* ***satisfied by*** $\left\langle \mathcal{D}, \Delta, \preceq^d \right\rangle$ ***given*** $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ *if* $\mathbb{M} = \bigcup_j \mathbb{M}_j$.

The inclusion $\mathbb{M} \subseteq \bigcup_j \mathbb{M}_j$ states what one intuitively expects from satisfying a set of intentions: that every interpretation agrees with the given intentions, or put differently, that every inference drawn from a narrative and a listener's background knowledge and beliefs entail the given intentions. In formalizing this idea, the use of the union $\bigcup_j \mathbb{M}_j$ allows general intentions, including disjunctive ones.
The inclusion $\mathbb{M} \supseteq \bigcup_j \mathbb{M}_j$ states that intentions are satisfied in *all* possible ways, exactly as stipulated by the author. Thus, the listener should not draw stronger inferences than those intended. Therefore, if $\mathsf{intend}(\mathtt{holds}(F_1, S_1) \vee \mathtt{holds}(F_2, S_2))$ is in $\mathcal{P}_i$, and if all interpretations satisfy only the first disjunct, then the intention is not satisfied.
Definition 12 makes explicit the fact that a default domain may satisfy only a subset of the intentions associated with a narrative. Indeed, the listener, who's background knowledge and beliefs are encoded in the default domain, does not have access to the author's intentions, and the manner in which the former interprets a narrative might not agree with what the latter might have intended. *We emphasize that our aim is not to suggest that the listener is actively trying to satisfy the intentions of the author, but only to identify the extent to which this satisfaction happens serendipitously.*
In closing this section, we point out that our approach is able to make sense even of situations where an author has conflicting intentions for a story. In this case, the conflicting intentions can be understood as the author wishing for the story to be purposefully ambiguous, so that different listeners will draw different inferences. This can be accommodated through our framework's ability to group intentions into subsets (each of them internally conflict-free) and expect that at least one of these subsets will be satisfied.

3.4. Abstract Structure

In considering the requirement of a shared structure at some level of abstraction of two narratives to be declared similar (cf. *(AS)*), one could debate on what operations on a narrative lead to appropriate abstractions. Instead of considering a specific, and ultimately ad hoc, set of such operations, we suggest two general abstraction criteria: Syntactic abstraction results by dropping some of a story's events, facts, or orderings of those. Semantic abstraction results by changing a story to admit (set-theoretically) more interpretations.

Definition 13 (Abstraction)**.** *Consider a default domain $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$. A discourse $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ is an **abstraction** of a discourse $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ if any of the following conditions hold: (i) $\mathcal{C}_1 \subseteq \mathcal{C}_2$, $\mathcal{S}_1 \subseteq \mathcal{S}_2$, $\preceq_1^s \subseteq \preceq_2^s$ (**syntactic abstraction**); (ii) every interpretation of $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ is an interpretation of $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ (**semantic abstraction**); or (iii) there exists a discourse $\langle \mathcal{C}_3, \mathcal{S}_3, \preceq_3^s \rangle$ that is an abstraction of $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, and of which $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ is an abstraction w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$.*

Two points worth noting: First, semantic abstraction is domain dependent, since what might count for one listener as abstraction (e.g., replace "December" with "Winter") might not be so for another (e.g., an inhabitant of Australia). Second, any two narratives always have at least one common abstraction w.r.t. any default domain: the empty narrative.

3.5. Checking for Similarity

To check for similarity between two given narratives, think of each narrative as giving rise to a graph, with each vertex corresponding to a discourse, and each edge connecting a discourse to one of its abstractions. Each narrative is at the root of its associated graph. The graphs share those vertices that correspond to common abstractions of the narratives.

The tug-of-war view of narrative similarity that was put forward in Section 1. is reflected in the following process: The listener, on the one hand, is trying to move from the roots towards the leaves of the graphs, until a shared vertex is found. The author, on the other hand, does not allow the listener to do so unrestrainedly, on the grounds that the discourses corresponding to certain vertices of each graph do not satisfy the intentions of the narrative at the root of that graph. Only if a balance can be reached between the two competing parties, one can call the two original narratives similar. This intuition is formalized in the next definition.

Definition 14 (Absolute Similarity)**.** *Consider a default domain $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$, two plausible narratives $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle, \langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, and two intended purposes $\langle \mathcal{P}_1, \Pi_1, \preceq_1^p \rangle, \langle \mathcal{P}_2, \Pi_2, \preceq_2^p \rangle$ for the two plausible narratives, respectively.*
*$\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ are **sceptically similar w.r.t.** $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ if there is a common abstraction $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ of $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, such that $\mathcal{P}_1$ and $\mathcal{P}_2$ are satisfied by $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ given $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$.*
*$\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ are **credulously similar w.r.t.** $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ if there is a common abstraction $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ of $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, such that some member of Π_1 and some member of Π_2 are satisfied by $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ given $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$.*
*In either case, $\langle \mathcal{C}, \mathcal{S}, \preceq^s \rangle$ is said to **support** the similarity of $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$.*

Sceptical similarity exists when all intentions of both narratives are satisfied in some common abstraction. Credulous similarity relaxes this stringent requirement so that only the most important intentions of the narratives need be satisfied. If not even those are satisfied, then the narratives are not said to be similar, as none of their common abstractions can be considered representative of the original narratives. Note that absolute similarity is *elaboration tolerant*: If two narratives are sceptically or credulously similar, then elaborating them with more details preserves their similarity.

Beyond a direct comparison of similarity between two narratives, a check for comparative similarity of two narratives against a target one is roughly this: For each of the two narratives, identify a narrative that is an abstraction of that narrative and the target narrative. Identify the abstraction among the two that satisfies more important (according to $\preceq_0^p$ below) intentions of the target narrative, while satisfying at least the important intentions of its original narrative.

Definition 15 (Relative Similarity)**.** *Consider a default domain $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ over a time-line $\langle \mathcal{T}, \preceq^t \rangle$, three plausible narratives $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle, \langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle, \langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, and three intended purposes $\langle \mathcal{P}_0, \Pi_0, \preceq_0^p \rangle, \langle \mathcal{P}_1, \Pi_1, \preceq_1^p \rangle, \langle \mathcal{P}_2, \Pi_2, \preceq_2^p \rangle$ for the three plausible narratives, respectively.*
*$\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ is **weakly more similar than** $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ **to** $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ **w.r.t.** $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ if there is a discourse $\langle \mathcal{C}_{10}, \mathcal{S}_{10}, \preceq_{10}^s \rangle$ that supports the credulous similarity of $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ and $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, and there is a set $\mathcal{P}_{10} \in \Pi_0$ that is satisfied by $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ given $\langle \mathcal{C}_{10}, \mathcal{S}_{10}, \preceq_{10}^s \rangle$, such that the following condition holds: for every discourse $\langle \mathcal{C}_{20}, \mathcal{S}_{20}, \preceq_{20}^s \rangle$ that supports the credulous similarity of $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ and $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, and for every $\mathcal{P}_{20} \in \Pi_0$ that is satisfied by $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ given $\langle \mathcal{C}_{20}, \mathcal{S}_{20}, \preceq_{20}^s \rangle$, it holds that $\mathcal{P}_{10} \not\preceq_0^p \mathcal{P}_{20}$.*
*$\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ is **(strictly) more similar than** $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ **to** $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ **w.r.t.** $\langle \mathcal{D}, \Delta, \preceq^d \rangle$ if $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ is weakly more similar than $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ to $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, but $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$ is not weakly more similar than $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$ to $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$ w.r.t. $\langle \mathcal{D}, \Delta, \preceq^d \rangle$.*

One may easily see certain properties of relative similarity: *(i)* It is possible that neither of two stories is more similar to the target story! This is the case if the two stories are, in some sense, too far away from the intended purpose of the target story. *(ii)* The story among the given two that is more similar to the target story is necessarily credulously similar to the target story. This property is essentially the contrapositive of, and explains, the previous property. *(iii)* A story that is sceptically similar to the target story is weakly more similar than any story to the target story; i.e., no story can be more similar than the former story to the target story.

4. Recognizing Narrative Similarity

We return in this section to our proposal for using the task of Recognizing Narrative Similarity as a yardstick for measuring progress in narrative understanding by machines. In the sequel we illustrate how notions presented herein can be used both to encode and reason about stories, and to identify which of two stories is more similar to a target one.

4.1. An RNS Task Instance

The first story is made up for the purposes of this work:

> *A: Have you had lunch? B: Yes. A: Too bad. I was thinking of taking you out for lunch. B: I guess,*

then, you haven't had lunch yet? A: No. B: Okay. Maybe I can take you out for coffee later.

The second story is Aesop's fable "The Fox and the Crane":

A fox invites a crane over for dinner, and serves soup in a shallow plate, making it impossible for the crane to eat. A few days later, it is the crane's turn to invite the fox, and serves soup in a vase-like plate, making it impossible for the fox to eat.

The target story is a (perhaps not that funny) anecdote:

A: Have you had lunch? B: Yes. A: Too bad. I was thinking of taking you out for lunch. B: I guess, then, you haven't had lunch yet? A: No. B: Too bad. I was thinking of taking you out for coffee.

For these stories, we expect that many humans would agree that: *(i)* Comparing them at a shallow level (e.g., by matching word occurrences, or syntactic structure) will identify the first story as the one more similar to the target story. *(ii)* Comparing them at a deeper level (e.g., theme) will identify the second story as the one more similar to the target story. We do not claim that everyone labels the triple identically, nor that the shallow / deep distinction is the only one to be considered when labeling the triple above. However, as per *(W4)*, we believe that with appropriate guidelines (as used, for instance, for the annotation of textual corpora (Palmer et al., 2005)), human annotators might usefully produce RNS triples that can serve their goal as suggested in this work.

Following *(W1)*, the stories in the RNS task are not given in natural language, neither to the machine aiming to solve the task, nor to the human annotators. Whether the formal representations given are "correct" with respect to their natural language representations above is immaterial, as the formal representations themselves are the sole objects of interest, and it is for those that one seeks to establish similarity.

Along with each story we provide the author's intentions. Depending on the variation of the RNS task that one may consider, such intentions may be provided as input to: *(i)* the human annotators and the machines aiming to solve the RNS task, *(ii)* the human annotators only, or *(iii)* neither. Even in the latter two of the proposed variations of the RNS task, where such intentions are not given to the machines, we expect that machines would need to infer them in sufficiently good approximation in order to solve the task. For the purposes of this work, and in line with *(W3)*, we shall assume that this latter task has been solved, and we shall focus on addressing the RNS task from that point onwards. For ease of reference, but also to capture the requirement that only what follows from the story can be prescribed to the actors, we name the actors in all stories by the constants "alice" and "bob" (but any other constants would also do).

First Story

The story's discourse is $\langle \mathcal{C}_1, \mathcal{S}_1, \preceq_1^s \rangle$, with $\mathcal{C}_1$ comprising

$\texttt{occurs}(ask(alice, bob, lunch), S_1)$
$\texttt{occurs}(reply(bob, yes), S_2)$
$\texttt{occurs}(say(alice, bob, think(alice, bob, lunch)), S_3)$
$\texttt{occurs}(ask(bob, alice, lunch), S_4)$
$\texttt{occurs}(reply(alice, no), S_5)$
$\texttt{occurs}(say(bob, alice, invite(bob, alice, coffee)), S_6)$

with $\mathcal{S}_1 = \{S_1, S_2, S_3, S_4, S_5, S_6\}$, and $S_i \preceq_1^s S_j$ if and only if $S_i, S_j \in \mathcal{S}_1$ and $i < j$. An intended purpose for the story is encoded by $\langle \mathcal{P}_1, \Pi_1, \preceq_1^p \rangle$, with $\mathcal{P}_1$ comprising

$\mathsf{intend}(\texttt{holds}(think(alice, bob, lunch), S_3))$
$\mathsf{intend}(\texttt{holds}(invite(bob, alice, coffee), S_6))$

with $\Pi_1 = 2^{\mathcal{P}_1}$, and with $\preceq_1^p$ being the subset relation.

Second Story

The story's discourse is $\langle \mathcal{C}_2, \mathcal{S}_2, \preceq_2^s \rangle$, with $\mathcal{C}_2$ comprising

$\texttt{holds}(fox(alice), S_1)$
$\texttt{holds}(crane(bob), S_1)$
$\texttt{occurs}(invite(alice, bob), S_1)$
$\texttt{occurs}(serve(alice, bob, soup), S_2)$
$\texttt{holds}(plate(soup, shallow), S_3)$
$\texttt{occurs}(time\text{-}lapse, S_4)$
$\texttt{occurs}(invite(bob, alice), S_5)$
$\texttt{occurs}(serve(bob, alice, soup), S_6)$
$\texttt{holds}(plate(soup, vase\text{-}like), S_7)$

with $\mathcal{S}_2 = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$, and $S_i \preceq_2^s S_j$ if and only if $S_i, S_j \in \mathcal{S}_2$ and $i < j$. An intended purpose for the story is encoded by $\langle \mathcal{P}_2, \Pi_2, \preceq_2^p \rangle$, with $\mathcal{P}_2$ comprising

$\mathsf{intend}(\texttt{holds}(cheats(alice, bob), S_3))$
$\mathsf{intend}(\texttt{holds}(cheats(bob, alice), S_7))$

with $\Pi_2 = \{\mathcal{P}_2\}$, and with $\preceq_2^p$ being the subset relation.

Target Story

The story's discourse is $\langle \mathcal{C}_0, \mathcal{S}_0, \preceq_0^s \rangle$, with $\mathcal{C}_0$ comprising

$\texttt{occurs}(ask(alice, bob, lunch), S_1)$
$\texttt{occurs}(reply(bob, yes), S_2)$
$\texttt{occurs}(say(alice, bob, think(alice, bob, lunch)), S_3)$
$\texttt{occurs}(ask(bob, alice, lunch), S_4)$
$\texttt{occurs}(reply(alice, no), S_5)$
$\texttt{occurs}(say(bob, alice, think(bob, alice, coffee)), S_6)$

with $\mathcal{S}_0 = \{S_1, S_2, S_3, S_4, S_5, S_6\}$, and $S_i \preceq_0^s S_j$ if and only if $S_i, S_j \in \mathcal{S}_0$ and $i < j$. An intended purpose for the story is encoded by $\langle \mathcal{P}_0, \Pi_0, \preceq_0^p \rangle$, with $\mathcal{P}_0$ comprising

$\mathsf{intend}(\texttt{holds}(lying(alice, bob), S_3))$
$\mathsf{intend}(\texttt{holds}(cheats(alice, bob), S_3))$
$\mathsf{intend}(\texttt{holds}(lying(bob, alice), S_6))$
$\mathsf{intend}(\texttt{holds}(cheats(bob, alice), S_6))$

with $\Pi_0 \subseteq 2^{\mathcal{P}_0}$ comprising sets that include the second and fourth intentions, and with $\preceq_0^p$ being the subset relation.

4.2. Interpreting the Stories

We present below a useful and plausible, but by no means definitive or unique, part of the domain knowledge and beliefs used to interpret the aforementioned stories. Following *(W3)*, we shall assume that this domain has been identified and is available to the machine, perhaps following an autonomous learning phase of text reading (Michael, 2009).

The default domain is $\langle \mathcal{D}, \Delta, \preceq^d \rangle$, with $\mathcal{D}$ comprising

$\texttt{static}(\mathit{lying}(X,Y) \rightarrow \mathit{cheats}(X,Y))$
$\texttt{static}((\mathit{serve}(X,Y,Z) \wedge \mathit{visiting}(Y,X) \wedge \neg\mathit{appropriate}(Z,Y)) \rightarrow \mathit{cheats}(X,Y))$
$\texttt{static}(\mathit{say}(X,Y,S) \rightarrow (S \oplus \mathit{lying}(X,Y)))$
$\texttt{static}((\mathit{think}(X,Y,Z_1) \wedge \neg\mathit{realizable}(X,Y,Z_1) \wedge \mathit{think}(Y,X,Z_2) \wedge \neg\mathit{realizable}(Y,X,Z_2)) \rightarrow \mathit{lying}(X,Y))$
$\texttt{static}((\mathit{plate}(\mathit{soup},\mathit{shallow}) \wedge \mathit{crane}(X)) \rightarrow \neg\mathit{appropriate}(\mathit{soup},X))$
$\texttt{static}((\mathit{plate}(\mathit{soup},\mathit{vase\text{-}like}) \wedge \mathit{fox}(X)) \rightarrow \neg\mathit{appropriate}(\mathit{soup},X))$
$\texttt{static}((\mathit{think}(X,Y,\mathit{lunch}) \wedge \mathit{answer}(Y,\mathit{lunch},\mathit{yes})) \rightarrow \neg\mathit{realizable}(X,Y,\mathit{lunch}))$
$\texttt{static}((\mathit{think}(X,Y,\mathit{coffee}) \wedge \mathit{answer}(Y,\mathit{lunch},\mathit{no})) \rightarrow \neg\mathit{realizable}(X,Y,\mathit{coffee}))$
$\texttt{causes}(\mathit{invite}(X,Y),\mathit{visiting}(Y,X))$
$\texttt{causes}(\mathit{ask}(X,Y,Q),\mathit{question}(Y,Q))$
$\texttt{causes}((\mathit{reply}(Y,A) \wedge \mathit{question}(Y,Q)),\neg\mathit{question}(Y,Q))$
$\texttt{causes}((\mathit{reply}(Y,A) \wedge \mathit{question}(Y,Q)),\mathit{answer}(Y,Q,A))$
$\texttt{causes}(\mathit{exogenous}(L),L)$ for any literal L
$\texttt{static}(\neg\mathit{lying}(X,Y))$
$\texttt{occurs}(\mathit{exogenous}(L),T)$ for any literal L and time-point T

with $\Delta \subseteq 2^{\mathcal{D}}$ treating only the last two clauses as defeasible, and with $\preceq^d$ giving preference to sets in Δ that maximize use of clause $\texttt{static}(\neg\mathit{lying}(X,Y))$[2] after first minimizing use of clause $\texttt{occurs}(\mathit{exogenous}(L),T)$.[3]

Regarding the expressivity of our proposed framework, the clause $\texttt{occurs}(\mathit{exogenous}(L),T)$ accounts for the listener's belief that certain things may change in the flow of a story even when no action in the story explicitly accounts for such change. On the other hand, such exogenous fluctuations need to be restricted to the extend possible. This kind of preference (essentially stating that spontaneous change is a very weak belief), is captured by $\preceq^d$ as defined above.

Using this default domain, we may interpret the stories. It can be verified that all intentions of each story are satisfied. As an illustration, consider the target story and an embedding mapping each state S_i to time-point i. By assuming $\neg\mathit{lying}(\mathit{alice},\mathit{bob})$ and $\neg\mathit{lying}(\mathit{bob},\mathit{alice})$ at 1, we infer $\mathit{question}(\mathit{bob},\mathit{lunch})$ at 2, $\mathit{answer}(\mathit{bob},\mathit{lunch},\mathit{yes})$ at 3, $\neg\mathit{realizable}(\mathit{alice},\mathit{bob},\mathit{lunch})$ at 4, $\mathit{question}(\mathit{alice},\mathit{lunch})$ at 5, $\mathit{answer}(\mathit{alice},\mathit{lunch},\mathit{no})$ at 6, $\neg\mathit{realizable}(\mathit{bob},\mathit{alice},\mathit{coffee})$ at 7, and, thus, that this cannot lead to an interpretation, as it would contradict our assumption. On the other hand, assuming $\mathit{lying}(\mathit{alice},\mathit{bob})$ and $\mathit{lying}(\mathit{bob},\mathit{alice})$ at 1 is consistent with an interpretation, and leads to the inference $\mathit{cheats}(\mathit{alice},\mathit{bob})$ and $\mathit{cheats}(\mathit{bob},\mathit{alice})$ at 1.

4.3. The Most Similar Story

Consider applying the following modifications to discourse $\langle \mathcal{C}_2,\mathcal{S}_2,\preceq^s_2\rangle$: include $\texttt{holds}(\mathit{cheats}(\mathit{alice},\mathit{bob}),S_3)$ and $\texttt{holds}(\mathit{cheats}(\mathit{bob},\mathit{alice}),S_7)$ (semantic abstraction); remove all remaining clauses and states other than S_3,S_7 (syntactic abstraction); rename the states S_3,S_7 to S_1,S_2 (semantic abstraction). Call the discourse resulting from this process $\langle \mathcal{C}_{2a},\mathcal{S}_{2a},\preceq^s_{2a}\rangle$, and note that it is an abstraction of $\langle \mathcal{C}_2,\mathcal{S}_2,\preceq^s_2\rangle$ w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$.

Consider applying the following modifications to discourse $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$: include $\texttt{holds}(\mathit{cheats}(\mathit{alice},\mathit{bob}),S_3)$ and $\texttt{holds}(\mathit{cheats}(\mathit{bob},\mathit{alice}),S_6)$ (semantic abstraction); remove all remaining clauses and states other than S_3,S_6 (syntactic abstraction); rename the states S_3,S_6 to S_1,S_2 (semantic abstraction). Observe that the resulting discourse is $\langle \mathcal{C}_{2a},\mathcal{S}_{2a},\preceq^s_{2a}\rangle$, and that it is an abstraction of $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$ w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$.

It follows that $\langle \mathcal{C}_{2a},\mathcal{S}_{2a},\preceq^s_{2a}\rangle$ is a common abstraction of $\langle \mathcal{C}_2,\mathcal{S}_2,\preceq^s_2\rangle$ and $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$. Note that renaming states is a meta-operation — since state names have no bearing — and, thus, renaming applies to the intentions of the discourses. Observe, then, that the intentions $\mathcal{P}_2 \in \Pi_2 \cap \Pi_0$ are satisfied by $\langle \mathcal{D},\Delta,\preceq^d\rangle$ given $\langle \mathcal{C}_{2a},\mathcal{S}_{2a},\preceq^s_{2a}\rangle$. In other words, w.r.t. this common abstraction $\langle \mathcal{C}_{2a},\mathcal{S}_{2a},\preceq^s_{2a}\rangle$, the listener's background knowledge and beliefs satisfy the important intentions of both stories. Hence, $\langle \mathcal{C}_2,\mathcal{S}_2,\preceq^s_2\rangle$ and $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$ are credulously similar w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$.

One can show that $\langle \mathcal{D},\Delta,\preceq^d\rangle$ does not satisfy $\mathcal{P}_2$ w.r.t. any (common) abstraction of $\langle \mathcal{C}_1,\mathcal{S}_1,\preceq^s_1\rangle$ (and $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$) w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$. Hence, $\langle \mathcal{C}_1,\mathcal{S}_1,\preceq^s_1\rangle$ and $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$ are not credulously similar w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$. Overall, it follows that $\langle \mathcal{C}_2,\mathcal{S}_2,\preceq^s_2\rangle$ is more similar than $\langle \mathcal{C}_1,\mathcal{S}_1,\preceq^s_1\rangle$ to $\langle \mathcal{C}_0,\mathcal{S}_0,\preceq^s_0\rangle$ w.r.t. $\langle \mathcal{D},\Delta,\preceq^d\rangle$.

5. Epilogue and Possible Sequels

Recognizing Narrative Similarity was put forward as a concrete task for pushing research on computational narrative understanding forward. As a first step towards building machines for the RNS task (and, perhaps, understanding how humans may already solve it), the notion of narrative similarity was investigated from a computational point of view. We briefly discuss next how our investigation herein relates to existing research in a number of different directions.

Starting with our choice to model the listeners (background knowledge and beliefs, and their use to interpret stories) by adopting a particular framework (Michael, 2010a), we note that our formalization of story similarity does not hinge on the details of this choice, and that other logic-based frameworks equipped with a notion of story interpretation (Hobbs et al., 1993; Mueller, 2009; Verheij, 2009; Bex et al., 2010) can be modularly substituted for what we have employed.

On the role of author intentionality in story understanding, our approach follows a middle road between the two prominent schools of thought: the one insisting that an author's intent is the *only* way to understand a narrative (Michaels and Knapp, 1982), or that it points to the *right* way among possibly many (Hirsch, 1967); and the other dismissing that view as a fallacy (Wimsatt and Beardsley, 1946), and decreeing that the author of a narrative is not relevant, but only the narrative itself and the listener's background knowledge and beliefs (Barthes, 1967). We approach narrative understanding as a purely interpretive process, but adopt an intentionistic view for the process of determining narrative similarity. Our intention satisfaction definition accommodates both pluralism and monism on how many interpretations an author's intentions may point to, and cleanly accommo-

[2] We are, thus, encoding the weak belief that one does not lie.

[3] Clauses with variables (denoted by capital letters) are shorthand for the set of their grounded instances. For the purposes of the present work this informal reading of such clauses suffices.

dates the case of having mutually inconsistent interpretations and prioritizing among them (Levinson, 2003).

Straightforward extensions of our framework can accommodate more or less intentionistic views of narrative by, respectively, filtering out interpretations not satisfying authorial intents in Definition 8, or by replacing the notion of intended purpose in Definition 10 with appropriate domain rules so that a listener will infer what she believes the intended purpose of the story to be, and use that for determining similarity. Irrespectively of one's philosophical stance on intentionism, however, in a pragmatic instantiation of the RNS task the intentions assumed by our framework should be either *(i)* explicitly given, or *(ii)* determined by the machine. Inferring such intentions is, we believe, the major research stepping stone towards building machines that succeed in the (latter and stricter form of the) RNS task.

We hasten to note that this latter task is not as arduous as it seems: If one follows the less intentionistic view discussed above, the problem reduces to one for which learning-based frameworks can be employed to provide both a semantics and a solution (Michael and Valiant, 2008; Michael, 2009).

Some prior work sought to avoid reference to authorial intentions, viewing the problem of narrative similarity as one of finding isomorphisms between representations of stories (Löwe et al., 2009). Such a view is implicitly present in the theory of plot units (Lehnert, 1981), suggesting that stories are similar if they have common plot units; and explicitly present in structure-mapping theory (Gentner, 1983), formulating analogy between stories as a relation-preserving[4] bijection between the story objects. In all these cases, the search for an isomorphism leads to brittleness of the definition of narrative similarity, as it makes it critically depend on the granularity of the employed representations; cf. the "clock in a scene" example (Löwe, 2010). The tug-of-war approach that we adopt, on the other hand, suggests a concrete strategy when formally representing stories: err on the side of including more details about a story, and let abstractions find the right granularity (cf. elaboration tolerance).

Our tug-of-war view of a listener as seeking to abstract stories and identify similarities echoes Schank's position that such abstractions are performed unconsciously by humans for comprehension, whereby stories one hears are mapped to those that one knows (Schank, 1990). Schank argues that we tell only stories we believe will be understood, and that communicating with someone with very different stories is difficult. This agrees with our tug-of-war view of an author as striving to have the story intentions satisfied. Communication is difficult exactly when the author's and listener's stories are not credulously similar; i.e., their common abstractions are so distant from the original stories that they no longer satisfy the important intentions of the stories.

In developing an actual RNS corpus, we suggest to: *(i)* Avoid making shallow resemblance positively or negatively correlated with similarity, as that can be easily picked up by machines and defeat the purpose. *(ii)* Include instances with both clear-cut and close-call answers, so that the depth of "understanding" can be measured. *(iii)* Employ both natural and artificial stories, and let research identify those properties of natural stories that can be exploited — and presumably are, by humans — in addressing the task. *(iv)* Represent stories in a formal language that avoids, as much as possible, any single modality and its idiosyncrasies, but include "debugging" information, such as their original source (e.g., text, pictures, recording), and the intended purpose of the stories as understood by the corpus designers.

For the last point above, we advocate representing stories as (partially-ordered) collections of events and facts, since this simple form facilitates their automated extraction from natural language corpora. By contrast, experiments done using the structure-mapping theory (Falkenhainer et al., 1989) on a version of the RNS task used stories that were represented in a rather intricate form, with no indication on how it could be extracted automatically for large scale experimentation.

Indeed, scalability in building and evaluating machines for story understanding is a major concern: "Yet it is still not known how to scale up a story understanding program so it can understand more than toy stories." (Mueller, 2002). We believe that the computational treatment of story similarity offered herein will aid in addressing the scalability concern, by allowing machines to employ existing frameworks and engage *autonomously* in the learning (Michael, 2008; Michael, 2010b; Michael, 2011) and reasoning (Kakas et al., 2011) processes necessary for understanding stories.

Success in developing machines for the RNS task will directly impact a number of areas. First, identifying similarity suggests a natural way to summarize stories, by finding ones that are similar but shorter than the original. Second, seeking similarities in stories around the world may help identify common grounds across cultures and help in cross-cultural negotiations. Third, story similarity offers a natural grouping / labeling of stories that can be used for indexing, storage, and retrieval. Employing story similarity as a metric may yield better clustering of news articles on the web, and better organization and search for literary or other works in library collections. Fourth, finding a story similar to a given one but with different shallow features may aid in author anonymization (Kacmarcik and Gamon, 2006), text steganography (Chang and Clark, 2010), or paraphrasing as a means for protection against copyright infringement.

An application deserving special mention — and even more so on Turing's centenary year — is using the RNS task as a form of a Turing Test, since machines that succeed in identifying story similarities could presumably be thought of as exhibiting certain cognitive abilities. Approaching the RNS task from this angle and associating it with a well-promoted competition (cf. the PASCAL RTE Challenge (Dagan et al., 2005)) could help in making both the ComMoN community more visible to sister communities, and the question of machine narrative understanding more broadly appealing.

A lot remains to be done in realizing the goal of building machines for story understanding, and we hope that many more stories *similar* (in our technical sense!) to the one told herein will help to make progress in this exciting endeavor.

6. Acknowledgments

The author would like to thank Mark Finlayson and the anonymous ComMoN'12 reviewers for useful feedback.

[4] It is worth noting that although the theory acknowledges that the isomorphism need not be perfect, it *identifies* which relations are to be preserved, which shifts but does not address the problem.

7. References

Roland Barthes. 1967. Death of the Author. *Aspen*, 5–6.

Floris J. Bex, Peter J. Van Koppen, Henry Prakken, and Bart Verheij. 2010. A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence. *Artificial Intelligence and Law*, 18(2):123–152, June.

Ching-Yun Chang and Stephen Clark. 2010. Linguistic Steganography Using Automatically Generated Paraphrases. In *Proceedings of the 11th Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL'10)*.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of the 1st Workshop on Recognizing Textual Entailment (RTE'05)*.

Brian Falkenhainer, Kenneth D. Forbus, and Dedre Gentner. 1989. The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41(1):1–63, November.

Dedre Gentner. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7(2):155–170, April–June.

Eric Donald Hirsch. 1967. *Validity in Interpretation*. Yale University Press.

Jerry R. Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as Abduction. *Artificial Intelligence*, 63(1–2):69–142, October.

Gary Kacmarcik and Michael Gamon. 2006. Obfuscating Document Stylometry to Preserve Author Anonymity. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL'06)*.

Antonis C. Kakas, Loizos Michael, and Rob Miller. 2011. Modular-E and the Role of Elaboration Tolerance in Solving the Qualification Problem. *Artificial Intelligence*, 175(1):49–78, January.

Wendy G. Lehnert. 1981. Plot Units: A Narrative Summarization Strategy. *Cognitive Science*, 5(4):293–387, October.

Jerrold Levinson, 2003. *The Oxford Handbook of Aesthetics*, chapter 1, pages 3–24. Oxford University Press.

Benedikt Löwe, Eric Pacuit, and Sanchit Saraf. 2009. Identifying the Structure of a Narrative via an Agent-Based Logic of Preferences and Beliefs: Formalizations of Episodes from CSI: Crime Scene Investigation. In *Proceedings of the 5th International Workshop on Modelling of Objects, Components, and Agents (MOCA'09)*.

Benedikt Löwe. 2010. Comparing Formal Frameworks of Narrative Structure. In *Proceedings of the 2nd Workshop on Computational Models of Narrative (ComMoN'10)*.

John McCarthy. 1959. Programs with Common Sense. In *Proceedings of the Conference on the Mechanization of Thought Processes*.

Loizos Michael and Leslie G. Valiant. 2008. A First Experimental Demonstration of Massive Knowledge Infusion. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*.

Loizos Michael. 2008. *Autodidactic Learning and Reasoning*. Ph.D. thesis, Harvard University, Massachusetts, U.S.A., May.

Loizos Michael. 2009. Reading Between the Lines. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*.

Loizos Michael. 2010a. Computability of Narrative. In *Proceedings of the 2nd Workshop on Computational Models of Narrative (ComMoN'10)*.

Loizos Michael. 2010b. Partial Observability and Learnability. *Artificial Intelligence*, 174(11):639–669, July.

Loizos Michael. 2011. Causal Learnability. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*.

Walter Benn Michaels and Steven Knapp. 1982. Against Theory. *Critical Inquiry*, 8(4):723–742, Summer.

Erik T. Mueller. 2002. Story Understanding. In *Encyclopedia of Cognitive Science*. Macmillan Reference.

Erik T. Mueller. 2006. *Commonsense Reasoning*. Morgan Kaufmann.

Erik T. Mueller. 2009. Story Understanding Through Model Finding. In *Proceedings of the 1st Workshop on Computational Models of Narrative (ComMoN'09)*.

Martha Palmer, Dan Gildea, and Paul Kingsbury. 2005. The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics*, 31(1):71–106, March.

Roger Schank. 1990. *Tell Me A Story: A New Look at Real and Artificial Memory*. Scribners.

Bart Verheij. 2009. Argumentation Schemes, Stories and Legal Evidence: A Computational Perspective. In *Proceedings of the 1st Workshop on Computational Models of Narrative (ComMoN'09)*.

William K. Wimsatt and Monroe C. Beardsley. 1946. The Intentional Fallacy. *Sewanee Review*, 54:468–488.

Which Dimensions of Narratives are Relevant for Human Judgments of Story Equivalence?

Bernhard Fisseni[1,2,3], **Benedikt Löwe**[1,3,4,5]

[1] Institute for Logic, Language and Computation, Universiteit van Amsterdam
Postbus 94242, 1090 GE Amsterdam, The Netherlands
`b.loewe@uva.nl`
[2] Fakultät für Geisteswissenschaften, Universität Duisburg-Essen
Universitätsstraße 12, 45117 Essen, Germany
`bernhard.fisseni@uni-due.de`
[3] Isaac Newton Institute for Mathematical Sciences
20 Clarkson Road, Cambridge CB3 0EH, United Kingdom
[4] Fachbereich Mathematik, Universität Hamburg
Bundesstraße 55, 20146 Hamburg, Germany
[5] Corpus Christi College, University of Cambridge
Cambridge CB2 1RH, United Kingdom

Abstract

We present an experimental approach to determining natural dimensions of story comparison. The results show that untrained test subjects generally do not privilege structural information. When asked to justify sameness ratings, they may refer to content, but when asked to state differences, they mostly refer to style, concrete events, details and motifs. We conclude that adequate formal models of narratives must represent such non-structural data.

1. Introduction

Traditional and current computational models of narrative (Rumelhart, 1980; Lehnert, 1981; Schank, 1982; Dyer, 1983; Turner, 1994; Pérez y Pérez and Sharples, 2004; Frank et al., 2003; Mateas and Stern, 2003; Mueller, 2004; Si et al., 2005) focus on structural aspects of the narrative in their representation: events, causal relations between events, temporal relations of events, agents of the narrative, spatial relations between agents and objects, etc.

In terms of the narratological distinction of *story* and *discourse* (cf., e.g., (Chatman, 1980)), the formal representation of the narrative is stressing the *story* over *discourse* (cf. (Young, 2007)). Several approaches strongly emphasize that abstracting from the *discourse* results will yield the *structural core of the narrative* that is used for storing the narrative in their memory, retelling the narrative, as well as decisions whether two narratives are the same. As a motivation for her *Plot Units*, Lehnert connects them to the cognitive representation of summaries in the mind of the reader: "When a person reads a narrative story, an internal representation of that story is constructed in memory." (Lehnert, 1981, pp. 293)

The most prominent examples of this approach are *Structure Mapping Theory* and its implemented version, the *Structure Mapping Engine* (Gentner, 1983; Falkenhainer et al., 1989). While technically not about narratives, but about analogy and analogical reasoning, its most iconic examples are the comparisons of narratives like *Karla the Hawk* and modifications (Gentner et al., 1993, p. 533) and the question whether human test subjects recognize structural analogy:

> Domains and situations are psychologically viewed as systems of objects, object-attributes, and relations between objects. ... These representations ... are intended to reflect the way people construe a situation. (Gentner, 1983, p. 156–157)

This emphasis on structural analogy is reflected in recent approaches to find formal representation systems for comparison of narratives (Löwe, 2010; Bod et al., 2011; Löwe, 2011; Bod et al., 2012). On the other hand, systems based on *Structure Mapping Theory* have been criticised on the basis of empirical results for ignoring salient features of narratives that are relevant for human judgments of story equivalence (cf. also the discussion in (Löwe, 2011, § 2)):

> We have shown that [the] lack of inclusion of emotive content [in Gentner's *Structure Mapping Engine*] has made it psychologically implausible. (Lam, 2008, p. 38)

A natural and much more general follow-up question is addressed in this paper, namely: *Which features of narratives are relevant for human judges of story equivalence?*

In other words: do untrained human subjects, confronted with the task of deciding whether narratives are "the same" (without further specification what is the precise meaning of this phrase), rely mostly on structural features, or do other features (that are traditionally counted as part of the *discourse*) play a role in these decisions as well?

Since we do not want to presuppose any particular narratological ontology of features and their classification, we use the vague term *dimension* to refer to the possible features of narratives that could potentially be used to distinguish narratives as similar or equivalent. Examples of potential *dimensions* are: (a) motifs and superficial aspects such as (features of) the setting, the inventory of characters, single events (not connections between events) stylistic similarity, but (b) also aspects story event structure. Other, more philological categories would be (c) the relationship between narrator, characters, reality, the "possible world" of the story (cf., e.g., (Martinez and Scheffel, 2009)). We consider it part of the goal of the research reported on in this paper to give a preliminary classification of the relevant *dimensions* as they occur in the empirical data.

In § 2., we give a description of three experiments that elicited story comparisons; in § 3., we discuss the results of the experiments and conclude with a list of ideas for future work.

2. Experimental Work

The experimental work approached the question how test subjects naturally talk about stories when comparing them. Hence, neither the structural nor the motif level were focused by the instructions. As stimuli, we used stories that were not specifically constructed for the purpose of the experiment, but at most slightly varied to introduce controlled differences.

The tasks were simple and the experiments were conducted as classroom experiments. Participants were students of German literature and language and were rewarded for their participation with chocolate. Test subjects received a questionnaire with instructions; these were also presented by the experimenter. Test subjects were given about 15 minutes (**Queneau I** and **Queneau II**) or 20 minutes (**Fairy Tales**) to perform the task. Only numbers for native speakers are reported, unless stated otherwise. In all experiments, test subjects were given ample opportunity to report difficulties and give commentaries.

2.1. Experiment *Queneau I*

Setup. Test subjects were given two of Queneau's *Exercises in Style* (Queneau, 1947), translated into German (Queneau, 1990) with a length between 7 and 12 lines. Queneau's work consists of 99 variants of a base story in which the narrator gets on a bus, witnesses an altercation between a man and another passenger, and then sees the same person later getting advice on adding a button to his overcoat. We selected the base variant (*notations*), the variant told in reverse temporal order (*rétrograde*), the variant in which the agents are replaced by botanical objects (*botanique*), and the variant in which offensive language is used (*injurieux*). In the following, we refer to the variants as **order**, **botan.** and **offens.**, respectively.

Each test subject was given the base variant and one of the other variants; the title of the stories and their provenance was not given.[1] Test subjects were asked to take the role of the editor of a story magazine, helping a colleague to make a decision with respect to a strict rule of the journal: not to publish the same story twice. The test subjects should determine whether the two stories given were the same, and should explain to their colleague why they reached this conclusion. It was varied whether the colleague himself had suggested that the stories were the same or not.[2]

[1] One test subject recognised the stories.

[2] This turned out to have no observable effect on the test subjects' judgment.

There were 65 test subjects overall, of these 59 native speakers of German, almost all in their first semester and most of them studying to become primary school teachers.

The responses of the test subjects were categorised *ex post*, and the frequency of the categories was reviewed; from the natural language descriptions used by the test subjects, 46 *labels* were constructed (most occurring very infrequently) and later grouped into eight *categories*: *content*, *details*, *imagery*, *order*, *structure*, *style*, *substitution*, and *theme*. For every story, there were categories that was expected to figure most prominently: *order* for variant **order**, *substitution* or *imagery* for variant **botan.**, and *style* for variant **offens.**

	order	**botan.**	**offens.**
Same story	17	8	10
Different stories	3	8	11
(no decision)	0	2	0
n	20	18	21
	order	**botan.**	**offens.**
Same story	4	4	3
Different stories	2	1	6
mentioned	6	5	9

Table 1: Sameness judgments and *expected categories* by test subjects (**Queneau I**). The upper table lists the judgments as *the same story* and *different stories* by variants. The lower table lists how many times the *expected categories* are mentioned as a factor of difference.

	details			**structure**		
sameness	**simil.**	**diff.**	*no dec.*	**simil.**	**diff.**	*no dec.*
order	2	1	0	0	0	0
botan.	1	3	1	2	1	0
offens.	1	2	0	0	0	0
	content			**theme**		
sameness	**simil.**	**diff.**	*no dec.*	**simil.**	**diff.**	*no dec.*
order	11	3	0	2	1	0
botan.	5	5	0	1	1	0
offens.	9	7	0	4	2	15

Table 2: Mention of *structure* and *details*, *content* and *theme* per story (**Queneau I**)

Results. In all cases, the *expected categories* are mentioned by a minority of test subjects (cf. Table 1): For the variant **order**, only six test subjects mention *order* as a factor of difference; for the variant **botan.**, *substitution* is mentioned by two people, *imagery* by five; for the variant **offens.**, *style* is mentioned by nine test subjects. This is particularly striking in the case of a difference of order, where the vast majority of test subjects considers the stories to be *the same*.

The categories *details* (place, time, location, etc.), *structure* (surface structure, deep structure, etc.), *theme* and *content* occur very rarely (cf. Table 2). *Content* and *theme* are used as an argument in favour of similarity in nearly all cases when they are mentioned (with two exceptions for *theme*). Only three people mention *structure* at all.

Difficulties. Test subjects do not generally formulate their answers clearly and assigning the categories to the descriptions requires interpretation of the intention of the test sub-

	order	botanic	offensive
same story	8	0	3
different story	0	2	5
no decision	0	0	7
n	20	18	21

Table 3: Test subjects mentioning *perspective* as a differing factor by decision regarding story similarity (**Queneau I**).

jects. As an example, we mention the use of the label "perspective" illustrated in Table 3 (only noted as a factor of difference). The numbers suggest that the test subjects have a very vague notion of perspective.

Interpretation. We interpret these data to show that structural factors are not the most important aspect with respect to which test subjects compare stories, if this is not triggered explicitly. We are surprised that the *expected categories* are not named more often.

2.2. Experiment *Queneau II*

The main change from the previous experiment was that we intended to increase the number of mentioned categories of comparison per test subject, with the expectation that this will increase the number of mentions of structural factors and the *expected categories*. Test subjects were asked to justify their decision regarding the sameness of the story by naming at least two "important aspects" with respect to which the stories were the same or differed. As in **Queneau I**, test subjects had to take the role of the editor; the additional layer of communicating to a colleague was removed. Stories and questionnaires remained the same, only that the order of different and same aspects was varied, without any effect. 41 test subjects, 37 of them native speakers, participated; most were in their second year and intending to become a teacher at a grammar or comprehensive school.

Results. Explicitly asking for more than one category had a strong effect: Nearly all subjects (30 out of 37) now mention single instances of events or details[3] regarding settings and characters in their lists (21), or textual details like text length (13), both as similarities or differences (often different categories for either side). The category *content* is again mentioned by about half the subjects (cf. Table 5), always as a similarity, and *theme* is again rather rare (cf. Table 5), but except for two cases in variant **offens.** it is mentioned as a difference. The *expected categories* are now mentioned by a majority of test subjects for each variant (cf. Table 4).

Only two test subjects formulate their observations regarding *order* identifying the temporal order of the second story;[4] the other four formulate in a way that it is not clear whether they correctly resolved the order of events, or assumed a reverse chronological order. Regarding variant **botan.**, most test subjects present the observation in very concrete terms ("humans and vegetables") rather than abstractly.

	order	botan.	offens.
Same story	6	2	7
Different stories	5	1	6
(no decision)	0	9	1
n	11	12	14
factor of...	**order**	**botan.**	**offens.**
Same story	0	0	0
Different story	6	8	10
mentioned	6	8	10

Table 4: Sameness judgments and *expected categories* by test subjects (**Queneau II**). The upper table lists the judgments as *the same story* and *different stories* by variants. The lower table lists how many times the *expected categories* are mentioned as a factor of similarity or difference.

	details			textual details		
sameness	**simil.**	**diff.**	*no dec.*	**simil.**	**diff.**	*no dec.*
order	5	2	0	1	2	0
botan.	1	5	1	1	4	0
offens.	3	3	1	4	2	0
	content			**theme**		
sameness	**simil.**	**diff.**	*no dec.*	**simil.**	**diff.**	*no dec.*
order	7	0	0	0	0	0
botan.	0	2	1	3	1	0
offens.	6	0	0	2	1	0

Table 5: Mention of *details*, *textual details*, *content* and *theme* per story (**Queneau II**)

The vast majority of "important differences" reported (several for nearly all test subjects) were details and motifs such as places and characters. A preference for reporting structural similarities could again not be confirmed.[5]

2.3. Experiment *Fairy Tales*

The preceding experiments used variants of a very short story with a very limited structure. As a next step, we aimed at testing story comparison with stories of greater structural complexity.

Stimuli. Each test subject was given two versions of the fairy tale *Die drei Federn* (*The three feathers*) of the Brothers Grimm. The base version was the short version from the first edition (Grimm and Grimm, 1812, No. 64, III) and the variants were versions of the significantly altered and longer version from the last edition (Grimm and Grimm, 1857, No. 63) altered in several ways (see below). In the story, a king sets tasks for his sons to complete; the *Dummling* (Stupid One) completes all of them, with magical help, while his brothers fail. This story was used because the two versions

[3]For reasons of space, we cannot give a complete breakdown of the data, but give only one example of a questionnaire (variant: **order**, decision: **same story**). The test subject mentions as aspects of similarity: *"Ort: Autobus, Gare Saint-Lazare; Zeit: Mittag; Personen: Mann mit Hut, Freund von jenem; Detail: Hut, Handlung, Mann im Bus, mit Hut"*. ("Location: bus, Gare Saint-Lazare; Time: Noon; Characters: Man with hat, friend of his; detail: hat, plot, man on the bus, with hat"); he or she mentions as aspects of difference: "[In] Gesch. 2 ist Zeit („heute Mittag“) genau erwähnt" ("In story 2, the time is mentioned precisely ('this afternoon')").

[4]We counted a mention of "temporal perspective" as an identification of the reversed order of narration.

[5]The following is an interesting quotation from one of the test subjects' answers: „Der Inhalt macht keine Geschichte aus; es kommt auf die Darstellungsweise und die benutzten Mittel an.“ ("Content does not determine a story; it is about the way of presentation and the [stylistic] devices used.")

were (to the experimenters) immediately recognisable as the same story, but were quite different in many ways: from details of the story to the concrete kind of the tasks.

There were four variants of the second version of the story:

Temp: the original version with a slight variation of the order to presentation: the outcome of the tasks was recounted before the details of the tasks.

Granularity: a version with lower granularity (similar to a summary),

End*: versions with a different ending, namely **End1:** one in which the end was just reversed (the brothers ruled, haggling until their death), **End2:** one in which the *Dummling* rules, but is remembered for bad governance and stupidity.

Setup. The instructions asked test subjects to report at least three "important aspects" with respect to which the stories were similar or differed. It was expected that test subjects would report a variety of differences, both structural and details. Test subjects were 38 students of German literature and language, most of them in their first year.

Results and interpretation. More than in **Queneau II**, test subjects overwhelmingly name details. The categories *content* or *structure* are very rare (six occurrences overall, three non-native speakers, mentioned as "similar" in all cases). There are also just two mentions of the category *style*.

	Temp	**Granularity**	**End1**	**End2**
Same story	12	7	3	2
Different stories	2	5	4	2
(*no decision*)	0	0	1	0

story is …	**Temp**	**Granularity**	**End1**	**End2**
Same story	3	1?	2	1*
Different stories	1	0	1	2
(*no decision*)	0	0	1	0

Table 6: Sameness judgments and mentions of the *expected labels* by test subjects (**Fairy Tales**). The upper table lists the judgments as *the same story* and *different stories* by variants. The lower table lists how many times labels corresponding to the actual manipulations are mentioned as a factor of difference. In the lower table, "?" marks an uncertain classification; "*" indicates that the end is mentioned, but the test subject wrongly claims that the ending is the same in both variants.

In the lower part of Table 6, we list how many test subjects recognized the actual manipulations of the stories. The data show that our manipulations do not generally result in the judgment that the stories are different. Differences between the stories are – according to the extension of Fisher's exact test for data as implemented in R (R Development Core Team, 2010) – at best marginally significant.

The vast majority of "important differences" reported (several for nearly all test subjects, while only few test subjects mention structural factors such as "course of action") were details and motifs such as places and characters. Relatively few test subjects mention the factors we manipulated in the story (cf. Table 6, lower half).

Seven people claim to know at least one of the stories; one of them claims to know both, clarifying: "→ same tale"; another modifies: "more or less", another: "parts of it, *Froschkönig*, *Aschenputtel*" (the Frog Prince, Cinderella), and also two of those who do not know the stories, say: "parts of it from other stories" or "the tale of *Aschenputtel*" (Cinderella). These remarks confirm that test subjects have a mixed motif-structure view on these tales.[6]

3. Discussion & Conclusion

We conclude that structural dimensions of stories are not a natural level of processing sameness judgments for untrained subjects. Different tasks trigger different reactions by test subjects: When asked to justify their actions, test subjects may refer to a vague notion of content, which arguably encompasses the event structure and causal links. However, when asked to produce many factors, references become much more concrete and less structural.[7] Details and motifs, linguistic features and other dimensions are also used by test subjects.[8]

We conclude that our experiments may be seen as evidence that either structural similarity does not suffice for sameness judgments or that the empirical grounding for formal models of narrative should not be based only on untrained and unfocused subjects. If a model of story similarity is to be cognitively adequate for untrained and unfocused subjects, it must allow selective access according to the goal of the comparison, and must be complemented by a model of story processing that determines which dimensions are focused.

Acknowledgements

The research in this paper was funded by the *John Templeton Foundation* (JTF) via the project *What makes stories similar?* (grant id 20565) and the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO) via the project *Dialogical Foundations of Semantics* in the ESF EuroCoRes programme LogICCC (LogICCC-FP004; DN 231-80-002; CN 2008/08314/GW). The authors acknowledge the financial support and the kind hospitality of the *Isaac Newton Institute for Mathematical Sciences* (programme *Semantics & Syntax*). The authors should like to thank Tim Kocher and Charlotte Wollermann (Duisburg-Essen) for giving them access to the test subjects.

4. References

Rens Bod, Benedikt Löwe, and Sanchit Saraf. 2011. How much do formal narrative annotations differ? A Proppian case study. In Charles Ess and Ruth Hagengruber, editors, *The computational turn: Past, presents, futures?, Aarhus University, July 4–6, 2011*, pages 242–245, Münster. MV-Wissenschaft.

[6] Consider also the following quotations from the questionnaires: „Die Geschichte hat den gleichen Kern, ist äußerlich aber unterschiedlich." ("The story has the same core, but is different in outer appearance"; judgment: different story, **Temp**) and „Die Intention bleibt auch bei leicht abgewandeltem Inhalt gleich (letzter Satz)." ("The intention stays the same with [slightly] altered content (last sentence)"; judgment: same story, **Temp**).

[7] It is also conceivable that a different educational background of the test subjects, could have a significant effect on the results.

[8] Motif and story indices (Uther, 2004; Thompson, 1955 1958) highlight that different levels of story comparison are philologically interesting.

Rens Bod, Bernhard Fisseni, Aadil Kurji, and Benedikt Löwe. 2012. Objectivity and reproducibility of Proppian narrative annotations. this volume.

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

Michael G. Dyer. 1983. *In-depth understanding: A computer model of integrated processing for narrative comprehension*. Artificial Intelligence Series. MIT Press, Cambridge MA.

Brian Falkenhainer, Kenneth Forbus, and Dedre Gentner. 1989. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 20:1–63.

Stefan L. Frank, Mathieu Koppen, Leo G. M. Noordman, and Wietske Vonk. 2003. Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, 27:875–910.

Dedre Gentner, Mary Jo Rattermann, and Kenneth D. Forbus. 1993. The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25:524–575.

Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170.

Jacob Grimm and Wilhelm Grimm. 1812. *Kinder- und Hausmärchen*. Realschulbuchhandlung, Berlin, 1st edition.

Jacob Grimm and Wilhelm Grimm. 1857. *Kinder- und Hausmärchen*. Dieterichsche Buchhandlung, Göttingen, 7th edition.

Samantha Lam. 2008. Affective analogical learning and reasoning. Master's thesis, School of Informatics, University of Edinburgh.

Wendy G. Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science*, 4:293–331.

Benedikt Löwe. 2010. Comparing formal frameworks of narrative structures. In Mark Finlayson, editor, *Computational models of narrative. Papers from the 2010 AAAI Fall Symposium*, volume FS-10-04 of *AAAI Technical Reports*, pages 45–46.

Benedikt Löwe. 2011. Methodological remarks about comparing formal frameworks for narratives. In Patrick Allo and Giuseppe Primiero, editors, *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*, pages 10–28, Brussel. KVAB.

Matias Martinez and Michael Scheffel. 2009. *Einführung in die Erzähltheorie*. C. H. Beck, München, 8th edition.

Michael Mateas and Andrew Stern. 2003. Integrating plot, character and natural language processing in the interactive drama *Façade*. In Stefan Göbel, Norbert Braun, Ulrike Spierling, Johanna Dechau, and Holger Diener, editors, *Technologies for Interactive Digital Storytelling and Entertainment. TIDSE 03 Proceedings*, volume 9 of *Computer Graphik Edition*. Fraunhofer IRB Verlag.

Erik T. Mueller. 2004. Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research*, 5(4):307–340.

Rafael Pérez y Pérez and Mike Sharples. 2004. Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA. *Knowledge-Based Systems*, 17:15–29.

Raymond Queneau. 1947. *Exercices de style*. Gallimard, Paris.

Raymond Queneau. 1990. *Stil-Übungen*. Suhrkamp, Frankfurt a.M.

R Development Core Team, 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien.

David E. Rumelhart. 1980. On evaluating story grammars. *Cognitive Science*, 4:313–316.

Roger C. Schank. 1982. *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge University Press.

Mei Si, Stacy C. Marsella, and David V. Pynadath. 2005. Thespian: Using multi-agent fitting to craft interactive drama. In Michal Pechoucek, Donald Steiner, and Simon Thompson, editors, *AAMAS '05. Fourth International Joint Conference on Autonomous Agents and Multiagent Systems 2005. Utrecht, Netherlands. July 25–29, 2005*, pages 21–28. ACM.

Stith Thompson. 1955–1958. *Motif-index of Folk-Literature: a Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Medieval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends. 6 volumes.* Indiana University Press, Bloomington, 2nd edition.

Scott Turner. 1994. *The creative process. A computer model of storytelling*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Hans-Jörg Uther. 2004. *The Types of International Folktales. A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson*. The Finnish Academy of Science and Letters, Helsinki.

R. Michael Young. 2007. Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies*, 8:177–208.

Story Retrieval and Comparison using Concept Patterns

Caryn E. Krakauer, Patrick H. Winston

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, Massachusetts
krakauer@mit.edu, phw@mit.edu

Abstract

Traditional story comparison uses key words to determine similarity. However, the use of key words misses much of what makes two stories alike. The method we have developed use high level *concept patterns*, which are comprised of multiple events, and compares them across stories. Comparison based on concept patterns can note that two stories are similar because both contain, for example, *revenge* and *betrayal* concept patterns, even though the words *revenge* and *betrayal* do not appear in either story, and one may be about kings and kingdoms while the other is about presidents and countries. Using a small corpus of 15 conflict stories, we have shown that similarity measurement using concept patterns does, in fact, differ substantially from similarity measurement using key words. The Goldilocks principle states that features should be of intermediate size; they should be not too big, and they should not too small. Our work can be viewed as adhering to the Goldilocks principle because concept patterns are features of intermediate size, hence not so large as an entire story, because no story will be exactly like another story, and not so small as individual words, because individual words tend to be common in all stories taken from the same domain. While our goal is to develop a human competence model, we note application potential in retrieval, prediction, explanation, and grouping.

Keywords: Goldilocks principle, story retrieval, intermediate features, concept patterns

1. Story comparison and precedent retrieval

Any full account of precedent-based reasoning must provide an account of how potentially relevant precedents are retrieved from memory. Considerable psychophysical research, reviewed in Finlayson and Winston (Finlayson and Winston, 2005), indicates that novices in a domain retrieve using superficial features, whereas experts in a domain retrieve retrieve using structure.

Finlayson and Winston showed how a range of behavior, from novice to expert, corresponds to an increase in the maximum chunk size considered by a matcher, starting with individual objects and ending with collections of objects and the relations among them. Thus, expert behavior corresponds to matching not on objects, nor on entire precedents, but on chunks of intermediate size, which led Finlayson and Winston to frame what they call the Goldilocks principle.

The Finlayson and Winston work was based on structure mapping theory (Falkenhainer et al., 1989; Gentner and Forbus, 1991), and thus requires computationally expensive graph matching. Our work, in contrast, is based on what we call *concept patterns*, which are reminiscent of plot units (Lehnert, 1981), and capture aspects of what we mean when we talk of, for example, *revenge* or *selling out*. As in Dehghani's work on analogy and moral decision making, retrieval is sensitive to known narratives (Dehghani et al., 2009).

In our implemented system, potential precedents are stored along with the concept patterns determined to lie within them. Then, at retrieval time, story-to-precedent matching is done by a fast dot-product computation on a concept-pattern vector derived from a story with concept-pattern vectors derived from potential precedents.

Thus, relative to the Finlayson-Winston work, our approach is fast and our concept patterns may span long chains of connected relations, while still retaining the flavor of retrieval based on intermediate features. [1]

Of course, once a potential precedent is retrieved, analysis begins, and judgments of similarity involve not just events and how they are arranged but also concept patterns and how they are arranged. Accordingly, we have begun to study the role of concept patterns in similarity judgments in general, not just in retrieval.

2. The Genesis substrate

We build on the Genesis System (Winston, 2011), a story understanding system that reads simple English, elaborates on what it reads by applying commonsense rules, and performs searches to detect concept patterns. At the commonsense level, Genesis notes, for example, that if you are killed, you become dead, and if I harm you, I harm your friends.

Concept patterns are higher level structures in which events are said to lead to other events, with possibly many intermediate events. We generally supply concept patterns in English, instructing Genesis directly, as in the following examples:

[1] Finalyson and Winston's thinking about intermediate features was originally inspired by Shimon Ullman's work on face finding in images (Ullman et al., 2002). He matched, for example, using features such as a nose and mouth combination, rather than, say, an eye or an entire face.

```
Start description of "Revenge".
xx is an entity.
yy is a entity.
xx is my friend.
xx's harming yy leads to yy's harming xx.
The end.

Start description of "Pyrrhic victory".
xx is an entity.
ll is an action.
zz is an entity.
xx's wanting ll leads to xx's becoming happy.
xx's wanting ll leads to zz's harming xx.
The end.
```

Concept patterns are chosen by the user, and are generally include two to three events. Users choose concept patterns based on the users' understanding of the stories. We are developing a method of automatically generating concept patterns, which is discussed later in the paper.

As stories are read by Genesis, commonsense rules are deployed, which have a tendency to connect the story's explicit events via causal relations. The search machinery that looks for satisfied *leads to* relations then exploits those causal connections to locate concept patterns. In our simple, abbreviated rendering of the plot in Shakespeare's Hamlet, that search machinery finds *Pyrrhic Victory*, *Leadership Achieved*, *Suicide*, and three instances of *Revenge*

3. Comparing stories using concept patterns

We compared stories on a concept level using three different methods, each of which serves a different purpose in story comparison. The methods are: comparing the number of concept patterns in common, noting the longest common substring of concept patterns, and weighted comparing usin concept pattern rarity.

Vector-angle Mode

Our first method compares the number of concept patterns in common for fast retrieval. We save story concept pattern counts in vectors. Then, using these vectors, our method calculates the angle between story vectors to determine similarity, with the metric varying between 0.0 and 1.0.

For example, the highest match for the Bay of Pigs Invasion by vector-angle is the Cambodia-Vietnam Invasion. Both conflicts have a *allied offense*, an *invasion*, and a *victory*. In both conflicts, a larger political entity supported a smaller group's invasion.

Order-sensitive Mode

Our second method takes into account the ordering of the concept patterns. For example, a *revenge* that is the result of a *betrayal* is different than a *betrayal* that is the result of a *revenge*. The importance of ordering can also be seen in the comparison of the American Revolution and Afghan Civil War, as shown in figure 1.

In the stories as provided, both the American Revolution and the Afghan Civil War contain *defense* and *allied support*, but they appear in different orders. In the American Revolution, the American people received allied support from France after Britain's attack. In the Afghan Civil War however, Russia gave allied support to Najibullah before the attack happened. In one, an ally came to the support of an already embattled nation. In the other, an ally helped stockpile weapons for an impending conflict. The ordering of two stories makes a difference in their overall similarity.

Rarity-sensitive Mode

The rarity of each concept pattern is also important in comparing stories. We recognize three variations:

- **Rare among a group of stories:** If a concept pattern is rare among a group of stories, it can be seen as more important when comparing similar stories. For example, when looking at a group of Disney-style fairy tales, two stories that have a *princess marries a prince* concept pattern, they do not seem as similar as two stories in which a concept pattern indicating *princess ditches the prince and marries a poor commoner*, because the ditch-the-prince concept pattern is rare.
- **Very common among a group of stories:** If a concept pattern is very common among some but not all stories, it may be useful for grouping stories. If a group of stories have concept patterns in common, but those concept pattern are much rarer among all stories, than that group of stories may make up a genre. For example, the "Disney-style fairy tale" genre may have concept patterns such as *princess and prince fall in love*, *villain causes prince and princess to be kept apart*, and *prince and princess live happily ever after*. If a new story is read with similar concept patterns, it may also be a Disney-style fairy tale.
- **Very common among most stories:** If a concept pattern is very common among most stories, then it is not particularly useful in deciding whether two stories are similar.

As an example of the influence of rarity, consider the rarities of the concept patterns in the American Revolution, with the concept rarities shown in table 1.

Concept Pattern	Rarity
Legal Disagreement	0.027
Invasion	0.167
Rebellion	0.069
Unwanted succession	0.042
Conflict	0.083
Allied Defense	0.014
Victory	0.208
Victory Defensive	0.069

Table 1: The rarity of concept patterns found in the American Revolution story. An example of the *victory*' pattern is the most common concept pattern, while an *allied defense*' is the most rare. Rarity is calculated by dividing the number of times a concept pattern appears by the total number of concept patterns seen.

The most common concept pattern is *victory*. A victory occurs in almost every story in the set, and so is very common. Because of this, a *victory* is a poor measure of similarity between these stories, but a very good indicator that the story

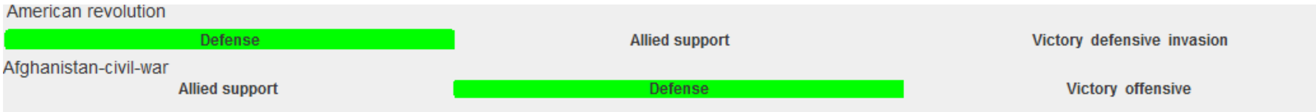


Figure 1: The in-order comparison of the American Revolution and the Afghan Civil War. While these stories both have defense and allied support, the two concept patterns appear in differing order. If order was not taken into account, then the two stories would share two in-common concept patterns. However, with order taken into account, the maximum sub-plot has only a length of one. The maximum sub-plot is *defense* which is highlighted in the figure.

is about a conflict. On the other hand, an *allied defense* is much more rare and therefore more important when measuring story similarity in the conflict domain.

We currently calculate rarity by dividing the number of times a concept pattern appears by the total number concept patterns seen.

4. Experimental results

We have run our system, in Vector-angle Mode, on 15 conflict summaries previously used in the work of Finlayson and Winston (Finlayson and Winston, 2005). These includes rebellions, wars, and political conflicts. Here, for example, is the American Civil War as provided to our system:

> *Start story titled "American Civil War". The United States is a country. The Confederacy is an entity. The Union is an entity. The Confederacy was a region of the United States. The Union was a region of the United States. The Union disliked the Confederacy because the Confederacy possessed slaves. The Confederacy left the United States because the Confederacy disliked the Union. The Confederacy left the United States because the Union possessed the Confederacy. The Union wanted the Confederacy to stay at the United States. The Union attacked the Confederacy because the Confederacy left the United States and the Union wanted the Confederacy to stay at the United States. The Confederacy attacked the Union because the Union attacked the Confederacy. The Union was stronger than the Confederacy. The Union defeated the Confederacy because the Union attacked the Confederacy and the Union was stronger than the Confederacy. The Union controlled the Confederacy because the Union defeated the Confederacy. The Union forced the Confederacy to return to the United States because the Union controlled the Confederacy and the Union wanted the Confederacy to stay at the United States. The end.*

Stories have been simplified mainly to get them through the front-end natural-language parser. Accordingly, the need for simplification will diminish as natural-language parsers improve.

Story simplification introduces the possibility of simplifier bias. From one simplifier's perspective, an attack might be recast as an *invasion* while from another, it might be described as a *counter-attack*.

If two interpretations are different enough, there may be a change in the analysis, but we view this as a feature, not a bug. If a simplifier thinks of two wars very differently, say one war was a *justified first strike* and another as an *unjustified invasion*, they would not be considered similar by the simplifier and likewise would not be considered similar by our system .

Figure 2 illustrates the differing results using concept-pattern vectors (top) and word vectors (bottom). Black represents a similarity score of zero and white represents a similarity score of 1.0, the maximum possible value.

When comparing the conflict stories on a word level, the difference between the similarity scores of most story pairs is small. Because all of the stories are on the same topic, they all share many keywords. Stories compared with themselves are white because the keywords are exactly the same, but when compared to other stories, the comparison scores are relatively low and do not change much from story to story. The mean and standard deviation for each method are as follows:

Method	Mean	Standard Deviation
Keyword	0.267	0.119
Concept Pattern	0.364	0.200

Table 2: The mean and standard deviation of similarity scores generated by each method. The standard deviation of story comparison by concept pattern is almost twice that of keyword comparison. Similarity scores are on a scale from 0.0 (not similar) to 1.0 (identical).

The Cambodian-Vietnam Invasion compared with the China War with Vietnam is an outlier. These two stories are two parts of an overall conflict, so the actors in both conflicts are the same.

We found that comparing stories using concept patterns performs in more congruence with our own interpretations. For example, the deviation of similarity score values is much higher than in keyword comparison on the fifteen conflict stories on which we ran experiments, just as we view story pairs as varying considerably in similarity. Following are three examples where concept pattern comparison finds similar stories but keyword comparison falls short.

- **American Revolution and the American Civil War:** Concept pattern comparison picks out the American Revolution and the American Civil War as being similar giving them a similarity score of 0.67, as they have several concept patterns in common (*unwanted succession*, *victory*, *conflict*, *legal disagreement*). This makes

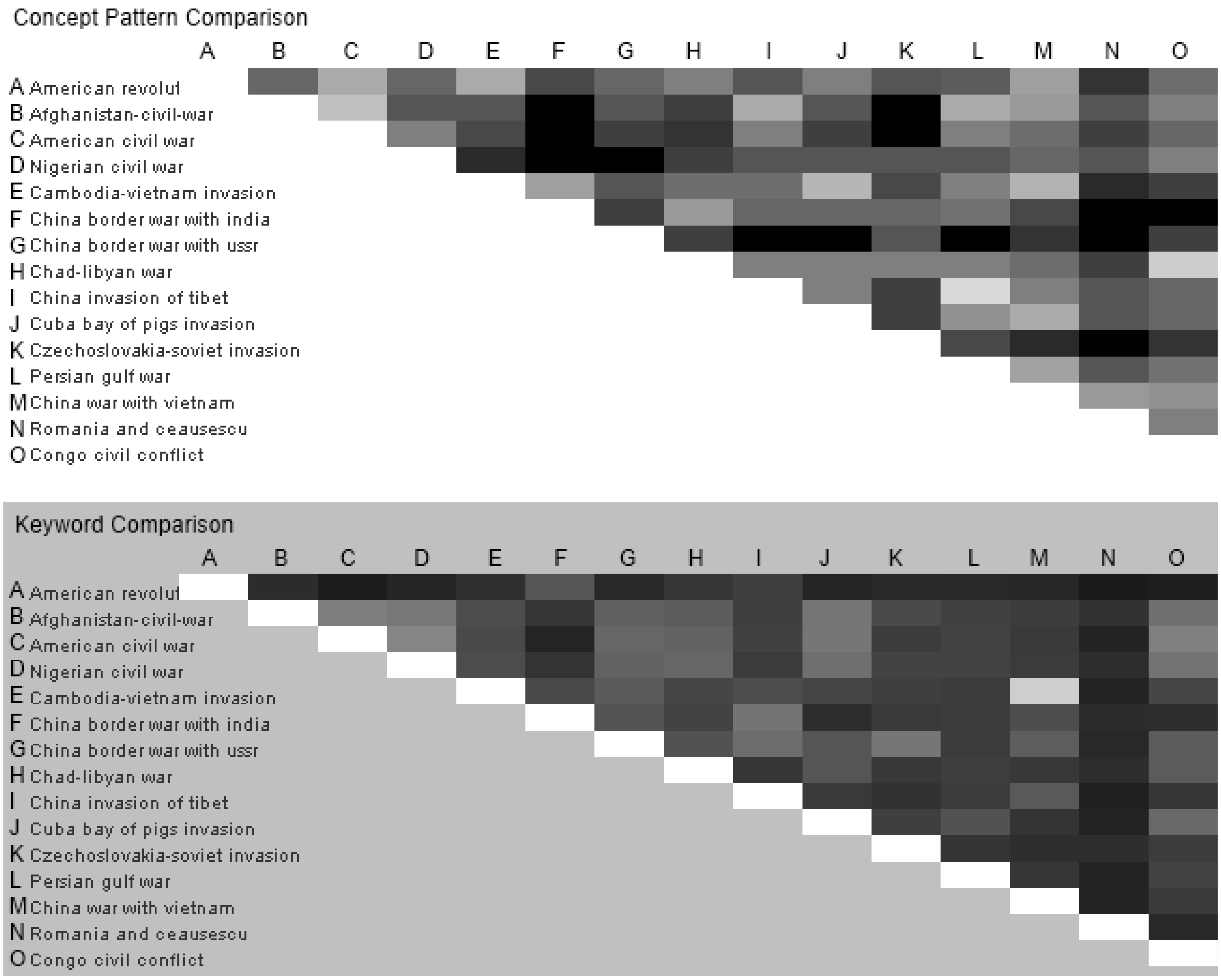


Figure 2: Top: similarity scores produced by concept patterns. Bottom: similarity scores from keywords. Comparison based on concept patterns has a greater diversity of scores and sensitivity to high-level structural matches even in the absence of low-level keyword correspondence. Similarity scores are created using the vector angle calculation.

sense, as both stories are about a part of a country rebelling from the main country over legal disputes (taxes in one case, slaves in the other). In the word comparison, these stories have a very low similarity score of 0.1 (as shown by the red). By using concept patterns to compare stories, more meaningful story comparison is performed.

- **China border War with India and the Cambodia-Vietnam Invasion:** Another example of the concept pattern comparison succeeding while the keyword comparison fails is the comparison between the China border War with India and the Cambodia-Vietnam Invasion. In both cases, two countries fought over an area of land (the Mekong Delta in the Cambodia-Vietnam conflict, and the Assam in the China-India conflict). The relevant concept patterns found are a land dispute along with two invasions (one by each country into the disputed region), which gives the comparison score of 0.71. The keyword comparison however, rates them as relatively unsimilar with a score of 0.26.

- **Afghanistan Civil War and the Czechoslovakia Soviet Invasion:** An example where keyword comparison has decided that two stories are similar, where in fact they are not, are the Afghanistan Civil War and the Czechoslovakia Soviet Invasion. Keyword comparison gives a score of 0.48, which is very high for keyword comparisons. However, the concept pattern comparisons give them a score of 0.0. The stories, while both involve the Soviet Union, are very different conflicts. In the Czechoslovakia Soviet Invasion, the Soviet Union invaded Czechoslovakia due to political reform. In the Afghanistan Civil War, the Soviet Union funded one side of a civil war, but did not actually attack. Thus, the two conflicts are quite different, which is shown by the concept pattern comparison.

5. Concept Pattern Generation

Our current work includes concept-pattern discovery directly from stories, circumnavigating the need to supply all concept patterns in English. It is in the same spirit as the concept-pattern discovery work of Mark Finlayson (Finlayson, 2012).

Our discovery process works by searching for concept patterns, consisting of two or three events in leads-to relations, common to two or more stories. In principle, there could be $O(n^3)$ such concept patterns in a story, where n is the number of events; in practice, there are far fewer, because only event pairs connected by causal chains qualify as potential concept patterns.[2] In addition, we filter out concept patterns that are too rare. We ignore concept patterns that only appear in a single story. This is reminiscent the approach taken by Chambers and Jurafsky in their work on unsupervised learning. (Chambers and Jurafsky, 2008) Due to their large amount of data, their system was preforming poorly. Accordingly, they eliminated rare occurrences of verb pairs, improving performance.

Once concept patterns in one story are computed, they can be compared to concept patterns appearing in one or more previously read stories In order for two concept patterns to be the same, they must align. In order to align, their structure must be the same, their events must be similar, and the concept pattern's actors must align. In order for two events to be similar, their actions must be similar. For example, *A invades B* is similar to *C attacks D*. Every word has a thread which is defined by WordNet. The thread for *invade* is {action, contend, attack, invade} while the thread for *attack* is {action, contend, attack}. Two words a similar if they are the same word, share a common parent, or if one of the words is a parent of the other. A "parent" is defined as the immediate parent to the word (so attack for invade, and contend for attack).

- Same Structure: Two concept patterns have the same structure if their leads to relations are the same. So $a \rightarrow b \rightarrow c$ has the same structure as $d \rightarrow e \rightarrow f$, but not $d \rightarrow e$ $d \rightarrow f$.
- Similar Events: Concept patterns can have different event types, as long as they are sufficiently similar. For example *punch*, *kill*, *insult*, and *murder* are considered sufficiently similar because they are all kinds of *harm*. Our system uses WordNet to determine if two words are similar in meaning.
- Aligned Actors: Two concept patterns can be aligned if the actors from each event correspond. So if the two concept patterns are: *a harms b leads to b harms a* and *c harms d leads to d harms c*, then the events align but *a harms b leads to b harms a* and *c harms d leads to e harms c* do not.

Below are examples of concept patterns generated by the system. The names were provided by us after the fact and are not known the the system.

- **Giving Aid (two events):**
  American revolution:
  - France helps America
  - France gives money to America

  Cambodia-vietnam invasion:
  - China helps Cambodia
  - China gives weapons to Cambodia
- **Revenge Attack (two events):**
  Afghanistan-civil-war:
  - Najibulla attacks Mujahideen
  - Mujahideen attacks Najibulla

  American civil war:
  - Confederacy attacks Union
  - Union attacks Confederacy
- **Wanting an entity to stay, and dislike between entities, leads to a defeat (three events):**
  Nigerian civil war:
  - Nigeria wants NigerianEast to stay
  - Nigeria defeats NigerianEast
  - NigerianEast dislikes Nigeria

  American civil war:
  - Union wants Confederacy to stay
  - Union defeats Confederacy
  - Confederacy dislikes Union
- **Wanting an invasion leads to an invasion, which is defeated (three events):**
  Cuba bay of pigs invasion:
  - UnitedStates wants exiles to invade Cuba
  - Exiles invade Cuba
  - Soldiers defeat Exiles

  China war with Vietnam:
  - Vietnam does not want China to invade Vietnam
  - China invades VietNam
  - Vietnam defeats China

Because there are many potential concept patterns in stories, care must be taken to only select the concept patterns that are meaningful when measuring story similarity. A concept pattern that only appears in just two stories is not likely to be important, as it can serve no role in demonstrating story similarity more generally. Likewise, a concept pattern that appears in all stories is not useful because it has no discriminatory power.

In our next step, we will attempt to use a mutual information metric to establish which of the candidate concept patterns are useful in story comparison.

6. Potential application

Our main goal in this work is to model human story retrieval, and in that connection, we are planning a series of psychological experiments. In passing, we note that our approach to similarity matching offers a promising approach to prediction, understanding, and grouping.

- *Retrieval and prediction:* By finding patterns in similar stories, the ending of a new story can be predicted by way of precedent. This is especially useful for understanding how a person from a culture different from our own will respond to a proposed course of action. If our system is loaded with stories that characterize a culture of interest, and is then presented with the beginning half of a course of action, its predictions may well be different from those predicted in the absence of those culture-characterizing stories. Suppose, for example, a person is presented with a story: "Charlie and Bob were friends.

[2] This idea emerge in a discussion that included Finlayson, fortuitously initiated by a fire drill in our building.

Charlie hit Bob in the face." If the person is then asked to predict the ending, his answer will depend on his culture and upbringing. By understanding people's reactions to situations, we can better predict the outcome of events.

- *Retrieval and explanation:* By finding a similar story, one better understood that a current story, explanations for events can be discovered. Consider, for example, the scenario: "Bob bought Jill flowers." Without any explanation for the action, a program would not understand the reasoning behind the action. By retrieving a similar story, the program may find an explanation for the action. If the similar story contained: "Mary bought Larry chocolates because Mary liked Larry." The program could extrapolate from the similar events that Bob may like Jill, causing him to give her a gift. By finding similar stories, unexplained events can be better understood.

- *Grouping:* By using concept patterns, stories can be grouped into categories. A group of concept patterns that are rare overall, but common among a group of stories may constitute a genre. For example, conflict stories may generally involve an attack and a victory, while fairy tales may involve falling in love and living happily every after. Grouping stories together helps to organize information, and can make story retrieval faster, because if stories are pre-grouped, a retrieval system only has to search in one or a few genres to find the most similar story.

7. Contributions

- We have implemented several mechanisms for story comparison based on concept patterns.

- We have shown, with a small corpus of 15 conflict stories, that retrieval based on concept-pattern vectors produces precedents more like those found by domain experts (structure) than those found by novices (superficial features).

- We have demonstrated, at an illustration-of-concept level, a mechanism that discovers concept patterns in story ensembles by searching for parallel event patterns.

The next step in the development of the program is to conduct studies in which human subjects are given stories and asked to compare them on a concept level. This will establish a ground truth of story similarity, and will allow better testing of our system's modeling fidelity

8. Acknowledgments

Research on Genesis has been supported, in part, by the National Science Foundation (IIS-0413206), the Office of Naval Research (N00014-09-1-0597), the Air Force Office of Scientific Research (A9550-05-1-0321), and the Defense Advanced Research Projects Agency (FA8750-10-1-0076).

9. References

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*.

Morteza Dehghani, Dedre Gentner, Kenneth D Forbus, Hamed Ekhtiari, and Sonya Sachdeva. 2009. Analogy and moral decision making. *New Frontiers in Analogy Research Proceedings of the Second International Conference on Analogy*.

Brian Falkenhainer, Kenneth D. Forbus, and Dedre Gentner. 1989. The structure mapping engine: Algorithm and examples. *Artificial Intelligence*, 41(1):1–63.

Mark A. Finlayson and Patrick Henry Winston. 2005. Intermediate features and informational-level constraint on analogical retrieval. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pages 666–671.

Mark A. Finlayson. 2012. *Learning Narrative Structure from Annotated Folktales*. Ph.D. thesis, MIT.

D Gentner and K Forbus, 1991. *MAC/FAC: A model of similarity-based access and mapping*.

Wendy Lehnert. 1981. Plot units and narrative summarization. *Cognitive Science*, 5(4):293–331.

Shimon Ullman, Michel Vidal-Naquet, and Erez Sali. 2002. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7).

Patrick Henry Winston. 2011. The strong story hypothesis and the directed perception hypothesis. In Pat Langley, editor, *Technical Report FS-11-01, Papers from the AAAI Fall Symposium*, pages 345–352. AAAI Press.

From the Fleece of Fact to Narrative Yarns: A Computational Model of Composition

Pablo Gervás

Universidad Complutense de Madrid
pgervas@sip.ucm.es

Abstract

From a given observable set of events, a large number of stories may be composed, by deciding to select or omit specific events, by restricting attention to smaller subsets of the overall setting, by focusing on particular characters, or by narrating the chosen events in different order. This particular task of narrative composition is not covered by existing models of storytelling or cognitive accounts of the writing task. This paper presents a model of the task of narrative composition as a set of operations that need to be carried out to obtain a span of narrative text from a set of events that inspire the narration. To provide guidance in structuring the task, an analogy is drawn between the narrative composition task and that of manufacturing textile fibres, with corresponding concepts of heckling the original material into fibres, then twisting these fibres into richer and better yarns. The model explores a set of intermediate representations required to capture the structure that is progressively imposed on the material, and connects this content planning task with a classic pipeline for natural language generation. As an indicative case study, an initial implementation of the model is applied to a chess game understood as a formalised set of events susceptible of story-like interpretations. The relationships between this model and existing models from other fields (narratological studies, cognitive accounts of writing, AI models of story generation, and natural language generation architectures) is discussed.

Keywords: Narrative generation, Theory of narrative, Representations, Natural language processing, Artificial intelligence, Cognitive science, Narratology.

1. Introduction

The task of composing a narrative based on a given set of events that have taken place has received little attention in terms of conceptual modelling. Efforts have been made to capture the structure of narratives as a finished product (by the narratology research community), to come up with a set of cognitive processes implied in the tasks of writing in general or of understanding narrative in particular (by the cognitive science community), to build models of how fictional plots are generated from scratch or of how discourse may be structured for a given plot (by the artificial intelligence community) and to construct functional architectures for generating text from conceptual data (by the natural language generation community). The task of putting together a narrative that conveys events that have already happened is related to all these aspects. It is also the kind of basic storytelling that people carry out in their everyday lives to communicate with one another, to convince, to inform, to remember the past, to interpret the present and to plan for the future.

Producing a model of this task is a significant challenge for various reasons. First, due to its complexity. The fact that it involves elements from various other tasks requires a certain familiarity with the different phenomena involved. The final product will need to have narrative structure, and the implied processes should at least be cognitively plausible. Second, because as a task it involves not just modelling the particular product resulting from the process (narrative structure), or of the processes themselves (narrative composition) but also requires a model of the input (the set of facts observed / remembered that constitute the source and the starting point for the composition). This element is different from the structures considered in cognitive accounts of narrative understanding (as it includes the events that may appear in the story but not necessarily the causal relations between them) and different from the set of events that are actually mentioned in existing stories (as the story will contain only a selection of all the events that actually happened, possibly filtered on the basis of a set of relevant causal relations postulated by the author). Such a representation of the input is implicit in the specification of the task, yet any attempt at computational modelling must start by representing it explicitly, as it will significantly influence the rest of the process.

The present paper attempts to address these challenges from an engineering point of view. The task is considered in terms of how its input and its expected output might be represented, and what processes might lead from one to the other. To manage the complexity of the problem, an analogy is established between the process of narrative composition and that of converting naturally occurring materials into lengths of line (whether yarns, thread or rope). This analogy is based on the similarities between the inputs (in both cases volumes of material of heterogeneous quality with no obvious unique linear presentation) and the outputs (for both a linear reorganization of selections of the same material, systematically structured to optimise certain desired qualities of the final result) of these processes. The analogy is discussed, and then further explored to provide a set of candidate subtasks for narrative composition that constitute a possible computational model for narrative composition. This analogy is taken in this paper to devise a computational model of the composition task which is then tested in a software implementation operating over a very simple representation of meaningful events involving a set of characters that interact over time in an elementary represented space. Chess provides a finite set of characters (pieces), a schematical representation of space (the board)

and time (progressive turns), and a very restricted set of possible actions. Yet it also allows very elementary interpretations of game situations in terms of human concepts such as danger, threat, conflict, death, survival, victory or defeat, which can be seen as interesting building blocks for story construction.
The paper reviews existing models related to the narrative composition task, outlines and discusses the analogy between narrative composition and the spinning of textile yarns, describes the proposed computational model, presents the case study for narration of chess games and finishes with discussion and conclusions.

2. Previous Work

A number of models of related tasks and elements arising from different fields of research are reviewed in this section to provide background material for the discussion. Due to the breadth of fields under consideration, exhaustive review in any one of them is beyond the scope of the paper. An attempt has been made in each case to gather here the set of elementary concepts in each field that are relevant for the understanding of the arguments in the paper.

2.1. Narratology

According to many theorists, narrative has two components: what is told (what narrative is: its content, consisting of events, actions, time and location), and the way it is told (how the narrative is told: arrangement, emphasis / de-emphasis, magnification / diminution, of any of the elements of the content). These have been named different ways by different researchers, story and discourse, *histoire* and *discours*, *fabula* and *sujzet*. There are alternative analyses that postulate different subdivions. Even between theories that agree on having just two levels of analysis there seem to be many subtleties that cast doubt on whether the same thing is meant by the different words. This presents a serious obstacle for researchers from the computational field trying to address the treatment of stories in any form. In order to avoid ambiguity, we will restrict our analysis here to three levels of conceptual representation of a story, and refer to these as the *story* (the complete set of what could be told, organised in chronological order of occurrence), the *plot* (what has been chosen to tell, organised in the order in which it is to be told) and the *narrative* (the actual way of telling it).
Narratologists, who specialize in the study of narrative, consider the concept of *focalization* (Genette, 1980) as the way in which a narrator restricts what he is telling about a particular scene to what might have been perceived by someone present in that scene. This may be one of the characters if the scene is told in the first person, or the narrator himself as if he had been present (if the story is told in the third person). This has an interesting implication in the fact that, through focalization, narrative discourse (and thereby the structure of stories) is influenced by the perception of space: events that take place simultaneously in different locations that cannot be perceived at the same time (this may be different cities but also different neighbouring rooms separated by a wall) usually require different narrative threads.

2.2. Cognitive Accounts of Writing

Flower and Hayes (Flower and Hayes, 1981) define a cognitive model of writing in terms of three basic process: planning, translating these ideas into text, and reviewing the result with a view to improving it. These three processes are said to operate interactively, guided by a monitor that activates one or the other as needed. The planning process involves generating ideas, but also setting goals that can later be taken into account by al lthe other processes. The translating process involves putting ideas into words, and implies dealing with the restrictions and resources presented by the language to be employed. The reviewing process involves evaluating the text produced so far and revising it in accordance to the result of the evaluation. Flower and Hayes' model is oriented towards models of communicative composition (such writing essays or functional texts), and it has little to say about narrative in particular. Nevertheless, a computational model of narrative would be better if it can be understood in terms compatible with this cognitive model. An important feature to be considered is that the complete model is framed by what Flower and Hayes consider "the rhetorical problem", constituted by the rhetorical situation, the audience and the writers goals.

2.3. Cognitive Accounts of Narrative Comprehension

Although this paper is concerned with modelling the process of narrative composition, it is indirectly affected by models of narrative comprehension in as much as the results of composition must be suitable for comprehension. Narrative comprehension involves progresive enrichment of the mental representation of a text beyond its surface form by adding information obtained via inference, until a situation model (representation of the fragment of the world that the story is about) is constructed (van Dijk and Kintsch, 1983). A very relevant reference in this field is the work of (Trabasso et al., 1989), who postulate comprehension as the construction of a causal network by the provision by the user of causal relations between the different events of a story. This network representation determines the overall unity and coherence of the story.

2.4. Story Telling

Storytelling efforts in AI have focused on two different tasks: that of building fictional plots from scratch and that of structuring appropriate discourse for conveying a given plot
The importance of causal relations in narrative comprehension has led to AI models of plot generation that rely heavily on the concept of planning. Many existing storytelling systems feature a planning component of some kind, whether as a main module or as an auxiliary one. TALE-SPIN (Meehan, 1977), AUTHOR (Dehn, 1981), UNIVERSE (Lebowitz, 1983), MINSTREL (Turner, 1993) and Fabulist (Riedl and Young, 2010), all include some representation of goals and/or causality, though each of them uses it differently in the task of generating stories. An important insight resulting from this work (originally formulated by (Dehn, 1981) but later taken up by others) was the distinction between goals of the characters in the story or goals of the author.

A less frequently modelled aspect but also very relevant is emotion, which clearly plays a heavy role in the appreciation of narrative. The MEXICA storytelling system (Pérez y Pérez, 1999) takes into account emotional links and tensions between the characters as means for driving and evaluating ongoing stories. The system evaluates the quality of a partial draft for a story in terms of the the rising and falling shape of the arc of emotional tensions that can be computed from this information.

With respect to the task of building an appropriate discourse for rendering a given plot, significant efforts have been carried out in the domain of cinematic visual discourse. These include work on use of flashback and foreshadowing to produce surprise (Bae and Young, 2008) and automatic generation of camera placements over time to define a visual discourse that best fits the plot to be rendered (Jhala and Young, 2010). Both of these efforts rely on a planning based approach to narrative, with plots represented as plans.

2.5. Natural Language Generation

The general process of text generation takes place in several stages, during which the conceptual input is progressively refined by adding information that will shape the final text (Reiter and Dale, 2000). During the initial stages the concepts and messages that will appear in the final content are decided (*content determination*) and these messages are organised into a specific order and structure (*discourse planning*), and particular ways of describing each concept where it appears in the discourse plan are selected (*referring expression generation*). This results in a version of the discourse plan where the contents, the structure of the discourse, and the level of detail of each concept are already fixed. The *lexicalization* stage that follows decides which specific words and phrases should be chosen to express the domain concepts and relations which appear in the messages. A final stage of *surface realization* assembles all the relevant pieces into linguistically and typographically correct text. These tasks can be grouped into three separate sets: *content planning*, involving the first two, *sentence planning*, involving the second two, and surface realization. An additional task of *aggregation* is considered, that involves merging structurally or conceptually related information into more compact representations ("Tom fled. Bill fled." to "Tom and Bill fled." or "the boy and the girl" to "the children"). Aggregation may take place at different levels of the pipeline depending on its nature.

3. Narrative Composition from an Engineering Point of View

The type of narrative that we want to address in this paper involves a linear sequential discourse where only a single event can be told at any given point. Yet reality is not like that. Events to be reported may have happened simultaneously in physically separated locations, and constitute more of a cloud than a linear sequence, a volume characterised by 4 dimensional space time coordinates. Composing a narrative for such an input involves drawing a number of linear pathways through that volume, and then combining these linear pathways (or a selection thereof) together into a single linear discourse. This type of linear pathway is sometimes referred to as a *narrative thread*. The analogy between narrative and the production of textiles is pervasive. A narrative of real or fictitious adventures is sometimes referred to as a *yarn*, which is formally a continuous strand of twisted threads as used in weaving or knitting. To spin is to draw out and twist (fibres) into thread but also to relate a tale, or to provide an interpretation of something in a way meant to sway public opinion.

Before we can take this analogy further we need to consider how thread is produced. An initial material (wool, cotton, hemp, sisal...) is selected to provide the starting fibres. These materials occur as volumes of fleece, cottonseed, or stems or leaves of plants. A process is applied to transform these volumes into sets of linear fibres. This is known in various ways but we will refer to it as *heckling*. These fibres are then spun into longer *yarns*, which are continuous lengths of interlocked fibres, suitable for use in the production of textiles, sewing, crocheting, knitting, weaving, embroidery and ropemaking. Several yarns may be twisted together to form a *strand*. Strands may themselves be further twisted together to produce rope.

The elementary description of the problem of narrative composition given above matches the task of converting a volume of fleece into strands of wool. An original mass with no clear linear structure is first processed into a set of linear fibres, some of which are later combined together into more complex elements (yarns, strands, rope) which are still linear but collectively exhibit properties not present in individual fibres (perform better than individual fibres under certain parameters). The assumption underlying this analogy is that the reality (as a set of facts) from which the narrative is to be drawn plays the role of the initial volume of fleece.

As a research tool, this analogy poses the following questions:

1. what is the narrative equivalent for a fibre, and what is the corresponding process of heckling
2. what is the narrative equivalent for a yarn, and what is the corresponding process of spinning
3. what is the narrative equivalent for a strand, and what is the corresponding process of twisting

Not all of these are equally important, but trying to answer them all provides rich picture of narrative composition as a computationally modelable task. To bound the problem at the other end, let us assume that the narrative equivalents of a rope would be large-scale narrative works such as novels, capable of bearing heavy loads of expectation of narrative quality. As both the product and the specification of the requirements it should fulfill are beyond the scope of a paper such as this, we will restrict our present endeavour to the level of yarns, understood as spans of narrative reduced in size and complexity yet already exhibiting the overall formal properties that we desire of narratives (multi-threaded and involving a certain complexity in terms of chronological and spatial relative differences between threads). Thus they would be good candidates to represent small everyday narratives and the elementary abilities of narrative composition. Nevertheless, it should be kept in mind that these yarns or narrative spans would be susceptible of further integration into larger narrative elements.

4. A Computational Model of Narrative Composition

The task of heckling can be related to the identification of appropriate focalization decisions for conveying a given material. Focalization, understood as the decision of which character the narration should follow, and how much of the environment around him at each point should be conveyed to the reader of the narrative, heckles the perception of reality into individual fibres (one for each possible focalizer character) that are linear and sequential in nature. For each character involved in the set of events to be conveyed, a possible focalization fibre can be drawn. In contrast with the physical fibres of textiles, different elements of the material (locations, objects, characters, events...) may feature simultaneously in more than one fibre. This difference is not considered problematic, and it will allow the model to represent important features of narrative, such as the possibility of including multiple perpectives of a given event.

However, it introduces the need for a narrative-specific fibre selection subtask at this stage. As fibres will have redundant information, and very different coverage of the set of events to be reported, an important challenge during narrative composition will be to select the most promising set of fibres from a given inspirational set of events.

The tasks of spinning and twisting are conceptually very similar, involving as they do very similar operations (combining thin linear segments in to thicker ones). At this point the analogy with textiles stops being useful. The structure of textile yarns or strands is significantly different from the structure of narrative in terms of its threads. Textile yarns actually develop in three dimensional space, with fibres spiralling around one another along the length of the yarn. Narrative threads must be combined into a linear sequence of discourse with each fibre taking centre stage of the discourse for a while, then stopping and leaving room for another to take its place, which may later stop and allow for a return to the initial fibre. Thus the combination of narrative fibres needs to take a different form that takes into account this difference in structure.

As a first approximation, two basic operations are considered:

Grafting two separate narrative fibres converge at given point in the narrative, and from then only a single fibre exists

Splicing two (or more) separate narrative fibres are combined into a single discourse sequence by snipping each one of them into smaller fragments of similar relative size and interleaving them in an appropriate order to form a single discourse that switches back and forth between them, covering their whole length

Grafting is a simpler operation than splicing. Grafting occurs when fibres for two different focalizers converge at a point in the discourse and one of them does not continue beyond that (that particular focalization does not appear any further in the discourse). This may occur because the focalizer character for one of the fibres need not appear anymore in the narrative to be told (either because he/she is dead or no longer plays a significant role in the story or because their view on the story is already covered by other focalizations considered more desireable). This type of structure occurs when a secondary narrative thread is inserted into a primary one to provide explanation for certain aspects of it (how someone else happened to be there, why particular characters react in certain ways...).

Splicing is the fundamental operation for handling multi-threaded narratives. It involves the following operations:

1. identifying for each fibre a set of potential break points where the narrative flow may be interrupted to switch to a different thread
2. identifying pairs of origin-target break points in different fibres such that abandoning one fibre at the origin break point and retaking the other at the target break point results in a desireable discourse

The identification of potential breakpoints must rely on various criteria that are either specific to the domain being narrated, or determined by various advanced effects such as suspense or narrative tension. This issue will not be addressed generically here but only for a specific case study below.

The pairing of origin-target breakpoints is also too specific or too complex to be addressed generically here. However, it is important to point out that the relative displacement (in terms of time and space) between the moment and the location associated with the two breakpoints plays a very important role. Other parameters that may need to be considered are the relative size of fibre fragments that results from snipping at a given breakpoint, and conceptual relations that occur between the elements (characters, events, locations, time moments...) described in each fibre around the breakpoint.

This introduces an additional aspect that needs to be modelled. Whenever two fibres have been spliced together, some kind of contextualization must be added at the target breakpoint, to provide the reader with an idea of how the fibre begin retaken relates to the fibre just abandoned. Expressions of the kind “Two days earlier, at headquarters in London,...” play this role, usually occurring at the beginning of new paragraphs or new chapters dealing with new focalizations.

4.1. Fibres and Heckling

An *event* is something that happens with a potential for being relevant to a story. Events occur at a specific location at a given moment of time. They may have preconditions and postconditions. In order to have a generic representation, each event is considered to have an associated *event description* that allows both a descriptive and a narrative component. The descriptive component of an event description contains predicates that describe relevant preconditions. The narrative component of an event describes the action of the event itself, but it may also contain additional narrative predicates describing the effect of the action. For instance, if someone moves from one place to another, elements at the original location may be left behind, and elements in the final location appear. These are not considered separate events but included in the description of the moving event.

A *fibre* is a sequence of events that either involve or are seen by a given character. It represents a focalized perception of the world. The extent of information that is included in the events for a given fibre is defined by the range of perception that is being considered and by the presence of any obstacles to perception in the surrounding environment. It may also be affected by the direction in which the focalizer is facing, or where he is focusing his attention. As these further refinements require advanced authorial decisions, we decide that the standard representation should include all the information within perceptive range, and leave the decision of whether to mention it to a later stage of composition.

The task of *heckling* involves establishing the range of perception, tracking the set of all possible characters involved in the events to be narrated, and for each character constructing a fibre representation that includes descriptions of all the event that the character initiates, suffers or perceives. These descriptions will appear in the fibre in chronological order.

4.2. Yarns and Twisting

A *yarn* would constitute the most elementary type of narrative. It should hold a sequence of fragments of fibre, possibly coming from different fibres. The event descriptions within a given fibre fragment will appear in relative chronological order, but transition to a different fibre fragment in the sequence may involve a shift in chronology. This allows for the representation of phenomena such as flashbacks or flashforwards. Transition to a different fibre fragment may also imply a change in focalization. Together with the change in chronology, this allows for the representation of alternating narration as a single linear discourse of simultaneous narrative threads.

A yarn is obtained by combining together fibres. This can be done in several ways:

- grafting a secondary fibre onto a primary one
- splicing together two (or more) fibres of equal importance
- dividing a fibre in a number of fragments and recombining them in a different order

The criteria to be applied in each case would need further study. This study should take into consideration existing work on narrative and rhetorical effects, as well as domain dependent information and issues such as purpose of the narration or author goals (and in general, the rhetorical problem as described by Flower and Hayes).

Nevertheless, some basic indications have been provided above. Further considerations are made in section 5. for the particular case of narratives told about a game of chess.

4.3. From Yarns to Spans of Narrative Text

To obtain yarns involves composition at a reasonably abstract conceptual level. This may be said to correspond to the planning process of Flower and Hayes' model, or to a content planning task in terms of the classic natural language generation pipeline model. However, the translation process of the Flower and Hayes' model (from these structural plans for the narrative to actual text) involves a further set of elementary operations. Some of these operations[1] are described here.

4.3.1. Contextualization

Once a specific structure for the narrative has been decided upon in terms of yarns, an important task is to establish appropriate contextualizations after each transition into a new fibre fragment. When changes occur at a certain point in the discourse to the location, the narrative time, or the focalization, these changes need to be indicated in some ways to avoid confusion. In the given representation, these changes occur only at the transition between fibre fragments included in a yarn. When the structure of a yarn becomes fixed, it must be traversed in the correct sequence, adding any required contextualizations at the appropriate places. These contextualizations may take different forms:

- inclusion of temporal expressions or temporal discourse markers to indicate changes in chronology (usually with respect to the narrative time holding at the end of the previously narrated fragment)
- inclusion of spatial expressions to indicate changes in location
- relying on combinations of the above and possibly additional discourse features to allow the reader to infer changes in focalization

As the last bullet point indicates, these mechanisms can be fairly complex, and a detailed analysis is beyond the scope of this paper.

4.3.2. Setting Narrative Parameters

The groupings of event descriptions within a yarn provided by fibre fragments constitute very good candidates for the assignment of narrative parameters such as which person to use in narration or verb tense. Because transitions between fibre fragments will signal changes in focalization or chronology, spans of discourse corresponding to single fibre fragments are likely to have similar overall values for person and tense. In the case of tense, the same information used for contextualization (relative shift in chronology from the previous fibre fragment to the current one) will play a significant role in establishing correct values for tense.

They are also likely to be significant in determining layout information (for instance, forcing the introduction of paragraph or even chapter breaks to match transition into new fibre fragments).

[1]The set of operations addressed here is not intended to be exhaustive. Many more would probably arise if the task were addressed in detail.

4.3.3. Filtering and Contextualizing Descriptive Information

An additional operation that may be considered at this stage is filtering the information available in the event descriptions in the yarn according to the desired authorial decisions. Once the author has decided whether the narration is to take place in the first or the third person (or the second), it may be relevant to flag part of the descriptive information contained in the yarn for specific events, so that it does not get realized. Such information would for instance be things that the focalizer character cannot actually see or has not noticed even though they are within perceptual range. This would correspond to a content determination task in terms of the classic natural language generation pipeline.

Also, the actual form that spatial descriptions are going to take may need to be refined, changing from an absolute perspective with respect to the overall space to referential expressions relative to the position and/or the orientation of the focalizer. This would correspond to a referring expression generation task in terms of the classic natural language generation pipeline.

4.3.4. Realization

Further operations at a more basic level of natural language generation would be required.

If a fluent natural text is desired, a stage of aggregation should be considered, for instance to pack together descriptive statements of similar structure, or replace enumerations of relative position of several individuals with generic description of the position of the whole set using an appropriate collective noun.

Additional stages of referring expression generation may be required to replace some of the fully specified references with pronominal references. These pronominal references (and their placement) must be adequately constructed to ensure the resulting discourse remains understandable and no unnecessary ambiguity is introduced.

A stage of lexical choice can be introduced, to explore the possibilities of using different lexical terms for recurring occurrences of the same concept. Depending on the desired style for the resulting narrative, this can be a strong requirement (for more literary styles) or an encumbrance (for texts that value referential precision more highly than aesthetic value). Again, the rhetorical problem and/or constraints on the writing task should be considered here to provide decision criteria.

5. A Case Study: Narratives from a Chess Game

To provide a preliminary benchmark for the various intuitions outlined in the rest of the paper the simplest approximation to a case study that could be conceived is described in this section. This is done by considering a chess game as a very simple model of a formalised set of events susceptible of story-like interpretations. Chess provides a finite set of characters (pieces), a schematical representation of space (the board) and time (progressive turns), and a very restricted set of possible actions. Operating on simple representations of a chess game in algebraic notation, exploratory solutions for the tasks of content selection and content planning are explored based on a fitness function that aims to reflect some of the qualities that humans may value on a discourse representation of a story.

A basic software implementation has been written that reads a description of a chess game in algebraic notation (see Table 1) and builds for it the kind of representations that are described above. The aim of this exercise is to consider broadly what particular domain dependent criteria may be applicable at each of the decision points outlined for the generic case in the previous sections.

Each individual chess piece taking part in the game is considered a character. Perception range is defined as the small space of 3 x 3 squares of the boad that constitutes that immediate surroundings of each piece at any given moment.

Events are triggered by pieces moves. Whenever a piece moves, this constitutes an event for the piece itself, for any other piece captured during the move, and for any other piece that sees either the full move, the start of the move or the conclusion of the move.

Fibres for each of the pieces are built by collecting event descriptions for those moves that they are involved in or they see. The same event may get described differently in different fibres depending on the extent to which the corresponding focalizer is involved in it.

Initial trials with this set up for simple realization of single fibres uncovered a number of conceptual problems. Because chess was used as a context, and no specific measures had been taken, the resulting narratives are not understood as being focalised as they are intended. Chess as a concept tends to invoke unconsciously focalization over the complete board.

Two main problems were found. First, there was no way of identifying from the rendering of the story fragment for each fibre who the story "was about" (the focalizer). Second, focalization was not clear, because the text gave no indication of what was seen at each stage. So the reader, identifying the whole thing as a description from a chess game, assumed focalization to be over the whole board.

To address the first problem, an introductory sentence was added to each fragment presenting this problem, stating who the focalizer is and where she is ("The black queen was four squares north of the centre of the board."). Also, descriptions of events in which the focalizer does not actually take part are translated in indirect form ("The black queen saw the white left bishop appearing ahead.").

To address the second problem, a brief description of what can be seen is also added in the cases where focalization has changed or was previously unknown, to establish the actual range of perception.

Additionally, chronological information in terms of moves or turns in the game was translated into a more intuitive temporal framework by considering that each move corresponds to a day, and appropriate temporal expressions are generated based on that premise. In this way, temporal expression become "two weeks earlier" rather than "14 moves earlier". This significantly reduces the perceived awkwardness.

Spatial information throughout was reformulated either in terms of cardinal points of the compass ("north", "south",...) or relative to the focalizer ("right" "left"

1. e4 c5	16. Bxe2 Be6
2. Nf3 d6	17. Rfd1 Rfd8
3. d4 cxd4	18. Bc5 Rd5
4. Nxd4 Nf6	19. b4 a5
5. Nc3 g6	20. Bf3 Rxd1+
6. Be2 Bg7	21. Rxd1 e4
7. Be3 O-O	22. Bxe4 Bxc3
8. O-O Nc6	23. Bxc6 Rc8
9. h3 d5	24. b5 Bxa2
10. exd5 Nxd5	25. Bd4 Bb4
11. Nxd5 Qxd5	26. Be5 Be6
12. Bf3 Qc4	27. b6 Rxc6
13. Nxc6 bxc6	28. b7 Rb6
14. c3 e5	29. Rd8+
15. Qe2 Qxe2	1-0

Table 1: Algebraic notation for an example chess game

"ahead" "behind"...). The first option was used for describing piece movements, for describing the initial position of pieces at the beginning of the story, or for describing the relative position of the location where a new fibre fragment starts with respect to the location where the previous fibre fragment had ended.

As the case study is intended primarily for basic trial of the intuitions, only the simplest composition operation has been implemented. This constitutes the grafting of the life fibre of a captured piece into the fibre of the capturing piece. This was carried out using the following criteria:

- split the fibre for the capturing piece at the point of the attack
- start the yarn with the first of the resulting fibre fragments
- then add the full fibre for the captured piece (from beginning to attack, but omitting the actual death)
- then add the second fibre fragment obtained from the fibre for the capturing piece (starting with a recapitulation of the attach followed by the death of the captured piece)

The process of dealing with the redundant information present in both fibres, and distributing it appropriately over the resulting fibre fragments (before and after the break, and between the fibres with different focalizers) has for the time being been resolved empirically and would be in need of further study.

An excerpt of an example rendering of the narrative span for a yarn obtained by grafting the fibres for two chess pieces is given below. This constitutes the story of the confrontation between the black and white queens. The yarn starts by narrating the life of the black queen from the beginning of the game, and follows her closely until the beginning of her attack on the white queen. It then tracks back to tell the story of the white queen from the beginning of the game to that same point. Then it describes the outcome of the attack. It finishes by telling how the story of the black queen ends (she does not live long to enjoy her triumph).

> The black queen was four squares north of the centre of the board. The third black pawn was to the right. (...) The black queen saw the third black pawn leaving to the right. (...) Three days later, the black queen moved southeast. The third white pawn remained behind. (..) The black queen saw the white queen appearing ahead. The black queen attacked the white queen.
>
> A month earlier three squares northwest, the white queen was three squares south of the centre of the board. (...) The white queen saw the black queen arriving. The black queen attacked the white queen.
>
> The white queen died. The black queen saw the white right bishop arriving. The white right bishop attacked the black queen. The black queen died.

6. Discussion

In appraising the proposed model it is important to consider that it is based on a number of hypotheses as to what the starting data on which it operates are. Although efforts have been made to start from the simplest possible representation of input data describing the set of events on which the final narrative must be based, it is possible that different initial assumptions might have led to different characteristics in the model.

Along these lines, it may be worth considering that any author facing the task of narrative composition would likely be operating from a memory of the set of events in question. This memory would mediate the task in two different ways. First, it may be remembered in incomplete or incorrect form. This would result in a narrative not matching the inspiring events through no conscious decision of the author or no explicit operation in the composition process. Second, the set of events as remembered may be interpreted by the author in different ways during the composition process. This task of interpretation would affect many of the decision points described in the model, mainly through the effect of the rhetorical problem and the constraints on the writing tasks, but possibly in further ways that have not been contemplated.

A number of related efforts exist to automatically derive narratives from sport games (Allen et al., 2010; Lareau et al., 2011; Bouayad-Agha et al., 2011). These efforts operate on input data in the form of statistics on a given game, and produce texts in the manner of newspaper articles covering similar games. These efforts, for instance, are heavily mediated by the set of data they start from, which arises from purpose specific abstraction and filtering of the set of real world events, driven by the purpose of the desired story.

From the point of view of narratological theory, the model as described captures a number of important features, such as concepts of focalization and chronology, and relates them closely to the input data, the decision criteria, and the computational processes being modelled. It also provides representation for issues such as person, tense and narrative time which are important characteristics of narrative as studied by narratologists.

The set of events before processing might be considered to represent the narratological concept of *story* as defined in section 2.1.. Another possibility might be that the story be the set of fibres obtained after heckling, or the subset of those fibres that get selected for inclusion in the final result. On this issue, the model presented in this paper has the merit of uncovering the degree of vagueness in existing description of these concepts in narratology. The concepts of fabula or story plot as a plan used by AI storytelling systems constitute a different more elaborate representation than the input considered in this paper. To obtain this refined representation it would be necessary to enrich the input with the causal inferences that link together all the events and then to select a subgraph of the resulting causal network to be used as driving plot, ommitting those events in the graph that are not included in that subgraph. This process will be considered in further work.

The structure proposed for yarns as it stands is very plain, as a result of the general aim of solving each problem with the simplest possible mechanism. However, as soon as more complex problems are addressed with the same framework, richer representation structures are bound to be needed, for instance to address the role of causal relations in the selection process. Existing narratological theories and causal models will surely be of use in extending and refining the model.

This paper addresses the task of obtaining a discourse representation from the representation of the input to a narrative composition process based on real life events. Although a representation that captured details of causal relations between events, along the lines of the causal network model (Trabasso et al., 1989), may play a significant role in this process at a deeper level, the work described here constitutes a first approximation that tackles relatively simple structural issues dealing with time and space, not causality. Yet it is clear that further work should address the role of causality in the various decision processes identified. This may present a significant obstacle, as difficult pragmantic inferences will have to be made to interpret the causal structure underlying observed events. This task is far from trivial.

Existing work on AI models of storytelling using planning approaches may be of great assistance in that endeavour. Planning features prominently both in the storytelling literature and in Flower and Hayes' account of writing. The planning applied in most storytelling systems involves drawing causal networks that connect events to one another, in order to provide a guiding line through a story or to suggest additional events that might be added to it. Similar solutions might be applied in order to identify potentially interesting connections between events in different fibres. For instance when events in one fibre can be causally related to events in another. Such connections could provide a very strong basis for ways of structuring the fibres into high quality yarns. The concept of author goals as postulated by Dehn relates very closely to the generic descriptions of the rhetorical problem and the constraints on the writing tasks discussed by cognitive accounts of writing. Both of these issues deserve attention in further work on this model.

Modifications of temporal ordering as considered in (Bae and Young, 2008) could inform the task of dividing an fibre into a number of fragments and recombining them in a different order. Additional options of introducing temporally displaced (possibly partial) versions of a given fragment (as well as or instead of the original fragment as told in its chronological position) must be considered as further work on the model.

The use of the tension arc for a story in the MEXICA system to evaluate the quality of a partial draft points the way towards very plausible and effective criteria both for selecting fibres during the planning stage and for evaluating yarns during the reviewing stage. These possibilities should be explored in further work, once a sufficiently rich representation is found, capable of representing emotion. As it is, the current prototype based on chess games applies very simple criteria for fibre selection, based on identifying fibres with the highest number of piece captures, under the assumption that captures constitute more emotionally charged events than other moves. The criteria for selecting break points, and for redistributing information around graft points also take into account intuititive criteria to sustain and maximise tension.

From a cognitive point of view, the set of operations postulated for the task of narrative composition aligns reasonably well with the processes described by Flower and Hayes. In terms of Flower and Hayes' model, heckling the original material into fibres, fibre selection and twisting fibres into yarns would constitute specific operations of the planning process. Contextualization, setting narrative parameters and realization would constitute operations of the translation process. The model for narrative composition as described in the present paper constitutes a very simple initial description of the task as a one-pass attempt. In more refined versions, the task should be addressed in a cyclic way, involving additional processes of evaluation and revision (the reviewing process of Flower and Hayes' model, currently not represented in the model), and allowing interaction between the various processes as controlled by a monitor.

Finally, a brief note on the use of chess games as indicative case study. Chess games present the advantage of having very clear temporal and spatial constraints, and constituting at heart a sketchy representation of one of the most dramatic settings for human experience: war. In that sense, it provides a good ground for simple experiments, and it is not intended as a contribution but as an illustrative example of the operation of the model of sufficient simplificity to be describable within the size restrictions of a paper such as this. Three aspects are identified as problematic with the chess domain. First, adequate representation of opportunities and threats in a chess game involves some serious representational challenges (Collins et al., 1991). Although significant progress may be made on this point, the effort invested is unlikely to lead to compelling narratives or narratives that bring insight on the task of narrative composition. Second, its extreme simplicity requires a number of additional operations (conversion of chronology to a framework in terms of days, or the introduction of compass points for specifying spatial directions) that may be introducing noise into the experiment. Third, the chronological structure of a chess game is in truth purely sequential, in contrast with the sets of events that would be considered when narrating from real life events. This has not been considered a serious obstacle in as much as focalization breaks up the set of events into separate views corresponding to different fibres, and the numbering or moves in the game provides a good indication of relative chronology of events within any given fibre and across fibres. Yet it also introduces unnecessary complexity when computing chronological allignment between paired break points, for instance, or in determining when jumps in time between successive event descriptions in a fibre should be flagged in the discourse. For these reasons, it is considered advisable to explore further investigation of the suitability of the model when applied to case studies in other domains that are richer in terms of their representation of time and space and that may lead to more compelling narratives with a stronger human interest. The search for alternative domains is made difficult by the need to obtain for them reliable records of all events, both relevant and irrelevant to the story that may be told. Only if this condition is satisfied can the corresponding problem be considered equivalent to the human task that we want to model.

7. Conclusions

The model presented in this paper constitutes a first approximation to a computational model of the task of narrative composition. It draws upon an analogy with textile manufacturing well-based on popular culture. This analogy has provided a break down into subtasks that has lead to interesting insights in terms of specific knowledge-based operations that need to be carried out during composition. These operations relate reasonably well with structural features of narrative as described in literary studies, such as focalization and chronology. They can also be correlated to the set of processes described in cognitive accounts of writing. Finally, they allign and integrate well with generally accepted task divisions for natural language generation. Additionally, a preliminary implementation over a simple indicative case study based on narrating chess games has shown the feasilibity of the approach in practical terms, as well as uncovering a number of elementary issues that arise from particular properties of the chosen domain. Overall, it seems fair to assume that the model might constitute a good starting point for further work both in terms of refining the model and extending the implementation to more complex case studies in other domains.

As specific research lines deserving attention, it is worth listing investigation into the role of causality in the various decisions processes, criteria for selecting interesting fibres to use in a composition, definition of procedures for trimming selected fibres to retain only the parts of them that justify their inclusion, and implementation of solutions for summarizing fragments of fibres that are uninteresting but relevant to the overall structure of the yarn. The model and the implementation should be extended to address in more detail the challenge of identifying appropriate breakpoint at which to splice fibres together, and the task of splicing together more than two fibres.

8. Acknowledgements

The author thanks the anonymous reviewers that contributed important insights to the final version of the paper. This research is partially funded by research grant (TIN2009-14659-C03-01) from the Spanish government.

9. References

N. D. Allen, J. R. Templon, P.S. McNally, L. Birnbaum, and K.Hammond. 2010. Statsmonkey: A data-driven sports narrative writer. In *Computational Models of Narrative: AAAI Fall Symposium 2010.*

B.-C. Bae and R. M. Young. 2008. A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Proc. ICIDS 2008.*

N. Bouayad-Agha, G.Casamayor, and L. Wanner. 2011. Content selection from an ontology-based knowledge base for the generation of football summaries. In *Proc. ENLG 2011*, pages 72–81.

G. Collins, L. Birnbaum, B. Krulwich, and M. Freed. 1991. Plan debugging in an intentional system. In *Proc. IJCAI 1991*, pages 353–359.

N. Dehn. 1981. Story generation after tale-spin. In *Proc. IJCAI 1981*, pages 16–18.

L. Flower and J.R. Hayes. 1981. A cognitive process theory of writing. *College Composition and Communication*, 32(4):365–387.

G. Genette. 1980. *Narrative discourse : an essay in method*. Cornell University Press.

A. Jhala and R. M. Young. 2010. Cinematic visual discourse: Representation, generation, and evaluation. *IEEE Trans. on Comp. Int. and AI in Games*, 2(2):69–81.

F. Lareau, M. Dras, and R. Dale. 2011. Detecting interesting event sequences for sports reporting. In *Proc. ENLG 2011*, pages 200–205.

M. Lebowitz. 1983. Story-telling as planning and learning. In *Proc. IJCAI 1983*, volume 1.

James R. Meehan. 1977. Tale-spin, an interactive program that writes stories. In *Proc. IJCAI 1977*, pages 91–98.

R. Pérez y Pérez. 1999. *MEXICA: A Computer Model of Creativity in Writing*. Ph.D. thesis, The University of Sussex.

E. Reiter and R. Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

M. Riedl and Michael Young. 2010. Narrative planning: Balancing plot and character. *J. Artif. Intell. Res. (JAIR)*, 39:217–268.

T. Trabasso, P. vand den Broek, and S.Y. Suh. 1989. Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12:1–25.

Scott R. Turner. 1993. *Minstrel: a computer model of creativity and storytelling*. Ph.D. thesis, University of California at Los Angeles, Los Angeles, CA, USA.

T. A. van Dijk and W. Kintsch. 1983. *Strategies of Discourse Comprehension*. New York: Academic Press.

“Is this a DAG that I see before me?”
An Onomasiological Approach to Narrative Analysis and Generation

Michael Levison, Greg Lessard

School of Computing, French Studies
Queen’s University, Kingston, Ontario, Canada
levison@cs.queensu.ca, greg.lessard@queensu.ca

Abstract

We present a framework for the analysis of literary texts by means of a semantic representation based on the use of directed acyclic graphs which may be threaded in various ways to represent elements of plot, character perspective, narrative sequencing and setting. The model is illustrated by application to a simple fairy tale and to a Sherlock Holmes story. We argue that it is possible to represent in this way, in a manner accessible to non-computer scientists, the high-level dependencies which underlie a text as well as particular characteristics of literary texts, including the use of various recurring narrative sequences. We provide examples of the functional representation used, of the graphical representations achieved and the results obtained when the semantic representations are used to drive a natural language generator.

Keywords: directed acyclic graphs, narrative structure of literary texts, natural language generation

1. Background

The formal representation of narrative structure[1] is a broad and complex field with many overlapping perspectives. We will begin by outlining briefly how the approach adopted here is related to some previous research.

Firstly, to use a longstanding but infrequently used term, our approach is **onomasiological**, in that it concerns itself with the meaning of a text independently of its realization in some natural language. It differs in this way from **semasiological**[2] analyses which take some particular instantiated text and seek to establish how its formal devices (words, syntax, etc.) give rise to the meaning which it carries, as, for example, in the case of Hobbs et al. (1993).

Secondly, it makes use of a detailed system of semantic representation which makes possible fine semantic distinctions. An example will perhaps illustrate this. Trabasso and Van den Broek (1985) present a short text containing the sentence “He lifted the boat up with a stick and found a turtle on top of it” and the subsequent sentence “Mark had always wanted Sally to see a turtle.” In their analysis, Trabasso and Van den Broek consider the second of the two sentences to be causally linked to the first, through the turtle. Note however that the meaning of the word *turtle* is not the same from one sentence to the next. In the first sentence, the word designates an individual (the specific turtle) while in the second it refers to an intensional perspective (some member of the class of turtles, but not necessarily this particular turtle). As we will see later in this paper, the relation between an individual and an intensional class cannot easily be construed as causal, or if it is, it represents a very special sort of causality. However, only a fine-grained representation of meaning can show this.

Thirdly, as we will see in more detail below, we represent the relations between the semantic elements of a text by means of directed acyclic graphs (DAGs). This has a number of advantages. Unlike models based on linear sequences or phrase structure grammars (such as Mandler and Johnson (1977)), nodes of the DAG may depend on more than one antecedent, a necessary trait in dealing with complex texts.[3] In addition, unlike models like Rhetorical Structure Theory (Mann and Thompson, 1988) which rely on labelled arcs, our representation limits its characterization of links strictly to dependency, which provides a higher level of abstraction. This approach has strong similarities to work such as Riedl and Young (2006), Cheong (2007) and others, which also makes use of DAGs. In some of those cases, though, the application of the model to video games or computer-generated narratives leads to links among the nodes of the graph being expressed in terms of user choices—a concept not found in literary texts—and having the potential for cyclicity, also absent from typical literary texts.

Fourthly, the majority of studies of narrative structure are typically based on short passages created for use in an experimental context. In contrast, we analyze existing, relatively complex, literary texts. In this paper, we focus our attention on mystery stories, specifically a story by Sir Arthur Conan Doyle. Such texts may be seen as sitting midway along the continuum of literary complexity. They are more complex than simple folktales which rely primarily on plot and where stock characters have constant traits.[4] But they are less complex than some novels in which characters evolve over the course of the narrative. However, as literary texts, mystery stories do bring into play the no-

[1] We will not discuss here the relations between **narrative** and other forms of discourse like **description**, **report**, **information** or **argument**. For discussion, see Smith (2003).

[2] For a discussion of the distinction between semasiology and onomasiology, see Baldinger (1964).

[3] See for example Black and Bower (1980) for a discussion of the weaknesses of Finite State Grammars and Phrase Structure Grammars in capturing complex narratives.

[4] Thus, a wicked witch is always wicked, the hero is always brave and good, etc. See Propp (1968).

tion of **genre**, that is, a set of constraints and expectations understood by authors and readers, and, as we will see below, access to libraries of **topoi**, or recurring narrative sequences, both relatively rarer in simple constructed texts. In addition, literary texts of good quality avoid the criticism addressed by Magliano and Graesser (1991) that experimenter-generated texts are often "disjointed, pointless and bizarre". Of course, there is a price to be paid for the use of existing literary texts: it is harder to control all variables, which makes it more difficult to measure with precision psycholinguistic phenomena such as retention over time or analysis of the effects of links on processing, and computational modelling becomes more complex.[5]
Finally, in light of our goal of studying literary texts, we have sought to provide a system of representation usable not just by computer scientists but also by linguists or literary specialists. To achieve this, the system of semantic representation chosen is designed to be relatively intuitive and to generate not just textual but graphical output. The examples presented below illustrate this goal.

2. Story, Threading and Narrative

Following on earlier work, including analyses by the Russian formalists, Gérard Genette distinguished three perspectives on a narrative text: **histoire**, that is, the overall collection of information which underlies a narrative, **narration**, that is, the action of recounting the elements of the histoire, and **récit**, that is, the product of a particular telling.[6] This distinction has given rise to a long series of literary debates that we will not explore here.[7] One also finds terminological diversity. So, for example, some, like Riedl (2004) and Cheong (2007), use the terms **fabula** and **sjuzhet** for the distinction between histoire and récit. In what follows, we will use the terms **story** for the underlying information, **threading** for the action of recounting some or all of the elements of the story, and **narrative** for the product of a particular threading.

2.1. Story

Our central thesis is that the essentials of a **story** can be captured, at different levels of granularity, by a series of directed acyclic graphs, or DAGs. A directed acyclic graph consists of a set of nodes, connected by unidirectional paths (technically called edges). That the graph is acyclic implies that no sequence of paths loops back to a node already visited.[8] The paths of a DAG represent a type of relation between the nodes known as a partial ordering.
In a story DAG, the nodes represent pieces of meaning, each at varying levels of granularity. Each node is denoted by a semantic expression in the form of a function call. The formalism used to express these function calls is described in Levison et al. (in press). We will provide simplified examples of it below.

[5]See however Zwaan and Radvansky (1998), Magliano and Graesser (1991) and Wiebe (1994) for examples of work based on literary passages.

[6]See Genette (1983) for a clear exposition of the issue.

[7]For an overview, see Culler (2002).

[8]As we noted earlier, this is true for essentially all literary texts. It may not be true for computer-generated narratives.

The paths of the DAG may be viewed in one of two ways. If they are downwards, as shown in Figure 1, they may be understood as representing the relation: `X is a prerequisite for Y` (i.e. a logical antecedent). If they are reversed, to point upwards, the partial ordering represented by the DAG may be described as dependency: `Y depends on X`.[9]
The DAG at the finest level of granularity carries the complete meaning of the story while the DAGs at coarser levels can be thought of as plot, abstract, synopsis, and précis at various levels of detail. We may expect that the nodes of the coarser DAGs may be developed into finer DAGs in the more refined forms. This has some psychological plausibility: for example Black and Bower (1979) provide evidence of the encapsulation by higher levels of lower ones in their work on episodes as chunks in memory.[10]

2.2. Threading and Narrative

A thread is a traversal of some or all of the nodes of a DAG. It does not necessarily follow the paths of the DAG.[11]
Some threads correspond to narratives. The semantic expressions represented by the nodes on the thread then carry the meaning of the narrative, and might be expressed in some particular natural language by a natural language generator.[12] Other threads might be chronological, showing alternative temporal orderings of events which respect the dependencies of the story DAG. Thus, in a story DAG composed of nodes A, B, and C, where B and C both depend on A but are independent of each other, one chronological threading may have B occur before C, while another threading may have C occur before B. In fact, some texts provide sufficiently imprecise chronological signals that it is impossible to determine which of B or C is chronologically earlier.[13]
Narrative threading may take place along many different dimensions. For example, events may be presented such that earlier events in the story DAG are recounted first, or alternatively, so that later events are recounted first, with 'flashbacks' to earlier events. Another sort of threading is based on the point of view of different characters. So, a murder mystery might be told from the point of view of the detective, of an observer (think of Holmes' Watson), or even of the murderer, although this would reduce the mystery. A third type of threading is based on degree of knowledge, or who knows what when. Thus, in the omniscient narrator framework, the thought processes of characters are made

[9]Strictly speaking, the relation should be 'depends directly on', i.e. in a single step. The relation 'depends on' is really the multi-step extension, its transitive closure.

[10]We will not discuss here various mechanisms for the manipulation of DAGs such as **Hierarchical State Transition networks** (Black and Bower, 1980) or the **IPOCL** planner (Riedl, 2004).

[11]Our notion of **thread** is similar to the concept of **narrative chain** defined by Chambers and Jurafsky (2009) as "a partially ordered set of narrative events that share a common actor". Note however that we allow for multiple dimensions of threading. See below.

[12]We have concealed here some complexities associated with natural language generation, such as the calculation of pronoun choice and some issues of adverbial choice.

[13]Think of the use of expressions like "meanwhile..." in a novel.

visible to the reader, while in a first person narrative, the thoughts of other characters remain invisible.

3. A basic illustration

Let us illustrate the story/narrative distinction by means of a simple fairy tale. We begin with a narrative (threading of the DAG) recounted in chronological order:

> Once upon a time, in the kingdom of Lobelia, there lived a pretty princess called Goldilocks. Her father, the King, facing a severe budget deficit, decided to levy a heavy tax, which became known as a Witch Tax, on the licences required by those who cast spells. In retaliation, the wicked witch of the East kidnapped Goldilocks, intending to hold her to ransom until the hated tax was repealed. (...) Prince Charming finally defeated the witch and rescued the princess, who immediately fell in love with him, and they agreed to marry.

The story of this narrative might be represented at a very coarse level by the DAG shown in Figure 1, where the functions have been given names which suggest their meanings and the paths represent the dependency relations among nodes.

Thus, for `ww_of_east` to kidnap `Goldilocks`, it is important for Goldilocks to be the daughter of the king, for `ww_of_east` to be a member of the coven, and for the king to have offended the coven. For the prince to rescue the princess, there must be a prince and something to rescue her from. In short, the rescue depends on the kidnapping and the princely introduction, while the kidnapping depends on three other eventualities, and so on. These are independent of a temporal (or chronological) antecedent ('X must take place before Y') and a textual antecedent ("Prince Charming married the Princess. This was the culmination of his brave rescue when the Princess was kidnapped by the wicked witch.")

Another possible threading of this story DAG is represented by the following passage, which begins at the end of the sequence of events:

> On a bright sunny Spring day, with crowds of people lining the route, the Prince and Princess were driven to their wedding ceremony in a handsome (!) carriage. Crowds of people lined the route to see the prince and princess on their wedding day. It was the culmination of a turbulent series of events. Prince Charming had first met the beautiful Princess Goldilocks when he rescued her after her kidnapping by the wicked witch of the East. Her rich and grasping father, King of the Lobelians, had offended the Society of Witches, who had decided to retaliate by holding the Princess to ransom until the King repealed his new Witch Tax. (...)

4. A more complex example

To illustrate some of the computational machinery of our model, we will turn now to an example drawn from the realm of detective fiction. Although more complex than simple fairy tales, this genre retains a relatively high level of structure and constraint.[14] So, for example, the reader of a typical example of the genre will be presented with some mystery, the detective will be confronted with the elements to be explained, the detective will reconnoitre the scene of the crime, come to some conclusion and then resolve the mystery.

The Conan Doyle short story *The Red-Headed League* provides a good illustration of this 'grammar'. In this story, Watson and Holmes are visited by a pawnbroker, Wilson, who recounts that, on the advice of his assistant, Spaulding, he had applied for, and been given, a particularly easy job which consisted in copying the Encyclopaedia Britannica each morning. The job was offered by a League devoted to the betterment of red-headed men, of which Wilson is one. However, after two months of work, Wilson is surprised to read an announcement on the door of the office where the copying takes place that the position has abruptly ended. Puzzled, he consults Holmes. Holmes asks several questions, discovers that Spaulding has recently been hired and works for less than the usual amount. Holmes visits the pawnbroker's shop, asks some directions of Spaulding and then examines the surrounding neighbourhood. He then departs and recounts to Watson and the police that Spaulding is in fact Clay, a criminal, who is intent on robbing a bank situated near the pawnbroker's shop. Holmes, Watson, the police and a banker hide in the dark near the bank vault and intercept Clay and his accomplice as they emerge from the tunnel which they have dug between the basement of the pawnbroker's shop and the bank during the periods when Wilson was absent copying the encyclopedia.

One part of the story DAG which underlies this narrative is the plot to rob the bank planned and implemented by Clay. This may be represented as shown in Figure 2.

Of course, each of the functions which make up the nodes of the DAG shown in Figure 2 is itself composed of more detailed functions. The interactions between these functions may be quite complex. We will illustrate this with three examples.

4.1. Aliases

In the Read-Headed League, the criminal John Clay takes the alias Vincent Spaulding when he works for Wilson, the pawnbroker. Wilson is unaware of the alias, so for him, Spaulding is Spaulding. Holmes, on the other hand, has previous knowledge of John Clay and after asking directions of Spaulding becomes aware of the alias. We may represent this state of affairs by Figure 3. In effect, the characters retain their separate identity in parts of the DAG which are not dependent on the alias. Thus, the nodes involving Wilson occur in the area of the DAG denoted by `P`, while any nodes relating to Clay but independent of the alias are in the area denoted by `Q`. Nodes which reveal the alias to a specific character or a group of characters occur in `R`, and the nodes which make use of the alias are dependent on these.

[14] For discussion, see Scaggs (2005).

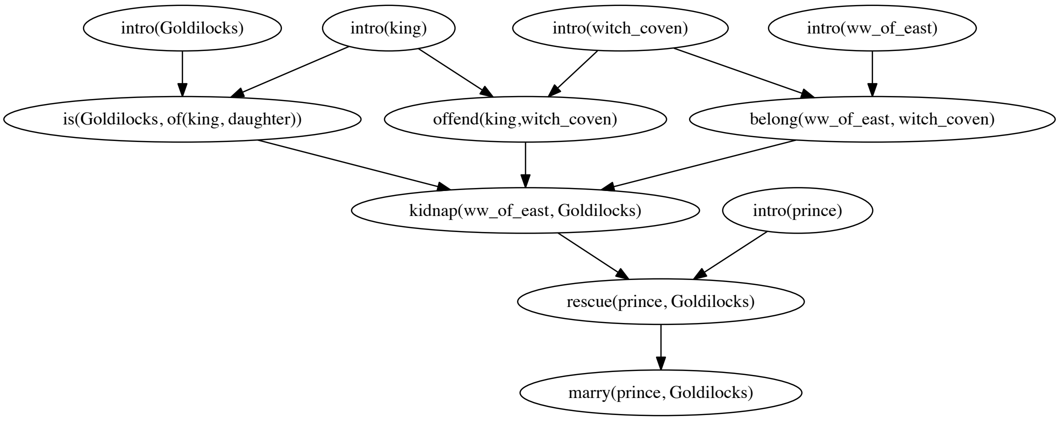


Figure 1: A coarse DAG representing the meaning of the fairy tale.

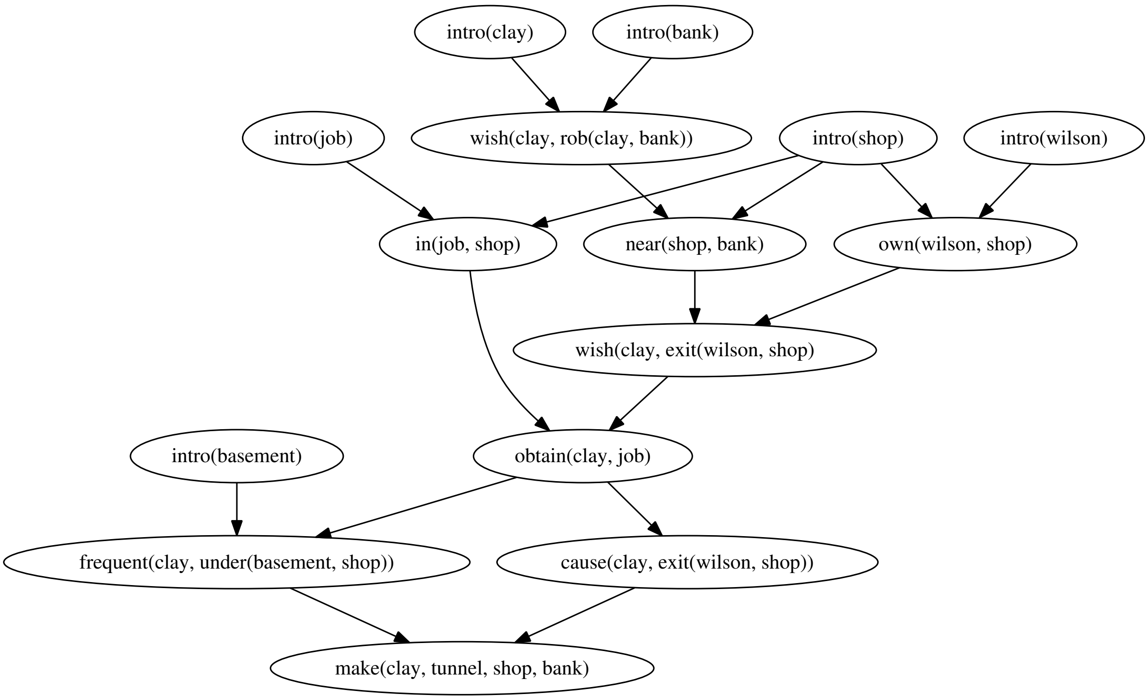


Figure 2: The DAG underlying the plot to rob the bank.

Of course, for the mystery to remain, the narrative thread must not visit the revelation nodes too early.

4.2. Point of view

Much narrative fiction also relies the use of varying points of view on some fictional world. Thus, different characters in a novel may differ in what they know, when they know it, and the attitude they take with respect to what they know. This may be represented by means of different narrative threadings of a story DAG. We will illustrate this state of affairs here by considering the point of view of Wilson in the Red-Headed League, which may be represented functionally as shown in Table 1.[15] The mystery which Wilson presents to Holmes is merely his perspective on the plot at a particular moment in time.

It is interesting to note, in passing, that in the telling of the story produced by Conan Doyle, it is never stated that Holmes explains the mystery to Wilson.

4.3. Topoi

The use of an alias is not peculiar to the Red-Headed League. Instances may be found in other Holmes stories

[15]We have simplified some elements of this formalism to enhance readability. Note also that John Clay has an accomplice, Archie, known to Wilson by the alias Ross.

```
own(wilson, shop)
employ(wilson, spaulding)
give(wilson, spaulding, remuneration[little])
go[often](spaulding, under(basement, shop))
need(wilson, money)
show(spaulding, wilson, advertisement)
say(advertisement, UNSPEC, seek(rhl, has(man[intension], qual(hair, red)))
meet(wilson, ross)
is(ross, of(secretary, rhl)
employ(ross, wilson, copy(wilson, encyclopedia))
require(ross, wilson, remain[every.morning](wilson, office))
copy[for.time.2.months](wilson, encyclopedia)
suspend[sudden](ross, wilson, copy(wilson, encyclopedia))
become(wilson, surprised)
```

Table 1: The functions behind Wilson's point of view.

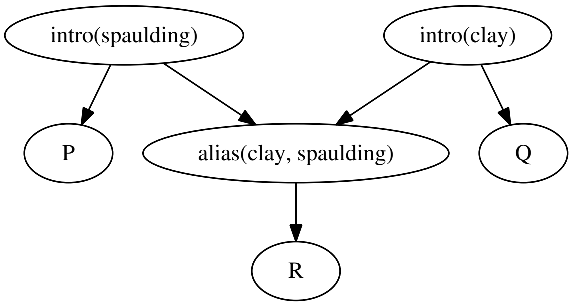


Figure 3: The `alias` DAG.

— Holmes himself has notable skill in disguise — and throughout literature. In the Red-Headed League, no reason is given as to why Clay needs an alias in his dealings with Wilson. In murder mysteries, however, aliases are created implicitly. Thus, following a murder, there comes into existence "the murderer", which is really an alias for some other character, perhaps "the butler", and unveiling the alias is the principal goal of the detective.

In light of this, we may think of the `alias` function as being drawn from a 'library' which will be available, in more or less richness, to authors and to readers.[16] More generally, we may think of a story DAG not as a creation *ex nihilo* in all its parts, but rather as the assemblage of a collection of often previously existing elements. Some genres like detective fiction or Harlequin Romances have a high level of re-use of topoi, while others do not. In fact, it could be argued that the presentation of fictional worlds, which are radically underspecified (no novel provides more than a sketch of most details)[17] relies crucially on such reuse of topoi.

[16]See Lessard et al. (2004) and de Rosnay et al. (2006) for discussion. Note also that the concept of the topos has links to the notion of **scripts**, proposed by Schank and Abelson (1977) and reused by others like Raskin (1985).

[17]See Pavel (1986) for details.

5. Elaboration and Interpolation

Operations of elaboration and interpolation, analogous to those described by Thomas (2010) for semantic trees, but rather different in detail, may be applied to DAGs as well, allowing a more detailed DAG to be derived from a coarser one. For example, the node: `kidnap(ww_of_east, Goldilocks)` may well be elaborated into a DAG of its own, describing how Goldilocks was walking on the river bank near the castle with her pet dog, when `ww_of_east` grabbed the dog and used it to lure the Princess into her cottage; while: `rescue(prince, Goldilocks)` might involve the prince visiting a wise magician to obtain a cape of invisibility, to use while he crept up to the cottage and carried the drugged Princess to safety.

There are several variant types of interpolation. A new node may be inserted between the prince's introduction and the rescue, in which the king writes to the prince begging him to rescue his daughter. This node will depend not only on the kidnapping, but on both the king's and the prince's introductions. Another node may be interpolated between the request and the rescue, in which the prince searches to locate the heroine, and so on. The second of these is an interpolation in an existing path; the first creates a node which depends on two unrelated ones, and calls for two new paths to be created.

These operations highlight a distinction between the DAGs proposed here and the semantic trees discussed by Thomas. In the latter, paths convey no meaning beyond the fact that the child-node is part of a refinement of the parent. There is no implication of dependence. The tree retains the meaning of the text at all levels of refinement, but no connection is made between related nodes in separate parts of the tree.

By contrast, in the elaboration of a story DAG, if the introduction of Goldilocks is refined to mention her pet dog and her penchant for walking along the river bank, some nodes in the elaboration of the kidnapping might be dependent on these. In other words, when a node in a DAG is elaborated, its paths from and to other nodes may need to be refined – restricted to only a few of its nodes.

This raises two questions. An important feature of a semantic expression is its ability to 'hide' details of a subexpression within a function, in effect a black-box environment, allowing it to be understood easily at different levels. Is this property inhibited by the difference between DAGs and semantic trees? Not necessarily, but a DAG at a finer level will need supplementary information to indicate nodes which should be coalesced to form a coarser one; while paths entering or leaving individual nodes of the group must now be considered as entering or leaving the group itself. We prefer at this time to conceive of the story as embodying a set of DAGs at different levels of gradation rather a single all-embracing one, but that is simply a choice.[18]

And, does it matter? Probably not. The eventual purpose of representing the meaning of a story by a DAG is to have it yield threaded tours. Each of these is itself a semantic expression, and has the hierarchical property, though at the expense of suppressing the dependencies. The dependencies, of course, make an important contribution to the threading process. In the narrative, however, the author may not mention them explicitly, but rather, leave them to the readers' imagination.

An example may help to illustrate this. In Shakespeare's Romeo and Juliet, the star-crossed lovers use friars as their means of communication. At a crucial moment, a message is delayed by the Plague. In a modern version, the lovers would probably communicate by cell-phone, and Romeo might be unable to listen to his voice-mail because of an electrical storm. In the meaning of the story, the breakdown of communications depends on some external, but significant, event. In a narrative, the author might just mention the event as an aside, at a place not specifically related to the crucial message: "It was a dark and stormy night."

6. From semantic representation to textual generation

Of course, the representations we have seen thus far exist at the semantic level. In order for them to be instantiated as text, some system of generation is required. Various solutions have been proposed to this problem, among them Meehan (1977) and Callaway and Lester (2002). We propose here two models for generation, one simple, the second more complex.

6.1. Trotter

At the most basic level, Thomas (2010) as part of his thesis, created a piece of software called **trotter** which reads in a semantic lexicon, including type signatures, formal and natural language equivalents for each function, and then processes some function call or set of function calls to produce either the full spelling out of the functions themselves or some basic natural language representation.

The starting point for this process is the specification of a **semantic lexicon** of which Table 2 provides an excerpt.

[18]Although the storage of a complex tree or DAG structure in a computer presents no problem, a graphical representation in hard-copy is impractical at any but the coarsest level. To display a diagram on a screen to visualize a story or narrative in fine detail might be achieved with software which can zoom in on a small area of the structure.

Each item in the semantic lexicon contains three fields separated by vertical bars: the **type specification** of the semantic entity, its **functional representation**, and a spelling out of the meaning in an informal natural language string.

So, for example, the semantic expression `holmes` belongs to the class of entities, requires no arguments and has a natural language representation of "Holmes". On the other hand, the expression `arrest` is a function which takes two entities as arguments and returns an `action`. The arguments of `arrest` are given the designators `[en1]` and `[en2]`. More mnemonically friendly designators might have been chosen, such as `[arrester]` and `[arrestee]`, and synonyms are also possible. The coarse natural language representation of the function is "`[en1]` arrests `[en2]`", where the items in square brackets are replaced by the natural language strings (field 3) of the two entities passed to the `arrest` function.

On the basis of a semantic lexicon, **trotter** outputs either a formal functional representation based on the contents of fields 1 and 2 of the semantic lexicon, or a rough natural language output based on field 3 of the semantic lexicon. So, for example, the portion of the Red-Headed League in which Holmes reconnoitres the pawnshop and establishes the identity of Spaulding is represented formally as follows:

```
examine(holmes(),shop()),
find(holmes(),near(shop(),bank())),
seek(holmes(),spaulding()),
see(holmes(),clay()),
find(holmes(),
  between(tunnel(),shop(),bank())))
```

while the informal spelling out in natural language looks like this:

Holmes examines shop,
Holmes finds shop is near bank,
Holmes seeks Spaulding,
Holmes sees John Clay,
Holmes finds tunnel is between shop and bank

6.2. VINCI

Of course, output of **trotter**, while it provides a sense of the gist of a text, fails to capture linguistic phenomena like agreement, inflection, tense and so on. A more complex linguistic representation may be obtained by using the sequence of semantic expressions produced as input to a natural language generation system. In our own case, we make use of the VINCI NLG, which we have developed over a number of years.[19] So, for example, the semantic representation of a fairy tale like the one presented earlier in this paper, after processing by VINCI, gives rise to the following output in English:[20]

> Once upon a time there was a king called Midas who lived in a castle. He was rich and vain.

[19]For details, see www.cs.queensu.ca/CompLing.

[20]These examples are drawn from Levison and Lessard (2004).

```
holmes ::  entity|holmes()|Holmes
bank ::  entity|bank()|bank
red ::  quality|red()|red
arrest ::  entity,entity -> action|arrest(en1, en2)|[en1] arrests [en2]
succeed ::  entity,action -> action|succeed(en1,act1)|[en1] succeeds at [act1]
intension ::  entity -> entity|intensional(en1)|a member of the class of [en1]
for ::  action,entity -> quality|for(act, time)|[act] lasts for [time]
reason ::  action,action -> action|reason(act1,act2)|[act1] in order that [act2]
```

Table 2: Some items from the semantic lexicon for the Red-Headed League.

> The king had a daughter, a princess named Marie, who was beautiful. The king warned Marie not to go out of the castle. The princess disobeyed the king. She left the castle. A sorcerer called Merlin lived in the woods. He was evil. The sorcerer kidnapped the princess. Nearby there lived a woodcutter who was named Axel. The king sought the help of the woodcutter. The woodcutter went to look for the fairy godmother. The fairy godmother passed Axel a magic sword. Axel searched for the sorcerer. The woodcutter killed the sorcerer with the magic sword. The woodcutter rescued the princess. The woodcutter and the princess got married and lived happily ever after.

In fact, since the semantic representation is itself language-independent, output is possible in more than one language. Thus, the same semantic representation which gave rise to the previous text also generates a parallel text in French:

> Il était une fois un roi qui s'appelait Midas et qui vivait dans un beau château. Il était riche et vain. Le roi avait une fille, une princesse qui s'appelait Marie et qui était belle. Le roi interdit à Marie de quitter le château. La princesse désobéit au roi. Elle quitta le château. Dans la forêt il y avait un sorcier qui s'appelait Merloc. Il était méchant. Le sorcier enleva la princesse. Aux alentours vivait un prince qui s'appelait Coeur de Lion et qui était beau. Le roi demanda l'aide du prince. Le prince chercha la bonne fée. La bonne fée donna une épée magique au prince. Le prince chercha le sorcier. Coeur de Lion utilisa l'épée magique pour tuer le sorcier. Le prince libéra la princesse. Le prince épousa la princesse et ils eurent beaucoup d'enfants.

7. Conclusions and Future Work

The concepts and examples discussed in this paper have focused primarily on elements of plot in a limited class of narratives, but we believe that the results obtained suggest that a model of literary texts based on an onomasiological perspective, which distinguishes the story from the narrative and which represents the former by means of a DAG and the latter by means of a threading of the DAG, is interesting from both the literary and the computational perspective. Among other things, it offers a framework for empirically verifying different literary narrative models through the computational representation of narrative structures at both the abstract level and, with the addition of natural language generation, at the textual level. In essence, the goal is to pass from the literary text to its abstract representation at various levels, and back again.

We believe that there is potential for interaction between the approach presented here, whose primary goal is the modeling of existing literary texts, and research aimed at the synthesis of narratives. Despite the fact that, as Riedl and Young (2006) point out, there are differences between what they call linear narratives found in novels and branching narratives found in video games, we believe that the synthesis of fabula and in particular the degree of **postdictability**, to use the term proposed by Kintsch (1980) to refer to the reader's sense that everything fits together, would be enriched by access to 'libraries' of formally represented fabula drawn from existing texts. This follows on from the call by Mandler and Johnson (1977) for a "whole earth catalogue" of frequent semantic sequences, the use of libraries of existing human-generated stories by Pérez y Pérez and Sharples (2001), and more recently the use of existing 'vignettes' as models for stories (Riedl, 2010).

Of course the question is how to assemble this corpus of structures. Parsing of existing texts is certainly one option—see for example Chambers and Jurafsky (2009)—but parsers still lack the degree of sophistication inherent in literary analyses. A number of projects in humanities computing are beginning to fill this gap, notably in the capture and representation of topoi,[21] but both these approaches are limited to the capture and representation of isolated topoi and fail to show the interplay of topoi within texts. Our goal is to fill this gap by providing a representation system accessible to and understood by literary specialists and linguists in order to build libraries of text-level semantic representations of structure.

In addition, as we have shown elsewhere (Lessard and Levison, 2005), narrative descriptions in literary texts tend to follow particular patterns. For example, the description of a room may begin in the foreground, then move to the background, may 'scan' a group of people, and so on. These may be seen as particular threadings of a DAG. According to Zwaan and Radvansky (1998), for example, this struc-

[21] See the **SatorBase** project (www.satorbase.org) and work by the **TopoSCan** project (Lessard et al., 2004)

turing is not fortuitous. To borrow their term, **situation models** along dimensions of time, foregrounding, space, causality and other dimensions, play a crucial role in text comprehension. The study of various models of threading offers, we argue, advantages similar to those provided by the construction of libraries of DAGs in helping us to understand this process.

8. Acknowledgements

The research described here has been made possible by a Standard Research Grant from the Social Sciences and Humanities Research Council of Canada.

9. References

K. Baldinger. 1964. Sémasiologie et onomasiologie. *Revue de linguistique romane*, 28:249–272.

J.B. Black and G.H. Bower. 1979. Episodes as chunks in narrative memory. *Journal of Verbal Learning and Verbal Behavior*, 18(3):309–318.

J.B. Black and G.H. Bower. 1980. Story understanding as problem-solving. *Poetics*, 9:223–250.

C. Callaway and J. Lester. 2002. Narrative prose generation. *Artificial Intelligence*, 139(2):213–252.

N. Chambers and D. Jurafsky. 2009. Unsupervised learning of narrative event chains. In *Proceedings of ACL/HCL*.

Y.G. Cheong. 2007. *A computational model of narrative generation of suspense*. Ph.D. thesis, Computer Science, North Carolina State University.

J.D. Culler. 2002. *The pursuit of signs: semiotics, literature, deconstruction*. Cornell University Press, Ithaca, N.Y.

É. Fromet de Rosnay, G. Lessard, S. Sinclair, F. Rouget, M. Vernet, É. Zawisza, L. Blumet, and A. Graham. 2006. À la recherche des topoï romanesques: le project TopoSCan. In D. Maher, editor, *Tempus in Fabula: Topoï de la temporalité narrative dans la fiction d'Ancien Régime*, pages 21–32. Presses de l'Université Laval.

G. Genette. 1983. *Nouveau discours du récit*. Éditions du Seuil, Paris.

J.R. Hobbs, M.E. Stickel, D.E. Appelt, and P. Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

W. Kintsch. 1980. Learning from text, levels of comprehension, or: why anyone would read a story anyway. *Poetics*, 9(87–98).

G. Lessard and M. Levison. 2005. Representation of spatio-temporal relations in human and machine-generated texts. In *New Directions in Cognitive Linguistics Conference*, pages 64–65, Brighton, UK.

G. Lessard, S. Sinclair, M. Vernet, F. Rouget, E. Zawisza, L.-É. Fromet de Rosnay, and É. Blumet. 2004. Pour une recherche semi-automatisée des topoï narratifs. In P. Enjalbert and M. Gaio, editors, *Approches sémantiques du document électronique*, pages 113–130. Europia, Paris.

M. Levison and G. Lessard. 2004. Generated narratives for computer-aided language teaching. In Lothar Lemnitzer, Detmar Meurers, and Erhard Hinrichs, editors, *COLING 2004 eLearning for Computational Linguistics and Computational Linguistics for eLearning*, pages 26–31, Geneva, Switzerland.

M. Levison, G. Lessard, C. Thomas, and M. Donald. in press. *Semantic Representation of Natural Language*. Continuum, London.

J.P. Magliano and A.C. Graesser. 1991. A three-pronged method for studying inference generation in literary text. *Poetics*, 20:193–232.

J.M. Mandler and N.S. Johnson. 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9:111–151.

W.A. Mann and S.A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

J.R. Meehan. 1977. Tale-spin, an interactive program that writes stories. In *Proceedings of the 5th international joint conference on Artificial intelligence - Volume 1*, IJCAI'77, pages 91–98, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

T. Pavel. 1986. *Fictional Worlds*. Harvard University Press, Cambridge, MA.

R. Pérez y Pérez and M. Sharples. 2001. Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, 13(2):119–139.

V. Propp. 1968. *Morphology of the Folktale*. University of Texas Press, Austin, TX.

V. Raskin. 1985. *Semantic mechanisms of humor*. Reidel, Dordrecht.

M. Riedl and R.M. Young. 2006. From linear story generation to branching story graphs. *IEEE Computer Graphics and Applications*, May/June:23–31.

M. Riedl. 2004. *Narrative Generation: Balancing Plot and Character*. Ph.D. thesis, Computer Science, North Carolina State University.

M. Riedl. 2010. Case-based story planning: Creativity through exploration, retrieval, and analogical transformation. *Minds and Machines*, 20(4):589–614.

J. Scaggs. 2005. *Crime fiction*. Routledge, London.

R.C. Schank and R.P. Abelson. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawerence Erlbaum Associates, Hillsdale, NJ.

C.S. Smith. 2003. *Modes of discourse: the local structure of texts*. Cambridge University Press, Cambridge.

C. Thomas. 2010. *The Algorithmic Expansion of Stories*. Ph.D. thesis, School of Computing, Queen's University.

T. Trabasso and P. Van den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24:612–630.

J.M. Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.

R. Zwaan and G. Radvansky. 1998. Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–185.

Automatically Learning to Tell Stories about Social Situations from the Crowd

Boyang Li, Stephen Lee-Urban, Darren Scott Appling, and Mark O. Riedl
School of Interactive Computing; Georgia Institute of Technology
Atlanta, Georgia 30032 USA
E-mail: {boyangli, lee-urban, darren.scott.appling, riedl}@gatech.edu

Abstract

Narrative intelligence is the use of narrative to make sense of the world and to communicate with other people. The generation of stories involving social and cultural situations (eating at a restaurant, going on a date, etc.) requires an extensive amount of experiential knowledge. While this knowledge can be encoded in the form of scripts, schemas, or frames, the manual authoring of these knowledge structures presents a significant bottleneck in the creation of systems demonstrating narrative intelligence. In this paper we describe a technique for automatically learning robust, script-like knowledge from crowdsourced narratives. Crowdsourcing, the use of anonymous human workers, provides an opportunity for rapidly acquiring a corpus of highly specialized narratives about sociocultural situations. We describe a three-stage approach to script acquisition and learning. First, we query human workers to write natural language narrative examples of a given situation. Second, we learn the set of possible events that can occur in a situation by finding semantic similarities between the narrative examples. Third, we learn the relevance of any event to the situation and extract a probable temporal ordering between events. We describe how these scripts, which we call plot graphs, can be utilized to generate believable stories about social situations.

Introduction

Storytelling, in oral, visual, or written forms, plays a central role in various types of entertainment media, including novels, movies, television, and theatre. The prevalence of storytelling in human culture may be explained by the use of narrative as a cognitive tool for situated understanding (Bruner 1991; McKoon & Ratcliff 1992; Gerrig 1993; Graesser, Singer & Trabasso 1994). This *narrative intelligence* (Mateas & Sengers 1999) is central in the cognitive processes employed across a range of experiences, from entertainment to active learning. It follows that computational systems possessing narrative intelligence may be able to interact with human users naturally because they understand collaborative contexts as emerging narrative and are able to express themselves by telling stories.

In this paper we consider the problem of creating and telling stories that involve common social situations. Most stories are about people (or objects and animals that behave like people in some way). Characters in generated stories should respect social and cultural norms, and perform common tasks in socioculturally acceptable ways. For example, during a trip to a restaurant, a character should perform actions that meet readers' expectation of what should happen in a restaurant. Further, to generate a love story in which a boy asks a girl out to a date at the movies, a system should know when it is okay for the boy to hold the girl's hand or when to try for a kiss. To omit these elements or to use them at the wrong time invites failures in believability or breakdowns in communication.

The generation of believable stories requires extensive knowledge that captures common social and cultural activities. Unfortunately, social and cultural models are notoriously hard to model by hand. For example, a simple model of restaurant behaviour uses 87 rules (Mueller 2007). A simulation game about attending a prom (McCoy et al. 2010) required 5,000 rules to capture the social dynamics associated with that situation.

As an alternative to production rules, one may consider employing *scripts* (Schank and Abelson 1977), a form of procedural knowledge that describes how common situations are expected to unfold, thus capturing social and cultural norms. A script about visiting a restaurant, for example, would encode the typical progression of events (entering, being seated, reading a menu, paying the bill, etc.). Many story generation systems make use of manually coded script-like knowledge, such as cases or hierarchical task libraries (e.g. Meehan 1976; Lebowitz 1987; Turner 1994; Perez y Perez & Sharples 2001; Cavazza, Charles, & Mead 2002; Gervas et al. 2005; Swanson & Gordon 2008; Riedl 2010; Li & Riedl 2010; Hajarnis et al. 2011). However, the effort required to manually code script-like information becomes a significant bottleneck. As a result, most story generation systems to date are restricted to a small number of hand-authored knowledge structures and can thus only operate within the bounds of a limited micro-world for which knowledge has been provided.

Automatically acquiring sociocultural knowledge can open up story generation systems to a wider repertoire of possible stories and domains. In this paper, we propose an approach for learning script-like knowledge from *crowdsourced* narrative examples. *Crowdsourcing* replaces a dedicated expert who solves a complicated problem with many members of the general public, or *workers*, each solving a simple problem (cf. Howe 2006, Quinn & Bederson 2011). In our case, we request each worker to provide a short real-world example of a common situation for which we wish to learn a script. For example, we may ask workers to describe an experience of a restaurant visit. Workers then tell stories in natural language that include typical events for that situation. Crowdsourcing thus provides a means for rapidly acquiring a highly specialized corpus of examples of a given situation, significantly simplifying the subsequent learning. Our initial results suggest that robust knowledge structures can be learned from small corpora containing only about 40 worker responses.

Our automated approach simultaneously learns both the events that comprise a situation and the typical ordering of these events from the crowdsourced narratives. By leveraging the crowd and its collective understanding of social constructs, we can learn a potentially unlimited range of scripts regarding how humans generally believe real-world situations unfold. We seek to apply this script-like knowledge to the generation of believable stories that involve common social situations or the direct engagement of virtual characters in social behaviors.

Background and Related Work

This section reviews story generation systems and discusses their reliance on hand-coded knowledge structures. We compare our crowdsourced approach for the acquisition of script-like knowledge to previous knowledge acquisition techniques and highlight its strengths and weaknesses.

Story Generation

Automated story generation systems search for a novel sequence of events that meet a given communicative objective, such as to entertain or convey a message or moral. The most common approaches to story generation are planning and case-based reasoning.

Planning-based story generation systems (Meehan 1971; Lebowitz 1987; Cavazza, Charles, & Mead 2002; Riedl & Young 2010; Li & Riedl 2010; Ware & Young 2011) use a causality-driven search to link a series of primitive actions to achieve a goal. The knowledge structures are usually too lean to fully represent common social scripts. Some story generation systems (cf., Lebowitz 1987; Cavazza, Charles, & Mead 2002; Li & Riedl 2010) attempt to enrich the generation process with hierarchical scripts that capture common ways of solving goals and performing tasks. System designers typically handcraft these hierarchical scripts.

Case-based story generators (Turner 1994; Perez y Perez & Sharples 2001; Gervas et al. 2005; Swanson & Gordon 2008; Riedl 2010; Hajarnis et al. 2011) attempt to construct novel stories by reusing prior stories, or cases. Sociocultural norms can be "baked into" the prior cases. Most case-based story generators to date have relied on hand-coded cases and stories, with two exceptions of note. First, the system described by Hajarnis et al. (2011) learns cases from human storytellers who enter stories via a custom interface. Cases can only be expressed in terms of a known set of possible actions, and are thus limited to a given micro-world. Second, SayAnything (Swanson & Gordon 2008) constructs new stories from fragments of stories mined from online blogs. This is a promising approach, although reliably selecting and reusing appropriate narrative fragments in the correct context remains an open problem. In contrast, our approach starts with a smaller number of crowdsourced stories specifically aimed at a particular situation that we wish to tell stories about, reducing the need to reason about context.

Script Knowledge Acquisition

Work on commonsense reasoning has sought to acquire propositional knowledge from a variety of sources. LifeNet (Singh & Williams 2003) is a commonsense knowledge base about everyday experiences constructed from 600,000 propositions asserted by the general public. According to Singh and Williams, this technique tends to yield spotty coverage. Gordon et al. (2011) describe an approach to mine causal relations from millions of blog stories. These systems do not attempt to create script-like knowledge representations; it is not clear how this knowledge would be used to generate novel stories. Open Mind Experiences (Singh & Barry 2003; Singh, Barry, & Liu 2004) is a database of stories and has been proposed as a means to generate new stories (Liu & Singh 2002).

Script-like knowledge can also be acquired from large-scale corpora with the goal of applying knowledge learned to the task of understanding news stories (Girju 2003; Bean & Riloff 2004; Brody 2007; Chambers & Jurafsky 2009; Kasch & Oates 2010). These systems attempt to find correlations between events appearing in these stories. In particular, the technique by Chambers and Jurafsky (2009) attempts to identify related event sentences and learn partially ordered *before* relations between events. While these works are intended to further natural language processing goals, such as script recognition, the learned scripts are general in nature and thus can be applied to a range of problems including story generation.

While corpus-based script learning can be very powerful, it also suffers from two limitations. First, the topic of the script to be learned must be represented in the corpus. Thus, it might be difficult to learn the script for how to go on a date to a movie theatre from a news article corpus. Second, given a topic, only the relevant events from the corpus should be extracted and irrelevant events should be excluded whereas a general corpus will have many irrelevant events that must be filtered. Ideally, one has a specialized corpus for each situation one wishes to learn a script for, but such specialized corpora rarely exist.

Crowdsourcing can be used to rapidly acquire a specialized corpus by paying, or otherwise incentivizing, a number of untrained human workers to provide examples of the topic in narrative form. With proper instructions, a crowd of amateurs can collectively create a specialized corpus from which high-quality scripts can be learned. The corpus will contain only relevant data and relatively complete examples of situations. In addition, the corpus may be specialized for any target domain. That is, crowdsourcing provides a means for rapidly acquiring a highly specialized corpus of examples of a given situation, which may significantly simplify subsequent learning.

Crowdsourcing usually breaks up a complex problem into a number of simpler subproblems to make them easily solvable for ordinary workers. Hence, crowdsourced results must still be filtered, aggregated, and summarized in an automated fashion to create a complete solution. This collaborative human-AI approach has been used to train spell checkers (Lasecki et al. 2011), teach robots to perform tasks (Butterfield et al. 2010; Chernova, Orkin, and Breazeal 2010), construct learning materials (Boujarwah, Abowd, and Arriaga 2012), and tackle other challenging problems.

Jung et al. (2010) extract procedural knowledge from eHow.com and wikiHow.com where humans enter how-to instructions for a wide range of topics. Although these resources are sufficient for humans, for computational systems, the coverage of topics is sparse (very common situations are missing). Further, instructions in these websites tend to use complex language, conflate instructions and recommendations, and involve complex and nuanced conditionals.

In the *Restaurant Game,* Orkin and Roy (2009) use traces of people in a virtual restaurant to learn a probabilistic model of restaurant activity. *The Restaurant Game* as a playable interactive system has an *a priori* known set of actions that can occur in restaurants (e.g., sit down, order, etc.) that were programmed in advance. Users select actions to perform to recreate restaurant-going experiences, which the system then uses to learn probabilistic event ordering knowledge. Our work is similar to this, except our approach also learns the primitive events from natural language narrative texts, in addition to temporal orderings between events.

Crowdsourcing Narrative Examples

To learn a script for a particular, given situation we use a three-step process. First, we query crowd workers to provide linear, natural language narratives of the given situation. After some time, a small, highly specialized corpus of examples is acquired. Second, we identify the salient events in these narratives. This is in contrast with Orkin and Roy (2009), where the set of possible actions are known in advance. Third, we identify the order of these events. The second and third step work together to extract a script as a graph from the crowd-supplied narratives. As workers are not experts in knowledge representation, we do not ask workers to author script graphs directly; we believe that for lay workers, providing step-by-step narratives is a more intuitive and less error-prone means of conveying complex information than manipulating complex graphical structures.

In the crowdsourcing stage, to facilitate the subsequent learning of events and their ordering, our system includes precise instructions to the anonymous workers. First, we ask workers to use proper names for all the characters in the task. This allows us to avoid pronoun resolution problems. We provide a cast of characters for common roles, e.g., for the task of going to a fast-food restaurant, we provide named characters in the role of the restaurant-goer, the cashier, etc. Currently, these roles must be hand-specified, although we envision future work where the roles are extracted from online sources of general knowledge such as Wikipedia. Second, we ask workers to segment the narrative such that each sentence contains a single activity. Third, we ask workers to use simple natural language; specifically we ask them to use one verb per sentence and avoid using compound sentences. Throughout the remainder of the paper, we will refer to a segmented activity as a *step*. Figure 1 shows two fragments of narratives about the same situation.

Story A	**Story B**
a. John drives to the restaurant.	a. Mary looks at the menu.
b. John stands in line.	b. Mary decides what to order.
c. John orders food.	c. Mary orders a burger.
d. John waits for his food.	d. Mary finds a seat.
e. John sits down.	e. Mary eats her burger.
f. John eats the food.	…
…	

Figure 1. Example crowd-sourced narratives.

Once a corpus of narrative examples for a specific situation is collected from the crowd, we begin the task of learning a script. In our work, a script is a set of *before* relations, $B(e_1, e_2)$, between events e_1 and e_2 signifying that e_1 occurs before e_2. These relations coincide with causal and temporal precedence information, which are important for narrative comprehension (Graesser, Singer, and Trabasso 1994). A set of *before* relations allows for partial orderings, which can allow for variations in legal event sequences for the situation. The tasks of learning the main events that occur in the situation and learning the ordering of events are described in the next sections.

Event Learning

Event learning is a process of determining the primitive units of action to be included in the script. By working from natural language descriptions of situations, we learn the salient concepts used by a society to represent and reason about common situations. We must overcome several challenges:

1. The same step may be described in different ways.
2. Some steps may be omitted by some workers.
3. A task may be performed in different ways and therefore narratives may have different steps, or the same steps but in a different order.

Our approach is to automatically cluster steps from the narratives based on semantic similarity such that clusters come to represent the consensus events that should be part of the script. Each step in a narrative is a phrase that may or may not be semantically equivalent to another step in another narrative. There are many possible ways to cluster sentences based on semantic similarity; below we present the technique that leverages the simple language encouraged by our crowdsourcing technique. First, we preprocess the narratives to extract the core components of each step: the main verb, the main actor, and the verb patient if any. Second, we identify the semantic similarity of each step using semantic gloss information from WordNet (Miller 1995). Finally we cluster steps in order to identify the core set of events.

Semantic Similarity

We use the Stanford parser (Klein & Manning 2003) to identify the actor, verb, and the most salient non-actor noun for each step. The most salient non-actor noun is identified using a rule-based approach. Once we have these components, the similarity between two corresponding components is computed as follows. For a pair of words (verbs or non-proper nouns), we obtain their similarity using the WordNet Gloss Vector technique (Patwardhan & Pedersen 2006). The WordNet Gloss Vector technique uses the cosine similarity metric to determine the similarity [0,1] for any two weighted term vectors for the desired synsets. To apply this technique, we need the appropriate WordNet synset for each verb or noun; we the Pedersen and Kolhatkar (2009) word-sense disambiguation technique to identify the best WordNet synset.

The similarity between two steps thus is computed as a weighted sum of the following elements:

- Semantic similarity of verbs
- Semantic similarity of nouns
- The difference in event location

Event location—a step's location as the percentage of the way through a narrative—helps disambiguate semantically similar steps that happen at different times, especially when a situation is highly linear with little variation. For example, when going to a movie theatre, one will "wait in line" to buy tickets and then may "wait in line" to buy popcorn. While both activities share semantic information, they should be considered distinct events.

Event Clustering

We model event learning as the clustering of steps, making use of the semantic information computed above. The resultant clusters are the events that can occur in the given situation.

Event clustering is performed in two stages. In the first stage, we make initial cluster assignments of steps from different narratives using shallow information. For each pair of steps from the same narrative, we record a *no-link* constraint, prohibiting these two steps from being placed into the same cluster. For each pair of steps from different narratives that have identical verbs and nouns, we record a *must-link* constraint, requiring that these two steps be placed within the same cluster. From this information, we produce an initial assignment of steps to clusters that respects all constraints.

In the second stage, we iteratively improve the cluster quality through the application of the k-Medoids clustering algorithm. The k-Medoids makes use of similarity between steps, as discussed above. We automatically set the similarity score to 1.0 if there is a *must-link* constraint between steps and 0.0 if there is a *no-link* constraint between steps.

The k-Medoid clustering algorithm requires *k*, the number of total clusters, to be known. We use a simple technique to sample different values for *k*, starting with the average narrative length, searching for a solution that minimizes intra-cluster variance while maximizing the extra-cluster distance.

Table 1. Crowd-sourced data sets.

Situation	Num. stories	Mean num. steps	Unique verbs	Unique nouns
Fast food	30	7.6	55	44
Movie date	38	10.7	71	84

Experiments and Results

To evaluate our event learning algorithm, we collected two sets of narratives for the following situations: going to a fast food restaurant, and taking a date to a movie theatre. While restaurant activity is a fairly standard situation for story understanding, the movie date situation is meant to be a more accurate test of the range of socio-cultural constructs that our system can learn. Table 1 shows the attributes of each specialized corpus.

For each situation, we manually created a gold standard set of clusters against which to calculate precision and recall. Table 2 presents the results of event learning on our two crowdsourced corpora, using the MUC6 cluster scoring metric (Vilain et al. 1995) to match computed cluster results against the gold standard. These values were obtained using parameter optimization to select the optimal weights for the clustering similarity function. The ideal weights for a given situation, naturally, depend on language usage and the degree to which variability in event ordering can occur. Table 2 shows how each portion of our algorithm helps to increase accuracy. Initial cluster seeding makes use of shallow constraint information. The semantic similarity columns show how phrase expansion improves our clusters. Event location further increases cluster accuracy by incorporating information contained in the implicit ordering of events from the example narratives. For each set of results, we show the average precision, recall, and F1 score for the best weightings for verb, noun, and event location similarity components.

Noting the differences between data sets, the movie date corpus has a significantly greater number of unique verbs and nouns, longer narratives, and greater usage of colloquial language. Interestingly, the movie date corpus contains a number of non-prototypical events about social interactions (e.g., *Sally slaps John.*) that appear rarely. This greater number of clusters containing few steps has a negative effect on recall values; a larger number of narratives would ameliorate this effect by providing more examples of rare steps. By crowdsourcing a highly specialized corpus, we are able to maintain precision in the face of a more complicated situation without

Table 2. Precision, Recall, and F1 Scores for the restaurant and movie data sets.

Situation	Gold std. num. events	Initial seed clusters			Semantic similarity			Semantics + Location		
		Pre.	Recall	F1	Pre.	Recall	F1	Pre.	Recall	F1
Fast food restaurant	21	0.780	0.700	0.738	0.806	0.725	0.763	**0.814**	**0.739**	**0.775**
Movie theatre date	56	0.580	0.475	0.522	0.725	0.580	0.645	**0.763**	**0.611**	**0.679**

restricting worker ability to express their conception of the salient points of the situation.

Improving Event Clustering with Crowdsourcing

While we believe that our event learning process achieves acceptably high accuracy rates, errors in event clustering may impact overall script learning performance (the effects of clustering errors on script learning will be discussed in a later section). To improve event-clustering accuracy, we can adopt a technique to improve cluster quality using a second round of crowdsourcing, similar to that proposed by Boujarwah, Abowd, and Arriaga (2012). Workers are tasked with inspecting the members of a cluster and marking those that do not belong. If there is sufficient agreement about a particular step, it is removed from the cluster. A second round of crowdsourcing is used to task workers to identify which cluster these "un-clustered" steps should be placed into. According to Boujarwah (personal communication), the multiple rounds of crowdsourcing required $110 for a single script, linearly increasing with situation complexity. Crowdsourcing is often used to improve on artificial intelligence results (von Ahn 2005) and we can increase clustering accuracy to near perfect in this way. However, in the long term our goal is minimize the use of the crowd so as to speed up script acquisition and reduce costs.

Plot Graph Learning

Once we have the events, the next stage is to learn the script structure. Following Chambers and Jurafsky (2009) we learn *before* relations $B(e_1, e_2)$ between all pairs of events e_1 and e_2. See Figure 2 for a visualization of a script as a graph. Chambers and Jurafsky train their system on the Timebank corpus (Pustejovsky et al. 2003), which uses temporal signal words. Girju (2003) uses causal signal words. Because we are able to leverage a highly specialized corpus of narrative examples of the desired situation, we can avoid reliance on signal words and instead probabilistically determine ordering relations between events directly from the narrative examples. The result of this process is a script-like structure similar in nature to a *plot graph* (Weyhrauch 1997), a partial ordering of events that defines a space of possible event sequences that can unfold during a given situation. Not only is a plot graph similar to a script, but it is also a data structure that has been used for AI story generation (Weyhrauch 1997; Nelson & Mateas 2005; Roberts et al. 2006; Sharma et al. 2010).

Initial Script Construction

Script construction is the process of identifying the plot graph that most accurately captures the most information out of the set of crowdsourced narratives. Each possible *before* relation between a pair of events is a hypothesis (i.e. $B(e_1, e_2) = true$ or $B(e_2, e_1) = true$) that must be verified. For every pair of events e_1 and e_2, we count the observation of evidence for and against each hypothesis. Let s_1 be a step in the cluster representing event e_1, and let s_2 be a step in the cluster representing event e_2. If s_1 and s_2 appear in the same input narrative, and if s_1 appears before s_2 in the narrative, then we consider this as an observation in support of $B(e_1, e_2) = true$. If s_2 appears before s_1 in the same narrative, this observation supports $B(e_2, e_1) = true$.

Figure 2. An example plot graph, adapted from Chambers and Jurafsky (2009).

The probability p_h of a hypothesis h equals k/n, where n is the number of observations and k is the observations that support h. Considering that the probability is only an estimate of the real world based on limited observations, we also estimate its confidence (cf. Wang 2009); a probability computed based on a small number of observations has low confidence. Without assuming prior distributions for orderings between arbitrary events, we use the imprecise Dirichlet model (Walley 1996) to represent this uncertainty. Suppose we have s additional observations whose values are hidden, the most optimistic estimate of the probability occurs when all hidden observations support hypothesis h, yielding an upper bound $p_h^+ = (k+s)/(n+s)$. Similarly, the most pessimistic estimate is $p_h^- = k/(n+s)$. Thus, the confidence in a probability is $c_h = 1 - (p_h^+ - p_h^-) = 1 - s/(n+s)$, where s is a parameter

We select relations for the plot graph in which the probability and confidence exceed thresholds $T_p, T_c \in [0,1]$, respectively. T_p and T_c apply to the entire graph and provide an initial estimate of the best plot graph. However, a graph that better explains the crowdsourced narratives may be found if the thresholds could be locally relaxed for particular relations. Below, we introduce a measure of plot graph error and an algorithm for iteratively improving the plot graph to minimize the error.

Plot Graph Improvement

Since a plot graph encodes event ordering, we introduce an error measure based on the expected number of interstitial events between any pair of events. The error is the difference between two distance measures, $D_G(e_1, e_2)$ and $D_N(e_1, e_2)$. $D_G(e_1, e_2)$ is the number of events on the shortest path from e_1 to e_2 on the graph (e_1 excluded); this is also the minimum number of events that must occur between e_1 and e_2 in all legal totally ordered sequences consistent with the *before* relations of the plot graph. In contrast, $D_N(e_1, e_2)$ is the normative distance from e_1 to e_2 averaged over the entire set of narratives. For each input narrative that includes sentence s_1 from the cluster representing e_1 and sentence s_2 from the cluster

$\boldsymbol{Q}$:= all of events (e_1, e_2) where e_2 is reachable from e_1 or unordered
Foreach (e_1, e_2) $\in \boldsymbol{Q}$ in order of decreasing $D_N(e_1, e_2) - D_G(e_1, e_2)$ **do:**
  $\boldsymbol{E}$:= all events such that for each $e_i \in \boldsymbol{E}$, $D_G(e_1, e_i) = D_N(e_1, e_2) - 1$
  Foreach $e_i \in \boldsymbol{E}$ **do:**
    If edge $e_i \rightarrow e_2$ has probability and confidence less than T_p, T_c and will not create a cycle if added to the graph **do:**
      Strengthen the edge by adding one observation in support of it
    If $e_i \rightarrow e_2$ has probability and confidence greater than T_p, T_c and adding $e_i \rightarrow e_2$ to the graph decreases *MSGE* **do:**
      Add $e_i \rightarrow e_2$ to the graph
Return graph

Figure 3. The plot graph improvement algorithm.

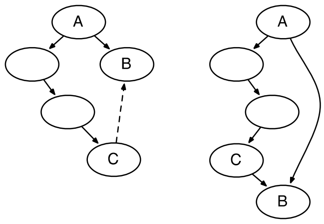


Figure 4. Compensation for errors between pairs of events.

representing e_2, the distance (i.e. number of interstitial sentences plus one) between s_1 and s_2 is $d_N(s_1, s_2)$. $D_N(e_1, e_2)$ is thus the average of $d_N(s_1, s_2)$ over all such input narratives. The mean squared graph error (MSGE) for the entire graph is:

$$MSGE = \frac{1}{|P|} \sum_{e_1, e_2 \in P} \left(D_G(e_1, e_2) - D_N(e_1, e_2) \right)^2$$

where P is the set of all ordered event pairs (e_1, e_2) such that e_2 is reachable from e_1 or that they are unordered.

We utilize this error measure to improve the graph based on the belief that D_N represents the normative distance we expect between events in any narrative accepted by the plot graph. That is, typical event sequences in the space of narratives described by the plot graph should have $D_G(e_1, e_2) \approx D_N(e_1, e_2)$ for all events. A particularly large $|D_N(e_1, e_2) - D_G(e_1, e_2)|$ may indicate that some edges with low probability or confidence could be included in the graph to make it closer to user inputs and reduce the overall error.

We implement a greedy, iterative improvement search for a plot graph that reduces mean square graph error (Figure 3). For each pair of events (e_1, e_2) such that e_2 is reachable from e_1 in the plot graph of directed edges, we search for all events E such that if $e_i \in E$ were the immediate predecessor of e_2 then $D_G(e_1, e_2)$ would be equal to $D_N(e_1, e_2)$. If there is a possible edge from e_i to e_2 (i.e., at least one observation that supports such an edge) then we strengthen the edge hypothesis by one observation. This intuition is illustrated in Figure 4 where the edge (dashed arrow) from event C to event B was originally insufficiently supported; adding the edge to the graph creates the desired separation between events A and B. This process repeats until no new changes to graph structure can be made that reduce the mean square graph error.

We find this approach to be effective at reducing graph error when T_p is set relatively high (> 0.5) and $T_c \approx$ 0.4. A conservative T_p initially discards many edges in favor of a more compact graph with many unordered events. A moderate T_c allows the improvement algorithm to opportunistically restore edges to the graph.

Table 3. Error reduction for both situations.

Situation	Error before Improvement		Error after Improvement		Avg. Error Reduction
	Avg.	Min.	Avg.	Min.	
Fast food	4.05	1.23	2.31	0.85	42%
Movie date	6.32	2.64	2.99	1.88	47%

Experiments and Results

Figure 5 shows plot graphs learned for the fast food restaurant and movie theatre date situations. These plots were learned from the gold standard clusters under the assumption that we can achieve near perfect clustering accuracy with a second round of crowdsourcing. The event labels are English interpretations of each event based on manual inspection of the sentences in each event. For clarity, some edges are omitted from the figure that do not affect the partial ordering. Rare events, such as *Sally slaps John* are excluded from the graphs because their clusters contain too few sentences and thus do not meet our probability and confidence thresholds.

Some statistics about the two graphs are shown in Table 3. Over 128 sets of different parameter settings, we found that iterative graph improvement led to an average error reduction of 42% and 47% for the fast-food restaurant and movie data situations respectively. The asterisks in Figure 5 indicate edges that were added during graph improvement. Note that it is not always possible to reduce graph errors to zero when there are plausible ordering varations between events. For example *choose menu item* and *wait in line* can happen in any order, introducing a systematic bias for any graph path across this pair. In general we tend to see ordered relations when we expect causal necessity, and we see unordered events when ordering variations are supported by the data.

Discussion and Future Work

There are several ways in which errors during event learning (i.e., clustering) can impact plot graph generation. First, steps may be improperly clustered, thus introducing observations of ordering relations between otherwise unrelated events, possibly causing cycles in the plot graph. If the number of improperly clustered sentences is relatively small, these relations have low probability and confidence and will be filtered out. Second, two distinct events may be merged into one event, causing ordering cycles in which all edges have high probability and confidence. When this happens, it is possible to eliminate the cycle by choosing an event to split into two. We select the event cluster in the cycle with the highest inter-cluster variance in the belief that high inter-cluster variance indicates that there is a natural split of sentences into two clusters. Third, an event may be split into two clusters unordered relative to each other. This creates the

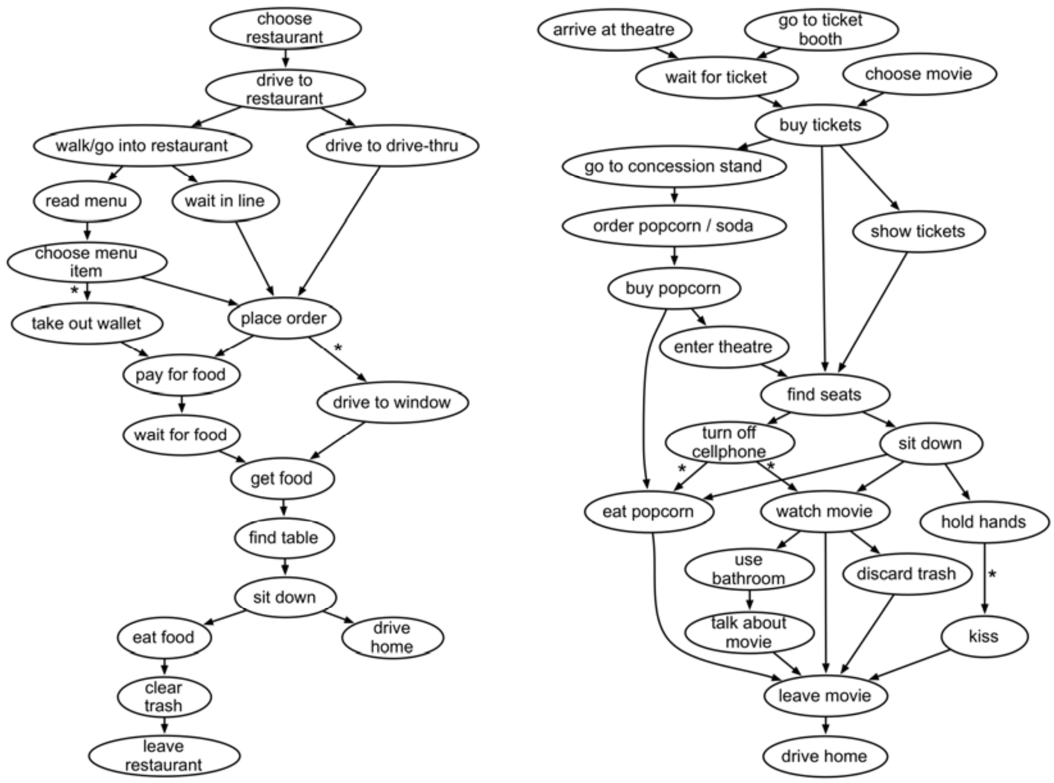


Figure 5. Plot graphs generated for the restaurant situation (left) and movie date situation (right).

appearance that an event must occur twice in any story generated from this script.

Closely inspecting Figure 5, we note that *before* relations includes causal sufficiency and mere temporal precedence as well as strict causal necessity. For example, before *placing an order* at a fast-food restaurant one can *wait in line* or *drive to drive-thru* but not both. Thus, both are sufficient for *placing an order*. The crowdsourced corpus for the restaurant is split relatively evenly between walk-in and drive-thru narratives, implying two main variations to the situation (this also accounts for the unordered *leave restaurant* and *drive home* events). Future work will be necessary to distinguish causal and temporal relations as well as necessity versus sufficiency. We believe this can be accomplished by more fully leveraging correlations (e.g. mutual information) between events. As with the event learning phase, it is always possible to ask crowd workers to provide causal information with questions about causal counterfactuals, a technique adapted from Trabasso and Sperry (1985).

Toward Story Generation

To an extent, the plot graph learned as described above grants narrative intelligence to a computational process. A model of common social situations—in the form of a plot graph—captures common beliefs of how those real-world situations unfold. A computational system must also be able to act on this narrative intelligence in order to: (a) tell a story about a sociocultural situation, (b) tell a story in which a common social situation occurs, or (c) directly engage in a social situation in a virtual world. Fortunately, the plot graph representation facilitates story generation and interactive execution (cf., Weyhrauch 1997; Nelson & Mateas 2005; Sharma et al. 2010; Roberts et al. 2006).

A plot graph defines a space of totally ordered event sequences that are believed to be “legal” ways for a given situation to unfold. By virtue of the way we learn the plot graph from human-provided examples, the knowledge structure generalizes across the most common ways in which the given situation manifests. Within the space of legal stories, we may consider different possible storytelling goals: the most prototypical story, the most unusual story, the most surprising, etc. According to Bruner (1991), interesting stories are those that deviate from the norm in some way.

The plot graph representation was originally used to determine what was possible for a user to do in an interactive fiction game. Although these systems are meant to provide narrative structure to games, we can view these systems as story generation systems when the interactive component is removed. To generate a story using a plot graph, a system must search for and select one totally ordered sequence from this set (Weyhrauch 1997; Nelson & Mateas 2005; Roberts et al. 2006; Sharma et al. 2010). To date, algorithms that use plot graphs have used the same set of heuristics to find sequences that reduce cognitive burden and reduce flailing, including:

- Location flow—events in the same location should occur together.
- Thought flow—events that are conceptually related should occur together.
- Motivation—a measure of whether plot points are motivated by previous plot points.

Other heuristic functions are used as well.

Generation of stories from *learned* plot graphs requires a slightly different approach. The plot graph describes a social situation that is relatively well constrained, so the only question that remains is how prototypical should the resultant story be. We define *typicality* as a function of the likelihood of events (nodes) and of specific sub-sequences (node-link-node sequences). By varying the inclusion of nodes and links according to their likelihood while respecting the before relations, we can generate stories that are legal but with arbitrary typicality within the norm.

We have a wealth of probabilistic information to draw from as a consequence of how we learn the plot graph, including:

- Typicality of events—the probability of an event being part of a situation, $P(e)$.
- Typicality of event orderings—the probability that a given ordering occurs, $P(e_1 \rightarrow e_2 \mid e_1 \wedge e_2)$.
- Adjacency—the probability that two events should occur immediately adjacent to each other, $P(e_1 * e_2 \mid e_1 \wedge e_2)$.
- Co-occurrence—the probability that any two events have been observed in the same crowdsourced story, $P(e_1 \wedge e_2)$.

The most prototypical story that can be generated from a given plot graph, for example, may be defined as inclusion of the *n* most probable events, ordered according to the most probable before relations between those *n* nodes. We can generate more interesting stories about the same situation by finding a legal sequence with (a) an unlikely event, such as kissing (kissing occurs in ~10% of crowdsourced examples); (b) likely events that occur in an unlikely ordering; (c) non-adjacent events that are typically adjacent; (d) pairs of events that have low co-occurrence; or (e) omission of an event that frequently co-occurs with a present event. We intend to investigate the effects of each of the above hypotheses on story novelty in order to develop tunable heuristics for the generation process.

Story generation from sociocultural plot graphs reaches full expressivity once we are able to differentiate links in the graph as denoting causal necessity or simple temporal precedence; this provides the richest variation among legal stories from which to choose a specific story or guide a virtual character's behavior. Once we differentiate between causal necessity and precedence, the story generation process can be performed using standard search techniques such as A*, forward or backward search, genetic algorithms and Monte Carlo methods.

Conclusions

Crowdsourcing provides direct access to humans and the ways in which they express experiential knowledge. A crowdsourcing approach has advantages over general corpus based learning: filtering irrelevant information, segmentation, and control of natural language complexity. Our approach capitalizes on these advantages by learning the primitive events from the segmented natural language and learning ordering constraints on these events directly from the crowd-sourced narrative examples.

Plot graph learning overcomes one of the primary bottlenecks in acquiring sociocultural knowledge required for effective generation of believable stories. While future work remains to tease out the full expressive power of automatically learned plot graphs, our approach makes it possible for a computational system to extend its narrative intelligence beyond a single, hand-crafted micro-world.

One of the strengths of our approach is the way in which we can leverage shared social constructs acquired directly from humans. Our approach learns the events that make up common situations directly from the language people use to describe those situations; event ordering captures shared social and cultural understanding based on people's descriptions of experiences. Thus, in addition to learning scripts for story generation, our system also learns a functional form of socio-cultural knowledge that could be applied to other computational narrative intelligence tasks such as story understanding.

Believable story generation requires in-depth understanding of the rich social situations that humans recognize and participate in everyday, yet this sort of experiential knowledge is rarely possessed by intelligent computational systems. A human-AI collaborative approach in which humans naturally convey experiential, social, and cultural knowledge to an intelligent system can overcome many of the hurdles to human-level AI problems.

Acknowledgements

The authors gratefully acknowledge the support of the U.S. Defense Advanced Research Projects Agency (DARPA) for this effort.

References

Bean, D., Riloff, E. (2004). Unsupervised learning of contextual role knowledge for coreference resolution. In *Proceedings of the 2004 HLT/NAACL Conference*.

Boujarwah, F., Abowd, G., Arriaga, R. (2012). Socially computed scripts to support social problem solving skills. In *Proceedings of the 2012 Conference on Human Factors in Computing Systems*.

Brody, S. (2007). Clustering clauses for high-level relation detection: an information-theoretic approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18, pp. 1-21.

Butterfield, J., Osentoski, S., Jay, G., Jenkins, O.C. (2010). Learning from demonstration using a multi-valued function regressor for time-series data. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*.

Cavazza, M., Charles, F., Mead, S. (2001). Planning characters' behaviour in interactive storytelling. *Journal of Visualization and Computer Animation*, 13, pp. 121-131.

Chambers, N., Jurafsky, D. (2009). Unsupervised learning of narrative event chains. In *Proceedings of ACL/HLT 2009*.

Chernova, S., Orkin, J., Breazeal, C. (2010). Crowdsourcing HRI through online multi-player games. In *Proceedings of the 2010 AAAI Fall Symposium on Dialog with Robots*.

Gerrig, R. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press.

Gervás, P., Díaz-Agudo, B., Peinado, F., Hervás, R. (2005). Story Plot Generation based on CBR. *Journal of Knowledge-Based Systems*, 18, pp. 235-242.

Girju, R. 2003. Automatic Detection of Causal Relations for Question Answering. *Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering—Machine Learning and Beyond*.

Gordon, A., Bejan, C.A., Sagae, K. (2011). Commonsense causal reasoning using millions of personal stories. In *Proceeding of the 25th Conference on Artificial Intelligence*.

Graesser, A., Singer, M., Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, pp. 371-395.

Hajarnis, S., Leber, C., Ai, H., Riedl, M.O., Ram, A. (2011). A case based planning approach for dialogue generation in digital movie design. In *Proceedings of the 19th International Conference on Case Based Reasoning*.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14.06, June 2006.

Jung, Y., Ryu, J., Kim, K.-M., Myaeng, S.-H. (2010). Automatic Construction of a Large-Scale Situation Ontology by Mining How-to Instructions from the Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(2-3), pp. 110-124.

Kasch, N., Oates, T. (2010). Mining script-like structures from the web. In *Proceedings of the NAACL/HLT 2010 Workshop on Formailism and Methodology for Learning by Reading*.

Klein, D., Manning, C. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*.

Lasecki, W.S., Murray, K.I., White, S., Miller, R.C., Bingham, J.P. (2011). Real-time crowd control of existing interfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology*.

Lebowitz, M. (1987). Planning stories. In *Proceedings of the 9th Annual Conference of the Cognitive Science Society*.

Li, B., Riedl, M.O. (2010). An offline planning approach to game plotline adaptation. In Proceedings of the 6th Conference on Artificial Intelligence *and* Interactive Digital Entertainment.

Liu, H., Singh, P. (2002). MAKEBELIEVE: using commonsense knowledge to generate stories. In *Proceedings of the 18th National Conference on Artificial Intelligence*.

Mateas, M., Senger, P. (1999). Narrative intelligence. In *Proceedings of the 1999 AAAI Fall Symposium on Narrative Intelligence*.

McCoy, J., Treanor, M., Samuel, B., Tearse, B., Mateas, M., Wardrip-Fruin, N. (2010). Comme il Faut 2: a fully realized model for socially-oriented gameplay. In *Proceedings of the 3rd Workshop on Intelligent Narrative Technologies*.

McKoon, G., Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, pp. 440-466.

Meehan, J (1976). *The Metanovel: Writing Stories by Computers*. Ph.D. Dissertation, Yale University.

Miller, G. (1995). WordNet: a lexical database for english. *Communications of the ACM*, 38(11), pp. 39-41.

Mueller, E.T. (2007). Modelling space and time in narratives about restaurants. *Literary and Linguistic Computing*, 22(1), pp. 67-84.

Nelson, M., Mateas, M. (2005). Search-based drama management in the interactive fiction Anchorhead. In *Proceedings of the 1st Conference on Artificial Intelligence and Interactive Digital Entertainment*.

Orkin J., Roy, D. (2009). Automatic learning and generation of social behavior from collective human gameplay. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*.

Patwardhan, S., Pedersen, T. (2006). Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In *Proceedings of the EACL Workshop on Making Sense of Sense*.

Pedersen, T., Kolhatkar, V. (2009) WordNet:: SenseRelate::AllWords – a broad coverage word sense tagger that maximimizes semantic relatedness. In *Proceedings of the ACL 2009 Conference*.

Pérez y Pérez, R., Sharples, M. (2001). MEXICA: a computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, 13, pp. 119-139.

Pustejovsky, J., Hanks, P. Saurí, R., See, A., Gaizauskas, R., Setzer, A. Radev, D. Sundheim, B., Day, D. Ferro, L., Lazo, M. (2003). The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*.

Quinn, A.J., Bederson, B.B. (2011). Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of The ACM SIGCHI Conference on Human Factors in Computing Systems.*

Riedl, M.O. (2010). Case-based story planning: creativity through exploration, retrieval, and analogical transformation. *Minds and Machines*, 20.

Riedl, M.O., Young, R.M. (2010). Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research*, 39, pp. 217-268.

Roberts, D.L., Nelson, M.J., Isbell, C.L., Mateas, M., Littman, M.L. (2006). Targeting specific distributions of trajectories in MDPs. In *Proceedings of the 21st National Conference on Artificial Intelligence*.

Schank, R., Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates.

Sharma, M., Ontañón, S., Mehta, M., Ram, A. (2010). Drama Management and Player Modeling for Interactive Fiction Games. *Computational Intelligence*, 26, pp. 183-211.

Singh, P., Barry, B. (2003). Collecting commonsense experiences. In *Proceedings of the 2nd International Conference on Knowledge Capture*.

Singh, P., Barry, B, Liu, H. (2004). Teaching machines about everyday life. *BT Technology Journal*, 22(4), pp. 227-240.

Singh, P., Williams, W. (2003). LifeNet: a propositional model of ordinary human activity. In *Proceedings of the Workshop on Distributed and Collaborative Knowledge Capture*.

Swanson, R., Gordon, A. (2008). Say Anything: a massively collaborative open domain story writing companion. In *Proceedings of the 1st International Conference on Interactive Digital Storytelling*.

Trabasso, T. and Sperry, L. (1985). Causal relatedness and importance of story events. Journal of Memory and Language, 24:595-611.

Turner, S. (1994). *The Creative Process: A Computer Model of Storytelling*. Lawrence Erlbaum Associates.

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., Hirschman, L. (1995). A Model-Theoretic Coreference Scoring Scheme. In *Proceeding of the 6th Conference on Message Understanding* (MUC6).

von Ahn, L. (2005). *Human Computation*. Ph.D. Dissertation, Carnegie Mellon University.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1), pp. 3-57.

Wang, P. (2009). Formalization of evidence: a comparative study. *Journal of Artificial General Intelligence*, 1, pp. 25-53.

Ware, S., Young, R.M. (2011). CPOCL: a narrative planner supporting conflict. In *Proceedings of the 7th Conference on Artificial Intelligence and Interactive Digital Entertainment*.

Weyhrauch, P. (1997). *Guiding Interactive Fiction*. Ph.D. Dissertation, Carnegie Mellon University.

Prototyping the Use of Plot Curves to Guide Story Generation

Carlos León, Pablo Gervás

Escuela Politécnica Superior, Instituto de Tecnología del Conocimiento
Universidad Alfonso X el Sabio, Universidad Complutense de Madrid
cleonazn@uax.es, pgervas@sip.ucm.es

Abstract

Setting objectives for automatic story generation is needed for a story generation system to produce content. Among the potentially useful methods, curves defining the evolution of specific features of a narrative that evolve along time are particularly appropriate because they focus on the evolution of those features and are easy to create, modify and understand by human users. In this paper we propose a theoretical definition of curve-based story generation, its relation to existing story generation algorithms and how this theory can be applied to new systems.

Keywords: computational narrative, curves, story generation

1. Introduction

Existing computational story generation systems are not autonomous enough to set their own generation objectives. These objectives must be therefore given by the programmer or by the user for the generation system to be able to produce satisfying content. In general, existing systems are able to generate stories in some particular domain which is usually defined in terms of rules or previous stories. Since the concrete definition restricts what can be generated, it can be considered to be a general way of constraining what the automatic story generation system can produce. However, this is far from general and a more specific objective-setting mechanism is more useful.

Among current objective-setting techniques, narrative curve specification is one of the most story-specific for generation. Narrative curves can be defined in general as constraints for some variables or aspects of a story that evolve along the story timeline. The classic *tension arc* is a good example of this definition of a narrative curve (Aristotle, 1974; Tobias, 2012; Zagalo et al., 2004). It very well depicts how the reader's perception of the evolution of the story evolves over time in terms of its conflicts and resolutions.

In this sense, curves offer a powerful and intuitive way of representing how a story should progress. From a human point of view, curves are a very intuitive representation. They are easy to understand, create and adjust to the user's objectives. Additionally, from the perspective of computation many mathematical techniques are available to handle them, which makes it easy to create an implementation for it.

This paper proposes a general definition of narrative curves it terms of their generic aspects and their relation with the computational properties that are interesting for curve-based narrative generation. In particular the application of the proposed model to existing automatic story generation systems has been studied in order to give an example of how this theoretical approach can serve to make a taxonomy of story generation systems that use curves.

1.1. Setting Objectives in Existing Story Generation Systems

Most existing story generation systems are able to set the generation objectives in several ways. Story generation systems based on planning usually define the generation objectives through a set of logic constraints that the planner must use as objective (Lebowitz, 1985; Riedl and Young, 2006). Some systems also take into account the characters' goals. For instance, Meehan's Tale-Spin (Meehan, 1976) performs storytelling by letting virtual characters satisfy their goals. Some systems even go beyond that and consider the duality of both kind of objectives (author's and character's), like MINSTREL (Turner, 1992).

Story grammars have also been used for story generation (Rumelhart, 1975; Lang, 1999). Story grammars do not have an explicit way of setting objectives, however the overall generation objectives can be considered to be hard-coded in the grammar itself. These systems can of course filter out those generated stories that are not wanted by the user (which is a way of establishing objectives) but these do not drive the generation process itself.

Several existing story generation systems also use plot curves to drive the generation process. MEXICA (Pérez y Pérez, 1999) uses the so called *tension* to represent *love*, *emotion* and *danger*. These three features are finally used as a single value (the *tension* itself). This tension is basically a list of discrete values that define a set of restrictions that the *emotional links* between the characters must fulfill.

Barros and Musse address tension arcs in Interactive Fiction (Barros and Musse, 2008). This system uses non-decreasing, discrete curves to represent dramatic tension. The system includes a definition of narrative tension based on the discovery of clues by the player. The interactive play is driven by minimizing the distance between the actual tension and an objective tension curve.

Stella (León and Gervás, 2011) drives the generation of a step-by-step story generation algorithm by trying to adjust the generation to a set of user-defined curves. Stella allows ad-hoc definition of curves and lets the user decide what they represent. In order to make this option possible, Stella does not make any assumption on the semantics of the curves, therefore forwarding the definition of such semantics to the programmer of the domain definition.

2. Defining Objective Curves

Narrative curve-based generation can be carried out in several ways. Since a curve can represent any variable and since the way in which time is handled in stories is not trivial, many aspects of how curves are used must be taken into account. The current model assumes for simplicity that curves have only two dimensions. The x axis represents the value of the variable that the curve is defining and the y axis represents the evolution of the story over time. Section 5. discusses how more dimensions could be added, but for clarity we will be using this simplification.

2.1. Values for Curves

Being a mathematical object, a narrative curve can represent any variable and thus any numerical value. For this explanation we are assuming real numbers. The semantics for the value that the curve takes depends on the computational system that uses it. No constraints are set on these variables because imposing restrictions on the kinds of value that the curve can represent would only limit the expressiveness of the model.

For instance, a curve could represent *danger*. Intuitively this would be representing the amount of perceived danger for the protagonist of the story by the audience. This value could vary in the range $[-1, 1]$, $[0, 1]$ or $(-, +)$, for example (the number of ranges is infinite).

Of course every option has a number of implications. If a finite range is used, the values are limited and there is a notion of *maximum* or *minimum* is the values that the variable can take. It could also be the case that negative values represent the opposite semantics, i.e. the protagonist is not in a dangerous situation.

All these options are not critic by themselves, and this model assumes that the computational system using the curves must define the required semantics and therefore the most appropriate definition of the variables. As can be seen, this model tries to be as general as possible in terms of the types of curves that can be defined.

2.2. Types of Curves

Having chosen a curve as the definition of the evolution of some aspect of the story along time, the generation must try to create a story whose corresponding evolution is similar enough to the objective curve. When a curve is defined and it is meant to drive the generation process in a story generation algorithm, it must be decided how to make the story match the curve.

While the generation is taking place, the story that is being generated must be implicitly or explicitly assigned a curve that must match the objective curve. However this is not straightforward because there is a number of possibilities for comparing curves beyond the pure numerical comparison. This types of matching are semantical and imply an interpretation of the curve.

This model proposes three levels of decision that define 6 types of curve-based generation:

- *Matching type*: the way in which the matching is performed is important. Curves can be matched by a strict comparison of its corresponding sequence or by a relative matching in which only the relative changes are compared.
  - absolute – every point in the objective curve corresponds to a computed value for the story, and it corresponds to one exact point in the story curve. Both points must have the same y value. This kind of curve gives a strict control on the generated curve, but on the other hand the length of the story must be known beforehand or it the generation must be expected to produce a story which is exactly the same length as the objective curve.
  - relative – in a relative matching type only the order of changes matters. The story curve and the objective curve have to match (within some error threshold), but only the relative position of the values in the curve is important. This kind of curves permit a looser definition but require a more complex computation.
- *Time*: handling time in narratives is usually very complex and it is subject to a deep study in the Narratology community. Regarding curves and computational generation, a simple division in which physical and relative changes of time has been made in order to avoid complicated details.
  - physical – this type assumes that the y axis in the curve specifies absolute time (in whatever unit). This means that the physic timeline of the story (the time at which events take place in the story) has to match the values of the curve at its corresponding time. It has to be kept in mind that this kind of time matching does not strictly use real time, but concrete *time units* that can be defined by the story generation system.
  - changes – these curves can be matched against the sequence of variable changes and not against the time in which they happen. Again, this is a loose definition and while it lets the generation produce stories in a less restrictive way, it requires a more sophisticated algorithm.
- *Level*: so far the theoretical definition of curves has assumed the existence of a general definition of story. While this is hard to study, in general the literature on Computational Models of Narrative shows that most systems have some sort of division between *plot* and *discourse*. The discussion of the narratological definitions is way beyond the focus of this paper, but plot can be defined as the list or graph of events taking place in the story world, and discourse as the ordered and filtered representation of events in a linear way, pretty much in a form that somehow resembles a textual narrative.

  According to this, curves can be used for driving the story plot, the discourse, or both. Hence this division:

- story plot – these curves represent the evolution of the plot or the underlying, logic sequence of actions and events in the story.
- discourse – curves can be matched again the discourse and not the story itself. These curves take the discourse into account. The difference between these two types of lines is discussed in Section 5.1.

A more complex model based on the proposed division could be designed. A model in which a curve could be absolute from the beginning and then the matching could switch to relative matching at some point could be defined and implemented, but the model would only become more complex and not necessarily more general.

3. Curve-Based Generation

After the objective curve has been defined, a story generation system must use this information as a partial definition of the objectives. At this point the story generation system has a curve Γ. The system must also have a domain-specific function $\Phi :: S \rightarrow [(\text{partial story}, val)]$ yielding, for every partial story, a value for some particular variable like *love* or *danger* in any comparable domain (real numbers, for example).

The generated content must correspond to a curve that matches the objective curve Γ according to any selection of the matching type (as described in Section 2.2.). This curve is generated by Φ.

At some stage of the generation the story generation algorithm must decide what to include in the partial story. This can be formally described as follows: the partial story is a list of events $S = [e_1, e_2, e_3, \cdots, e_n]$, and the event e_{n+1} can be chosen from a set of potential candidates $C = \{c_1, c_2, c_3, \cdots, c_n\}$.

This selection must be done according to the objective curve Γ. Since given the definition the story generation system includes the domain-specific definition of Φ, it is possible to compute $\Phi(S + c_i)$ for every $i \in [1, |C|]$. This creates a set of partial curves among which it is possible to choose the best candidate.

The assumption of the possibility of the comparison of curves is supported by several mathematical approaches to curve comparison (Buchin et al., 2009; Cui et al., 2009). We can therefore consider that several mathematical solutions for computing the level of matching exist in the literature. A review of these algorithms is obviously beyond the scope of this paper.

According to the proposed definition, we can assume that as long a matching function between two curves is provided, a computational system can use it to generate a story of which the corresponding curve matches the objective curve.

In this way the generation can be automatically driven by the objective curve Γ. The quality of this generation obviously depends partially on the quality of the domain definition of the Φ function. The implications of this are discussed in Section 5.

3.1. Error Threshold

The value of the computed value for a partial story $\Phi(S + c_i)$ will probably be slightly different from the corresponding value of the objective curve Γ at best. This enforces the inclusion of an *error threshold* in the theoretical model.

This error threshold depends on the particular instance of the generation and sets the allowed difference between the objective curve and the value of the Φ function at some particular stage of the generation. The bigger the error threshold the less strict the matching to the objective curve. This additional variable helps to constrain the generation by setting the level of similarity that must exist between the objective curve and the curve corresponding to the story.

4. Application of the Theoretical Model to Existing Story Generation Systems

The proposed theoretical definition of curve-driven story generation can be used to classify existing story generation systems.

MEXICA (Pérez y Pérez, 1999) performs story generation by matching curves following an *absolute* matching type. Regarding time, it uses the *changes* pattern and it takes places at story level.

The study can also be applied to Interactive Fiction. The system by Barros and Musse (Barros and Musse, 2008) uses curves that are matched in an *absolute*, *physical* way and at *discourse* level.

Stella (León and Gervás, 2011) generates stories by using curves that it tries to match by *relative* comparison. The time matching is *physical* and the curve-based generation is carried out at story level. Stella admits curves representing any variable and not only *tension*.

As the proposed model evolves, the possible taxonomy would be richer and it would probably make it possible to create families of story generation systems. Expanding the model so that it considers more fundamental aspects of curves is planned as part of the future work.

5. Discussion

The proposed theoretical system for driving automatic story generation by the use of plot curves is not enough by itself. Defining objectives is much more complex than the representation of some aspects of the final story. Objectives can be required to be able to enforce a specific action or event to occur in the story. While a careful design could let plot curves be that specific, they are most likely not the best option for that.

In that sense, plot curves can only prototype a certain set of features of the desired story. If a certain event must be present in the resulting story, a logic condition on the objective state of the state space search would be a easier option (assuming that the story is created by searching). In general, the most expressive definition of generation objectives would probably contain several structures: curves, length constraints, a set of events to be included and so on.

Along this paper it has been assumed that curves are two-dimensional. While this can be the case, it makes sense to devise a system in which curves can have more than two dimensions. Curves could actually be represented by planes or hyperplanes, thus letting the generation system take several non-independent values into account. This would clearly improve the coverage of several narrative aspects, for instance the relation between love and tragedy if those were concrete variables.

If instead of having a curve that represents a value (like *love*) against time, we had a curve that evolves over time and also depends on the number of characters, an additional dimension would be added to the Φ function. There is no theoretical aspect in the proposed model that prevents from doing that, but an implementation of the new Φ function would be more complex. Every dependency between love and the number of characters would have to be coded in order to have a perfect function. This is probably not impossible, but it requires more effort. While this expansion of the current model is beyond the scope of the paper, further work contemplates it.

Section 2.2. describes a classification of curves. The authors are aware that other classifications are possible. This proposal focuses not on the general aspects of curves and their appropriateness for representing data, but on those aspects of automatic story generation that must be represented in most cases. This list has been designed based on an analysis of the literature and on the author's experience, but it is possible to think of a different taxonomy.

Additionally, a deep understanding of curves has the potential to help in defining standard plot lines for various curve configurations. This could automate the process of story generation by letting the user choose among a predefined and possibly modifiable set of typical plot lines. Such an option could then be used by an automatic story generation system for producing classical plots in a very straightforward way from the user's point of view.

The proposed theory is not complete by itself in a specific Artificial Intelligence system. It assumes that the Φ function can be defined in a domain-specific manner. While this is true in general, it must be taken into account that this definition can be very complex. For example, a narrative generation program usually must be able to handle a certain set of semantic symbols that represent some domain. Every implementation of this theoretical model should therefore be able to provide an implementation of Φ able to yield a value for every symbol. Additionally, the impact of some specific symbols will be context dependent. The implementation of the domain dependent part of the model is therefore assumed to be very complex.

5.1. Story Curves and Discourse Curves

The difference between story curves and discourse curves deserves special attention. When a story is transmitted in some or other way, the discourse reorders and filters events from the original plot, and the sequence of events being told matches time in a different way. This has a major implication for a story generation algorithm: if a curve to be matched represents a discourse, the story generation system must be able to predict the discourse, the generation would be wrong otherwise.

A different option would be to use two different types of generator. One of them would create the story plot by ordering the events sequentially and the other one would focus or creating a discourse based on the original plot. Both in this option and with a story generator aware of the discourse generation, the set of functions for the different variables would be radically different.

The set of Φ functions for story generation have to care about a strict temporal ordering and with strong focus on causality and effects. How the *danger* in a story evolves depends on the rules defining the domain for a certain story. In this case the "real danger" is taken into account no matter the perception of it by an audience.

On the other hand, those Φ functions defined for a discourse do not have to take time into account necessarily. In this kind of generation, the *danger* is not in the story but in the discourse, and therefore the importance of the audience is high. The protagonist of the story could be safe in some scene, but if the set up of that scene gives the feeling that the protagonist is about to die, the curve for the *danger* value would have to go up.

As a conclusion, the implementation of the domain-dependent definition of the functions defining curves depends on whether the curve will be matched against a discourse or against a story plot.

6. Conclusion

In this paper a theoretical definition of plot curves to drive story generation in a automatic story generation system has been presented. The relation of this model with existing story generation algorithms has been explained and the benefits and drawbacks have been discussed.

So far the model has proven to be valid in programmed prototypes and to create a simple taxonomy for story generation systems. More work has to be done in order to demonstrate the general validity of the proposed theory.

As previously discussed, the model can be refined. Currently the research on this model is focusing on trying to discover general aspects of curves regarding their semantic properties in order to lessen the amount of effort needed to create a domain-specific implementation of every variable. While some systems base the curve description on the single idea of *narrative tension*, the authors think the generation can be much richer if several variables guiding a narrative are taken into account.

7. Acknowledgements

This research is funded by the Ministerio de Investigación, Ciencia e Innovación through the project NOVA: TIN2009-14659- C03.

8. References

Aristotle. 1974. *Poética*, volume 8. Colección Biblioteca Románica Hispánica IV.

Leandro Motta Barros and Soraia Raupp Musse. 2008. Towards consistency in interactive storytelling: Tension arcs and dead-ends. *Comput. Entertain.*, 6:43:1–43:17, November.

Kevin Buchin, Maike Buchin, and Yusu Wang. 2009. Exact algorithm for partial curve matching via the Fréchet distance. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA09)*, pages 645–654.

M. Cui, J. Femiani, J. Hu, P. Wonka, and A. Razdan. 2009. Curve matching for open 2d curves. *Pattern Recogn. Lett.*, 30:1–10, January.

R. Raymond Lang. 1999. A declarative model for simple narratives. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*, pages 134–141. AAAI Press.

M Lebowitz. 1985. Storytelling as Planning and Learning. *Poetics*, 14:483–502.

C. León and P. Gervás. 2011. A top-down design methodology based on causality and chronology for developing assisted story generation systems. In *Proceedings of the 8th ACM conference on Creativity and cognition*, C&C '11, pages 363–364, New York, NY, USA. ACM.

J. Meehan. 1976. *The Metanovel: Writing Stories by Computer*. Ph.D. thesis, Yale University.

R. Pérez y Pérez. 1999. *MEXICA: A Computer Model of Creativity in Writing*. Ph.D. thesis, The University of Sussex.

M. Riedl and R. Young. 2006. Story Planning as Exploratory Creativity: Techniques for Expanding the Narrative Search Space. *New Generation Computing, Computational Paradigms and Computational Intelligence. Special Issue: Computational Creativity*, 24(3):303–323.

D. E. Rumelhart. 1975. Notes on a schema for stories. *Representation and Understanding: Studies in Cognitive Science*, pages 211–236.

R.B. Tobias. 2012. *20 Master Plots: And How to Build Them*. F+W Media.

Scott Turner. 1992. *MINSTREL: A Computer Model of Creativity and Storytelling*. Ph.D. thesis, University of California at Los Angeles, Los Angeles, CA, USA.

Nelson Zagalo, Anthony Barker, and Vasco Branco. 2004. Story reaction structures to emotion detection. In *Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, SRMC '04, pages 33–38, New York, NY, USA. ACM.

Simulating Plot: Towards a Generative Model of Narrative Structure

Graham Alexander Sack
Columbia University
English & Comparative Literature Department
E-mail: gas2117@columbia.edu

Abstract

This paper explores the application of computer simulation techniques to the fields of literary studies and narratology by developing a model for plot structure and characterization. Using a corpus of 19th Century British novels as a case study, the author begins with a descriptive quantitative analysis of character names, developing a set of stylized facts about the way narratives allocate attention to their characters. The author shows that narrative attention in many novels appears to follow a "long tail" distribution. The author then constructs an explanatory model in NetLogo, demonstrating that basic assumptions about plot structure are sufficient to generate output consistent with the real novels in the corpus.

Keywords: computer simulation, narrative theory, narratology, agent-based modelling, theory of the novel, plot, character

1. Introduction

Although computer-based analysis remains a minority pursuit in literary criticism, it has gained particular traction over the past 25 years within the subfields of stylistics and authorship attribution. Studies in this area generally utilize statistical analysis of word frequencies to identify similarities and differences in authorial style (see Burrows "Delta" 2002). The study that follows draws inspiration from this body of research by counting the frequency and co-occurrence of a generally ignored sub-class of common words: character names.[1] However, my approach and intentions differ in two crucial respects from previous studies.

First, rather than style, this paper is concerned with plot and characterization, two areas about which computational analysis has had little to say. As critic Franco Moretti has argued, plot is the crucial element that must be quantified if computational methods are to gain traction in mainstream literary criticism. This paper is an effort to do so.

Second, the overwhelming majority of prior computational studies in literary criticism have been descriptive—counting and classifying the surface features of a text. This study, however, is focused on generative models. Although I make use of descriptive analysis, the intent is to motivate a computer simulation that I will show is sufficient to reproduce several key stylized facts about actual narratives.

This paper is divided into two parts:

Part 1 uses descriptive quantitative analysis to develop a set of stylized facts about plot and characterization based on a corpus of sixty 19th Century British novels.

Part 2 develops and reports the results from a computer simulation of narrative structure.

[1] Character names are often regarded as noise and excluded from authorship and stylistics analysis because they are not consistent across texts.

2. Descriptive Analysis

2.1 The "Long Tail" in Narrative Attention

In *The One vs. The Many* (2003), literary critic Alex Woloch repositions the questions of plot and characterization with which narratologists and formalists have traditionally been concerned in terms of the concept of "narrative attention." Woloch announces his intention to…

> redefine literary characterization in terms of [a] distributional matrix: how the apportioning of attention to any specific individuals is intertwined with the narrative's continual apportioning of attention to different characters who jostle for limited space within the same fictive universe (Woloch, 13).

Woloch argues that "narrative attention" in novels (and, by extension, in narratives generally) is a scarce resource that authors must choose how to allocate amongst the characters populating their stories. "Attention" is a broad term that may encompass a variety of effects including the *frequency* of representation as well as its *intensity* or *memorability*. Woloch uses "attention" as an index of character development that is not strictly reducible to traditional concepts such as point-of-view or focalization.[2]

Taking a cue from Woloch, this paper begins by applying quantitative rigor to the qualitative concepts of "distribution" and "apportioning of narrative attention." While recognizing that "attention" may accumulate in a variety of ways, for the sake of this study I will adopt the simplifying assumption that the distribution of name mentions (an observable metric)[3] can be used as an

[2] Woloch offers Mr. Jingle in Dickens' *The Pickwick Papers* as an example of a minor character with few appearances whose distorted speech patterns nevertheless draw disproportionate attention away from the novel's weak protagonist.

[3] For each novel, a list of potential character names was generated using the Stanford NER, consisting of all proper names occurring ≥ 5 times. Error rates for the unsupervised

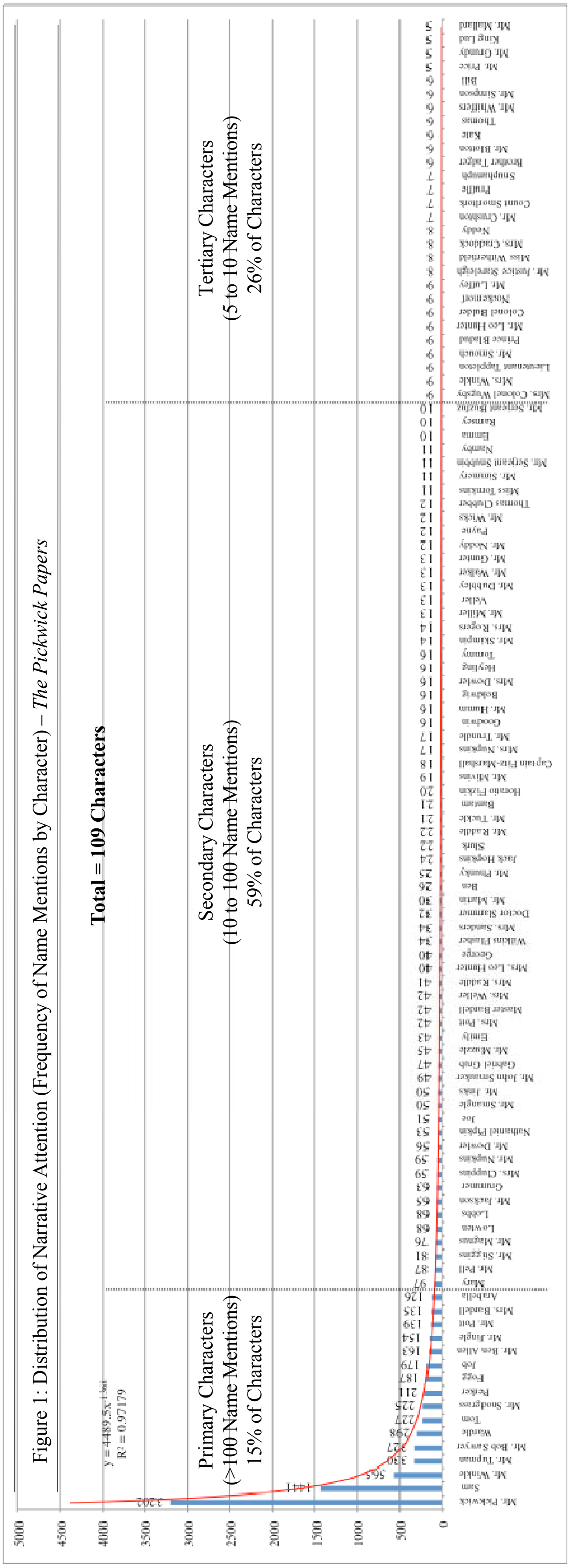
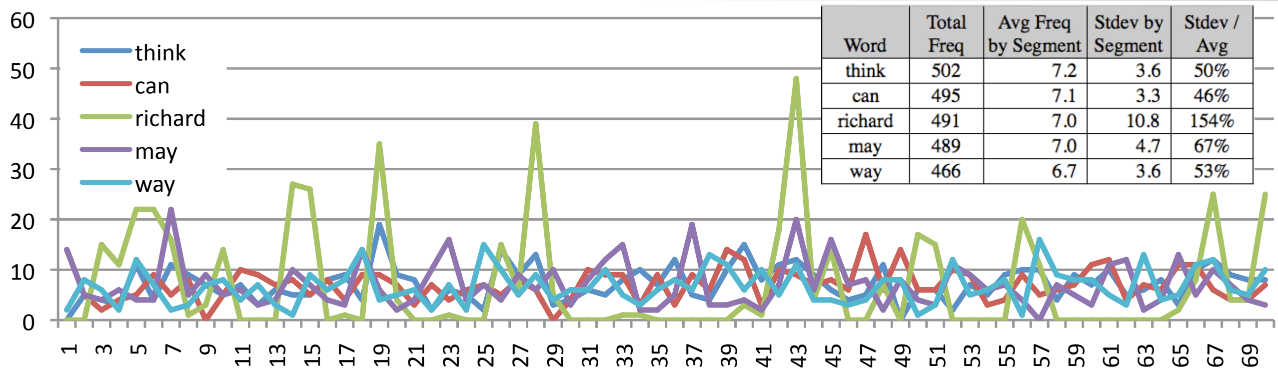


Figure 1: Distribution of Narrative Attention (Frequency of Name Mentions by Character) – *The Pickwick Papers*

Table 1: Goodness of Fit by Novel

Title of Novel	Power Law	Expon-ential	Logari-thmic	Linear
A Study in Scarlet	0.796	0.896	0.815	0.576
Adam Bede	0.978	0.831	0.715	0.315
Alton Locke	0.923	0.723	0.385	0.144
Annals of the Parish	0.957	0.773	0.401	0.152
Aurora Floyd	0.964	0.912	0.809	0.449
Barchester Towers	0.963	0.878	0.786	0.395
Belinda	0.953	0.923	0.732	0.368
Bleak House	0.911	0.962	0.441	0.173
Castle Rackrent	0.930	0.982	0.928	0.704
Daniel Deronda	0.957	0.929	0.776	0.383
David Copperfield	0.966	0.889	0.281	0.079
Deerbrook	0.896	0.967	0.827	0.460
Doctor Thorne	0.946	0.927	0.724	0.357
Dracula	0.869	0.870	0.884	0.693
East Lynne	0.954	0.941	0.759	0.366
Emma	0.939	0.959	0.804	0.436
Hard Cash	0.990	0.835	0.524	0.181
Henry Esmond	0.958	0.900	0.473	0.188
History of Pendennis	0.991	0.830	0.523	0.170
In the Year of Jubilee	0.919	0.937	0.790	0.470
Jack Sheppard	0.954	0.924	0.772	0.383
Jane Eyre	0.963	0.811	0.316	0.098
Jude the Obscure	0.970	0.822	0.754	0.406
Lady Audley's Secret	0.947	0.937	0.711	0.365
Little Dorrit	0.866	0.987	0.816	0.460
Mansfield Park	0.933	0.954	0.820	0.460
Mary Barton	0.940	0.937	0.778	0.450
Middlemarch	0.963	0.865	0.787	0.372
New Grub Street	0.913	0.951	0.886	0.613
North and South	0.911	0.935	0.696	0.367
Oliver Twist	0.869	0.968	0.843	0.503
Our Village	0.888	0.664	0.575	0.291
Paul Clifford	0.949	0.900	0.850	0.495
Persuasion	0.865	0.987	0.939	0.654
Phineas Finn	0.894	0.965	0.765	0.400
Pride and Prejudice	0.898	0.985	0.893	0.575
Sybil	0.951	0.923	0.794	0.399
Tess / d'Ubervilles	0.978	0.831	0.638	0.308
Ambassadors	0.876	0.960	0.859	0.615
Bride of Lammermoor	0.973	0.904	0.762	0.399
Egoist	0.846	0.986	0.947	0.684
Heart of Mid-Lothian	0.964	0.920	0.708	0.328
Mill on the Floss	0.952	0.938	0.784	0.404
Moonstone	0.880	0.980	0.940	0.649
Richard Feverel	0.941	0.954	0.800	0.434
Pickwick Papers	0.972	0.895	0.426	0.140
Picture of Dorian Gray	0.901	0.931	0.909	0.746
Portrait of a Lady	0.912	0.955	0.839	0.502
Return of the Native	0.927	0.897	0.826	0.526
Sign of the Four	0.988	0.861	0.772	0.440
Jekyll and Hyde	0.802	0.958	0.928	0.936
Tenant of Wildfell Hall	0.907	0.879	0.366	0.134
Way We Live Now	0.939	0.955	0.765	0.373
Wings of the Dove	0.876	0.990	0.936	0.680
Woman in White	0.961	0.939	0.802	0.441
Tom Brown	0.980	0.790	0.466	0.181
Vanity Fair	0.980	0.795	0.622	0.235
Villette	0.964	0.881	0.415	0.158
Waverley	0.979	0.884	0.662	0.296
Wuthering Heights	0.930	0.943	0.736	0.412
Average	**0.931**	**0.907**	**0.721**	**0.406**

Figure 2: Word Frequency by 5000 Word Segment (*Bleak House*)



Word	Total Freq	Avg Freq by Segment	Stdev by Segment	Stdev / Avg
think	502	7.2	3.6	50%
can	495	7.1	3.3	46%
richard	491	7.0	10.8	154%
may	489	7.0	4.7	67%
way	466	6.7	3.6	53%

instrumental variable for the distribution of "narrative attention" (a latent, unobservable variable).[4] For dramatic rather than narrative plot structures this instrument could be modified—for example, for a film or TV series one might measure screen time, while for a play one might measure the number of lines that a character speaks. While far from exhaustive, this metric is adequate to reveal a variety of compelling patterns.

By way of example, Figure 1 depicts the statistical distribution of character name mentions in Charles Dickens' *The Pickwick Papers.* The result is striking—109 characters organized into what one might term "the long tail": a small set of central characters represented by the spike on the left followed by a steep drop off to a long but shallow tail consisting of dozens of characters who are mentioned fewer than 10 times. Mr. Mallard, Mr. Price, Mr. Grundy, Bill—even a reader exceptionally well-versed in this novel is unlikely to recognize these names or remember the existence of these characters; and indeed, that seems to be the point. The characters at the far end of "the long tail"—which roughly correspond to what Woloch calls "minor minor characters" (Woloch, 116)—exist to be forgotten. The large volume of such characters is inseparable from the paucity of name mentions: readers experience them as a depersonalized mass rather than as individuals, as narrative scaffolding, on the border between character and landscape. Beyond the right edge of distribution lie even deeper levels of obscurity: de-individualized choral characters, anonymous strangers, unnamed servants.

Table 1 shows the goodness for fit for power law, exponential, logarithmic, and linear curves against the character name distributions for sixty 19th Century British novels. The data shows that the distribution of narrative attention in most novels from the period approximates either a power law or exponential distribution, implying that the "long tail" is a common pattern in novelistic form.

A wide range of phenomena are also known to follow a long tail: wealth distribution, website hits, and online books sales, for example, all obey a power law. The data for the novels sampled suggests that character name mentions and, by extension, narrative attention, are similarly distributed. That the distribution of attention within a novel should closely resemble the distribution of wealth within a nation is a provocative fact that calls for explanation.

One answer may be that the long tail in narrative attention is merely a special case of Zipf's law, which states that word frequencies in a large corpus follow a power law. Since character names are a subset of the words in a novel (accounting for ~2-4% of all word occurrences on average), it may seem intuitive that they too should follow a power law. But there are a few problems with this explanation.

First, although character name mentions in nearly all of novels in the sample follow a long tail, they do not all follow a power law: names in many novels lack the sharp peak typical of power laws and are better approximated by an exponential distribution (Table 1).

Second, character names are not distributed across a text in the same way as other classes of words. The frequency of common vocabulary words is relatively consistent across all segments of a text: high frequency words like "of," "and," and "the" are high frequency everywhere. The prevalence of character names, on the other hand, varies substantially. For example, of the 250 most frequent words in Dickens' *Bleak House*, 19 are character names and 231 are common vocabulary words. If the text is divided into 5000 word segments, the frequency of the typical common vocabulary word varies from segment to segment with a normalized standard deviation of 60%. For character names, the standard deviation across segments is 214%. Figure 2

lists varied widely (8%-54%) making it necessary to vet them by (1) cross-checking entries against reference works on literary characters or (2) manually checking names in the body of the texts. Place names and other sources of noise were removed and character name variations merged into single entities. Character name counts do not include anaphora.

[4] Critic William Gass has argued for such a radically linguistic theory of character: "A character, first of all, is the noise of his name… to create a character is to give meaning to an unknown X; it is absolutely to define; and since nothing in life corresponds to these X's, their reality is borne by their name. They are where it is" (Gass, 49-50). Following Gass, this study assumes that where a character's name is present, that character is present.

provides a clear picture of the difference: the most frequent name in *Bleak House* is "Richard" (a reference to the character Richard Carstone). "Richard" appears roughly the same number of times as the words "think," "can," "may," and "way," but it has 2-3 times the standard deviation. This difference reflects the fact that high frequency vocabulary words are determined by an author's style, which, at least for 19th Century novels, tends to be fairly consistent across a text, while character name prevalence is determined by the plot, which varies substantially. The distribution of attention in novels, then, is best approached by looking at how characters are instantiated on a scene-by-scene basis in the plot.

2.2 How Narrative Attention Accumulates

To better understand the long tail distribution, it is helpful to do an inspection of the way narrative attention accumulates over the course of a novel. I begin by using a word frequency analysis program (the Intelligent Archive) to divide each novel into 5000 word segments (ignoring chapter breaks) and I then count the number of times that each character is mentioned in each segment. The result is a set of time-varying "character prevalence vectors" that can be graphed to provide a visualization of plot and character development. I graph the name mentions for the top 25 characters in each novel on both (1) a segment-by-segment basis and (2) a cumulative basis.

Consider two representative cases: Jane Austen's *Pride and Prejudice* and Charles Dickens' *Bleak House*.

Pride and Prejudice provides a base-case for the way narrative attention accumulates over time in novels. As Figure 3(a) shows, Elizabeth Bennett dominates narrative attention in *Pride and Prejudice*: she is named ~800 times, twice that of the next most mentioned character. The remainder of the dramatis personae fall off in development gradually, with no sharp breaks or discontinuities. Figure 3(b) shows the attention paid to each character in each 5000 word segment of the novel. Elizabeth (represented by the dark blue line) is the dominant presence in almost every segment of the novel. The secondary cast is represented episodically by a succession of peaks: Jane (segment 2), Darcy and Charles Bingley (segment 3), etc. Narrative attention cycles through these secondary characters, returning to each every 2-4 segments to allocate a "peak." Two tiers of characters emerge: Elizabeth, the consistent, primary object of narrative attention and a secondary cast of 6-10 characters, who occupy background positions in the narrative with occasional moments of foregrounding. While Elizabeth is the source of narrative consistency, it is via this process of rotation through secondary figures that the novel generates a sense of plot development and variety. Figure 3(c) offers a cumulative perspective on this process. One notes the near-perfect linearity of Elizabeth's development and the relative straightness of all the other paths. The linearity of Elizabeth's path reflects the extreme consistency of the narrative attention devoted to her: Elizabeth's name is mentioned roughly 25-40 times in nearly every one of the novels 24 5,000 word segments so that her cumulative appearance by the nth segment is roughly n times her appearance in the first. Moreover, we note that the lines do not cross in the cumulative diagram. The relative position of each line indicates the corresponding character's ranking in terms of overall narrative importance. Elizabeth is 1st, Darcy is 2nd, Jane is the 3rd, etc. The fact that the lines do not (or rarely ever) cross means that these rankings never change. The structure of character development is static: the characters that are marked as narratively important in the first several chapters remain so throughout the remainder of the novel. Likewise, characters initially assigned to minority positions will never change their place in the narrative order of things. Narrative attention in entirely predictable: once a secondary character, always a secondary character.

Dickens' mid-Victorian multi-plot behemoth, *Bleak House*, provides a striking contrast. *Bleak House* consists of 69 5,000 word segments, features an enormous cast of characters (81 by my count), and mixes first and third person point of view. Looking at figure 4(b), we note the obvious differences from *Pride and Prejudice*: narrative attention is distributed as a dizzying series of disconnected, sharp peaks with no overarching source of consistency: characters appear for a segment or two and then step out of frame. The development of attention devoted to the primary characters in *Bleak House* proceeds in a manner analogous to that of the secondary characters in a single-plot novel such as *Pride and Prejudice*, that is, through an organizing logic of rotation. The novel cycles through its enormous cast characters episode by episode, developing them in fits and starts. The wavy, plateauing paths in figure 4(c) are symptomatic of this episodic pattern of development: a character receives a burst attention and then is ignored for a half dozen segments until there is another burst of attention. Moreover, there is a thorough confusion of narrative ranking, evidenced by the innumerable crossings and re-crossings of the narrative paths. The status of characters in *Bleak House* is constantly shifting as they are upgraded and downgraded in terms of narrative importance: characters that appear in the background of narrative attention in one segment may step into the foreground in another. It is impossible to predict who the primary characters will be by the novel's end based on the allocation of attention at the novel's beginning. *Bleak House* likewise lacks a high peak, with name mentions dropping off very gradually. As Table 1 shows, it is best fit by an exponential distribution rather than a power law.

Pride and Prejudice and *Bleak House* represent two poles in the temporal dynamics of narrative attention—one adhering strictly to a logic of consistency and predictability and the other to a logic of variety and unpredictability. Most other novels fall between these poles and their graphs appear as linear combinations of the contrasting temporal processes represented.

Figure 3: Narrative Attention - *Pride and Prejudice*

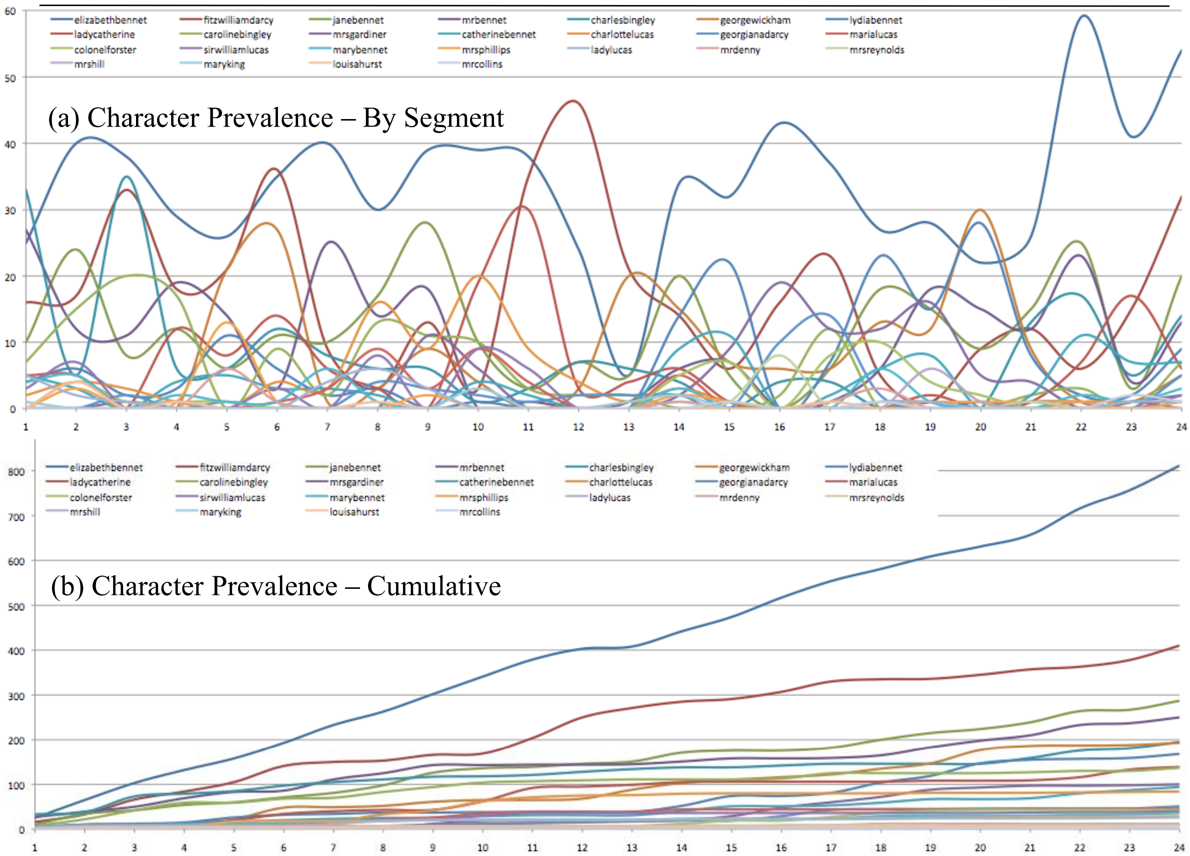
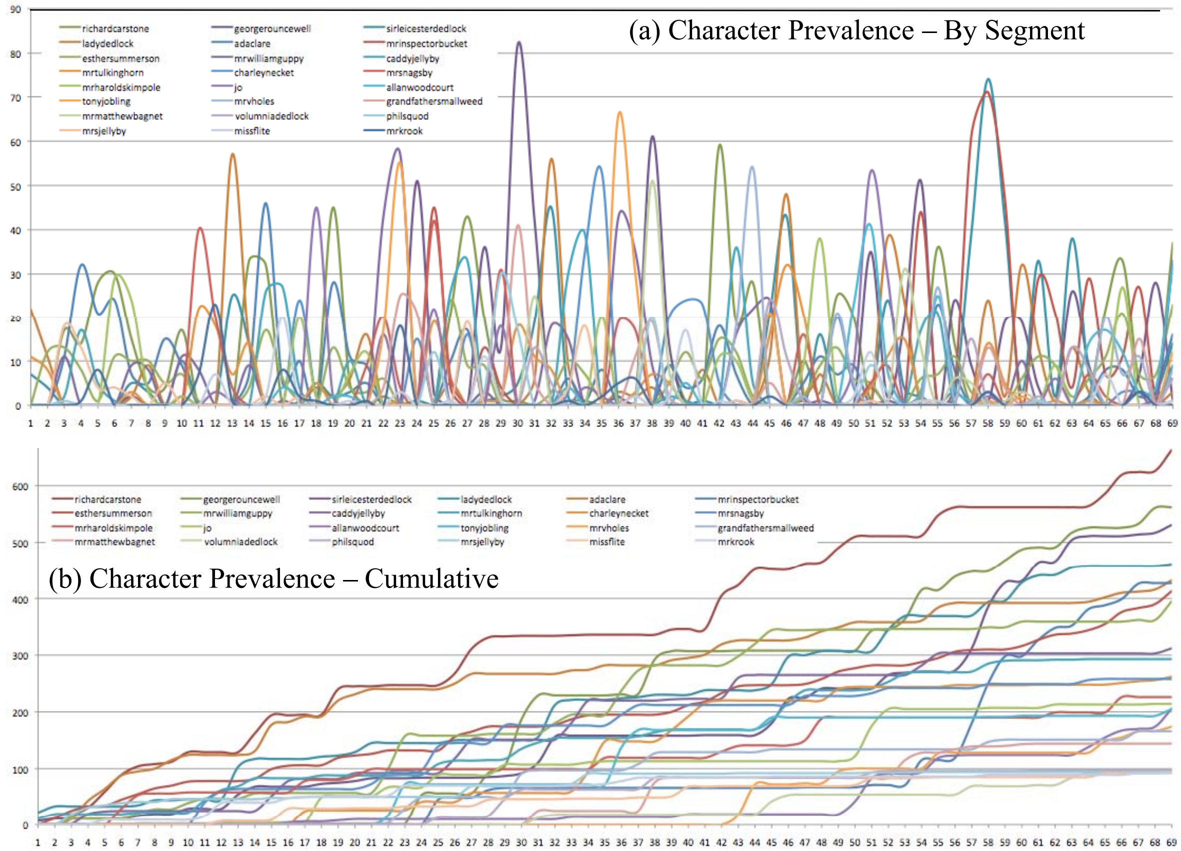


Figure 4: Narrative Attention - *Bleak House*



3. Generative Models

3.1 Simulating Narratives

Computer simulation techniques can play a valuable role in elucidating the dynamics behind narrative attention and plot described above. There are a number of potential approaches.

Characters in a narrative could be treated as independent agents in an agent-based model (ABM). Features of the narrative's structure, such as the distribution of narrative attention, would then be understood as an emergent property of rule-based character interactions. Character-agents might pursue pre-specified motives (e.g., to get married, to solve a murder); alternatively, Woloch's proposition that characters compete for scarce narrative attention could be represented by an objective function that characters seek to optimize. The dynamics of the system would be impacted by starting conditions related to a narrative's form and genre, such as the size of a novel's cast, character development conventions (e.g., whether minor characters are fixed in subordinate roles or may become the center of dramatic action in a subplot or parallel plot), and plot development protocols (e.g., linear vs. episodic plot structure, single vs. multiple plots, number and relation between subplots). Different starting conditions and rules of interaction would produce different distributions in narrative attention, which could be calibrated against actual novels to provide a better understanding of what parameters (character number, plot structure, etc.) drive structure. This would help literary critics and narratologists to situate extant authors, genres, and national and historical traditions within the range of narrative possibilities.

Such an approach treats narratives as self-organizing complex adaptive systems (CAS). Versions of this character-driven or "emergent narrative" approach have been used in a virtual reality context to generate interactive narratives (Aylett 1999, 2000; Cavazza 2007). One drawback of this approach is that it downplays the role of the author by making characters entirely self-directed. The "author," under this rubric, is present only in the starting conditions pre-specified by the choice of parameters: he is entirely non-interventionist. Although evidence certainly exists to support this version of authorship—Henry James, for example, speaks of the autonomy of his characters in the prefaces to *Roderick Hudson* and *The Portrait of a Lady*—this approach is at odds with the intuition most of us have that novels are meticulously crafted objects that undergo extensive revision; nor does this model seem adequate to describe narrative forms in which the consistency and believability of character behavior is sacrificed to other concerns, as in agit-prop political fiction.

A more realistic simulation that accounts for authorial intervention might model a narrative as the interaction between two levels of agency: an author-agent and a set of character-agents. Character-agents would pursue motives, while the author-agent would intervene to optimize an objective function related either to aesthetic criteria ("Is there sufficient conflict?"), narrative interest ("Is the plot too simple or too complex?"), or thematic content ("Does the narrative illustrate the desired themes?"). Author-driven computer models have been implemented in MINSTREL (Turner 1994) and MEXICA (Perez y Perez & Sharples 2001) for the purpose of original story generation. Moreover, criteria have been developed for measuring and modelling story novelty, conflict, dramatic arc, and suspense (see Perez y Perez et al 2008, Ware & Young 2010, O'Neill & Riedl 2011). It may be possible to adapt and repurpose this body of narrative generation AI research to facilitate hypothesis testing in literary studies.

Yet another modelling approach is to use a system dynamics sensibility, eschewing author and character agency in favor of a structuralist approach that envisions narrative as composed of sub-structures with combinatorial rules, akin to "story-grammar" narrative generation systems such as ProtoPropp (Gervás, Díaz-Agudo, Peinado, & Hervás, 2005). By way of illustration, it is this approach that I will focus on in the remainder of this paper. My central concern will be to construct an explanatory model of narrative structure using a few basic assumptions.

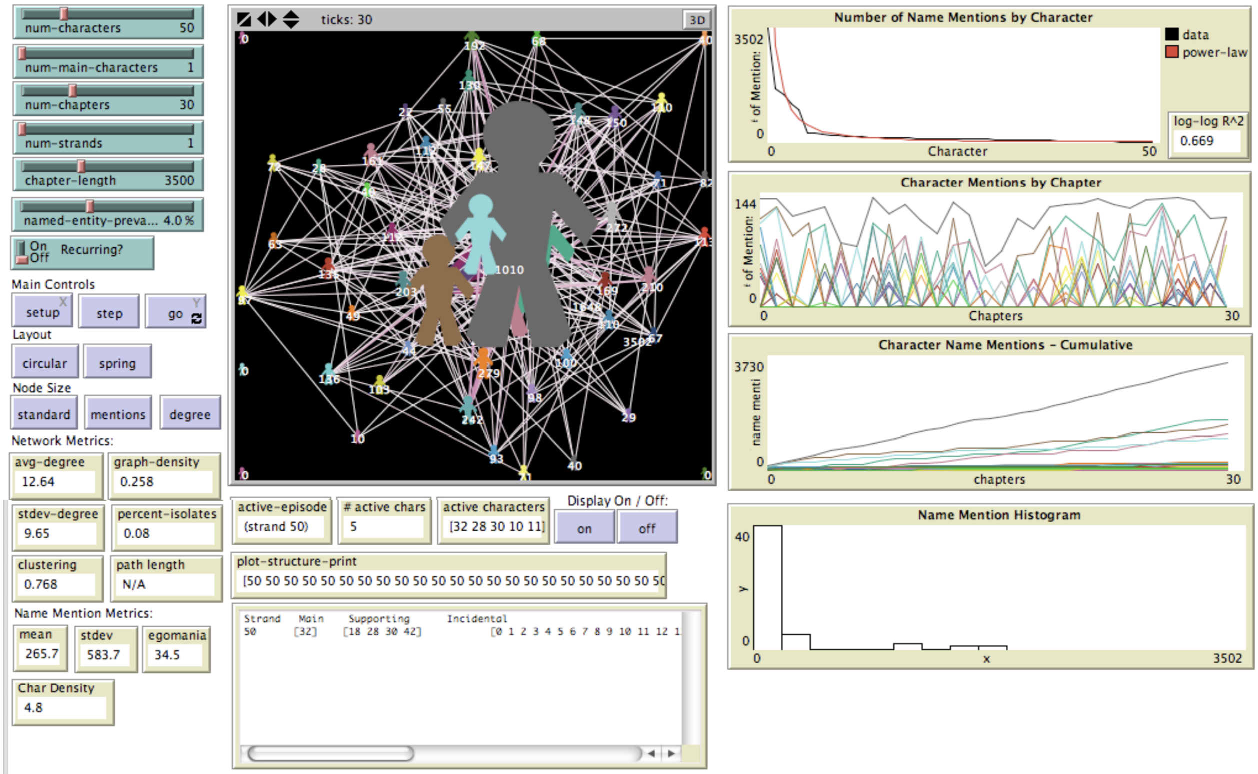
3.2 Assumptions

I begin by assuming that a plot structure is composed of a set of interwoven "plot strands." For a concrete example one might think of the plot structure of a serialized novel such as *Bleak House* or a television series like *The Wire*. Such narratives generally have multiple plot strands (in TV parlance, referred to as an "A plot," "B plot," "C plot," etc.). Each plot strand is instantiated in scenes. A plot structure, then, consists of a particular realized sequence of scenes. For example, if there are three plot strands (A, B, C), one possible plot structure might be A, B, A', C, B', A" while another might be B, A, C, B', A'. I further assume that plot strands interweave, alternating with one another such that the same strand cannot be instantiated in two consecutive scenes, and that each strand must be instantiated as a scene at least once in a plot structure.

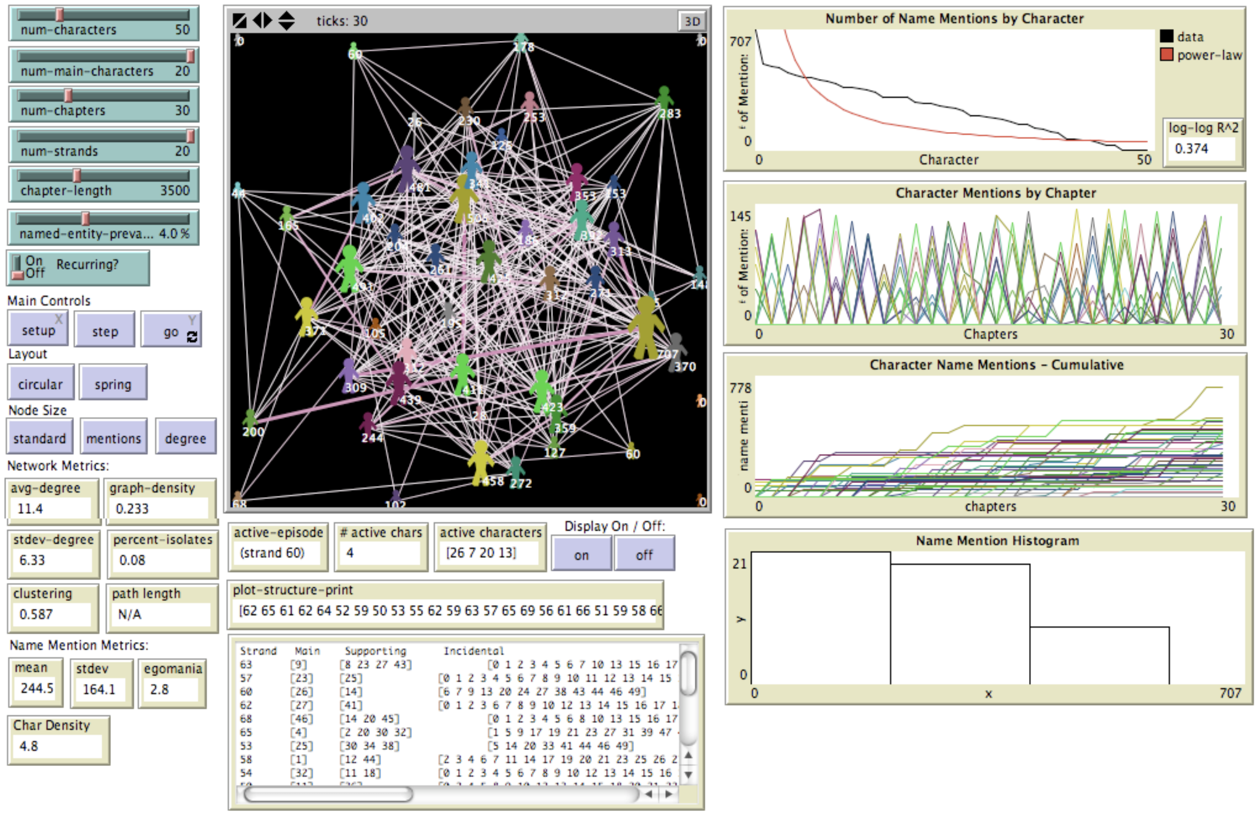
I next assume that each strand has an internal hierarchy consisting of main characters, supporting characters, and incidental characters who appear in its scenes. These characters occupy different levels of importance to the plot and therefore receive varying levels of narrative attention. For modeling purposes, consistent with the data I have gathered for novels, I assume that "narrative attention" can be measured instrumentally by the number of times that a character's name is mentioned. Main characters are assumed to be the primary focus of a plot strand and therefore must appear in all of a strand's scenes and receive the greatest level of narrative attention. Supporting and incidental characters may or may not appear in any given scene

Figure 5: Examples of Model Output

(a) # of main characters = 1; # of plot strands = 1



(b) # of main characters = 20; # of plot strands = 20



and receive less attention than main characters.

3.3 Methodology

NetLogo was used to implement this model. The user specifies the number of characters, plot strands, and scenes. At set-up, the model generates (1) a character hierarchy for each strand consisting of main, supporting, and incidental characters, and (2) a random plot sequence consistent with the combinatorial rules specified above. The model then progresses sequentially through the plot, instantiating each strand as a scene in the predetermined order. The model is stochastic and each time a strand is instantiated as a scene, three things happen:

1. A list of characters is randomly selected to appear in the scene from the strand's hierarchy.
2. A quantity of narrative attention (measured by name mentions) is randomly allocated to each character. The total amount of attention available is fixed by chapter length and name prevalence, which are user specified. As a result, attention is a scarce resource and allocation is a zero-sum game, consistent with Woloch's thesis.
3. To represent character interactions, a weighted undirected link is formed between each pair of characters appearing in a scene. The link is weighted as a random overlap between the number of name mentions of each character it links.

The model generates output in several formats: (1) time-plots of the scene-by-scene and cumulative number of name mentions assigned to each character, (2) an overall distribution of narrative attention along with a measure of the fit of this distribution against power law and exponential functions, and (3) a social network diagram and network metrics describing the character interactions.

3.4 Results

Although simplistic in its assumptions, this simulation is sufficient to reproduce a number of the salient features of narrative attention in the novels sampled.

If the number of plot strands and main characters are set low—corresponding to a narrative that is tightly focused around one or a few characters in a single story line—the results closely resemble those observed for a Bildungsroman such as *Pride and Prejudice*. See figure 5(a). The cumulative diagram output by the model depicts discrete, non-intersecting trajectories similar to those we saw for Austen's now, reflecting consistency in character development and rigidity in the rankings of narrative importance. The distribution of attention across the characters fits a power law with a high R-squared.

If the number of plot strands and main characters are set high—corresponding to a narrative focused around a large ensemble of characters across many subplots or parallel plots—the results closely resemble those observed for a sweeping social problem novel such as *Bleak House*. See figure 5(b). The model reproduces the many-peakedness of the scene-by-scene diagram and the plateau or wave shape of the lines in the cumulative diagram (indicative of limited bursts of narrative attention rotating through the large cast) as well as the many crossings of the character development trajectories (indicative of the unpredictable / shifting status of the characters in terms of importance to the plot).

Figure 6: Parameter Sweep of Model Output

Constants: # of characters = 50; # of scenes = 30

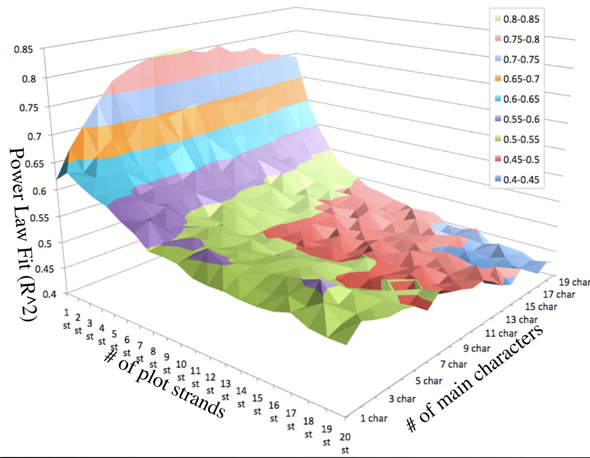


Figure 6 shows a sweep of the model's output in parameter space. The z-axis is the average goodness of fit of a power-law distribution. The x-axis represents the number of main characters (from x = 1 to x = 20) and the y-axis the number of plot strands (from y = 1 to y = 20). The number of characters is held constant at 50 and the number of scenes is held constant at 30. The model is run 40 times for each (x,y) pair, for a total of 16,000 runs. As the graph shows, the distribution of narrative attention fits a power law well for a low number of plot strands. As the number of plot strands increases, the fit erodes, particularly if the number of characters is increased along with the strands.

4. Conclusion

The simulation that I have developed is intentionally simplistic: I have modelled plot structure and characterization only in terms of combinatorial rules for plot strands. I have not attempted to give any internal sophistication to characters (such as motives), nor have I attempted to represent anything in terms of thematic or generic content. Nevertheless, this simple model is sufficient to generate results directionally consistent with the way actual novels allocate narrative attention. I have not shown that this assumption is necessary, merely that it is sufficient, and there are a number of other models that may be capable of generating similar results, such as the agent-based models of character interaction I outlined above.

This has been intended as both a methods paper and a case study. The author hopes that it has offered an example of the way that simulations can empower computational literary criticism to move beyond the description of surface features to the testing of hypotheses about plot, character, and narrative structure.

5. References

Aylett, R. "Narrative in virtual environments — towards emergent narrative." In Mateas, M., & Sengers, P. (Eds.), *Narrative Intelligence: Papers from the AAAI Fall Symposium* (Technical Report FS-99-01). AAAI Press, Menlo Park. 1999.

Aylett, R. "Emergent narrative, social immersion and storification". In *Proceedings of the 1st International Workshop on Narrative and Interactive Learning Environments*. 2000.

Burrows, John. "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing*. Vol. 17, No. 3, 2002. 267 – 287.

N. Dames, D. Elson, K. McKeown, 2010. "Extracting Social Networks from Literary Fiction." *Proceedings of 48th Annual Meeting of Assoc. for Computational Linguistics*, Sweden.

D. Easley, J. Kleinberg. "Power Laws and Rich Get Richer Phenomena." *Networks, Crowds, and Markets.* Cambridge UP, 2010.

Finkel J., Grenager T., & Manning C. "Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling." *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics* (ACL 2005), pp. 363-370.

Gass, William. "The Concept of Character in Fiction." *Fiction and the Figures of Life*. New York: Vintage Books, 1971.

Gervás, P., Díaz-Agudo, B., Peinado, F., & Hervás, R. "Story plot generation based on CBR." Journal of Knowledge-Based Systems, 18 (4–5). 2005.

James, Henry. Preface to *Roderick Hudson* and *The Portrait of a Lady*. *Henry James: Literary Criticism.* New York: The Library of America, 1984.

Moretti, Franco. "Network Theory, Plot Analysis." *New Left Review*. Vol. 68. Mar-Apr 2011.

O'Neill, Brian & Riedl, Mark. "Toward a Computational Framework of Suspense and Dramatic Arc." *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction*. 2011.

Perez y Perez, R., & Sharples, M. "MEXICA: A computer model of a cognitive account of creative writing." *Journal of Experimental and Theoretical Artificial Intelligence*, 2001. 13.

Turner, S. R. (1994). *The Creative Process: A Computer Model of Storytelling.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Ware, S. & Young, R. M. "Modeling Narrative Conflict to Generate Interesting Stories." AAAI, 2010.

Wilensky, U. 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern Univ. Evanston, IL. http://ccl.northwestern.edu/netlogo/.

Woloch, Alex. *The One vs. the Many*. Princeton: Princeton UP, 2003.

A Choice-Based Model of Character Personality in Narrative

Julio César Bahamón, R. Michael Young

Liquid Narrative Group, North Carolina State University
Raleigh, NC, USA
jcbahamo@ncsu.edu, young@csc.ncsu.edu

Abstract

The incorporation of interesting and compelling characters is one of the key components of effective narrative. Well-developed characters have features that enable them to significantly enhance the believability and overall quality of a story. In this paper we present preliminary research on the development of a computational model aimed at facilitating the inclusion of compelling characters in narrative that is automatically generated by a planning-based system. The model centers on the use of an intelligent process to express character personality. In this model, personality is operationalized as behavior that results from choices made by a character in the course of a story. This operationalization is based on the Big Five personality structure and results from behavioral psychology studies that link behavior to personality traits. We hypothesize that the relationship between choices and the actions they lead to can be used in narrative to produce the perception of specific personality traits in an audience.

1. Introduction

The automatic generation of character behavior in Interactive Narrative (IN) is an area where much work is still possible. Characters are an essential part of narrative; their features and nuances can add to the complexity of a story and its discourse. The presence of compelling characters that have distinct and well-defined features is a principal contributor to the effectiveness of narrative. Effective characters enable the audience to form a clear mental model of their beliefs, desires, intentions, and morality. This understanding of the characters can lead to a better understanding of the entire story and thus to a more compelling delivery of its content or message.

Characters with a well-defined behavior are a powerful asset in narrative composition; however, the incorporation of character behavior can greatly increase the burden placed on the IN author. As narrative delivery mechanisms such as digital games and virtual environments become more complex and detailed, the effort necessary to create characters that have distinct features increases in its difficulty. Additionally, providing users with higher degrees of agency also results in the need for characters that better adapt to user choices and to changing conditions in the story world.

This paper presents an approach for the incorporation of compelling characters in automatically generated narrative. The approach is based on the development of a computational model that enables characters to have distinct and well-defined personalities. In this model, character personality is founded on the hypothesis that character choices that lead to character actions in a story can significantly influence a character's perceived personality.

The goal of this research is not to create a model that fully recreates all known personality types but rather one that enables the representation of a general subset with enough detail to elicit a predictable cognitive response by the audience. The model could enable authors to achieve specific goals such as ensuring that audiences can clearly differentiate evil characters from non-evil ones based on how they behave, i.e. characters are defined by their choice of actions.

Results from this work most directly apply to systems used to create IN due to the reduction of authorial burden and increased creative freedom that may be provided. Additionally, focusing on the effect that choices have on personality perception in the context of narrative can help us advance models of story comprehension and more significantly develop methods to automatically generate stories that purposely affect such comprehension.

2. Background and Related Work

Planning-based narrative generation focuses on the use of AI planners to automatically generate stories that are interesting and coherent (Young, 1999). One of the principal motivations for work in this area is the importance of storytelling in human culture. Humans use stories to describe, understand, and relate events (Mateas and Sengers, 1998). Additionally, computer-generated narrative can be applied to various domains where it can assist in knowledge transfer (e.g. training simulations, activity visualizations, instructional videos).

Considerable effort has been dedicated by AI researchers to the development and improvement of techniques, algorithms, and architectures to enable the application of the problem solving capabilities of AI planners to the automatic generation of narrative that is both interesting and coherent (Riedl and Young, 2010; Riedl and Young, 2003).

In the area of IN, the ability to generate character behavior that adjusts in response to user actions or changing story conditions has not been fully addressed by researchers. Although models have been developed to direct character interactions (Riedl and Stern, 2006) and compose stories based on predefined character models (Lebowitz, 1984), none of these focus specifically on controlling character behavior within the context of a story. Furthermore, these models do not directly address the goal of eliciting in the audience the perception of specific personality types. Finally, the character models that are addressed by existing research efforts have focused on a specific subset of character actions: utterances in dialog (Mairesse and Walker, 2007; Reed et al., 2011). Our research focuses on another

class of actions –physical actions– and the role that this class plays in the construction of the mental model that the reader forms when experiencing a story.
The solution introduced in this paper relies on the operationalization of creative writing principles for the automatic generation of stories. While there are a number of principles that are relevant to the automatic support of the writing process, our work focuses on the importance of character personality for the development of fictional characters.

3. The Concept of Character in Narrative

Characters are an essential component of narrative (LaPlante, 2007; Chatman, 1978). The importance of characters becomes more apparent when we consider the critical role they play in the composition of a story. Characters are vital for the realization of crucial story elements such as events and dialog (Morrison, 2010; LaPlante, 2007; Chatman, 1978).
In order for characters to contribute to the effectiveness of a story they should be well-defined. Among the factors that contribute to the effective definition of a character we include: physical attributes, talents, emotions, beliefs, and personality. Characters that portray these factors in an interesting and believable manner are considered round; characters that fall short of this expectation are considered flat.

3.1. Personality and Its Importance to Character Development

Personality is a key component of what makes a narrative character round. A character's personality can make it more believable and compelling, consistent yet capable of surprising the audience. A character's lack of personality can create the perception of being flat, thus detracting from the story and reducing its effectiveness.

3.2. The Relevance of Actions

Characters play an essential narrative role as agents of change in a story. According to narrative theory, characters can be the recipients or originators of change (Chatman, 1978; Morrison, 2010). Change can result both from a character's actions and its reaction to the actions of others or story events, i.e., characters can act and be acted upon. It follows from this principle that actions are one of the main techniques used by creative writers to define and describe fictional characters (LaPlante, 2007; Bulman, 2007; Morrison, 2010).
The research described here focuses on actions as one of the key elements that define personality in the audience's mental model of the story. The central problem we address is the selection of character actions taking into account their properties such as goals, beliefs, and moral traits. Additionally, we consider that characters can be shaped by their reaction to story events, in particular the effect that such events can have on the choices they make.

3.3. Choice and the Expression of Personality

Considering the structure of a story, specifically plot points where branching occurs (Barthes and Duisit, 1975), we intuitively expect instances when actions follow a choice. For example, in *The Iliad*, Aquilles must choose whether to help the Greeks in the Trojan War. We deduce that choices made by characters can have a direct impact in determining the actions they perform. Furthermore, we argue that choices may be linked to specific personality traits. This idea is supported by research in behavioral psychology that has found correlation between people's actions and their personality (Mehl et al., 2006; Funder and Sneed, 1993).
We posit that the link between choice and personality can be used in narrative to enable the perception of specific personality traits. An audience that is made aware of the existence of multiple choices that are available to a character will form an opinion of such character's personality based on (1) the choices made and (2) the causal chain of events or circumstances that precede the choices.
We have identified two specific story aspects where choice and character personality intersect:

1. Stories can be constructed to include choices that express a character's personality, i.e. characters make choices that are consistent with their assigned personality traits. For instance, an *agreeable* character only makes choices that result in honest behavior.
2. Stories can be constructed to include events that justify or explain a choice that does not agree with the character's personality. This type of structure may be used to show more complex or surprising characters. For instance, an *agreeable* character makes a choice that results in dishonest behavior after multiple attempts to engage in honest alternatives.

For this research we initially focus on aspect (1), under the assumption that once it is computationally modeled, aspect (2) will be an extension that can be derived from it. Our approach addresses choice as a character-centric event that can be directly linked to a character's personality traits. We use choice as the means to express a character's personality.

3.3.1. A Narrative Example

Consider the following story with two alternate endings.

> *Amos is a farmer whose son has fallen gravely ill. His son may die unless he undergoes a very expensive surgery. Amos must obtain a large amount of money soon if he expects to save his son. Amos considers his options for obtaining the money, such as asking his friends for help, getting a loan from the credit union, selling his only tractor on eBay, or even robbing the local bank.*

Alternate Ending 1:

> *After careful consideration, he decides to sell his only tractor, even though that means that his work at the farm will be much harder from now on. Selling the tractor provides enough money to pay for the operation and save his son.*

Alternate Ending 2:

> *After careful consideration, he decides to grab his shotgun and goes to local the bank. He robs the bank and obtains enough money to pay for the operation and save his son.*

Both story endings have the same set of choices available to the character, Amos. However, the choice made in ending 1 shows a clear attempt to resolve the crisis through what could be characterized as honest behavior. In contrast, in alternate ending 2 the character engages in behavior that could be characterized as dishonest. We contend that it is the specific choice made by the character, when considering the available alternatives, that characterizes his behavior as honest or dishonest. This distinct choice is what enables the expression of a specific personality trait.

4. A Computational Model of Personality

Our approach is to create a computational model that enables the representation of distinct character personality traits in the context of a story world. The model is intended to provide authors with the ability to create story characters that have a rich set of behaviors that can be adjusted based on authorial goals and in response to story events or user interaction. The model works under the premise that narrative characters can be distinguished or classified by visible manifestations of their personality, i.e. their choices, actions, and dialog.

We focus on the creation of an intelligent process that enables the automatic generation of behavior that matches personality traits assigned to story characters. In this model, personality is expressed in the form of actions linked to choices. Actions are determined by individual elements such as goals, beliefs, moral traits, and personality. Characters are further shaped by their reaction to external events or the effect that these have on them. For example, let us consider a character's reaction to an aggression; an agreeable character may respond with forgiveness whereas a non-agreeable character may respond with revenge.

4.1. The Big Five Personality Structure

In order to design a planning-based story generator that can create characters with distinct personalities, we are developing a computational model of behavior based on personality traits. Our model uses the Big Five personality structure defined by Goldberg (1990). This structure provides a taxonomy for the classification of personality using the following five factors:

1. Extroversion
2. Agreeableness
3. Dependability
4. Emotional Stability
5. Culture (or Openness).

Within each classification there are distinct bi-polar personality traits, e.g. honesty vs. dishonesty.

Each factor is linked to specific personality traits that can be mapped to a set of behavioral manifestations. According to results obtained by social psychologists Mehl et al. (2006) and Funder and Sneed (1993), there is high correlation between personality traits and specific, observable, behaviors. Study results indicate that witnessing a certain behavior can elicit the perception of a personality trait associated with it.

4.2. Computational Approach

A simplistic approach to the development of a computational model of behavior would be to annotate the actions in the action library of a planning-based story generator (e.g., Riedl and Young (2003)) with specific personality traits. Actions are chosen by the planner during narrative generation using the annotations as part of a filter mechanism. However, this approach would not adequately achieve our purpose for several reasons. First, it requires a labor-intensive process. Every time actions are added to the planning library, it is necessary to update their annotations to indicate the specific personality traits to which they apply. More significantly, actions may need to be further annotated to indicate every situation in which it is appropriate to use them. When we consider that there could be many situations that justify or preclude the validity of an action it is evident that the work for the author would increase exponentially.

The ideal solution would use a declarative approach, in which a character's properties are used to dynamically determine the set of actions that he or she should perform. Such a method would scale to complex domains and generalize to IN applications beyond simple test cases or academic story generators.

4.3. Intelligent Action Selection

Our process aims at enabling the intelligent selection of actions considering the context in which they execute and without requiring extensive hand-annotation of actions in the planning library. In this approach, the execution context determines the appropriateness of actions for specific story characters. For example, the action $Kill(actor, target)$ may only be appropriate for an agreeable character if the context indicates either that he or she is behaving in self defense or that the target is an evil enemy who must be defeated. However, the same action may always be valid for a disagreeable character.

The process selects actions after analyzing the current execution context and evaluating the space of possible story plans. The execution context is derived from the current state of the story world, the properties of the characters and other actors in the story, and the set of open goals that are yet to be achieved in the plan. Among the specific story elements analyzed by the action selection process we include:

1. The causal chain of events that precedes the plot point where an action is needed.
2. The character's personality traits.
3. Previous actions that the character has performed.
4. Future actions that the character may perform.
5. The character's relationships, e.g. friends, enemies.
6. The set of choices that the character has already made.
7. An evaluation of the past and possible future consequences of choices made by the character.

Information obtained from the analysis of the execution context is used to advise the planning process on the selection and placement of actions, to produce desired behaviors. To this effect, we are currently working on a mapping between observable behaviors and personality traits using empirical results from social psychology (Funder and Sneed, 1993; Jackson et al., 2010). The objective is to operationalize the mapping as a set of plan structure characteristics that when present result in specific character behavior.

4.4. The Choice Process in a Planning Context

Our initial model of character choice is based on modifying the process used by a least commitment planning algorithm, such as POP (Weld, 1994), to select actions. Choice occurs after an open goal has been selected from the agenda and before a new action is added to the plan.

Two factors are considered to select the set of actions considered by the choice process: (1) the action must be relevant, i.e. one of its effects establishes a result that accomplishes the goal and (2) the action can be performed by the character, i.e. the value that represents the character who executes the action can be bound to the parameter in the action used to designate the principal actor. We assume that the planner's data structures and knowledge representation will be modified to enable reasoning about who or what performs an action.

***CHOICE* process pseudo code:**

Given a character that performs the action (C), the effects that the action must produce (F), the library of domain actions (L), and the current plan (P)
$A =$ the set of actions in L that establish F as an effect
$A_T =$ trim A by removing the actions for which C is not a principal performer of the action
$A_R =$ invoke the $RankActions$ function (see below)
while $A_R \neq empty$ **do**
    $A_e =$ select the top action from A_R
    Remove A_e from A_R and add it to the plan
    Update the agenda and causal links
    Recursively invoke the planning process.
    if a plan P is found **then return** P
    else
        if $A_R = empty$ **then return** failure
        end if
    end if
end while

The *RankActions* function analyzes the effects of an action to measure their compliance with the personality traits of a character. It returns a list of actions in descending order of compliance.

Given a set of actions (A), a character who performs the action (C), a character repository (R), and the current plan (P)
for each action A_i in A **do**
    $Rank_i = 0$
    for each effect E_j produced by A_i **do**
        for each personality trait P_k assigned to C **do**
            if $Complies(C, E_j, P_k, R, P)$ **then**
                $Rank_i = Rank_i + 1$
            end if
        end for
    end for
end for

The $Complies$ function evaluates whether an effect is consistent with behavior associated with a personality trait. For this purpose we will develop a declarative representation that enables the use of an extensible library of mappings between behavior and personality traits. For example, if C_l is an agreeable character the effect $(dead\ C_m)$ is consistent only if C_m is not a friend of C_l and C_l has motive to eliminate C_m. On the other hand, if C_l is highly disagreeable the effect is consistent regardless of the relationship between the two characters or C_l's motives. The $Complies$ function uses information from the character repository and the current plan to inform the evaluation process.

5. Discussion and Future Work

5.1. Planning Algorithm Modifications

The next step in this effort is the development of an algorithm for the evaluation and placement of story actions. The algorithm must avoid the computationally intensive option of generating all the possible plans and then selecting those that exhibit the required character personality traits. The process must also guarantee that the story is coherent, i.e. added actions are part of a valid causal chain of events.

An analysis of the plan structure characteristics needed for the model indicates that a solution solely based on new constraints and heuristics may not be sufficient. Instead it is necessary to consider changes to the process used to construct the plan structure. The modified algorithm should enable operations such as: changing the ordering of actions currently in the plan, increasing or reducing action decomposition, changing or introducing a causal chain of events, and dynamically introducing behavior-related constraints. These modifications would facilitate the construction of plans that treat choice as a first-class object, .i.e. the plan is built to include choices that result in the desired character behaviors.

5.2. Proposed Evaluation

Essential to this research will be validating the claim that narrative generated using our model includes characters that have distinguishable personalities. The validity of the claim will be tested through user studies designed to measure whether the generated character behavior elicits in the audience the perception of the corresponding personality traits.

In order to have an environment conducive to experimentation, we also propose to incorporate our narrative model into a Mixed-Initiative Story Editor (Horvitz, 1999). The editor will enable non-expert users to create stories with the assistance of an intelligent user interface.

6. Conclusion

We have presented preliminary research aimed at the development of an intelligent mechanism that enables the automatic generation of narrative that elicits the perception of distinct character personalities without the need of a labor-intensive process. We use a solution based on a declarative approach, in which a character's properties and the story context are used to model the choices that determine the set of actions that he or she performs in the course of the story.

7. References

Roland Barthes and Lionel Duisit. 1975. An Introduction to the Structural Analysis of Narrative. *New Literary History*, 6(2):pp. 237–272.

Colin Bulman. 2007. *Creative Writing: a Guide and Glossary to Fiction Writing*. Polity, Cambridge, UK.

Seymour Benjamin Chatman. 1978. *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press, Ithaca, N.Y.

David C. Funder and Carl D. Sneed. 1993. Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64(3):479–490.

Lewis R. Goldberg. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216.

Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 159–166, New York, NY, USA. ACM.

Joshua J. Jackson, Dustin Wood, Tim Bogg, Kate E. Walton, Peter D. Harms, and Brent W. Roberts. 2010. What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness. *Journal of Research in Personality*, 44(4):501–511.

Alice LaPlante. 2007. *The Making of Story: a Norton Guide to Creative Writing*. W.W. Norton, New York.

Michael Lebowitz. 1984. Creating characters in a storytelling universe. *Poetics*, 13(3):171–194.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*, pages 496–503.

Michael Mateas and Phoebe Sengers. 1998. Narrative Intelligence. In *AAAI Fall Symposium*. AAAI Press.

James R. Meehan. 1977. Tale-spin, an interactive program that writes stories. In *Fifth International Joint Conference on Artificial Intelligence*, pages 91–98.

Matthias R. Mehl, Samuel D. Gosling, and James W. Pennebaker. 2006. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5):862–877.

Matt Morrison. 2010. *Key concepts in creative writing*. Palgrave Macmillan, Houndmills, Basingstoke, Hampshire; New York.

A. A. Reed, B. Samuel, A. Sullivan, R. Grant, A. Grow, J. Lazaro, J. Mahal, S. Kurniawan, M. Walker, and N. Wardrip-Fruin. 2011. A Step Towards the Future of Role-Playing Games: The SpyFeet Mobile RPG Project. In *7th Annual Conference on Artificial Intelligence and Interactive Digital Entertainment*.

Mark O. Riedl and A. Stern. 2006. Failing believably: Toward drama management with autonomous actors in interactive narratives. *Technologies for Interactive Digital Storytelling and Entertainment*, pages 195–206.

Mark O. Riedl and R. Young. 2003. Character-focused narrative generation for execution in virtual worlds. *Virtual Storytelling. Using Virtual Reality Technologies for Storytelling*, pages 47–56.

Mark O. Riedl and R. Michael Young. 2010. Narrative Planning: Balancing Plot and Character. *Journal of Artificial Intelligence Research*, pages 164–167.

Daniel S. Weld. 1994. An introduction to least commitment planning. *AI Magazine*, 15(4):27–61.

R. Michael Young. 1999. Notes on the use of plan structures in the creation of interactive plot. In *AAAI Fall Symp. on Narrative Intelligence*, pages 164–167.

Persuasive Precedents

Floris Bex[a], Trevor Bench-Capon[b] and Bart Verheij[c]
[a] School of Computing, University of Dundee, UK
[b] Department of Computer Science, University of Liverpool, UK
[c] Department of Artificial Intelligence, University of Groningen, The Netherlands
E-mail: florisbex@computing.dundee.ac.uk, tbc@liverpool.ac.uk, b.verheij@ai.rug.nl

Abstract

Stories can be a powerful vehicle of persuasion. We typically use stories to link known events into coherent wholes. One way to establish coherence is to appeal to past examples, real or fictitious. These examples can be chosen and critiqued using legal case-based reasoning (CBR) techniques. In this paper, we apply these techniques to factual stories, assessing a story about the facts using precedents. We thus show how legal reasoning in a CBR model is equally applicable to reasoning with factual stories.

Keywords: Stories, precedents, case-based reasoning, argumentation

1. Introduction

Stories can be a powerful vehicle of persuasion. They can be used, for example, to present evidence about "what happened" in a particular case in a coherent and believable way (Bex et al. 2010), or to convince others to follow a particular course of action (Bex and Bench-Capon 2010). A story does not persuade by imparting explicit rules like an argument does, but instead by exposing a coherent and plausible sequence of events. Thus, for example, we more readily believe a set of evidence if we can structure it using some coherent story (Bennett and Feldman 1981).

One way to establish the coherence of a story is to appeal to examples, real or fictitious. In previous work (Bex 2011, Bex and Verheij 2011), we argued that a story is coherent if it fits a *story scheme*, a generalised pattern of events akin to a script (Schank and Abelson 1977). Story schemes model the way things tend to happen in the world; for instance, the restaurant script lists the roles (customer, waiter) and sequence of events (ordering, eating, paying) for a typical restaurant visit. These abstract story schemes depend on *precedent stories*: the restaurant scheme we use is based on our experiences of restaurants.

In realistic argumentative contexts people will usually find it more effective to cite a precedent story rather than an abstract story scheme. As an example, suppose two people who know each other meet on a train: on one story it is a chance encounter, in another it is an arranged meeting. If both regularly use the train at similar times a chance meeting is entirely plausible. If they rarely use the train, or live elsewhere, it is less so. But citing a particular story can help, particularly a personal one: *you remember when you met Bill on the Rialto bridge? Neither of you knew the other was in Venice, but these coincidences do happen*. The object here is to establish from personal experience that the improbable actually does occur from time to time, so the coincidence is at least possible. An appeal to personal experience or an appeal to a well-known story is much more powerful than citing a story scheme for chance encounters: *A is at location L for reason RA - B is at location L for reason RB - RA and RB are unrelated - A and B meet*. The real story provides a unity to elements which would remain entirely disconnected in the abstract scheme.

Citing a similar story thus helps establishing coherence. Here, it is important that the current story and the supporting example be relevantly similar (Walton 2010). In AI, this similarity is usually enforced by requiring that each item in the precedent matches exactly one item in the target and vice-versa, and that if there is a correspondence between two statements, then there must also be correspondences between its arguments.

If we cannot find a precedent story which matches on enough facts, we can attempt to find a more general precedent (e.g. citing a story which contains a coincidence but says nothing about chance meetings). However, in such a case it is easier to reduce the force of the example by pointing to relevant differences. These distinctions can then be *emphasised* and *downplayed*.

The work in AI and cognitive science (see Gentner and Forbus for a comprehensive overview) has so far mainly focused on retrieving precedents, matching the current story to the precedent and calculating the degree in which a story and its precedent match according to some hard-coded principles. Work in legal case-based reasoning (CBR) (Aleven 1997, Ashley 1990) presents a more fluid approach, in which identifying relevant similarities and differences between legal cases becomes the subject of argumentation.

In this paper, we show how techniques for mapping common in AI can be combined with argumentative techniques for citing, emphasising and downplaying stories can be applied to factual (i.e. non-legal) stories.

2. Legal Case Based Reasoning

The leading legal CBR systems are HYPO (Ashley 1990) and CATO (Aleven 1997). We will base our approach to CBR on CATO. The key idea of CATO is that cases can be described as collections of *factors*, stereotypical fact situations that have legal relevance (e.g. in cases concerning trade secrets we have *information_disclosed_* *in_negotiations*, *plaintiff_took_security_measures*). The facts of the case determine whether particular factors are present or absent from a case. Typically a case will contain a number of factors, some favouring the defendant and some favouring the plaintiff, and the court will need to decide which set of reasons prevail.

Guidance on the relative strengths of sets of factors can be obtained from the precedent cases. If the combination presented in a case under consideration (the current case) has been found before, then it would be expected that the decision in the past case would be the decision in the current case. Normally, however, there will be no exact match and missing and additional factors will serve to *distinguish* the current case from the precedent. Equally some precedents may point one way and other the other, so providing counter examples.

CATO supports a three ply form of argument:

1. One side cites a precedent case (a case with factors in common with the current case) decided for their side;
2. Other side presents counter examples (cases with factors in common decided for the other side) and distinguishes the cited cases;
3. Original side may distinguish the counter example, and cite any additional reasons to support their side.

CATO recognises the following argument moves: Citing a case to a past case with a favourable outcome (Ply 1); Distinguishing a case with an unfavourable outcome (Ply 2); Emphasising the significance of a distinction (Ply 2); Downplaying the significance of a distinction (Ply 3); Citing a favourable case to emphasise strengths (Ply 3); Citing a favourable case to argue that weaknesses are not fatal (Ply 3); Citing a counterexample (Ply 2). In section 4 we further discuss these moves when we relate them to stories about the facts.

3. Stories and story schemes

Stories are finite sequences of facts, events or states of affairs that are assumed, at least for the moment, to have happened or existed. Stories are specific rather than general. Consider a simple example story about Tony (T), who killed Gordon (G) in a knife fight: *stabs(T,G) – stabbing_injured(G) – died(G)*. Story schemes are abstract scenarios, the structure of which is close to that of stories. Basically, a story scheme is a sequence containing narrative units (Propp 1968), which represent (sets of) generalized facts or types of facts: *has_motive(x) – stabs(x,y) – stabbing_injures(y) – dies(y)*. The narrative units thus represent what we call *story roles*, general roles that facts in a story can take.

A story can be matched to a story scheme by assigning the facts to their respective story roles, that is, matching the facts in the story to the relevant narrative units in the scheme (Bex 2011, Bex and Verheij 2011). This matching is similar to what in existing work in AI and cognitive science (Gentner and Forbus 2011) is called *mapping*: given a base case and a target case, a mapping consists of a set of correspondences, each linking a particular item in the base with a particular item in the target. Here, both the base and target are usually specific (instantiated) structures. We follow Schank (1986) in that we match specific stories to story schemes.

After matching, the coherence of the story is determined by checking whether the story has no "loose ends" (there are facts in the story that do not match a narrative unit in the scheme) and whether the story "has all its parts" (all the narrative units in the relevant scheme are matched by a fact in the story) (Bex 2010). For example, our example story does not complete the example scheme, as there is no fact that matches the narrative unit *has_motive(x)*.

4. Argument Moves and Precedents

CATO's cases are very similar to story schemes. Story schemes are clusters of abstract facts (narrative units) and cases are clusters of legally qualified abstract facts (factors). Hence, the three ply form of argument and the argument moves from CBR can be used in a factual situation as well as a legal situation.

To determine the coherence of a story, a precedent story is cited as the basis for the construction of a story scheme. After a precedent has been cited, variants of CATO's argument moves can then be used to argue about the differences and similarities between the precedent and the current story. Effectively, these argument moves are about whether the current story relevantly matches the story scheme based on the precedent.

Citing a precedent story: This move establishes a story scheme based on the precedent story, and then (implicitly) argues that the current story matches that scheme.

Distinguishing a precedent story: The precedent story and the current story will each contain elements beyond those required for matching a common story scheme. For example, the current story and the precedent may have the central action in common (stabbing), but may well differ as to the type of people involved. Such differences can be offered as reasons to argue that the current story does *not* match the scheme established by the precedent.

We can identify different kinds of distinction between stories (cf. Wyner and Bench-Capon 2007), depending on whether the current story is missing a fact required to make the story coherent, or has an additional fact (that the precedent lacks) which jeopardises the coherence of the story. In the first case, there is an *assumption* satisfied in the precedent which is not satisfied in the current case: this means that the current story is not complete, it does not "have all its parts". In the second case, there is a fact in the current story which supplies an *exception* to the story scheme.

Emphasising the significance of a distinction: this move accompanies a distinction and attempts to pre-empt any attempt to downplay; it seems as much rhetorical as logical.

Downplaying the significance of a distinction: Downplaying a distinction has variants according to the nature of the distinction. If the distinction is an unsatisfied assumption, it is necessary to point to some fact in the current story which can play a similar role, thus having the current story complete the story scheme after all. If the current story has what appears to be an exception, downplaying involves finding a fact in the current story that provides an exception to that exception.

Citing a case to emphasise strengths and *citing a case to argue that weaknesses are not fatal:* These two moves respond to a distinction and involve citing other stories which can serve as precedent stories. When it is argued that the current story misses an assumption, new precedent stories can help to show that this assumption is not vital to the coherence of the story. Against the second type of distinction one can cite further precedent stories matching the current story and containing the alleged exceptions, showing that it is possible to have this additional fact in a coherent story. These moves are essentially attempts to shift the story scheme relied on slightly. The difference between *emphasise strengths* and *weakness not fatal* seems to be largely rhetorical, focusing on the strengths or alleged weaknesses of the current story, respectively.

Citing a counterexample: This move involves citing a new precedent story that argues for a different story scheme. Counterexamples are used to demonstrate that there are alternatives, and so avoid tunnel vision.

5. An example of reasoning with precedents

Having looked at the individual moves, let us consider an example to show them in action. In our example, the observation to explain is that Tony killed Gordon in a knife fight. That Tony killed Gordon is not at issue: there were plenty of witnesses as it took place quite openly in a Glasgow street. But it is important to get a story establishing Tony's motive, as this will affect the sentence. Wilma and Bert are discussing the matter. Note how by citing a precedent (top) for the current story (bottom), they are establishing a possible story scheme (middle).

- ***Wilma***: 'Tony and Gordon were youths from the same neighbourhood; perhaps it was a gang thing, like *West Side Story*.[1]' In the precedent, Bernardo (B) of the Sharks gang kills Riff (R) of the Jets gang in a knife fight. Here, ***citing a precedent story*** attempts to establish that the motive was gang feud:

same_area(B, R) ↓match — *feud(B, R)* ↓match — *kills(B,R)* ↓match
same_area(x, y) — *feud(x, y)* — *kills(x,y)*
↑match *same_area(T, G)* — ↑match *feud(T, G)* — ↑match *kills(T,G)*

- ***Bert***: 'But Tony and Gordon are middle class kids, and whoever heard of middle class kids being in gangs?' Here Bert ***distinguishes*** by mentioning an **exception**: the story has an additional fact, that Tony and Gordon are middle class, that the precedent story lacks – Jets and Sharks are lower class immigrant gangs. Thus, Bert argues that the current story does in fact not match the scheme established by the West Side Story precedent, because being middle class is an exception to the rule that people from the same neighbourhood may be involved in a feud:

same_area(T, G) ∧ middle_class(T) ∧ middle_class(G) ***so*** *¬feud(T, G)*

This rule means that there can be no match between the current story and scheme 1.

- ***Wilma***: 'Maybe it was a family feud like in Romeo and Juliet, they were middle class.' This is an example of ***weaknesses not fatal***, citing a precedent story with a similar motive that does include the alleged exception: Romeo Montague (R) kills Tybalt Capulet (T) with a knife and the Capulets and Montagues were middle class:

R ∧ T middle_class, — *feud(R, T),* — *kills(R, T)*
↓match — ↓match — ↓match
x ∧ y middle_class, — *feud(x, y),* — *kills(x, y)*
↑match — ↑match — ↑match
T ∧ G middle_class, — *feud(T, G),* — *kills(T, G)*

- ***Bert***: 'The Capulets and Montagues were Italian, and vendettas are very Mediterranean, but this was Scotland.' Bert ***distinguishes*** by mentioning a ***missing assumption***, that only in Italy do middle class people get into feuds:

R ∧ T middle_class, — *R ∧ T Italian,* — *feud(R, T)*
↓match — ↓match — ↓match
x ∧ y middle_class, — *x ∧ y prone_to_feuding,* — *feud(x, y)*
↑match — ??no match — ↑match
T ∧ G middle_class — *feud(T, G)*

- ***Wilma***: 'But Tony and Gordon's families were supporters of football clubs involved in a notorious feud, Glasgow Rangers and Celtic FC' Wilma ***downplays*** the distinction ***by providing facts*** that can take the place of *x and y are Italian*:

x ∧ y prone_to_feuding
↑match
different_clubs(T, G)

- ***Bert***: 'But Tony was estranged from his family, as he was in a relationship with Gordon's sister.' Bert again ***distinguishes with exception***: once disowned and allied to the other family it is unlikely that Tony would continue the football feud.

different_clubs(T, G) ∧ involved_with_family(T, G) ***so*** *¬feud(T, G)*

- ***Wilma:*** 'But Romeo still killed Tybalt in *Romeo and Juliet*, even though he was involved with Tybalt's cousin Julia.' Wilma ***downplays*** by ***denying the exception***, pointing to her previous precedent.

R ∧ T Italian, — *involved_with_family(R, T),* — *feud(R, T)*
↓match — ↓match — ↓match
x ∧ y prone_to_feuding, involved_with_family(x,y), feud(x,y)
↑match — ↑match — ↑match
different_clubs(T,G),involved_with_family(T,G),feud(T, G)

[1] Most of our precedents will be taken from fiction. Everyone knows real examples of these schemes, but they know *different* examples: classic fiction provides a common cultural repository of stories.

- ***Bert***: ‘Maybe it was about Gordon’s sister. Perhaps Gordon started the fight, like in *Hamlet*, so Tony acted in self defence.’ Bert now changes main story scheme by citing a ***counter example***, in which Hamlet (H) defends himself after Laertes (L) attacks him because the latter blames Hamlet for the death of his sister, Ophelia (O).

$$\begin{array}{c} attacks(L,H) \wedge defends(H) \\ \downarrow_{\text{match}} \\ attacks(x,y) \wedge defends(y) \\ \uparrow_{\text{match}} \\ attacks(G,T) \wedge defends(T) \end{array}$$

- ***Wilma***: ‘But in Gordon’s case, his sister did not die so he would have less incentive to attack Tony.’ Wilma ***distinguishes*** by mentioning a ***missing assumption***, that the attacker’s sister died.

$$\begin{array}{cc} sister(O, L) \wedge dies(O) & attacks(L,H) \wedge defends(H) \\ \downarrow_{\text{match}} & \downarrow_{\text{match}} \\ sister(z, y) \wedge dies(z) & attacks(x,y) \wedge defends(y) \\ ??_{\text{no match}} & \uparrow_{\text{match}} \\ & attacks(G,T) \wedge defends(T) \end{array}$$

- ***Bert***: ‘In *Cavelleria Rustica*, no-one died but Alfio (A) still attacked Turrido (T) to protect his honour’ Bert argues that ***weakness not fatal***: and cites another precedent that also does not include the missing assumption (that the attacker’s sister died) but still matches the story scheme.

$$\begin{array}{c} attacks(A,T) \wedge defends(T) \\ \downarrow_{\text{match}} \\ attacks(x,y) \wedge defends(y) \\ \uparrow_{\text{match}} \\ attacks(G,T) \wedge defends(T) \end{array}$$

Note that the debate is not just there to satisfy curiosity. It matters legally which story is accepted. A fight mutually entered into (*West Side Story* and *Romeo and Juliet*) would be manslaughter, but a gangland killing would get a heavier sentence than a family feud in the current climate. Finally if we follow *Cavelleria Rustica* and the Laertes role of *Hamlet*, we can explain Tony’s role as self defence and he might even be acquitted.

6. Conclusions

In this paper, we have shown how reasoning with factual stories and story schemes can be modelled in the style of legal Case Based Reasoning models. It turns out that the precedent cases of CBR have a natural counterpart in factual reasoning: story schemes. The facts of the story can then be mapped to the elements of these story schemes (narrative units) in the same way as facts of a case can be mapped to the elements of cases (factors). This allows for the argumentative moves of CATO to be applied to factual stories, enabling moves like citation and distinction in discussions.

The link between CBR and stories allows for a more realistic way of discussing story coherence: precedent stories can be cited, obviating the need to explicitly model abstract story schemes. The argument moves then enable a natural dialogue concerning the facts of the story.

The current model thus specifies Walton’s (2010) Scheme for Argument from Analogy, which uses story schemes to determine the similarity between precedents cases/stories and the current case/story. This type of argument from Analogy is not just useful when talking about past events (as is the case in this paper), but also when trying to persuade someone to a particular course of action. We are more inclined to follow some course of action which has proven successful in the past. Thus, citing precedents in which success was achieved might convince someone to take the same course of action, provided their current situation is relevantly similar to the precedent.

Previous work in AI on general analogy (Gentner and Forbus 2011) captures the logic of analogy: it tells us what we require to state one and how to apply one. It does not, however, allow for argument moves *about* the analogical mappings. In contrast, the work in AI and Law (Ashley 1990, Aleven 1997) captures precisely these argument moves (e.g. analogizing and distinguishing) in a dynamic argumentation setting, whilst leaving precise mappings implicit. The framework for analogical case-based reasoning sketched in this paper therefore aims to capture both a precise matching and a possibility of argumentative discussion about this mapping.

The framework not only allows for the matching of factual stories to other factual stories (as in Schank 1986) or legal cases to other legal cases (as in HYPO and CATO), but also provides a way of matching factual stories to legal cases via legal rules (Bex and Verheij 2011). Furthermore, as discussed in (Bex 2011), the correspondences themselves, represented as legal or commonsense rules, can also be subject of argumentation. Thus, the framework is the first to capture all aspects of both factual and legal case-based reasoning in a single defeasible framework.

In this paper the theoretical foundations for factual precedent-based reasoning have been built. However, in order to make practical implementations feasible, a corpus of stories that can act as precedents is needed. These stories are ideally expressed in some common ontology to facilitate automatic processing. It is up to the young *Computational Narratives* field to tackle any problems concerning such a corpus and an ontology head-on, so as to provide an impetus not only to this research but to the entire field. Once a corpus is available, implementation of the CATO argumentation moves is relatively straightforward. In addition to the original version in (Aleven 1997), they have been implemented as a multi-agent system (Allen et al 2000) and using argumentation schemes (Bench-Capon 2012).

References

V. Aleven. *Teaching Case Based Argumentation Through an Example and Models*. PhD Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 1997.

M. Allen, T. Bench-Capon and G. Staniford. A Multi Agent Legal Argument Generator. In *Proceedings of the 11th International Workshop and Database and Expert System Applications*. IEEE Computer Society. (2000) 1080-86.

K. D. Ashley. *Modeling Legal Argument*. MIT Press, Cambridge, MA, 1990.

T. Bench-Capon. Representing *Popov v Hayashi* using Dimensions and Factors. Artificial Intelligence and Law 20:1. (2012) 15-36.

W.L. Bennett,. and Feldman, M.S. (1981) *Reconstructing Reality in the Courtroom: Justice and Judgment in American Culture*, Methuen–Tavistock, London.

F.J. Bex, *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*, Springer, Dordrecht, 2011.

F.J. Bex and T. Bench-Capon (2010) Persuasive Stories for Multi-Agent Argumentation. *Proceedings of the 2010 AAAI Fall Symposium on Computatonal Narratives. AAAI Technical Report FS-10-04*, AAAI Press, Menlo Park CA.

F.J. Bex, P.J. van Koppen, H. Prakken, B. Verheij, A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence. *Artificial Intelligence and Law* **18:2** (2010), 123–152.

F.J. Bex, B. Verheij. Legal Shifts in the Process of Proof. *Proceedings of the 13th ICAIL, Pittsburgh, USA* (2011), 11–20, ACM Press, New York.

D. Gentner and K.D. Forbus. Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2:3, (2011) 266-276.

P. McBurney and S. Parsons. Dialogue games for agent argumentation. In I. Rahwan and G. Simari (eds.): Argumentation in Artificial Intelligence. Springer, Berlin (2009), 261-280.

V. Propp, *The Morphology of the Folktale*, University of Texas Press, Austin (TX), 1968.

R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*, Lawrence Erlbaum, New Jersey, 1977.

R.C. Schank *Explanations Patterns: Understanding Mechanically and Creatively*, Lawrence Erlbaum, Hillsdale (New Jersey), 1986.

D. Walton, Similarity, precedent and argument from analogy, Artificial Intelligence and Law 18 (2010), 217-246.

A.Z. Wyner and T.J.M. Bench-Capon. Argument Schemes for Legal Case-based Reasoning. *Proceedings of the 20th Jurix, Leiden, the Netherlands* (2007), IOS Press, Amsterdam.

Integrating Argumentation, Narrative and Probability in Legal Evidence

Bart Verheij
Artificial Intelligence, University of Groningen
b.verheij@ai.rug.nl

Abstract

Reasoning on the basis of legal evidence has been analysed using three types of approaches: argumentative, narrative and probabilistic. As each type of approach has been defended as a complete account of evidential reasoning, it is natural to assume that there is an integrating perspective. It is here proposed that a logico-probabilistic argumentation theory can integrate argumentative, narrative and probabilistic approaches to legal evidence.

Keywords: legal evidence, argumentation, narrative, probability

1. Approaches to legal evidence

There exist three types of approaches to reasoning with legal evidence: argumentation approaches, narrative approaches and probabilistic approaches (Kaptein et al., 2009; Dawid et al., 2011). Of each type of approach there exist accounts that suggest a complete picture; nothing else seems to be needed. Argumentation approaches focus on arguments for and against what has happened in a criminal case, using reasons grounded in the available evidence. In narrative approaches, plausible stories are constructed as hypotheses about what has happened, and checked and compared on the basis of the evidence. In probabilistic approaches, numeric calculations aim at determining the probability that hypothesized events have happened given the evidence, and at updating probabilities in the light of new evidence.

As these types of approaches are superficially very different, but still have been defended as complete, the question arises whether there exists an overarching integrating perspective. I hold that such a perspective exists, and that a formalization can clarify the relations between argumentation, narrative and probabilistic approaches to reasoning with evidence. In section 2, the three perspectives are discussed in a way that is congruent with the integrating perspective sketched in section 3.

2. Argumentation, narrative and probability

Argumentative approaches (Anderson et al, 2005; Wignore, 1931) are strong in their analysis of complex structures of reasons pro and con. There can be arguments about the justificatory power of a reason (using Toulmin's pro-warrants or Pollock's con-undercutters; cf. Verheij, 2005). Figure 1 shows an argument that the suspect is punishable for a crime because of committing it, grounded in a witness testimony. The argument is attacked because the absence of the witness on the crime scene suggests that he is lying. Formal versions (Prakken 2004; Verheij, 2000) have been proposed, and have proven their analytic value. However, whereas Dung's seminal work (2005) provided a mathematically mature foundation of argumentation, the field now struggles with a confusingly large number of different semantics.

Narrative approaches (Wagenaar et al., 1993) take a synthetic perspective by focusing on the construction of hypothetical stories about what has happened (Figure 2). These stories are then compared in terms of their

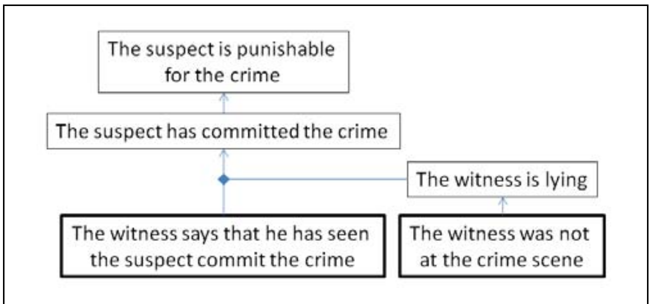


Figure 1: Argumentation

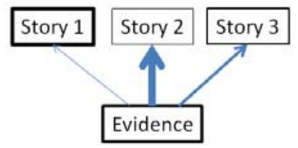


Figure 2: Narrative

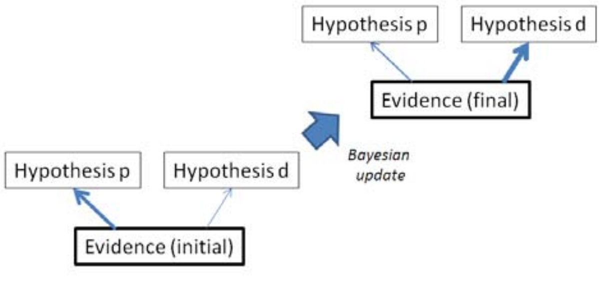


Figure 3: Probability

plausibility and matching with the evidence. In the figure, the different levels of plausibility and of matching have been indicated by lines and arrows of different width. Checking which elements of a story are supported and which not (evidential gaps; Bex, 2011) and determining the consequences of a story in order to for instance test an alibi (story consequences; Bex, 2011) are helpful investigative tools. The emphasis on the existence of different stories helps prevent tunnel vision. However, narrative approaches are more productive in the critical questioning of dubious cases (as in the work of Wagenaar and colleagues) than for decision making: how are plausibility and evidential matching to be determined, and how must they be compared, for instance when (as in

Figure 3) there is one story with high plausibility but a low match, and another with low plausibility and a high match? Also current formal grounding of narrative approaches is limited, but see Bex (2011) for a formalized hybrid argumentative-narrative approach.

In a probabilistic approach (Figure 3), numeric values are attached to the evidence and its support (expressed e.g. as a conditional probability) for the hypotheses proposed by the plaintiff (p) and defendant (d). Bayesian updating (using likelihood ratios) revises the evidentiary support values when new evidence is added. In the figure, the degree of evidentiary support is suggested by the width of the arrows; initially the evidence provides stronger support for the plaintiff's hypothesis, but finally the situation is reversed. An example of a probabilistic analysis is (Kadane & Schum, 1996). An advantage of probabilistic approaches is that they are well-founded in mathematical theory, but a limitation is that they assume more numbers than are available. It also happens that a probabilistic presentation of evidence leads to errors in court (Buchanan, 2007).

3. Integration by a logico-probabilistic argumentation theory

I claim that a new synthesis of logical and probabilistic techniques is needed, and that an argumentation perspective provides the key to such a synthesis. Therefore, I have initiated the development of a logico-probabilistic argumentation theory (LPAT), building on earlier work in mathematics and intelligent systems applied to legal decision-making. LPAT is a non-trivial synthesis of two seminal foundational theories, namely Kraus-Lehmann-Magidor preferential logic (1990) and Kolmogorov's classic probability axioms. In LPAT, qualitative, rule-based arguments have a quantitative interpretation. The numbers in such a quantitative interpretation can be objective (expressing frequencies) or subjective (expressing values and preferences). Argument strength is defined as a conditional probability. Stories about what can have happened become conclusions of arguments with the available evidence among their premises. So in LPAT there is no formal distinction between 'story conclusions' and other conclusions; stories only tend to consist of several elements. In LPAT, 'holistic' arguments from all premises to a final conclusion are formally connected to Wigmore-style 'analytic' arguments (1931), that consist of structured maps of premises, intermediate positions pro and con, and conclusions.

4. Conclusion

By the mixture of logical and probabilistic techniques, a logico-probabilistic argumentation theory has Bayesian Networks (Taroni et al., 2006; Hepler et al., 2007) as a central competitor. However, whereas Bayesian Networks are successful in data-modeling, additional tools (such as utilities) are required to model decision-making. LPAT addresses this issue by using logical techniques for decision-making and probabilistic techniques for data-modeling. In this way, LPAT may help alleviate the miscommunication between legal decision-makers and forensic data analysts that has leads to infamous judicial errors (Buchanan, 2007; Derksen & Meijsing, 2009).

After a period in which the causal metaphor associated with Bayesian Networks has had priority, it is time for a return to reasons, as formalized by a logico-probabilistic argumentation theory.

Acknowledgements

This position paper could not have been written without the cooperation and discussions with Floris Bex, Henry Prakken and Silja Renooij (cf. the NWO Forensic Science project; http://www.ai.rug.nl/~verheij/nwofs/).

References

Anderson, T.J., Schum, D.A. & Twining, W.L. (2005). *Analysis of Evidence*, second edition. Cambridge: Cambridge University Press.

Bex, F.J. (2011). *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Dordrecht: Springer.

Buchanan, M. (2007). Conviction by Numbers. *Nature* 445: 254-255.

Dawid, P., Twining, W. & Vasaliki, M. (eds.) (2011). *Evidence, Inference and Enquiry*. Oxford: Oxford University Press/British Academy.

Derksen, T. & Meijsing, M. (2009). The Fabrication of Facts: The Lure of the Incredible Coincidence. In H. Kaptein, H. Prakken & B. Verheij (eds.): *Legal Evidence and Proof: Statistics, Stories, Logic*, 39-70. Farnham: Ashgate.

Dung, P.M. (1995). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and n-Person Games. *Artificial Intelligence* 77: 321-357.

Hepler, A.B., Dawid, A.P. & Leucari, V. (2007). Object-Oriented Graphical Representations of Complex Patterns of Evidence. *Law, Probability & Risk* 6 (1-4): 275-293.

Kadane, J.B. & Schum, D.A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. New York, NY: Wiley.

Kaptein, H., Prakken, H. & Verheij, B. (eds.) (2009). *Legal Evidence and Proof: Statistics, Stories, Logic (Applied Legal Philosophy Series)*. Farnham: Ashgate.

Kraus, S., Lehmann, D. & Magidor, M. (1990). Nonmonotonic Reasoning, Preferential Models and Cumulative Logics. *Artificial Intelligence* 44: 167-207.

Prakken, H. (2004). Analysing Reasoning About Evidence with Formal Models of Argumentation. *Law, Probability & Risk* 3:33-50.

Taroni, F., Aitken, C., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*. Chichester: Wiley.

Verheij, B. (2000), Dialectical Argumentation as a Heuristic for Courtroom Decision-Making. In: *Rationality, Information and Progress in Law and Psychology*. Liber Amoricum Hans F. Crombag, 203-226. Maastricht: Metajuridica Publications.

Verheij, B. (2005). *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers.* The Hague: T.M.C. Asser Press.

Wagenaar, W.A., Van Koppen, P.J. & Crombag, H.F.M. (1993). *Anchored Narratives: The Psychology of Criminal Evidence*. New York, NY: St. Martin’s Press.

Wigmore, J.H. (1931). *The Principles of Judicial Proof or the Process of Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*, 2nd edition. Boston, MA: Little Brown.

Arguments as Narratives

Adam Wyner

Department of Computer Science, University of Liverpool
Ashton Building, Asthon Street, Liverpool, L69 3BX, United Kingdom
adam@wyner.info

Abstract

Aspects of narrative coherence are proposed as a means to investigate and identify arguments from text. Computational analysis of argumentation largely focuses on representations of arguments that are either abstract or are constructed from a logical (e.g. propositional or first order) knowledge base. Argumentation schemes have been advanced for stereotypical patterns of defeasible reasoning. While we have well-formedness conditions for arguments in a first order language, namely the patterns for inference, the conditions for argumentation schemes is an open question, and the identification of arguments 'in the wild' is problematic. We do not understand the 'source' of rules from which inference follows; formally, well-formed 'arguments' can be expressed even with random sentences; moreover, argument indicators are sparse, so cannot be relied upon to identify arguments. As automated extraction of arguments from text increasingly finds important applications, it is pressing to isolate and integrate indicators of argument. To specify argument well-formedness conditions and identify arguments from unstructured text, we suggest using aspects of narrative coherence.

1. Introduction

There are several lines of research on argumentation: argumentation schemes (Toulmin, 1958; Walton, 1996), abstract argumentation frameworks (Dung, 1995), and text analysis of arguments (Moens et al., 2007; Wyner et al., 2010). In this research context, it becomes increasingly important to identify the well-formedness conditions or properties of the instantiated argument patterns. For this purpose, we propose to apply aspects of *narrative coherence* to identify arguments from text; in other words, arguments are a *species* of narrative, and as such, we can not only consider corpora of argumentative texts as narrative, but also apply tools of narrative analysis to arguments.

In the following, we briefly outline the various strands of research on argumentation, propose a problem to be addressed, outline the means to begin to investigate this problem, and provide a use case with corpora.

2. Strands of Analysis

Argumentation schemes (AS) were developed in informal logic to represent a range of arguments found in ordinary conversation (Toulmin, 1958; Walton, 1996), where ASs are sterotypical (or normalised), defeasible patterns of reasoning from premises (and exceptions) to a claim. They emerged as part of the analysis of *fallacious* argumentation. For example, we have *Argument from Expert Opinion* where: *An individual is an expert in a domain, and the individual states that a proposition P is true, and P is a statement within the domain, therefore, P is true*; clearly, there are range of ways to critically examine the argument, so it is defeasible. ASs may be said to contrast with arguments that cannot be defeated: *Every man is mortal, and Socrates is a man, so therefore, Socrates is mortal.*

ASs have also been used in formal, computational approaches to argumentation (Bench-Capon and Prakken, 2010; Prakken, 2010; Atkinson et al., 2011): an AS is analysed in terms of its predicates and terms, a semantic model is given, contrasts between elements of the AS are interpreted as attack by other arguments, and the resulting set of arguments in their attack relations can be evaluated in an argumentation framework (Dung, 1995). A range of ASs have been proposed (Walton et al., 2008). Within this formal work, an important contrast remains between arguments made using Propositional and Predicate Logics and those made using ASs. The former are *strict* and can be abstractly stated *irrespective of the content of the propositions or predicates*. On the other hand, ASs are defeasible; it is not apparent that we can abstract from the content, particularly as the mode of critiquing the argument depends on the content in complex ways. For example, in the AS for *Practical Reasoning* about a course of action, whether one should or should not follow the proposed course of action rests on the possibility of *alternatives* to the given action and the consequences of those alternatives.

Another line of research investigates the discourse structure of arguments (Sporleder and Lascarides, 2006; Moens et al., 2007; Wyner et al., 2010), where argument indicators such as *supposing* or *therefore* are used to identify relevant textual passages that indicate elements of an argument. However, we know that textual identification and extraction of ASs is difficult, and there has been little reported success. In part, we claim, this is because the internal structure of ASs in textual terms is poorly understood. Beyond the identification of (sparse) argument indicators, what other features characterise an argument? In particular, what aspects of textual coherence and discourse structure apply (Webber et al., 2011)? In our view, it would be informative to approach the analysis of ASs in terms of narrative analysis since we can decompose the large and complex problem of AS identification into component issues that can be partly, but signficantly, addressed using current tools.

3. Problem Statement

Structural analysis of language has a long history and has applied virtually all aspects of language from phonology, morphology, syntax, semantics, and stories (Jakobson, 2002; Chomsky, 1965; Montague, 1974; Propp, 1928). A central issue is to account for *systematic linguistic phenomena* from the range of possibilities. For example, given a catalogue of lexical items, only some patterns appear as well-formed strings that represent noun phrases while others are unacceptable; the patterns go far beyond the normative, grammatical stipulations of grammar books. To account for such patterns, a common analytic strategy is deployed - we define a set of fundamental elements (or features in structured patterns) and conditions on their well-formed combination (as well as manipulations on the patterns). The conditions are induced from the data and may take highly abstract forms, e.g. binding constraints between pronouns and their antecedents. Such an account would take the form of a *grammar*, broadly conceived, which (ideally) covers all and only the well-formed strings of the corpus, or even better, accounts as well for strings not in the corpus, giving the account *predictive power*.

Turning to the topic at hand, we analogise the problem and analysis stated above to argumentation schemes, for which the analytic methodology has not yet been applied. To make the point concrete, we can create an *argument recognition task*. Suppose we sample 10 random paragraphs from different topics on Debatepedia, which is a wikipedia of debates that present both sides of an issue.[1] Given that each of these paragraphs is from an argumentative source, each of them present coherent arguments (to the author's best efforts); we call these the *argument paragraphs*. To this sample, we add 10 paragraphs, from 3 to 8 sentences long, where each sentence of each paragraph is selected randomly from other topics on Debatepedia; we call these the *non-argument paragraphs*. For example (not from this task), we routinely accept arguments of the following form (in the absence of other information):

> **An Argument**
> Suppose: Professor Hayes is an expert in Astrophysics; and
> Suppose: Professor Hayes states that the Andromeda galaxy is 2.7 million light-years from our galaxy; and
> Suppose: that Andromeda galaxy is 2.7 million light-years from our galaxy is an astrophyical statement; so
> Therefore, Andromeda galaxy is 2.7 million light-years from our galaxy.

However, the following is incomprehensible, even if it has a similar overall form of an argument.

> **A Non-argument**
> Suppose: six teenagers were arrested after a crime spree; and
> Suppose: it's traditional to have a Thanksgiving meal with a family; so,
> Therefore, earthquakes can be expected in San Francisco.

How do we explain the intuitive difference between the argument and the non-argument?

Thus, we have an analogy to the comparison between well-formed sentences and sentences constructed from random words. The question is: can we reliably, intuitively distinguish argument paragraphs from non-argument paragraphs in our corpus of 20 paragraphs? The extent to which we can suggests that we have some intuitive criteria by which we can 'recognise' an argument. We would then want to formalise and operationalised the analysis.

[1] Accessed March 30, 2012. http://debatepedia.idebate.org/

4. The Narrative Move

Our proposal is to apply discourse and narrative analytic concepts and tools to the analysis of ASs, trying to see what light such an approach sheds on the analysis of arguments. Among the questions to consider are:

- Is there a 'characteristic' temporal and aspectual structure, e.g. simple present tense in all sentences?
- Are arguments specified within ontological domains, and which classes and relations give rise to well-formed arguments?
- Do the individuals referred to in the argument bear a particular range of thematic or narrative roles?
- How do discourse indicators and propositional attitute verbs signal argument components?
- How do pronominal anaphor and ellipsis reinforce argument coherence?

Answering these questions, we can begin to formulate a 'data representation' suitable particularly for narrative coherence of arguments. By defining such a data structure, we would specify the linguistic elements that contribute to the construction of the argument, then use them in a semi-automated text analytic tool to identify arguments.

5. Towards a Tool

We propose to use the General Architecture for Text Engineering (GATE) tool for text analysis (Cunningham et al., 2002; Wyner et al., 2010; Wyner and Peters, 2010). Some of the textual elements above can be examined with existing GATE components (e.g. tense, discourse indicators, thematic roles, and pronominal anaphora) while others must be constructed (ontological structure and propositional attitudes). Our approach uses both manual and automated annotation interactively, where the tool automatically *highlights* relevant textual components that signal argument passages for manual annotation; identification of these passages with signalling indicators can then be used in the development of higher level argument identification. At the workshop, we intend to demonstrate the components of the tool and its application to interactively assist users in identifying argumentative passages in text.

6. Use Case

We propose the analysis of arguments in Debatepedia as our initial use case and corpus. The advantage of this data set is that it is already partially constructed, which can be useful in scoping the problems and identifying argument elements. We will apply manual and automatic annotation techniques to begin to address the questions raised above.

7. Acknowledgements

The author was supported by the FP7-ICT-2009-4 Programme, IMPACT Project, Grant Agreement Number 247228. The views expressed are the author.

8. References

Katie Atkinson, Trevor Bench-Capon, Dan Cartwright, and Adam Wyner. 2011. Semantic models for policy deliberation. In *Proceedings of ICAIL '11*, pages 81–90, Pittsburgh, PA, USA.

Trevor Bench-Capon and Henry Prakken. 2010. Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, 18(2):153–174.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press.

Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of ACL '02*, pages 168–175.

Phan Minh Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358.

Roman Jakobson. 2002. *Phonological Studies*. Mouton de Gruyter.

Marie-Francine Moens, Erik Boiy, Raquel Mochales-Palau, and Chris Reed. 2007. Automatic detection of arguments in legal texts. In *Proceedings of ICAIL '07*, pages 225–230, New York, NY, USA. ACM Press.

Richard Montague. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press.

Henry Prakken. 2010. An abstract framework for argumentation with structure arguments. *Argument and Computation*, 1(2):93–124.

Vladimir Propp. 1928. *Morphology of the Folktale*. University of Texas Press.

Caroline Sporleder and Alex Lascarides. 2006. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, 14(3):369–416.

Stephen Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Douglas Walton. 1996. *Argumentation Schemes for Presumptive Reasoning*. Erlbaum, Mahwah, N.J.

Bonnie Webber, Markus Egg, and Valia Kordoni. 2011. Discourse structure and language technology. *Natural Language Engineering*, December. Online first.

Adam Wyner and Wim Peters. 2010. Lexical semantics and expert legal knowledge towards the identification of legal case factors. In Radboud Winkels, editor, *Proceedings of JURIX '10)*, pages 127–136. IOS Press.

Adam Wyner, Raquel Mochales-Palau, Marie-Francine Moens, and David Milward. 2010. Approaches to text mining arguments from legal cases. In Enrico Francesconi, Simonetta Montemagni, Wim Peters, and Daniela Tiscornia, editors, *Proceedings of SPLeT '10, Lecture Notes in Computer Science*, pages 60–79. Springer.

Towards a Computational Model of Narrative Persuasion: A Broad Perspective

James Niehaus[1], Victoria Romero[1], Jonathan Pfautz[1], Scott Neal Reilly[1], Richard Gerrig[2], Peter Weyhrauch[1]

Charles River Analytics[1]
625 Mount Auburn St.
Cambridge MA, 02138

Department of Psychology[2]
Stony Brook University,
Stony Brook, NY 11794-2500

E-mail: jniehaus@cra.com, vromero@cra.com, jpfautz@cra.com,
snealreilly@cra.com, rgerrig@notes.cc.sunysb.edu, pweyhrauch@cra.com

This paper presents a preliminary view on the elements of persuasive narratives from a computational perspective. We argue for a broad perspective of narrative persuasion, drawing on existing literature from multiple disciplines. We present a brief, first-steps analysis of the possible narrative elements that may influence narrative persuasion. Finally, we consider how these elements may influence the formation of narrative corpora.

Keywords: narrative, persuasion, computational models of narratives

1. Introduction

This paper presents a preliminary view of the elements of persuasive narratives from a computational perspective. Our purpose is (1) to provide initial steps that may lead to the practical applications of computational models of narrative persuasion, including tools to construct, evaluate, and refine narrative arguments and (2) to recommend features for the construction of narrative corpora in support of this research.

Persuasive narratives have existed for thousands of years (e.g., Plato's *Republic* in 380 BC), and they exist in many forms, such as children's fables, ancient parables, religious texts, infomercials, war stories, political speeches and anecdotes. Despite their longevity and their prevalence in everyday life, little is known about which components, pieces, or elements of narrative influence their ability to persuade. Furthermore, this knowledge does not exist to understand narrative persuasion at a computational level, such that it may be understood or generated by a computer. Therefore, we present a brief, first-steps analysis of the possible narrative elements that may influence persuasion, drawing on existing literature from multiple disciplines. We also consider how these elements may influence the design and creation of narrative corpora to further narrative research.

For the purposes of this paper, we define persuasive narratives as the telling of temporally related information and events in an attempt to influence the emotions, attitudes, beliefs, or behaviors of the audience. The "telling" of these narratives may occur in many mediums and styles—such as novels, newspaper articles, films, puppet shows, comics, or speeches—and the "hearing" is the corresponding acts of observing or participating, such as conversing, reading, watching, or listening.

2. Elements of Narrative Persuasion

Narrative analysis in the humanities (narratology) has developed over more than 2,400 years since the foundational work of Aristotle's *Poetics*. The 20th century movement of narrative structuralism (e.g., (Barthes, 1966; Greimas, 1983; Labov & Waletzky, 1967; Propp, 1968)) codifies and extends observational wisdom about classical narrative forms found in narrative traditions, such as myths and folktales. More recent research includes semi-automated recognition of these types of narrative structures (Finlayson, 2009; Finlayson, 2011).

Narrative structures in this tradition include aspects of the plot, characters, settings, narrative time, narrative space, and motivations. The core elements of classical structuralist narrative theory capture insight from a variety of media, ranging from literature, myths, and folktales, to graphic novels and animation. However, these decompositions of narrative are not sufficient to capture the many elements of narrative that are both available and important to narrative persuasion. For example, a typical structuralist analysis of a narrative will not include demographics of the audience, which are vital to effective communication and persuasion. Therefore, we propose the classification and analysis of narratives by including less commonly considered features, such as the social setting of the audience (e.g., alone or in a group), the medium, the style (e.g., formal vs. informal language use), the narrative author, and the audience. An expanded narrative theory may be embodied in an ontology of narrative elements (where narrative is defined broadly) that captures these multiple facets of narrative. Table 1 displays some identified features of narratives that incorporate and extend beyond traditional structuralist views.

Abstract narrative theory alone may not be useful in many practical persuasion contexts without empirical evidence to show the effects of these narrative elements on audiences. A cohesive perspective describing the effects of narrative on the audiences' attitudes, emotions, and behaviors does not currently exist. There is extant research that specifically addresses the effects of narratives on audience's emotions, attitudes, beliefs, and behaviors, but this research is spread across disciplines (e.g., literature, linguistics, psychology, marketing, rhetoric, and media studies) with little communication between them. There is no theory or perspective uniting this work, neither are there common methodologies or approaches to research.

Some researchers are conducting empirical research on the effects of narratives on audiences, including psychologists (Appel & Richter, 2010; Gerrig, 1993;

Green, Strange, & Brock, 2002; Pennington & Hastie, 1992) and marketers (Adaval & Wyer, 1998; Dal Cin, Zanna, & Fong, 2004). Research in the psychology of narrative provides some insight into the empirical ramifications of narrative elements. Psychological studies have measured some aspects of how narratives affect the recall, comprehension, beliefs, attitudes, and affect of an audience (Gerrig, 1993). In particular, Gerrig (2011) provides evidence that a reader's preferences for the outcome of the plot influences both the reader's narrative experience and the resulting impact of the narrative on the reader. The sentences that prompt mental encoding and referencing of these preferences form "diversion points" in which the experience of readers diverges. If a model could predict with some accuracy the locations and ramifications of these diversion points based on known reader preferences, the model could provide some information on the narrative's persuasiveness with respect to an audience. As opposed to much of the low-level analysis of cognitive processes by psychologists, marketers such as Dal Cin, Zanna, and Fong (2004) often approach narrative studies from a higher conceptual level, invoking constructs such as narrative transportation and emotional investment, which, while valid and useful in many contexts, are less amenable to computational modeling because of the specificity needed by computation. Efforts to unify or correlate these and other diverse perspectives may provide deep insights into narrative persuasion.

Table 1: Possible feature categories that may influence narrative persuasion, with examples

Feature category	Feature Examples
Aristotelian elements	plot, spectacle, characters, and motivations
Jungian / Cambellian archetypes	the hero, the joker, the structure of the monomyth
Presentation	press release, oration, comic, radio message
Communication	told verbally by trusted individual, handed pamphlet by stranger
Audience	culture, demographic, and affiliation
Social situation	delivered in company of others, private communication
Genre	modern action film, Star Wars, Greek Mythology, Hadiths
Culturally specific preferences	sacrifices are honorable, *the joker* is received negatively
Impact	influences emotions, beliefs, and behavior

3. Narrative Persuasion Corpora

Corpora of narratives can be a valuable research tools to develop, empirically validate, and compare narrative theories and models. We propose two recommendations for narrative corpora to support computational studies in narrative persuasion. The first is to include a broad view of narrative that extends beyond internal narrative structures to the context of narrative, such as presentation, communication, audience, and social context in Table 1. The second is to include the representation of empirical data within the corpus. Researchers now use many techniques to study the effects of narrative from think aloud procedures, to recorded reading times, to neurological sensors such as fMRI or EEG. The ability to pair this empirical data with narrative structures and elements could be a valuable tool for exchanging research results and promoting active research.

4. Conclusions

While this paper presents a few possibilities and research directions to examine the elements of narrative persuasion from a computational perspective, we believe that this contribution only barely scratches the surface of a broad and deep area of research that is likely to continue for many years. To take early steps in this direction, we argue for an expanded theory of narrative persuasion that incorporates a diversity of features and empirical evidence as basis for computational models, and we argue that the construction of narrative corpora should include these considerations.

5. References

Adaval, R. & Wyer, R. S. (1998). The Role of Narratives in Consumer Information Processing. *Journal of Consumer Psychology,* 7(3), 207-245.

Appel, M. & Richter, T. (2010). Transportation and Need for Affect in Narrative Persuasion: A Mediated Moderation Model. *Media Psychology,* 13(2), 101-135.

Barthes, R. (1966). An Introduction to the Structural Analysis of Narrative. *New Literary History,* 6(2), 237-272.

Dal Cin, S., Zanna, M. P., & Fong, G. T. (2004). Narrative Persuasion and Overcoming Resistance. *Resistance and persuasion,*175-191.

Finlayson, M. A. (2009). Deriving Narrative Morphologies Via Analogical Story Merging. *New Frontiers in Analogy Research,*127-136.

Finlayson, M. A. (2011). *Learning Narrative Structure From Annotated Folktales.* MIT.

Gerrig, R. (2011). Individual Differences in Readers Narrative Experiences. *Scientific Study of Literature,* 1(1), 88-94.

Gerrig, R. J. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale Univ Pr.

Green, M., Strange, J., & Brock, T. J. (2002). Narrative Impact: Social and Cognitive Foundations. Taylor & Francis, Inc.

Greimas, A. J. (1983). *Structural Semantics: An Attempt at a Method*. Lincoln, Nebraska: University of Nebraska

Labov, W. & Waletzky, J. (1967). Narrative Analysis. In J. Helm (Ed.), *Essays on the Verbal and Visual Arts* (pp. 12-44). Seattle: University of Washington Press.

Pennington, N. & Hastie, R. (1992). Explaining the Evidence: Tests of the Story Model for Juror Decision Making. *Journal of Personality and Social Psychology,* 62(2), 189.

Propp, V. (1968). *Morphology of the Folktale*. Austin, TX: University of Texas Press.