7th Workshop on Computational Models of Narrative

CMN 2016, July 11–12, 2016, Kraków, Poland

Edited by

Ben Miller
Antonio Lieto
Rémi Ronfard
Stephen G. Ware
Mark A. Finlayson

OASICS

Editors

Ben Miller
Georgia State University
Department of English
Atlanta, Georgia, USA
miller@gsu.edu

Antonio Lieto
Department of Computer Science
University of Torino
Torino, Italy
lieto@di.unito.it

Rémi Ronfard
Centre de recherche INRIA
IMAGINE Laboratoire Jean Kuntzmann
Grenoble, France
remi.ronfard@inria.fr

Stephen G. Ware
Institute of Knowledge
Institute of Knowledge
College City
editorsb@cs.institute.org

Mark A. Finlayson
Florida International University
School of Computing and Information Sciences
Miami, FL, U.S.A.
markaf@fiu.edu

ACM Classification 1998
F.4.1 Knowledge Representation Formalisms and Methods, I.2: Artificial Intelligence, I.2.0 General/Cognitive simulation, I.2.7 Natural Language Processing, I.6 Simulation and Modeling, I.2.8 Problem Solving, Control Methods, and Search, I.2.11 Distributed Artificial Intelligence, J.4 Psychology, J.5 Literature

ISBN 978-3-95977-020-0

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-95977-020-0.

Publication date
October, 2016

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

License
This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0): http://creativecommons.org/licenses/by/3.0/legalcode.
In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:
- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/OASIcs.CMN.2016.0

ISBN 978-3-95977-020-0 **ISSN 1868-8969** **http://www.dagstuhl.de/oasics**

OASIcs – OpenAccess Series in Informatics

OASIcs aims at a suitable publication venue to publish peer-reviewed collections of papers emerging from a scientific event. OASIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Daniel Cremers (TU München, Germany)
- Barbara Hammer (Universität Bielefeld, Germany)
- Marc Langheinrich (Università della Svizzera Italiana – Lugano, Switzerland)
- Dorothea Wagner (*Editor-in-Chief*, Karlsruher Institut für Technologie, Germany)

ISSN 2190-6807

www.dagstuhl.de/oasics

Contents

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Preface

Welcome to the Seventh Workshop on Computational Models of Narrative. This year finds us colocated with the Digital Humanities 2016 conference (DH2016) for the first time in Kraków, Poland. This association was driven by a recognized need to bring together researchers in the humanities and in computer science toward common goals revolving around understanding and theorizing around narrative, and applying it to artifacts and topics of interest to the humanities.

The workshop's special focus is on how the computational modeling, analysis, and generation of narrative has affected approaches in the humanities for understanding narrative in or across textual, aural, or visual media. Themes represented at the workshop connected to the representation of narrative, connections between knowledge representation and narrative, the use of heuristics to handle complexity, incorporation of insights about human thinking, the use of narrative to organize information in the humanities, the relationship between top down and bottom up approaches for narrative understanding, or how narrative is seen to function differently depending upon the medium.

This year we received 16 submissions; of these two were declined. In keeping with our goal of inclusiveness, 14 papers were accepted, some on condition of revision. One paper was withdrawn during the revision period. The papers in this volume cover story planning, narrative structure, emotional and narrative annotation, story classifiers, machine learning of motifs, narrative queries, and narrative ontologies. This rich diversity of topics shared an overall concern with the necessity of developing computational and analytic frameworks for understanding and generating narrative.

The workshop program is complemented with a thought-provoking keynote talk by John Bateman, from the Institute for Transmedial Textuality Research at Bremen University, entitled "From Narrative to Visual Narrative to Audiovisual Narrative: the Multimodal Discourse Theory Connection". We are incredibly grateful to John for accepting the invitation and vigorously taking part in the many question and answer sessions that punctuated the workshop.

Last year's efforts to transition the organization of CMN to a committee continued, as the principal organizers for this year's workshop were Ben Miller, Rémi Ronfard, Antonio Lieto, and Stephen Ware. Mark Finlayson remained committed to the workshop series as the Series Chair.

Many thanks go to our collocated conference hosts, without whom this year's workshop would not have been possible. Our workshop venue was generously provided by Jagiellonian University, logistical support was provided by the organizers of the Digital Humanities 2016 conference and the Creative Media Industries Institute at Georgia State University, and supplemental funding was provided by the School of Computing and Information Sciences at Florida International University.

Ben Miller, Rémi Ronfard, and Mark Finlayson

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

List of Authors

George Ajíbádé
Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀, Ife, Nigeria
solajibade@yahoo.com

Adam A. Amos-Binks
North Carolina State University
Raleigh, NC, USA
aaamosbi@ncsu.edu

Valentina Bartalesi
ISTI-CNR
Pisa, Italy
valentina.bartalesi@isti.cnr.it

Andrea Bellandi
Istituto di Linguistica Computazionale "A. Zampolli" - CNR
Pisa, Italy
andrea.bellandi@ilc.cnr.it

Giulia Benotto
Istituto di Linguistica Computazionale "A. Zampolli" - CNR
Pisa, Italy
giulia.benotto@ilc.cnr.it

Corinne Brenner
Harmony Institute
New York, NY, USA
corinne.j.brenner@gmail.com

Joshua Eisenberg
Florida International University
Miami, FL, USA
jeise003@fiu.edu

Mark Finlayson
Florida International University
Miami, FL, USA
markaf@alum.mit.edu

Francesca Frontini
Istituto di Linguistica Computazionale "A. Zampolli" - CNR
Pisa, Italy
francesca.frontini@ilc.cnr.it

Catherine V Gale
Wingspan Productions
London, UK
cat@wingspanproductions.co.uk

Francesco Galofaro
Politecnico di Milano
Milan, Italy
galofaro_francesco@hotmail.com

Pablo Gervás
Universidad Complutense de Madrid
Madrid, Spain
pgervas@sip.ucm.es

Emiliano Giovannetti
Istituto di Linguistica Computazionale "A. Zampolli" - CNR
Pisa, Italy
emiliano.giovannetti@ilc.cnr.it

Raquel Hervás
Universidad Complutense de Madrid
Madrid, Spain
raquelhb@fdi.ucm.es

John Johnson
Harmony Institute
New York, NY, USA
jsj@harmony-institute.org

Fahad Khan
Istituto di Linguistica Computazionale "A. Zampolli" - CNR
Pisa, Italy
fahad.khan@unive.it

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson
OASICS Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Magdalena Maria Kubas
University of Warsaw
Warsaw, Poland
klara.klarnet@gmail.com

Carlos León
Universidad Complutense de Madrid
Madrid, Spain
cleon@ucm.es

Inderjeet Mani
Atelier Milford Paradise
inderjeet.mani@gmail.com

Carlo Meghini
ISTI-CNR
Pisa, Italy
carlo.meghini@isti.cnr.it

Daniele Metilli
ISTI-CNR
Pisa, Italy
daniele.metilli@isti.cnr.it

Philippe Mongin
ISTI-CNR
Jouy-en-Josas, France
mongin@greg-hec.com

Deborah Ninan
Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀,
Ife, Nigeria
jninan@oauife.edu.ng

Odetunji Ọdẹ́jọbí
Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀,
Ife, Nigeria
oodejobi@oauife.edu.ng

Marianne Reboul
Labex OBVIL, Sorbonne Paris-IV
Paris, France
marianne.reboul@free.fr

David Roberts
North Carolina State University
Raleigh, NC, USA
dlrober4@ncsu.edu

Jason Sherwin
Harmony Institute
New York, NY, USA
jason@harmony-institute.org

Anneke Sools
University of Twente
Enschede, Netherlands
a.m.sools@utwente.nl

Bernard Veldkamp
University of Twente
Enschede, Netherlands
b.p.veldkamp@utwente.nl

Sytske Wiegersma
University of Twente
Enschede, Netherlands
s.wiegersma@utwente.nl

W. Victor H. Yarlott
Florida International University
Miami, FL, USA
wvyar@cs.fiu.edu

R. Michael Young
North Carolina State University
Raleigh, NC, USA
young@csc.ncsu.edu

From Narrative to Visual Narrative to Audiovisual Narrative: the Multimodal Discourse Theory Connection

John A. Bateman

Bremen Institute of Transmedial Textuality Research, Bremen University, Bremen, Germany
`bateman@uni-bremen.de`

Abstract

Models of narrative have been proposed from many perspectives and most of these nowadays promote further the notion that narrative is a transmedial phenomenon: i.e., stories can be told making use of distinct and multiple forms of expressions. This raises a range of theoretical and practical questions, as well as rendering the task of providing computational models of narrative both more interesting and more challenging. Central to this endeavour are issues concerned with the potential mutual conditioning of narrative forms and the media employed. Methods are required for isolating narrative properties and mechanisms that may be generalised across media, while at the same time appropriately respecting differences in medial affordances. In this discussion paper I set out a corresponding approach to characterising narrative that draws on a fine-grained formal characterisation of multimodal discourse developed on the basis of both functional and formal linguistic models of discourse, generalised to the multimodal case. After briefly setting out the theoretical principles on which the account builds, I position narrative with respect to the framework and give an example of how audiovisual narratives such as film are accounted for. It will be suggested that a common anchoring in a well specified notion of discourse as an intrinsically multimodal phenomenon offers beneficial new angles on how narratives can be modelled, as well as establishing bridges between humanistic understandings of narrative and complementary computational accounts of narratives involving communicative goal-based planning.

1998 ACM Subject Classification I.2.7 Natural Language Processing

Keywords and phrases Narrative, computational modelling, discourse, multimodality

Digital Object Identifier 10.4230/OASIcs.CMN.2016.1

Category Invited Talk

1 Introduction

It has become relatively common to discuss models of narrative by drawing on various theoretical constructs and ideas proposed within the broad field of narratology. However, the narratologist David Herman [18, pp. 47–48] sets out what he describes as two 'great ironies' of narratological work on narrative as follows. First, although Barthes in his early considerations of accounts of narrative in the 1960s and 1970s had already called for a 'second linguistics' going beyond the limits of the sentence and structural linguistics at that time [4, p.83], narratology instead adopted many tenets of precisely that structural linguistics Barthes was urging moving beyond. And, second, despite the fact that many foundational works on narrative had begun by addressing oral storytelling, narratology has since been

© John A. Bateman;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 1; pp. 1:1–1:11
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

largely dominated by an orientation to written literary texts. This shaping of the field has, it will be suggested in this discussion paper, led to a systematic lack of attention to areas of textual description that are now required in order to track narratives as they move across 'media' and purposes – as in particular when we see narrative aspects being drawn on in the contexts of games, argument, biography and so on.

Less well considered is the fact that the last 20 years has seen within various branches of linguistics precisely that 'second linguistics' that Barthes was looking for. I will call this for current purposes 'discourse linguistics', although there are several contributing areas that use a range of nomenclatures spanning functional linguistics, formal linguistics, psycholinguistics, computational linguistics and more. The central unifying feature drawn on here is that we now have a rich foundation of discourse-related linguistic methods for dealing with connected texts 'beyond' the sentence. These have only been applied fragmentarily to the issues typically discussed in narratological approaches to narrative despite the possibilities suggested of beginning to anchor more abstract interpretative schemes in empirically-grounded linguistic research on texts and discourse. This suggests a sustainable point of possible interchange between more humanistic approaches to 'text' and more concrete linguistic-feature oriented approaches from linguistics.

A development of this kind gains considerable further motivation by the fact that both narratology and the computational modelling of narrative are seeking to move beyond verbal narratives. There is considerable discussion of what is labelled 'transmedial narrative' (e.g., [29, 30, 32]) on the one hand, while computational models and formalisms are being suggested not only for connected text but also for film and virtual cinematography, for combinations of image and text as in graphic novels, and in interactional contexts and interactive narrative. Whereas narratology still has rather limited experience in this direction, recent developments in the formal and functional characterisation of *multimodal discourse* can be directly applied. This is then the main focus of this position paper: showing how new accounts of multimodal discourse may serve as a powerful intermediate level of modelling that would connect abstract narrative concerns with concrete, and thereby modellable, features of narrative artefacts and performances.

The structure of the paper is consequently as follows. First I set out how a linguistically-inflected model of discourse can be applied multimodally. Second, I suggest how this may relate to narrative concerns. And finally, I set out some directions of current and future work that may take us further, including applications to interactive narrative and games.

2 From discourse linguistics to multimodal discourse analysis

The essential property of discourse (and texts) that required the move beyond sentence-oriented accounts is the very different mechanisms of meaning construction that become relevant. Whereas sentence meaning can largely be characterised in terms of compositional semantics operating within monotonic logics of various kinds, this does not work for texts. As long stated in text linguistics of almost all persuasions, a text is not a kind of 'super-sentence' but belongs to a different class of linguistic behaviour.

Consequences of this qualitative change can be seen even within sentences, however, as in the trivial examples:

1. She went to the park and played football.
2. She played football and went to the park.

Regular interpretations of these two sentences would suggest differing temporal sequencing of the events described, despite the fact that a logical conjunction makes no such commitments.

Moreover, the temporal interpretations are *defeasible*: that is, it may well be, in particular contexts of use, that no such temporal assertion was intended. Such variations in meanings are often shunted to areas of pragmatics and contextualisation, drawing on aspects of world knowledge and additional reasoning mechanisms.

'Classical' theories of text interpretation generally attempt to characterise this process of making meaning from texts as an inference process where particular propositions are derived from the logical propositional content of the utterances of the verbal text. These are then further developed by the application of cultural and background knowledge of various kinds. Problematic with this family of approaches is that it is generally difficult to characterise precisely *which* facets of background/cultural knowledge are required in any way apart from simply describing the text. Since as competent language users and text producers we generally know what texts mean, the 'analysis' comes to describe what we already knew the text means. This is not an adequate theory of meaning construction in texts for many reasons, and particularly not for any attempt to provide computational models.

Two objections that are particularly important are the following. First, pulling in information as required from world knowledge is a potential blackhole in terms of the extent of the reasoning that is required: this is not realistic in that hearers evidently come to interpretations very quickly and appear *not* to pull in arbitrary quantities of background knowledge – methods must be provided which indicate more specifically just what knowledge may be required and in what detail. Second, the interpretation process is in many respects backwards in that it assumes detailed world knowledge in order to understand the text. In fact, the text as text *asserts* world knowledge, i.e., it presupposes that certain relationships hold, regardless of whether the interpreter has the necessary knowledge already or not. Only by this means can text also function as constitutive of meanings and connections in a culture rather than simply echoing them.

Such implicit assertion of relationships occurs at various levels of generality and abstraction: commonly the implicit assertions of a text are sufficiently general, relating to issues of temporal order, causality, explanations, etc., that it is not necessary for an interpreter to proceed to further levels of testing *unless necessary for the specific communicative goals of the text*. This critical distinction between pursuing text understanding in terms of world knowledge and pursuing text understanding in terms of a text's implicit assertions is the basis underlying several current dynamic models of discourse. Within formal discourse semantics such *dynamic* properties of textual interpretation first received focused attention and formalisation with the development of inherently dynamic semantics, within which the semantics of an entire sequence of contributing blocks of information depends crucially on the ordering of that sequence (e.g., [20, 17]). This has been progressively extended since that time.

Currently the most prominent linguistic account of the dynamic properties of discourse semantics is that couched in Asher and Lascarides' Segmented Discourse Representation Theory (SDRT: [21, 2]). Similar emphases of the dynamic nature of discourse interpretation and production occur across most approaches to discourse, including both formal and functional linguistic schools (cf. [22]). The level of interpretation of discourse semantics that such models provide forms a mediating buffer between general world and cultural knowledge and the information that is directly required to interpret a text. Discourse interpretation then operates by constructing a semantic representation for each incoming discourse contribution, traditionally a sentence or utterance, which is then linked by means of discourse relations into a growing discourse structure. Discourse relations are defined so that both their applicability to particular semantic representations and the requirements they make of context are made explicit. They thus look both 'downwards' towards concrete linguistic forms (and their compositional semantics) and 'upwards' towards context. The requirements made of context define precisely the ways in which identifiable 'gaps' in interpretation are both created and resolved.

Important for us here is that it is possible to extend such a treatment of dynamic discourse semantics *multimodally* so that it can be applied to any similarly dynamic 'textual' artefact or performance, regardless of the modalities, or forms of expressive resources, that are employed – by these means we can draw connections across verbal texts and audiovisual discourses as realised in film as well as conjoined verbal-visual texts such as graphic novels and comics. In several papers, our group in Bremen has set out how such an extension of modelling capabilities can be applied to a diverse range of media (e.g., [5, 9, 11, 34]); a detailed introduction to the foundations of the approach is given in Bateman and Wildfeuer [10], while the particular characterisation of the *modes* of multimodality, and their relation to discourse semantics on the one hand and to more traditional notions of 'media' on the other, is given in Bateman [7, 8].

A brief example showing a discourse semantic view of what happens during the interpretation of a fragment of film will help clarify this position. It can quickly be established that viewers are making assumptions of meanings for such fragments over and above an appreciation of what happens in each shot and these assumptions show striking similarities to effects observed in verbal discourse interpretation requiring both dynamic interpretation and defeasible reasoning. A sequence decomposition of a fragment from Michelangelo Antonioni's *Blow Up* (1966) taken from towards the end of the film is shown in Table 1. During the entire fragment there is no dialogue or non-diegetic sound (simplifying our task); we only hear leaves being blown in the wind; frame numbers run from the beginning of the extract and are approximate. Although *Blow Up* is a classic film for the purposes of interpretation from many theoretical and analytic perspectives, including cultural and narrative concerns, our interpretation here will focus more precisely on how a well specified multimodal discourse analysis can pick out a film fragment's operation in terms of generating expectations. Such expectations are also driven partially by convention, which may then be intended in the film's composition in order to create particular opportunities for filmic meaning articulation.

Following the sequence through, we have the following interpretation; for readers unfamiliar with the fragment, the effect to be described can probably be gained (although significantly weakened) simply by examining the images presented in the table in sequence before reading the next paragraph.

First, we see the main character (simply named 'the Photographer') looking for something that he had formerly seen under a tree (shot 1). We focus in on him from above and behind (shot 2) as he continues to look on, now aware that what he is looking for will not be found. At the end of the shot, he looks up. We then see the branches of a tree, blowing in the wind (shot 3). At this point, viewers have a strong expectation concerning just what they are being shown: namely, what the Photographer is looking at, or seeing, when he looked up. This interpretation is now so standard and entrenched in our way of seeing film that viewers will make it without conscious deliberation – it is, in fact, for good perceptual reasons difficult *not* to make this interpretation. Shot 3 then continues with a pan downwards and to the right away from the branches towards the Photographer. When we reach the Photographer (frame 482), however, he is not looking up at the branches after all: he is looking slightly forwards and out of the frame towards the left. He then looks down and then back to the right, finally turning right (frame 721) and walking off in this direction, the camera tracking with him.

This sequence commonly evokes an impression of surprise in attentive audiences – although they may not be able immediately to articulate where this response originated. This

Table 1 Shot-by-shot breakdown of an example from Antonioni's *Blow Up* (1968). Shots are described in terms of the traditional distance continuum from long shot (LS), through medium long and medium close shots (MLS, MCS), down to head-and-shoulders shots (H&S).

Shot	Frame	Focus	Type	Description	Image
1 (6.66s)	0	P (David Hemmings)	MLS	crouching down, looking under a tree	
2 (6.08s)	159	P	MCS	looking around on the ground	
		P	MCS	looks up	
3 (3.54s)	306	tree branches	MLS	seen from below	
(3.79s)	391	branches	MLS	panning down-right ↘	
		branches background and P	MLS	panning down-right ↘	
(9.95s)	482	P	MCS	panning stops on Photographer	
	721	P	H&S	photographer turns and moves right, camera tracking with him	

demonstrates that more is occurring during the interpretation of the film than simply following what is happening. When the Photographer is seen looking to the left of the frame rather than up towards the tree in the middle of shot 3, the previous interpretation of the beginning of shot 3 as showing what he is looking at is thrown into doubt. Conventionally, a shot of branches from below following a shot of someone looking up will be interpreted as a point-of-view shot [12] even though the film never explicitly stated that this is what shot 3 was doing. What has happened is that the film's *structure* makes this claim. In other words, how the film has been structured leads viewers to the defeasible discourse hypothesis that the way that this shot is to be connected into the unfolding discourse is by means of a 'projection' [5], or mental perception, relationship.

This discourse interpretation then needs to be retracted when the pan down towards the right comes to an end showing that the Photographer is not looking at the tree at all. The situation here is even more complex, however, since we have additional cinematic conventions that come to bear. Actually there would be nothing 'in reality' preventing the beginning of shot 3 being related as a point-of-view shot and having the character look somewhere else during the pan downwards. *But this is not how point-of-view shots work.* An additional part of the convention is that during a point-of-view shot the character whose point-of-view is adopted cannot engage in independent action: they are constructed filmically as 'looking' and that is all. Antonioni plays with this convention in several of his films.

Slightly more formally, therefore, we have the following situation of interpretation. In the first two shots we have views of a single character that by virtue of filmic cohesion [33] of location, figure, posture, continuous sound and image properties (such as colour balance), contributes to a structure realising a narrative sequence with the shared topic 'the Photographer'. With the shift in shot 3, we hypothesise a perceptual relationship and so introduce an embedded topic structure subordinate to the point-of-view of the topic of the overall structure. All runs well with this hypothesised discourse structure until the character returns into shot: now we no longer have the option of maintaining an *embedded* topic structure because we appear to have rejoined the discourse structure that we started with. This raises a difficult task of reconciliation for the interpreter, which may well leave observable behavioural effects that can be subjected to empirical investigation. The only fully coherent single reading that combines all the possibilities is then that somehow the embedded topic structure came to an end without being cued explicitly in the filmic material – that is, the continuous pan does not provide any material support for this interpretation. To what extent viewers are willing to maintain interpretations for which there is no support will depend on several factors, including their attentiveness, familiarity with the medium, and willingness to overlook narrative gaps.

Taken together this demonstrates that the organisation of film is expectation generating, conventionalised, motivated (in that it commonly draws on perceptual routines) and violable. And all of these properties are, first, standard properties attributed to verbal discourse and the contributions made to such discourses and, second, well describable in terms of progressively applying defined discourse relations which constitute interpretative hypotheses concerning how the discourse is to be made sense of. In Bateman and Wildfeuer [10] we offer several examples of how discourse relations for various media can be defined; in Bateman [6] there is further discussion of how filmic discourse structures such as the example used here may be conventionalised into filmic idioms in similar ways to constructions in language.

3 From multimodal discourse to transmedial narrative

The discussion so far has opened up many points of contact with standard proposals within narratology. According to Ryan [30, p. 4], for example, narrative necessarily involves:

- construction of a storyworld, individuating agents, objects and their spatial arrangements,
- contingency, including accidents and the deliberate actions of agents,
- linkages between physical states and goals, emotions, intentions so as to produce coherence, motivation, closure and intelligibility.

Moreover, narrative is seen as a means of 'sense-making', of providing explanations. All of these facets were already identified in the brief description of filmic interpretation from a discourse semantic perspective in the previous section. And, indeed, many of the properties commonly associated with narrative can also be seen to be at work in other communicative situations or text types. Introducing topics, maintaining these and providing additional information to produce coherence are common features of most discourse types, not only narrative.

This opens up the main line of inquiry being suggested in this paper: to what extent might it be the case that we can offer more formalised views of some basic narratological constructs by modelling them in terms of discourse semantics? Thus, for example, we might see classic (although still hotly debated) categories such as 'focalisation' (cf. [14]) as *discourse achievements* brought about by hypothesising particular sequences of discourse relations.

This may bring several benefits for advancing the formal and computational modelling of narrative constructs. Consider again the example of the previous section and ask how it would be described in terms of focalisation. Although the relation between point-of-view and focalisation is itself subject to detailed theoretical discussions, we can for current purposes simplify somewhat and propose that we have at the beginning of shot 3 of the sequence a move towards 'internal focalisation', where our access to the storyworld is made relative to that of the Photographer character. This narratological status is then immediately deconstructed as the pan reveals the violation of the convention. *This lets us consider labels such as 'focalisation' in terms of the discourse hypotheses that are taken to hold moment-by-moment during the discourse interpretation of some sequence.*

Several authors in narratology and related studies have emphasised that it is possible to interpret almost anything as a narrative. The more relevant consideration is therefore how much evidence does an artefact or performance *itself* give that such an interpretation is relevant or intended. Wolf describes such features in terms of narrative cues [35]: the concern is then to what extent an artefact or performance provides *narrativisation cues* – the greater the number of cues, the more likely it is that that artefact/performance can be profitably interpreted as narrative. And, as Wolf emphasises, such cues can be found in many media, including paintings and other static depictive representations. By relating such narrative cues back to discourse semantics, we begin to have a link from rather abstract notions of focalisation, on the one hand, and more established mechanisms of discourse planning as commonly applied in the computational modelling of narrative in any case, on the other (e.g., [36, 3, 15]).

In short, we may conceive a formalised account of discourse semantics as an additional and complementary level of description that mediates between manipulable technical features of a medium and more abstract, narratively-centred interpretations and descriptions that would themselves be too far more removed from the details of expression to be drawn directly into planning or interpretation processes. Under this view, an appropriate planning of such narrative devices may be more readily achieved by positioning a level of discourse semantics between purpose and form – this would mean that increasing identification and empathy, increasing tension, achieving point-of-view, etc. all become *textual* goals to be achieved via discourse.

The planning process would not then need to relate high-level goals such as increasing tension or achieving a point-of-view shot to particular filmic expressive resources directly for planning purposes, but could work instead by means of posting intermediate discourse goals that have been established – primarily by empirical study – to have such desired effects as consequences. Such discourse goals may then range freely over all of the expressive resources that a medium provides (e.g., camera movements, lighting, dialogue, gesture and gaze, facial expression, temporal sequencing and so on), offering a far greater freedom of selection that is nevertheless related back to communicative goals.

4 Notes for the future

As a closing set of comments I will draw out some final considerations that could draw support from the line of inquiry suggested in the previous sections. These relate primarily to the drive to push accounts of narrative and its computational modelling beyond traditional narrative media such as verbal texts and narrative film and to take in the challenges of interactive narrative and computer-supported games. I will note two points of interest here related directly to the proposed role of discourse semantics.

First, the long discussion contrasting narrative-based interpretations of games and game-based interpretations of games (cf. [19, 27, 31, 1]) has shown that it appears untenable to 'reduce' games and game playing to narrative, although there do appear to be narrative elements at work. This demonstrates that it cannot be the case that a field such as 'narratology' provides a suitable foundation for both narrative and games. In contrast, we can well see both narrative and many game activities as involving the construction of coherent discourses, particularly when we adopt the position common to treatments of multimodal discourse that *actions* play a central role (e.g, [26, 13, 16]). Here, moreover, the ready relation of discourse back to interaction in language makes it even more relevant for interactive media such as gaming. This allows us to pursue accounts whereby narrative aspects and gaming actions do not stand in any conflict. Discourse modelling may show both being supported at the same time: some of the discourse relations and structures brought about may be contributing to sequences of actions or turn-taking and so on; others may be acting as narrativisation cues, supporting narrative interpretations of what is occurring.

Second, and related to the first, the idea of interactive narrative has itself run into a range of theoretical concerns that raise doubts whether the two poles – interaction, involving freedom of choice and action on the part of participants, and narrative, involving a generally authored narrative arc with plot points and resolutions – are compatible. If participants can 'do what they want' then there can be little guarantee that a satisfying narrative results (cf. [23, 28]). An anchoring in an approach enriched with more explicit treatments of discourse semantics can also be applied to this state of affairs in order to provide some conceptual clarifications with potential implications for implementation strategies.

Briefly sketched, models of dynamic discourse semantics show that it is not possible to contrast interactive narratives (and games) with more classical narratives on the basis that the former involves an active participant and the latter does not. Since discourse interpretation is *always* seen as the active formation of discourse hypotheses in context in order to provide interpretations, the interpreter is always highly active. It is then possible to consider both the source of interpretative cues and the various kinds of interpretations

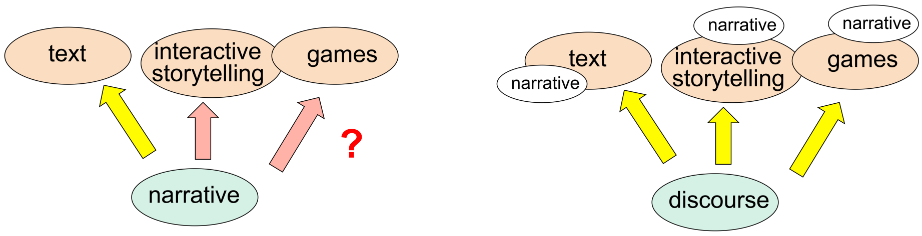


Figure 1 Using discourse semantics to offer a bridge across media.

that may be pursued more carefully. In particular, any contributions that are made on the part of a computational system supporting the interactive narrative can be seen in the light of the narrative cues that they offer. These cues need to be planned and reconciled with the events and states of knowledge that are unfolding within the narrative. They naturally provide a locus for inserting system-derived narrative arcs into the interaction without forcing participants either to notice them or to follow them: this then reflects a discourse that is making a narrativisation 'offer' to its participants – with respect to the management of these cues, it is the computational system that best fits the role of 'narrator'.

However, and *simultaneously*, each and every participant must also be seen as an active participant and a necessary component of that participation is the construction of discourse coherence for what is occurring. This discourse coherence can just as well include narrative interpretations that differ from those that may have been planned by the computational narrator. Cues for such narrative interpretations can (but do not have to) be included in the actions (construed broadly) of those participants. In this sense, there may be many narratives unfolding on the basis of the 'same' developing scenario – consolidation and synchronisation of these narratives can only then be an *interactive achievement* precisely as studied in multimodal extensions to conversation analysis (cf. [24, 25]). Again, what appears essential here is the recognition of a variety of discourse mechanisms and how these provide basic tools for 'getting the job of narration done' – even (or rather, especially) when there may be multiple simultaneous narratives being constructed interactively.

Taken together, the conception of discourse discussed here then might make it possible to bridge across different but arguably related domains, such as: narrative, interactive stories, games, since they can all (at least partially) be characterised as instances of discourse. This is suggested graphically in Figure 1. On the left-hand side we see depicted how many discussions in the field have proceeded, with narrative placed at a foundational level. In contrast, on the right-hand side we see the position advocated in this paper. In this view, it is discourse that provides a bridge across different communicative situations and media. Narrative is then just one of the kinds of the interpretations that the appropriate manipulation of discourse expectations and hypotheses can bring about.

Narrative is often considered from its own perspective – and this can no doubt be useful. But linguistically (and now multimodally), it is 'just one' among many forms that can be constructed via discourse. What remains to be done, therefore, is undertaking the considerable body of empirical work necessary in order to ascertain just how various media manage their expressive resources for the creation of such discourse.

References

1 Jonne Arjoranta. Narrative tools for games: focalization, granularity, and the mode of narration games. *Games and Culture*, pages 1–22, 2015.

2 Nicholas Asher and Alex Lascarides. *Logics of conversation*. Cambridge University Press, Cambridge, 2003.

3 Byung-Chull Bae, Yun-Gyung Cheong, and R. Michael Young. Toward a computational model of focalization in narrative. In *FDG'11: Proceedings of the 6th International Conference on Foundations of Digital Games*, pages 313–315, New York, NY, 2011. ACM.

4 Roland Barthes. Introduction to the structural analysis of narratives. In Stephen Heath, editor, *Image–Music–Text*, pages 79–124. Fontana, London, 1977 [1966]. originally published in French in *Communications*, 8.

5 John A. Bateman. Towards a *grande paradigmatique* of film: Christian Metz reloaded. *Semiotica*, 167(1/4):13–64, 2007.

6 John A. Bateman. Film and representation: making filmic meaning. In Wolfgang Wildgen and Barend van Heusden, editors, *Metarepresentation, Self-Organization and Art*, European Semiotics, pages 137–162. Lang, Bern, 2009.

7 John A. Bateman. The Decomposability of Semiotic Modes. In Kay L. O'Halloran and Bradley A. Smith, editors, *Multimodal Studies: Multiple Approaches and Domains*, Routledge Studies in Multimodality, pages 17–38. Routledge, London, 2011.

8 John A. Bateman. Methodological and theoretical issues for the empirical investigation of multimodality. In Nina-Maria Klug and Hartmut Stöckl, editors, *Sprache im multimodalen Kontext / Language and Multimodality*, number 7 in Handbooks of Linguistics and Communication Science (HSK), pages 36–74. de Gruyter Mouton, Berlin, 2016.

9 John A. Bateman and Karl-Heinrich Schmidt. *Multimodal Film Analysis: How Films Mean*. Routledge Studies in Multimodality. Routledge, London, 2012.

10 John A. Bateman and Janina Wildfeuer. A multimodal discourse theory of visual narrative. *Journal of Pragmatics*, 74:180–218, 2014. `doi:10.1016/j.pragma.2014.10.001`

11 John A. Bateman and Janina Wildfeuer. Defining units of analysis for the systematic analysis of comics: A discourse-based approach. *Studies in Comics*, 5(2):371–401, 2014. `doi:10.1386/stic.5.2.371_1`

12 Edward Branigan. Point of view in the fiction film. *Wide Angle*, 8(3):4–7, 1986.

13 Hans-Jürgen Bucher. Multimodales Verstehen oder Rezeption als Interaktion. Theoretische und empirische Grundlagen einer systematischen Analyse der Multimodalität. In Hans-Joachim Diekmannshenke, Michael Klemm, and Hartmut Stöckl, editors, *Bildlinguistik. Theorien – Methoden – Fallbeispiele*, pages 123–156. Erich Schmidt, Berlin, 2011.

14 Seymour Chatman. Characters and narrators: filter, center, slant and interest-focus. *Poetics Today*, 7(2):189–204, 1986.

15 Markus Eger, Cámille Barot, and R. Micheal Young. Impulse: a formal characterization of story. In Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard, editors, *Proceedings of 6th Workshop on Computational Models of Narrative (CMN 2015)*, pages 45–53, 2015.

16 James Paul Gee. *Unified discourse analysis: language, reality, virtual worlds, and video games*. Routledge, London and New York, 2015.

17 Irene Heim. File change semantics and the familiarity theory of definiteness. In C. Schwarze and A. von Stechow, editors, *Meaning, Use and Interpretation of Language*, pages 164–178. De Gruyter, Berlin, 1983.

18 David Herman. Toward a transmedial narratology. In Marie-Laure Ryan, editor, *Narrative across Media: The Languages of Storytelling*, pages 47–75. University of Nebraska Press, Lincoln, 2004.

19 Jesper Juul. A clash between game and narrative: a thesis on computer games and interactive fiction. Master's thesis, Institute of Nordic Language and Literature, University of Copenhagen, Copenhagen, February 2001[1999]. translated from Danish by the author. URL: `https://www.jesperjuul.net/thesis/`.

20 Hans Kamp. A Theory of Truth and Semantic Representation. In Jeroen A.G. Groenendijk, T.M.V. Janssen, and Martin B.J. Stokhof, editors, *Formal methods in the study of language. Part 1*, number 136 in Mathematical Centre Tracts, pages 277–322. Mathematisch Centrum Amsterdam, Amsterdam, 1981.

21 Alex Lascarides and N. Asher. Discourse relations and defeasible knowledge. In *Proceedings of the 29th. Annual Meeting of the Association for Computational Linguistics*, pages 55–63, Berkeley, California, 1991. Association for Computational Linguistics. URL: `http://acl.ldc.upenn.edu/P/P91/P91-1008.pdf`.

22 James R. Martin. *English text: systems and structure*. Benjamins, Amsterdam, 1992.

23 Michael Mateas. A preliminary poetics for interactive drama and games. In Noah Wardrip-Fruin and Pat Harrigan, editors, *FirstPerson: New Media as Story, Performance, and Game*, pages 19–34. MIT Press, Cambridge, MA, 2004.

24 Lorenza Mondada. The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65:137–156, 2014.

25 Lorenza Mondada. Multimodal resources and the organization of social interaction. In Andrea Rocci and Louis de Saussure, editors, *Verbal Communication*, pages 329–350. Mouton de Gruyter, 2016.

26 Sigrid Norris. Modal density and modal configurations: multimodal actions. In Carey Jewitt, editor, *The Routledge Handbook of multimodal analysis*, pages 78–90. Routledge, London, 2009.

27 Ken Perlin. Can there be a form between a game and a story? In Noah Wardrip-Fruin and Pat Harrigan, editors, *FirstPerson: New Media as Story, Performance, and Game*, pages 12–18. MIT Press, Cambridge, MA, 2004.

28 James Owen Ryan, Michael Mateas, and Noah Wardrip-Fruin. Open design challenges for interactive emergent narrative. In H. Schoena-Fog, L.E. Bruni, S. Louchart, and S. Baceviciute, editors, *Interactive Storytelling. Proceedings of the 8th International Conference on Interactive Digital Storytelling (ICIDS 2015)*, pages 14–26, Berlin, December 2015. Springer. URL: `https://games.soe.ucsc.edu/sites/default/files/ryanEtAl_OpenDesignChallengesForInteractiveEmergentNarrative.pdf`.

29 Marie-Laure Ryan, editor. *Narrative across Media: The Languages of Storytelling*. University of Nebraska Press, Lincoln, 2004.

30 Marie-Laure Ryan. On the Theoretical Foundations of Transmedial Narratology. In Jan Christoph Meister, Tom Kindt, Wilhelm Schermus, and Malte Stein, editors, *Narrative Beyond Literary Criticism*, pages 1–23. De Gruyter, Berlin, 2005.

31 Marie-Laure Ryan. From narrative games to playable stories towards a poetics of interactive narrative. *StoryWorlds: a journal of narrative studies*, 1:43–59, 2009.

32 Marie-Laure Ryan and Jan-Noël Thon, editors. *Storyworlds across Media. Toward a Media-Conscious Narratology*. University of Nebraska Press, Lincoln and London, 2014.

33 Chiao-I Tseng. *Cohesion in Film: Tracking Film Elements*. Palgrave Macmillan, Basingstoke, 2013.

34 Janina Wildfeuer. *Film Discourse Interpretation. Towards a New Paradigm for Multimodal Film Analysis*. Routledge Studies in Multimodality. Routledge, London, New York, 2014.

35 Werner Wolf. Narrative and narrativity: a narratological reconceptualization and its applicability to the Visual Arts. *Word & Image: A journal of verbal/visual enquiry*, 19(3):180–197, 2003.

36 R. Michael Young and Johanna D. Moore. DPOCL: a principled approach to discourse planning. In *Proceedings of the Seventh International Workshop on Natural Language Generation, Kennebunkport, Maine, USA, June 21-24, 1994*, Kennebunkport, Maine, USA, 1994.

Trailer Brain: Neural and Behavioral Analysis of Social Issue Documentary Viewing with Low-Density EEG*

Jason S. Sherwin[1], Corinne Brenner[2], and John S. Johnson[3]

1 **Harmony Institute, New York, NY, USA**
jason@harmony-institute.org
2 **Harmony Institute, New York, NY, USA**
3 **Harmony Institute, New York, NY, USA**

Abstract

The effects of social issue documentaries are diverse. In particular, monetary donations and advocacy on social media are behavioral effects with public consequences. Conversely, information-seeking about an issue is potentially done in private. We designed a combined free-viewing and rapid perceptual decision-making experiment to simulate a real scenario confronted by otherwise uninformed movie-viewers, i.e., to determine what degree of support they will lend to a film based on its trailer. For a cohort of subjects with active video-streaming (e.g., Netflix) and social media accounts (e.g., Facebook), we recorded electroencephalography (EEG) and behavioral responses to trailers of social issue documentaries. We examined EEG using reliable component analysis (RCA), finding reliability within subjects across multiple viewings and across subjects within a given viewing of the same trailer. We found this reliability both over EEG captured from whole-movie viewing, as well as over 5-second movie segments. Behavioral responses following trailer viewing were not consistent from first to second viewings. Rather, support choices both tended towards extremes of support/non-support and were made faster upon second viewing. We hypothesized a relationship between reliability behavioral metrics, finding credible evidence for it in this dataset. Finally, we found that we could suitably train a naive classifier to categorize production value and narrative voice ratings given to the viewed movies from RCA-based metrics alone. In sum, our results show that EEG components during free-viewing of social issue documentary trailers can provide a useful tool to investigate viewers' neural responses during viewing, when coupled with a post hoc behavioral decision-making paradigm. The possibility of this tool being used by producers and filmmakers is also discussed.

1998 ACM Subject Classification F.1.1 Models of Computation, I.6 Simulation and Modeling, J.3 Life and Medical Sciences

Keywords and phrases EEG, reliable components analysis, machine learning, documentary films

Digital Object Identifier 10.4230/OASIcs.CMN.2016.2

1 Introduction

Video media have the potential to manifest an almost infinite variety of effects on society, over a long timescale. This is because the conduits of such effects are human viewers, each bringing a different nervous system to bear on the sensory stimulus of a given video. Furthermore, both the sensorial complexity of visual-auditory stimuli and the culturally-laden content in

* This work was supported by the Rita Allen Foundation and the Harmony Institute.

© Jason Sherwin, Corinne Brenner, and John S. Johnson;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 2; pp. 2:1–2:21
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

such videos (e.g., visual and auditory semantics) make the connection between stimulus and nervous system response a challenging relationship to capture.

Social issue documentary trailers provide a useful stimulus set to study these effects because they are created to elicit action on the part of the viewer. Representative actions covering a range of economic, social and educational effects, can be quantified. Large user-bases of online video-streaming services (e.g., Netflix) allow instant access to more information on a given topic via viewing the full movie advertised in the trailer. Similarly, the large user-bases of social media networks (e.g., Facebook) allow support (e.g., a Facebook 'Like') of any shared content. Due to the currency market functionally being an even larger user-base than the previous two, and the frequency with which social issue documentaries are associated with requests for donations to specific actions or causes, monetary spending behavior becomes yet another measure to assess a trailer's effects. Consequently, the effects of social issue documentary trailers on a large population can be assessed on a subject group pulled from these overlapping user-bases.

Many studies have investigated potential relationships between behavioral and neural response to video media [9, 10, 12, 13]. For example, Kato et al. claimed changes in functional magnetic resonance imaging (fMRI) response of prefrontal and medial prefrontal cortices among subjects supporting a political candidate when viewing negative video advertisements about that candidate [18]. Aside from methodological concerns about this study (e.g., uncorrected p-values for fMRI blood oxygen-level dependent (BOLD) maps), this study considers how video media affects neural response when the decision to support a given candidate has already been made. In other words, there is no behavioral consequence of the video in this case. In another study of videos about political candidates, Zhang et al. found that neural response and computer vision analysis that revealed rhetorical gestures of the candidates were related [34]. However, as with Kato et al., this study recruited subjects on the basis of their prior support of the political party of either candidate (e.g., Democrat or Republic for the 2012 U.S. Presidential Campaign). Furthermore, there were no behavioral responses that followed movie-viewing and, even if there had been, such responses would be irrelevant because the election had already taken place.

The techniques of neuroimaging analysis for movie-viewing are also varied. Hasson et al. has shown that reliable component analysis (RCA) applied to fMRI is a viable tool for analyzing such data [12], finding relationships between RCA-based metrics and viewer/listener comprehension [11, 30], as well as time-resolved memory-encoding measured with both electrocorticography (ECoG) and fMRI [6, 10]. While some of these studies recorded behavioral responses (e.g., in the form of post hoc comprehension surveys [9, 13]), there was no element of support either for or against the viewed content embedded in these responses. In a variation of RCA that includes dimensionality reduction as a preliminary preprocessing step to recorded electroencephalography (EEG), Dmochowski et al. found that crafted video narratives (e.g., scenes from Hollywood feature films) elicited more reliable component activity in EEG than control videos [5]. In follow up work to this result, Dmochowski et al. found that such activity in a subject cohort (i.e., a small group) could be related to mass social media response for viewers (i.e., a large group) who naively watched the same video for the first time [3]. Both the studies of Hasson et al. and Dmochowski et al. occurred within precisely controlled and otherwise uncommon viewing environments (e.g., laying down in an fMRI machine or enclosed within a room shielded for radio frequency interference).

While there is no direct mapping yet between candidate narrative theories and neuroimaging analysis, RCA and related techniques provide a possible method for investigation. But RCA-based techniques frame the problem of narrative characteristics and content delivery on the sensor and/or voxel level, looking at the consistently reliable activity in that measurement. As Honey et al. and others have shown [14, 28], the flexible length of temporally receptive windows (TRWs) in the cortex mediates a hierarchical comprehension of complex temporal stimuli (e.g., movies, or video narratives). The extent to which these hierarchies of comprehension map to simple narrative content choices, such as narrative voice (e.g., participatory or expository) and production style (e.g., as manifested by high, medium or low budget productions), has yet to be uncovered.

Other work has shown that content characteristics manifest in audience comprehension of the viewed media and subsequent choices – explicitly or implicitly made – concerning that media. While RCA analyses have incorporated surveys of qualitative comprehension [12, 13, 10] and preference [3] post hoc, the impact of comprehension on decision time has not been investigated. Though largely ignored in previous work, the focus on decision time is important because it can be an indication of cognitive conflict when posed with alternative choices [19].

There has been some academic work on the topic of audience response to movie trailers, but not with the same rigor as seen for other video media. For example, Jerrick utilized a survey approach to gauge film trailer effectiveness in the U.S. college student market [17]. In another study, Findsterwalder et al. used post hoc interview transcripts as the basis for a qualitative analysis of movie trailer effectiveness in the New Zealand cinema market [8]. Neither of these studies examined audience response at the fine level of detail allowed by biometrics, such as EEG or fMRI. The Jerrick study did not analyze the effects of the movie trailer viewing on follow-up behavior with regards to having seen that video, while the Findsterwalder et al. study only did so through a qualitative analysis of interview transcripts.

In this paper, we describe how an experimental paradigm combining free-viewing of movie trailers and a follow-up decision-making task allows us to analyze the behavioral and neural responses elicited by an otherwise naive viewing. To do so in a more natural viewing environment, we use a wireless low-spatial density EEG outside of a radio-frequency-shielded environment to measure neural response during movie trailer viewing. By following each trailer viewing with an alternative forced choice (AFC) task whose consequences are linked to personal video-streaming and social media accounts, as well as a fixed monetary endowment, we gauge an immediate level of support for the viewed trailer that covers educational, social, and economic support categories of interest, respectively. We use proven techniques of RCA, as applied to EEG by Dmochowski et al. [5, 3, 17], to track reliable activity across both multiple viewings and/or subjects on both whole-movie and sub-movie (e.g., 1 second) time scales. Finally, we use regression and other machine learning techniques to link behavioral, neural and/or narrative features of the viewed movies.

2 Methods

2.1 Subjects

12 subjects were recruited for this study and data from 2 subjects were unusable due to EEG hardware malfunction. Of the 10 remaining subjects (5 male), the age range was 32.6 ± 2.0 years. Informed consent was obtained from each of them in accordance with the guidelines and approval of the IRB Solutions Institutional Review Board. Each subject was told they were to be paid $20 for their participation in this study. The requirements for participation were that each subject had at the time of the experiment an active movie-streaming account (e.g., Netflix) and a Facebook account to which he/she must log in at the experiment's start. Also, no subject should have viewed two or more of these movies in their entirety before recruitment to this experiment.

Table 1 Title, production value (PV) and narrative voice (NV) of all movies' trailers.

Movie No.	Title	Theor, PV	Alg. PV	Theor. NV	Alg. NV
1	A River Changes Course	high	mid-high	expository	expository
2	Cool It	mid	mid-high	participatory	participatory
3	Bag It	mid	mid-low	participatory	participatory
4	No Impact Man	low	low	participatory	part.-obs.
5	YERT	low	low	participatory	participatory
6	Carbon Nation	mid	mid-high	expository	expository
7	Fall and Winter	high	high	expository	expository
8	Fight for the Planet	low	low	expository	exp.-part.
9	Home	high	mid-high	expository	expository

2.2 Stimuli Overview

We presented movie trailers using Matlab and PsychToolbox [1] on a Dell Inspiron 15 (5000) 4th-Gen Core i7 laptop. We selected movie trailers from a database of over 435 films that constitutes the StoryPilot database [31]. The StoryPilot database catalogues social issue documentary films across a broad range of topics, noting such film characteristics as social issue topic, production value (PV), narrative voice (NV) and other metrics.

Researchers independently classified all films in this database for social issue content, PV and NV, among other metrics, based on predetermined criteria. NV values were based on a classification scheme developed by Nichols [24]. Of the entire database, inter-coder reliability was assessed for 79 films of the total sample coded by 2 or more researchers. Inter-coder reliability statistics for PV were 83.1% agreement, Cohen's kappa: 0.70; For NV, coding was resolved by discussion, so agreement was 100%.

From this sample, we selected nine (9) films on the social issue topic of the environment. This was done to control for any audience priority differences across different social issues. The other decision criteria for selecting these 9 films were to cover a range of low, mid, and high PVs, as well as both expository and participatory NVs, as logged in the StoryPilot database (i.e., the theoretic values for PV and NV in Table 1). The movies whose trailers were selected for viewing are listed in Table 1.

Table 1 also shows algorithmic values for PV and NV. These values were determined from re-running the inter-coder reliability analysis on the original coding values that had gone into previous discussion-based analysis. We ran this calculation to remove any potential subjective bias resulting from group discussion of the selected movies' NV and PV values. Across two coders for NV, we found Cohen's kappa of 0.63. For PV, we calculated Cohen's across three coders both with and without allowance for variation by a value of one (i.e., low to mid, or high to mid). Such an allowance is suitable for PV because it is a monotonic value scale, unlike NV. Without allowance, we found Cohen's kappa values were 1.0, -0.03, and -0.03 among the three unique coder pairings. With allowance, we found Cohen's kappa values of 1.0, 0.62, and 0.62. Therefore, inter-coder reliability on NV was quite high among our selected film trailers, while it was strong amongst PV, though there was some variability. Based on these inter-coder results, the nine selected films covered our criteria of mono-topical films that covered a broad range of NV and PV values.

After each viewing, each subject was presented with a prompt screen displaying 4 possible behavioral responses to the movie trailer.

Table 2 Behavioral response choices and linked icons.

Action	Linked Icon
Donate $1 to a cause related to the viewed trailer	
'Like' this movie on their Facebook page	
Add the full movie to their movie-streaming queue	NETFLIX
Indicate 'No Interest' in the trailer topic	N/A

2.3 Behavioral Paradigm

Subjects were shown a sequence of nine (9) unique movie trailers at an inter-stimulus interval (ISI) determined by the subjects' readiness. Subjects sat at a comfortable distance from the screen. The paradigm and data acquisition are illustrated in Figure 1.

In total, subjects viewed unique pseudo-randomized orderings of the trailer sequence twice. In particular, all trailer viewings were randomized, so that no subject would see the exact same ordering of movie trailers.

The prompt screen contained visual icons, each representing different possible ways of supporting or not supporting the previously viewed movie trailer (middle frame of Figure 1A). After each movie trailer, the subject was instructed to select from four possible actions in response to the previously viewed trailer as shown in Table 2.

Subjects logged into Facebook and movie-streaming accounts at the start of the experimental session to simulate actual consequences of support choices. Subjects were also alerted to this forthcoming choice, as well as to which keyboard buttons must be used to indicate his/her choice, at the start of the experimental session. Each subject was told to respond as quickly as possible. Choice-button relationships were pseudo-randomized for each choice screen to counter a possible habituation effect in choice selection.

2.4 Data Acquisition and Preprocessing

EEG data was acquired without electrostatic shielding using an Advanced Brain Monitoring X10 9-channel system (Carlsbad, California) with scalp electrodes arranged in the 10-20 System. Data was sampled at 256 Hz. A software-based 0.5 Hz high pass filter was used to remove DC drifts and a 30Hz low pass filter was used to isolate relevant EEG power bands. These filters were linear-phase to minimize delay distortions. Stimulus events, i.e., movie start, movie frame flips, movie end, choice screen and keyboard responses, were recorded on separate channels.

Two preprocessing steps were employed on the EEG data. First, each channel was z-scored. Second, to ensure that all subjects were on the same timeline for each movie, the movie frame

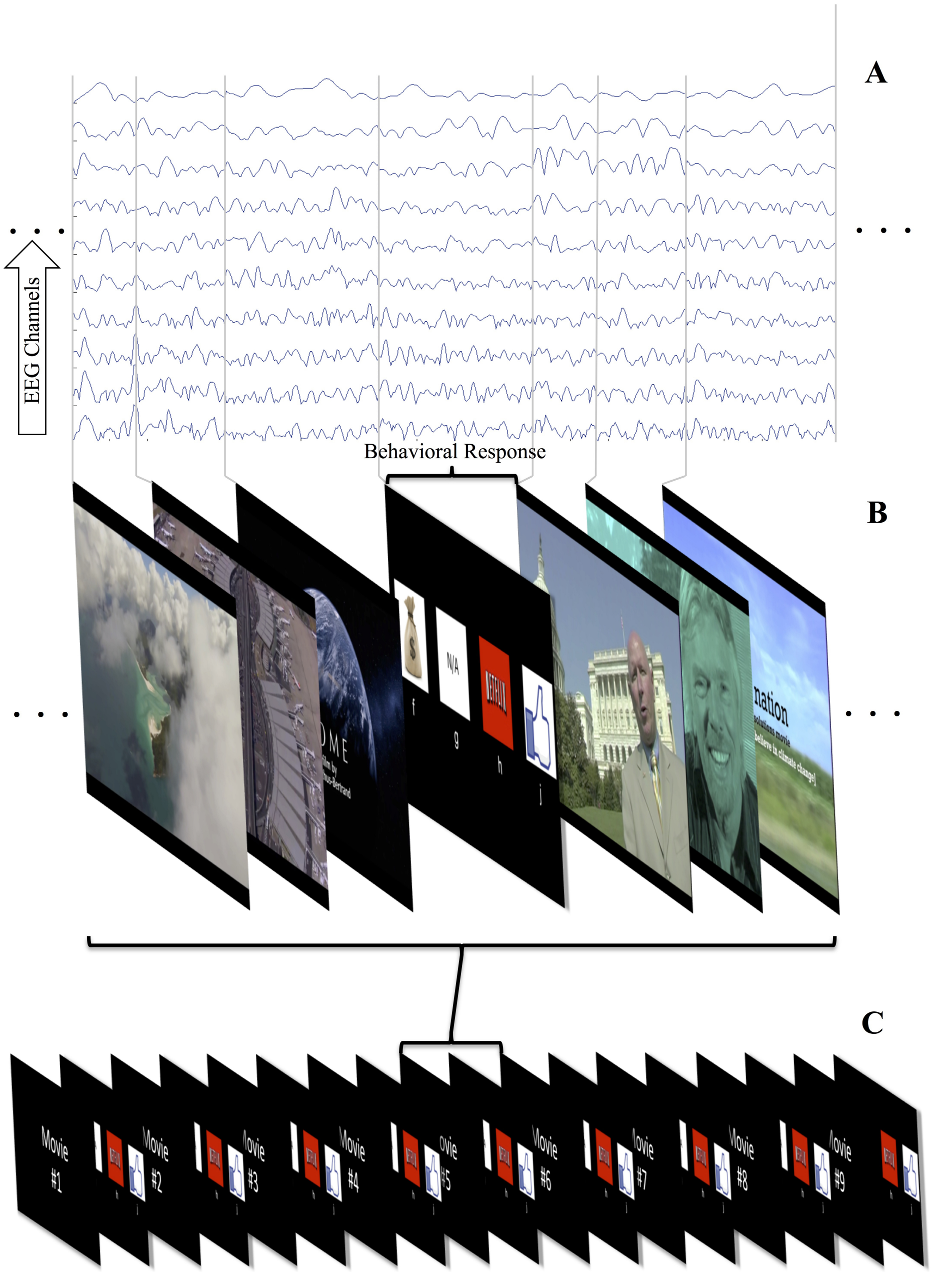


Figure 1 Paradigm overview. Frames of chosen movie trailers are shown (B) with events marked across all EEG channels (A). Horizontal ellipses before and after movie frames indicate previous and subsequent movie frames (B) across EEG time series (A). Each row of signal represents a different EEG channel from the n = 9 channels used here. Following each movie viewing, there was a prompt message for the subject to get ready and proceed (not shown) to the choice screen (middle frame of B). Response time was recorded from presentation of the choice screen as indicated. The subsequent movie then began playing. A total of nine (9) movies, each with prompt/response screens comprised one full block (C). Two blocks were run per subject in succession. Movie sequences were pseudo-random and so was choice key matching at each response screen (e.g., 'add to Netflix queue' could be in either of the four positions shown after each viewing).

flip events were used as a marker between each of which the EEG data was time-averaged. This resulted in one EEG measurement per frame flip. This method of time-locking EEG to movie stimuli counteracts slight potential variability in EEG hardware sampling rates that can cause substantial drift as recording enters multiple minutes. Also, the variability in the refresh rate of any PC video card is mitigated with this technique. We considered the use of Independent Components Analysis (via the 'fast ica' algorithm) to remove eye-blink and eye-movement artifacts, but saw no noticeable differences on the EEG data when this technique was inserted before z-scoring.

2.5 Behavioral Analysis

Choice values and response times (RTs) from prompt were analyzed for each movie. As a starting hypothesis, choice values were assigned numerical equivalents as listed previously to gauge choice variability upon repeat viewing. Additional analysis employed a coding scheme, whereby choice-values of 1 and 4, i.e., 'Donate $1' and indicate 'No Interest', were grouped together, while choice-values of 2 and 3, i.e., ' 'Like' this movie' and 'Add the full movie to queue', were also grouped together. Both this binary coding scheme and response times were analyzed with standard t-tests to gauge differences from first- to second-viewing.

2.6 EEG Analysis Overview

Our primary method for analyzing EEG was reliable component analysis (RCA) [5, 3, 4]. RCA is a technique used to analyze neural data acquired during presentation of continuous stimuli, for which there is no canonical locking event (e.g., as in standard P300 or oddball EEG paradigms). In general, RCA requires at least two neural data signals $X_1 \in \mathbb{R}^{DxT}$ and $X_2 \in \mathbb{R}^{DxT}$, where D is the number of channels and T is the number of time samples. RCA seeks a weight vector ($\mathbf{w}$) such that the resulting linear projections $\mathbf{y}_1 = X_1^T\mathbf{w}$ and $\mathbf{y}_2 = X_2^T\mathbf{w}$ have maximal correlation. Following the technique of Dmochowski et al. [4], we selected the weight vectors that yielded the three highest correlations after linear projection. These projections are the three most reliable components, henceforth to be known as 'the reliable components'. We applied this technique to different subsets of the overall neural data, both for whole movie-viewing and sub-clips.

2.7 EEG Analysis: RCA with Whole-Movie Viewing

To examine common neural activity between repeat viewings, we assumed X_1 and X_2 to be first- and second-viewings of the same movie trailer. In this manner, we calculated the reliable components for each subject viewing each movie trailer. We then validated the significance of these components by bootstrapping. In particular, we time-shuffled one of the signals before recalculating the reliable components, keeping the correlation value from each shuffle. We performed N = 10,000 such calculations, for each subject and each movie, so that the true correlation value determined from RCA could be compared to the distribution of values obtained from the time-shuffle-based calculation. A reliable component was deemed significant if the correlation value it yielded between X_1 and X_2 exceeded p = 0.05 (two-tailed, p = 0.025). Both negative and positive correlations are possible.

To examine common neural activity across subjects within the same viewing, a similar calculation was made, except all subjects' data signals were concatenated such that each subject's data was paired with every other subject's data for a given viewing (see Table 3). Color-coding in the table indicates telescoped blocks by which all subject combinations were formed to make X_1^{all} and X_2^{all}.

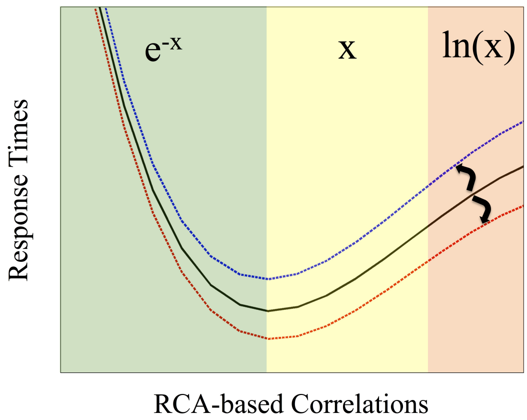


Figure 2 Hypothesized relationship between RCA-based correlations and response times. Three portions of variability of response times are shown in green, yellow and orange shading. The exponential decay portion (green) represents high response times when correlations are low, while the logarithmic portion (orange) represents mid-range response times when correlations are high. Lowest response times are hypothesized for the mid-range correlations connected to the extremes by a linear relationship (yellow) between dependent and independent variables. Variation of the constants multiplied by these three portions will produce variability of the curve as exhibited by the red and blue dashed curves around the central black curve.

Table 3 Illustration of concatenation for all-subject RCA.

$X_1^{all} \equiv$	1	1	…	1	2	2	…	2	…	8	8	9
$X_2^{all} \equiv$	2	3	…	10	3	4	…	10	…	9	10	10

The reliable components from the newly formed X_1^{all} and X_2^{all} were then calculated. As for individual subjects, the bootstrapping provided a threshold to evaluate each reliable component's significance.

2.8 EEG Analysis: RCA with Time-Resolved Viewing

To examine common activity between multiple viewings in a time-resolved manner, RCA was also applied to sub-clips of all trailers. In particular, sub-clips of X_1 (first viewing) and X_2 (second viewing) were taken from 150-frame ($\approx$5s at screen refresh rate) time intervals and subject to RCA calculation, including bootstrapping for significance. This 150-frame time interval began at trailer onset and then was moved forward in time by 30 frames ($\approx$1s) to do another RCA calculation with bootstrapping. This procedure was repeated until the whole of X_1 and X_2 were covered. A correction by false discovery rate (FDR) was applied to each interval's p-value, due to the repetition of this procedure for the number of overlapping 150-frame windows in each movie trailer. The result of this calculation is a time-resolved map of reliable components, each with a significance value against a computed null distribution, for each movie trailer and each subject.

2.9 Relating EEG to Behavioral Responses

We examined a possible link between EEG and behavioral metrics. In particular, we tested the hypothesis that response times vs. RCA-based correlations followed a trend illustrated in Figure 2.

We hypothesized this relationship as a corollary to previous work linking higher correlation values to greater comprehension of a viewed movie [10, 13]. Particularly, in the context of a decision to choose among support levels for a viewed documentary trailer, we reasoned that the corollary of this result is that extremely low and high correlation values would lead to delayed reaction times, while mid-range correlation values would lead to the quickest reaction times. Because correlation values have been linked monotonically to movie comprehension, we reasoned that the low correlations would produce generally higher RTs than high correlations, thus the shape of the hypothesized relationship in Figure 2.

2.10 Classifying Narrative Voice and Production Value from EEG

We also examined a link between the EEG data and the narrative features of each movie trailer. In particular, we trained a decision tree classifier [29] on the results of the RCA from whole-movie viewing. We used the results of whole-movie viewing because these metrics were calculated from both all time points of the trailers and all subjects who viewed the movie trailers. This approach allowed us to sample as many time points and viewers as possible to judge narrative content of a given movie.

3 Results

3.1 Behavioral Choices More Extreme and Faster Upon Second-Viewing

An overview of choice values for each subject and movie is shown in Figure 3. Not only does Figure 3 show a broad range of choice-values across most subjects and movies within each viewing (Figure 3C and Figure 3D), but it also shows the variability in the magnitude by which choice values fluctuate between first and second viewings (Figure 3A). Noting changed and unchanged choice-values with a binary coding scheme, we see in Figure 3B that the majority of behavioral responses were different from first to second viewing. The number of choice-value differences was different across movies (t-test, $p < 0.01$) and subjects (t-test, $p < 0.01$).

Not only did we find that choice-values changed from first to second viewing, but they also followed the trend shown in Figure 4A. In particular, choice-values on the whole went from mid-level support (e.g., 2's and 3's) to extreme-levels of support (e.g., 1's and 4's). We quantified this by grouping mid-level and extreme-level choice-values. We compared the number of mid- and extreme-level choices within each viewing, finding significant differences for each (first-viewing: t-test, $p < 0.01$; second-viewing: t-test, $p < 0.02$). We also compared the number of mid- and extreme-level choices across viewings, finding significant differences for each (mid-level: t-test, p « 0.01; extreme-level: t-test, p « 0.01).

We also examined response times once subjects were prompted for their choice-values. Figure 4B shows an overview of mean and standard error response times (RTs) for each movie. As the figure shows, first-viewing RTs were significantly greater than those of the second-viewing. We quantified this by grouping all RTs within each viewing and then comparing to those of the second viewing (first-viewing > second-viewing: t-test, $p < 0.02$).

3.2 EEG Analysis: Whole-Movie Viewing Shows Certain Films Having High RCA Correlation

We performed whole-movie RCA for each movie trailer and subject. Figure 5 shows correlation magnitudes' mean and standard errors across subjects within each movie.

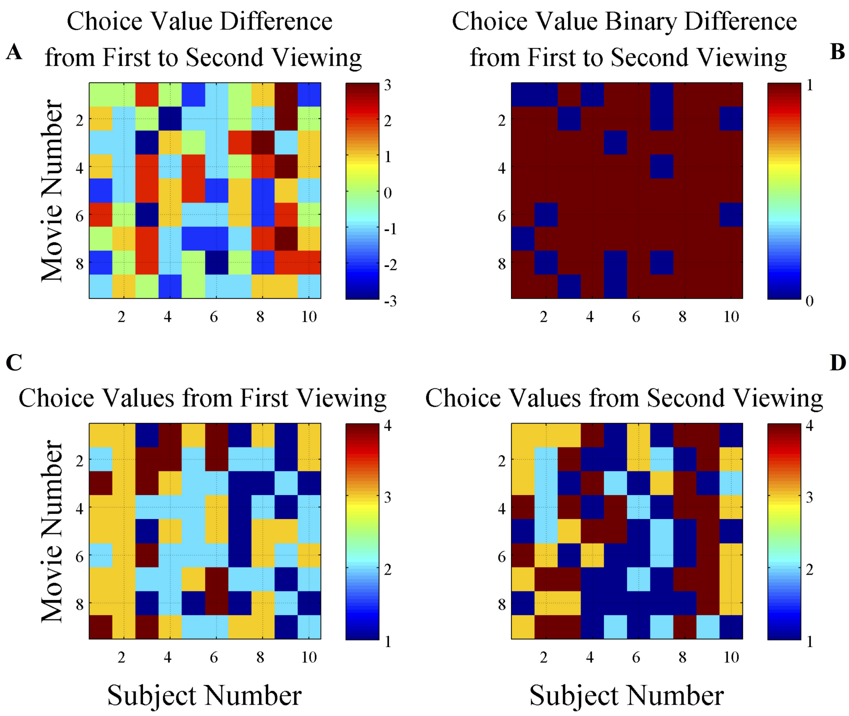


Figure 3 Overview of choice values by movie and subject number. The difference in choice value from first to second viewing (A) is shown with its own color scale covering [-3,3]. For instance, a choice value of 1 in the first viewing and a choice value of 3 in the second viewing would be a difference in choice value of 2 (e.g., Movie Number = 1, Subject Number = 3). The binary version of this plot is shown next to it (B) with its own color scale, indicating 1 when the choice value did not change and 0 when it did. The bottom row of matrices shows the actual choice values from first viewing (C) and second viewing, each with the same color scale covering [1,4]. Movie numbers are taken from Table 1.

From Figure 5, we see variability in each of the components' mean correlation magnitudes, while standard errors across movies within each component are consistent. Whole-movie RCA within each viewing also showed variability by movie trailer. Figure 6 shows an overview of significant (green) and insignificant (red) component/viewing pairings.

From Figure 6, we also see the variability of significant components for each movie. One movie ("Bag It") maintained significant correlations across subjects within both viewings for all components, while four movies had only one component insignificantly correlated across both viewing. All other movies had at least one component showing insignificant correlation in at least two component-viewing pairs.

3.3 EEG Analysis: RCA with Time-Resolved Viewing Shows Frame-by-Frame Audience Impact

We also tracked RCA correlation values at a time-resolution of 1s (Figure 7). Figure 7 shows the time course of significant windows for three components of both positive and negative correlations for a given movie and subject.

From Figure 7, we see that positive correlations (blue) dominate, but that negative ones (red) exist as well. We summarized these results across all subjects and movies by calculating the movie-length-normalized number of significantly correlated windows (Figure 8). Figure 8

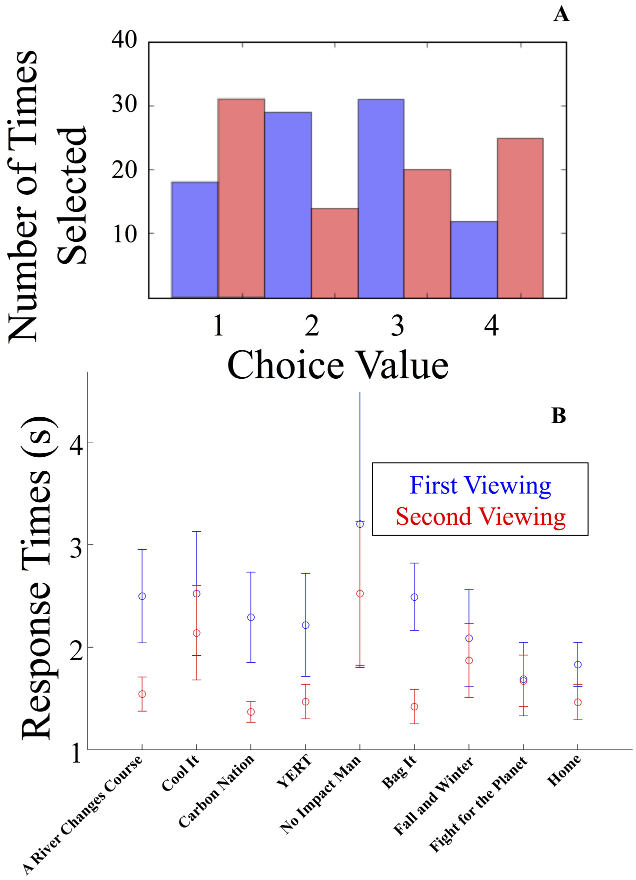


Figure 4 Overview of choice values across all movies and response times by movie. Viewing number is color-coded in the inset. Breakdown of choice values across all movies for first and second viewings is shown with the same color-coding (A). Mean response times and standard errors (B) are shown for first and second viewings. Error bars indicate standard error around the mean response time for the indicated movie. Across all movies, response times dropped from first to second viewing (t-test, $p < 0.02$).

shows mean and standard error number of significantly correlated windows for each movie, where the bar is color-coded to the movie trailer name. Positive and negative correlations are shown.

From Figure 8, we see that positive correlations are more frequent than the negative ones. But non-zero negative correlations were seen across all components, most notably from the first and third components, so they have been reported here, despite earlier work ignoring such results [5, 3].

3.4 RCA of EEG Relates to Behavioral Responses

We examined the relationship between EEG metrics obtained with RCA and behavioral responses (Figure 9). In particular, we tested the hypothesis that RCA correlation magnitudes would produce a relationship against response times (RTs) of the form, $y{=}Ae^{-x}{+}B\ln(x){+}Cx{+}D$. First and second-viewings are temporally dependent events, so we examined each scenario alone, also considering the difference in RTs from first- to second-viewing.

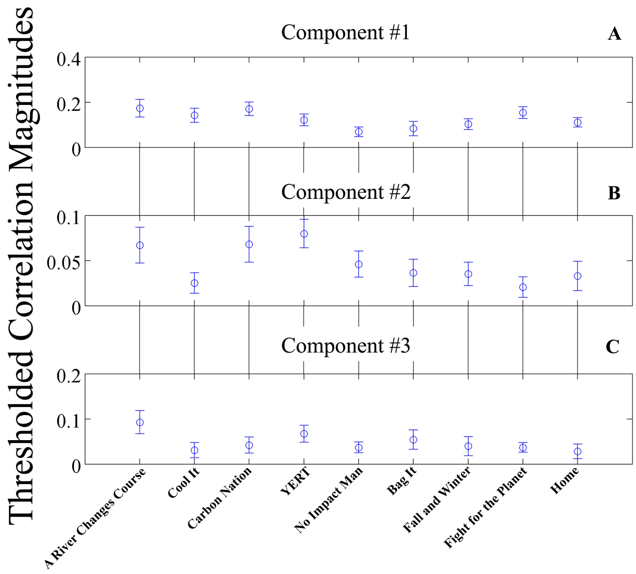


Figure 5 Correlation magnitudes across subjects for each movie. First (A), second (B) and third (C) components' mean correlation magnitudes are shown with error bars indicating standard error. Movie labels for the bottom plot (C) apply to corresponding plots above (A and B). Correlation values that were insignificant for a given subject-movie pair were zeroed out. Since magnitudes of significant correlation values are used, both positive and negative correlations from first- to second-viewing are represented here.

Figure 9A shows variation of first-viewing mean RTs with summed significant correlation magnitudes. High RTs are found for low correlation magnitudes. Low RTs are found for mid-range correlation magnitudes. And mid-range RTs are found for high correlation magnitudes. A similar trend is seen in Figure 9C, but here the mean difference in RTs is plotted. Also, differences in RTs are seen for both low and high correlation magnitudes. Both Figure 9A and Figure 9C show significant fits for the hypothesized variation of the behavioral with the neural metric. Figure 9B shows a trend in which second-viewing RTs fall off nearly monotonically with correlation magnitudes. At this sample size though, the fit is not significant.

3.5 Narrative Features of Trailers Classified from EEG

We also examined the relationship between EEG metrics obtained with RCA and the narrative features of the movie trailers, such as Production Value (PV) and Narrative Voice (NV), as shown in Table 1. As labels, we used the clear categories defined by theoretic NVs and PVs. We did this in order to have clear training labels that did not straddle multiple categories. The RCA values we used made a 6-component feature vector for each movie, in which there were two sets of 3-component correlation values, covering first and second viewings.

Using a decision tree classifier [29], we found mean classification scores across leave-one-out modeling as shown in Table 4. For the PV model, the error was 0.33 ± 0.56 (chance error = 0.66). For the NV model, the error was 0.11 ± 0.04 (chance error = 0.50). Generally, we found that whole-movie RCA-based correlation values (in fact, a sum over all components' correlation values) above a certain threshold was characteristic of Participatory film classifications, while those below were characteristic of Expository film classifications (see Figure 6 for visual representation of this, too).

As Table 4 shows when compared to Table 1, the NV classifier is able to classify nearly all movies correctly, making only one mistake ("Fight for the Planet" incorrectly classified

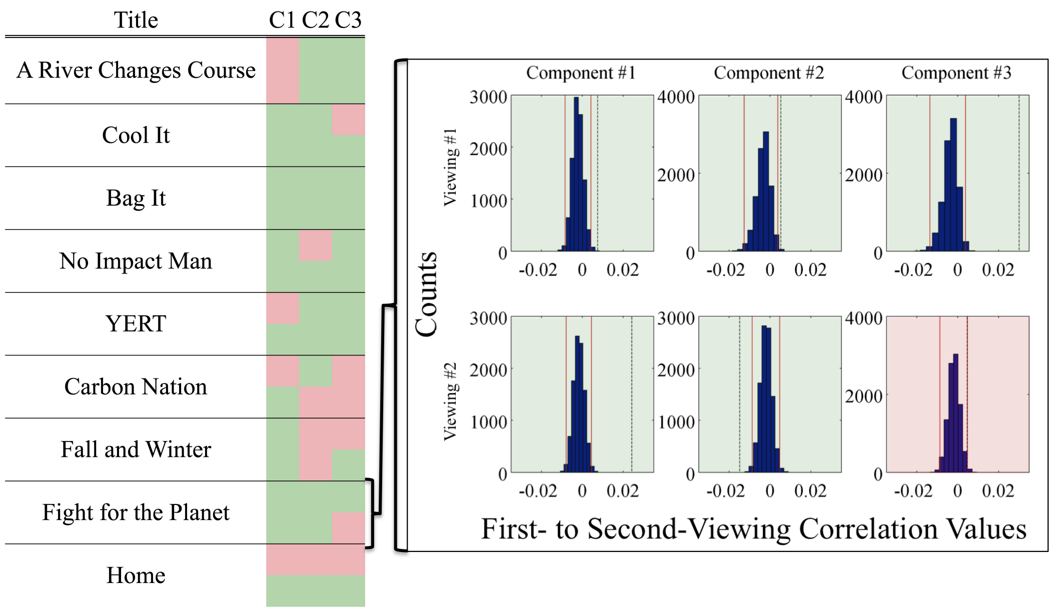


Figure 6 Overview of all movie trailers' whole-movie correlations across all subjects. Inset shows correlation values of each component and viewing against null distributions for all subjects' EEG on indicated movie trailer (e.g., Fight for the Planet). The top row indicates correlation values across subjects for the first viewing, while the bottom indicates those of the second viewing. Each column represents the indicated component number. The x-axes of each plot indicates correlation values and the y-axes indicate counts obtained from permutation testing. Blue bars indicate histograms from permutations (N = 10,000) and vertical red lines on either side of the distributions indicate p = 0.025 thresholds (two-tailed). Vertical black dashed lines indicate actual correlation values. Green and red shading of each plot indicate whether the actual correlation values are significant (green) or non-significant (red). The summary table shows significant (green) and insignificant (red) component/viewing combinations for each movie, with C1, C2 and C3 representing the first, second and third components.

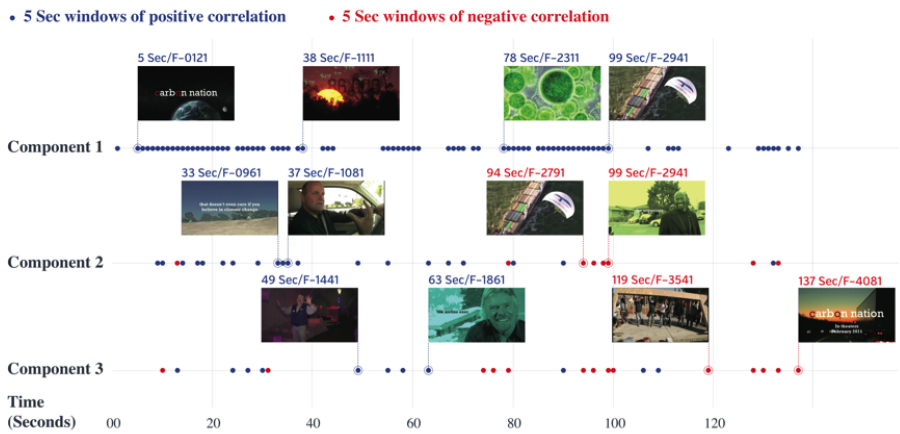


Figure 7 Example of time-resolved RCA for given subject watching a movie trailer (e.g., Carbon Nation). Each row shows 5-sec windows that are positively (blue) and negatively (red) significant from first- to second-viewing. Each indicated frame shows time in seconds (Sec) and frame number (F).

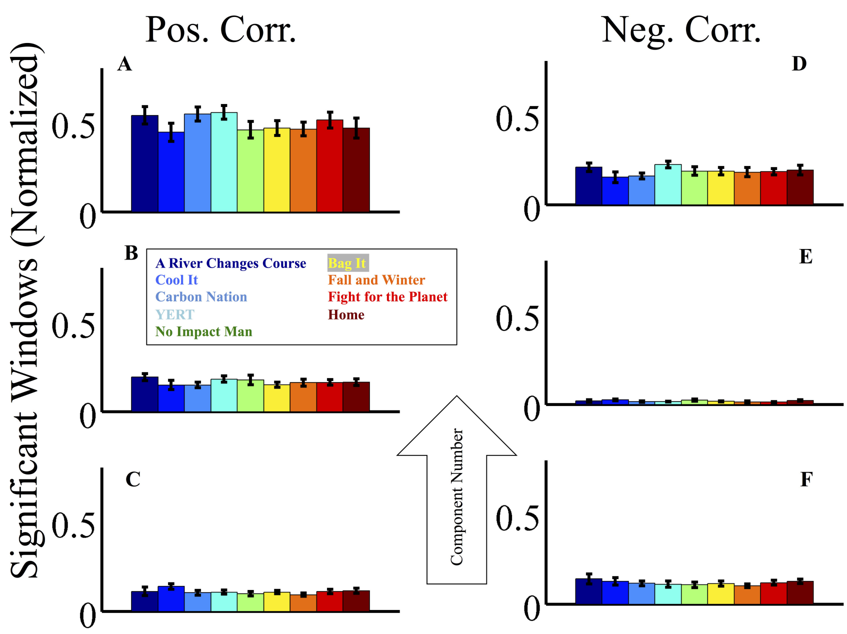


Figure 8 Number of significantly correlated 5-second windows normalized by movie length for each movie. The height of each bar indicates the mean number of windows and error bars indicate the standard error over all subjects. Each movie is color-coded according to the legend in the inset. Positive correlations of the first (A), second (B), and third (C) components are shown beside negative correlations of the first (D), second (E) and third (F) components from subjects' EEG.

more often as Participatory than Expository). When we examine this movie more closely in Table 1, we see from the algorithmic determination of NV that there was some coder disagreement as to the Narrative Voice. Although not an equal split in disagreement in the classifier score, the score for Participatory NV is not absolute (i.e., 1.00, as it is for other film trailers).

Also, comparing Table 1 and Table 4 for PV, we find significantly above chance performance for the classifier (chance accuracy = 0.33). Examining the errant classifications in Table 4, we find that the decision tree scores reflect the variability seen in the inter-coder reliability of PV, and the summarized algorithmic PV values of Table 1. For instance, as an example of a correct classification, "A River Changes Course" was classified with highest score to be a high PV (0.74), though algorithmic determination of PV is mid-high. The classification score for mid PV was the second highest for this film trailer (0.25). Alternatively, as an example of incorrect classification, "Bag It" was classified as low PV (0.90), though algorithmic determination of PV is mid-low. The classification score for mid PV was the next highest classification score (0.09). Similar trends were seen with other films.

4 Discussion

In this study, we identified neural correlates and behavioral metrics that differentiate viewer response amid a subject cohort that first watches a social issue documentary trailer and then is tasked with choosing a level of support/interest for that video's topic. This task simulates an important aspect of video media viewing, whereby consumers are able to instantly exercise behavioral response to such content via social media, video-streaming or economic means. Differences in neural and/or behavioral activity across movies that displayed a range of

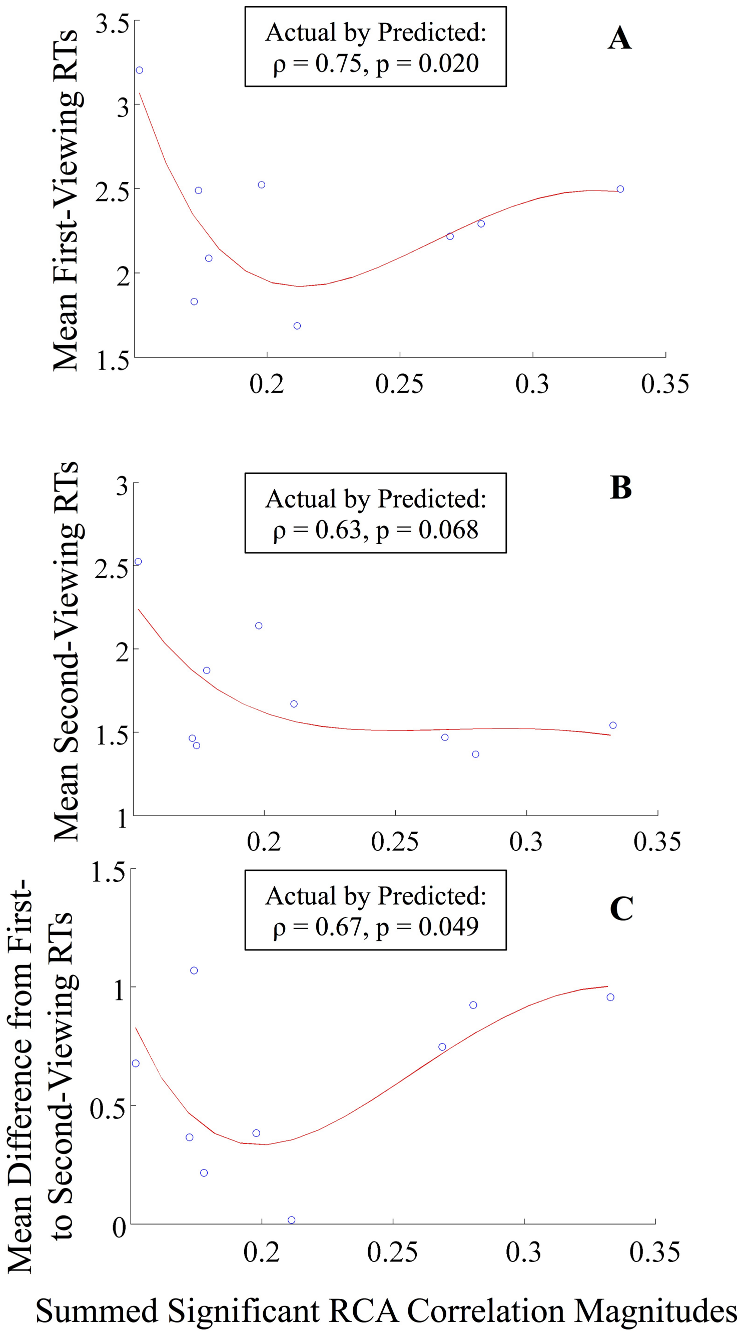


Figure 9 Variation of response times (RTs) with RCA correlation magnitudes. Blue dots represent each movie. Red lines indicate curve-fits for each plot (A-C). All curve-fits are of the form $y{=}Ae^{-x}{+}B\ln(x){+}Cx{+}D$. Each curve-fit is shown with actual vs. predicted correlation values and p-values. Variation of first-viewing RTs (A) was found to be significantly fit (ρ = 0.75, p = 0.020). Variation of second-viewing RTs (B) showed a trend, though non-significant at our p = 0.05 threshold (ρ = 0.63, p = 0.068). Variation of the mean difference from first- to second-viewing RTs (C) was found to be significantly fit (ρ = 0.67, p = 0.049).

Table 4 Decision tree classifier outputs for Production Value (PV) and Narrative Voice (NV), using RCA whole-movie viewing metric inputs. Winning scores are highlighted in **bold**.

	Decision tree PV scores			Decision tree NV scores	
Title	Low	Mid	High	Expository	Participatory
A River Changes Course	0.01	0.25	**0.74**	**1.00**	0.00
Cool It	**0.97**	0.02	0.02	0.29	**0.71**
Bag It	**0.90**	0.09	0.01	0.15	**0.85**
No Impact Man	**0.71**	0.28	0.02	0.33	**0.67**
YERT	**0.72**	0.26	0.02	0.34	**0.66**
Carbon Nation	0.01	0.00	**0.99**	**1.00**	0.00
Fall and Winter	0.01	0.42	**0.57**	**1.00**	0.00
Fight for the Planet	**0.67**	0.31	0.02	0.02	**0.98**
Home	0.01	0.45	**0.54**	**1.00**	0.00

production values and narrative styles manifested themselves in a combination of temporally precise and whole-movie neural components, and choice value selection and response times. Furthermore, we showed that the neural components from viewing could predict above chance production value (PV) ratings and narrative voice (NV). Below, we discuss these results within the context of previous approaches to capturing neural and behavioral response to video media, especially audience reaction to movie trailers. Finally, we conclude by proposing that the system developed for this study is a viable platform to analyze production choices in movie trailer content creation.

4.1 Faster and More Extreme Behavioral Choices Relate to Social Issue Behaviors

Our behavioral observations demonstrate the tremendous variability in support choice a viewer has for a given film trailer. But we found that these choices shift towards extremes of support/non-support in the aggregate (via more donations and disinterested support choices on second-viewing) and that they are made faster upon second viewing. The increase in decision speed fits within expectations established by brand recognition literature. For instance, MacDonald and Sharp [22] replicated a classic study by Hoyer and Brown [15] in which subjects' product choices were gauged by response time for products having a brand name either previously known or unknown to the subject. They found that response times to the known brand were significantly quicker than those to the unknown, inferring that more decision effort is exercised in the case of unknown brands' product characteristics, thus causing longer decision times. Applying this model to each film trailer, the support choice decision upon second viewFigure 4ing is made faster concerning its brand (Figure 4B), due to the first viewing having occurred already, i.e., each subject has already obtained greater familiarity with the movie brand as communicated by its trailer.

A possibly unintended correlate of this faster response time is that, on a population level, the support level is pushed to the extreme. The shift in choices from Facebook 'Like' and movie-streaming 'Add to Queue' to 'Donate' and 'Not Interested' (Figure 4A) supports our hypothesis that these choices can be arranged in the cardinal order shown in Table 2. Previous work on the relationship between social media activism and economic activism supports this cardinality. For instance, Qualman coins a term socialnomics to describe the phenomenon by which social media movements are transformed into movements with economic and other consequences, rather than the reverse [26]. The relationship between 'Add to Queue' and 'Not Interested' can be understood in the context of brand recognition effects on response times. Decisions of 'Not Interested' have a lower time cost than 'Add to Queue' decisions, since the latter will require further time investment to see the movie in its entirety. Finally, the relationship between a social action (e.g., Facebook 'Like') vs. a secretive one (e.g., 'Add to Queue') can be understood in the context of secret ballot behavior in political elections. Nichter considers the influence of economic incentives in voting behavior when a ballot is cast in secret, thus ensuring a barrier between the entity purchasing the vote and the one casting it [25]. A starting point for this analysis is that voting behavior differs by whether that vote is done in private or public. For instance, in the case of a Facebook 'Like', a vote is publicly cast in support of a given movie trailer or the movie itself. Conversely, there is no public involvement in the choice to 'Add to Queue' because the decision is only recorded in a given movie-streaming account. Using the paradigm and system described in this study, the consequences of public versus private actions could be studied in further detail, though more precise manipulation of public vs. private design variables would be needed.

4.2 RCA Metrics Relate to Narrative Features Previously Thought To Be Qualitative

Our whole-movie trailer and time-resolved RCA results demonstrate that the experimental paradigm designed for this study reveals consistent neural response with a subject cohort that is relatively small in size when compared to other forms of audience testing. While we found variability in mean component correlations between first and second viewing across movie trailers, there was no significant difference in standard errors across the subject cohort (Figure 5). We found similar variability in the mean number of significant 5-second EEG correlations, but also no difference in their standard errors across the subject cohort (Figure 8). With no other participatory criteria than possession of both movie-streaming and Facebook accounts, the uniformity in within-movie neural response for such a small subject population (compared to those of market surveys and focus groups) shows that consistent neural response for a large population can be obtained from a relatively small sample size, provided that subjects are chosen within paradigm-relevant constraints. Such consistency is somewhat of a confirmation of previous results obtained by Dmochowski et al. [3], in which a subject cohort on the order of 20 subjects whose neural response was collected and analyzed were predictive of certain behavioral responses of millions of social media users.

One of the most compelling aspects of this study is the link the neural components provide to production value (PV) and narrative voice (NV) characteristics of each film trailer. For PV, we found a significantly above-chance ability to classify the quality of movie production from only the neural components across the population. Even more compelling, for NV, we found a near-perfect accuracy in classifying Expository vs. Participatory NV from whole-movie viewing RCA calculated across the population of subjects. These results indicate that a population-level neural indicator for audience perception of PV and NV may exist in the brains of viewers and, more impressively, be measurable with EEG. A possible reason for the neural basis of PV discrimination is not as simple as one for NV, but there is corroborating neural literature to provide a starting basis for explaining this result. First, for PV discrimination, viewers used in this study have had extensive exposure to the produced films that are standard fare of modern media. In that prolonged exposure, an implicit understanding of high- and low-quality production develops, as it does in music, theater and other temporal forms of media [2, 7, 32, 33, 21]. Due to this accumulated experience, the subject population implicitly is able to grade any viewed movie segment from this experiential context. The extent to which that context impacts the slow-wave variations measured in the EEG is not revealed by this paper, but it is probable that the comprehension measurable with RCA is impacted by the attention demands either mediated or inhibited by high, low or medium PV [16, 23].

The classification of NV from the whole-movie viewing RCA alone also has potential roots in current neural literature. It is widely believed that the default mode network (DMN) is linked to autobiographical narrative and self-awareness [27]. While the DMN activity is prominent in a non-viewing situation, its balance with networks involved in decision-making and active attention is also widely established [27, 20], providing an overall energy balance in activity between the various networks of the human brain. Due to the Participatory PV being a first-person telling, it is possible that the whole-movie viewing RCA incorporates a population-level measurement of DMN activity, either indirectly or directly, that is measurable with the EEG and ultimately classifiable. The finding that Participatory films tended towards having generally higher correlation values than Expository films could indicate that audience comprehension is augmented from this NV choice. While further work is needed to establish this hypothesis, for instance with MRI-based imaging and analysis methods, or expanding the number of films analyzed in this manner, the possibility of an audience-wide neural measurement of first-person engagement opens new avenues in the choice of NV available to content producers, potentially moving media/narrative creation techniques beyond current theoretical constructs and narrative classification schema (e.g., [24]).

4.3 RCA Metrics' Relation to Behavioral Decision Time May Index Movie Narrative Comprehension

Another of the most compelling aspects of this study is the potential link between behavioral and neural response. The aim of accessing nervous system response to social issue video media is to get closer measurement of the circuits executing decisions whose behavior is often unpredictable and difficult to understand. For instance, we found such variability in the utter inconsistency of choice values for a given subject and movie pairing (Figure 3). The experimental testing of our hypothesis concerning RCA-based correlation metrics and response times (hypothesized relationship: Figure 2; actual relationship: Figure 9) shows that, even with a small set of movie trailers (e.g., nine) within a given social issue (e.g., environment), a consistent relationship emerges between neural metrics previously shown to index media comprehension and behavioral metrics connected to cognitive conflict of choice. Figure 9A shows that the low RCA-based correlation, i.e., inconsistent neural response and hence possibly low video comprehension, occurs before a support choice that takes a comparatively long time to execute. Considering this result from the perspective of brand recognition and cognitive conflict [22], this relationship could mean that movie trailers for which response times are high are not communicating their message in a way that encodes into the nervous system of the viewer, leading to confusion upon presentation of the choice screen immediately following the trailer's conclusion. Furthermore, this trend remains upon second viewing (Figure 9B), hence making the average drop in RTs due to repeated viewing not strong enough to counteract the lingering choice conflict that follows viewing (see Figure 9C).

Considering previous work linking RCA-based metrics to comprehension, the relationship between medium/high RCA-based correlations and RTs could point to interplay between surety of choice and invoked understanding in the audience. First-viewing RTs rise to a plateau for high RCA-based correlations, while second-viewing RTs trend towards a sink (see Figure 9A and Figure 9B). No firm conclusions can yet be drawn on this difference without a larger stimulus set, i.e., more movie trailers, to confirm or disprove. A preliminary conclusion could be that reliable neural activity upon first viewing instigates a consideration of the appropriate level of support once the movie ends which would be bounded by the penalty of time spent on decision. The bounding creates the rise to the plateau. Such a penalty does not arise for cases of low RCA-based correlation because cognitive effort is still being spent on video comprehension. In the case of second viewing though, no such time penalty exists: comprehension is high and so anticipation of the target choice precedes presentation of the choice screen, thereby reducing RT. Finally, the mid-range RCA metrics demonstrate a possible 'sweet spot' in viewer certainty, where a sufficient balance between consistency and variability of neural response provokes the fastest RTs. Further research is needed to gauge exactly why such a valley in RTs exists as a function of RCA-based correlation metrics.

Acknowledgements. This work was supported by a grant from the Rita Allen Foundation and the Harmony Institute. Joanna Raczkiewicz contributed to experimental design discussions and Sher Chew designed the presentation of data shown in Figure 7. Julian Saliani assisted with decision tree analysis.

References

1 D. H. Brainard. The Psychophysics Toolbox. *Spatial Vision*, 1997.

2 Simon Carlile. Psychoacoustics. In *The Sonification Handbook*, pages 41–61. Logos Verlag, 2011.

3 J. P. Dmochowski, M. A. Bezdek, B. P. Abelson, J. S. Johnson, E. H. Schumacher, and L. C. Parra. Audience preferences are predicted by temporal reliability of neural processing. *Nat Commun*, 5:4567, 2014. doi:10.1038/ncomms5567.

4 J. P. Dmochowski, A. S. Greaves, and A. M. Norcia. Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage*, 2014. URL: http://arxiv.org/pdf/1407.6110.pdf, doi:10.1016/j.neuroimage.2014.12.078.

5 J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra. Correlated components of ongoing EEG point to emotionally laden attention – a possible marker of engagement? *Front Hum Neurosci*, 6:112, 2012. doi:10.3389/fnhum.2012.00112.

6 M. M. Farbood, D. J. Heeger, G. Marcus, U. Hasson, and Y. Lerner. The neural processing of hierarchical structure in music and speech at different timescales. *Front Neurosci*, 9:157, 2015. doi:10.3389/fnins.2015.00157.

7 Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: Facts and models*. Springer, 2007. arXiv:arXiv:1011.1669v3, doi:10.1007/978-3-540-68888-4.

8 J. Finsterwalder, V. G. Kupppelwieser, and M. de Villiers. The effects of film trailers on shaping consumer expectations in the entertainment industry – A qualitative analysis. *Journal of Retailing and Consumer Services*, 2012. doi:10.1016/j.jretconser.2012.07.004.

9 O. Furman, U. Hasson, L. Davachi, Y. Dudai, and N. Dorfman. They saw a movie: long-term memory for an extended audiovisual narrative. *Learn Memory*, 14(6):457–467, 2007.

10 U. Hasson, O. Furman, D. Clark, Y. Dudai, and L. Davachi. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, 57(3):452–462, 2008. doi:10.1016/j.neuron.2007.12.009.

11 U. Hasson, A. A. Ghazanfar, B. Galantucci, S. Garrod, and C. Keysers. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci*, 16(2):114–121, 2012. doi:10.1016/j.tics.2011.12.007.

12 U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach. Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640, 2004. doi:10.1126/science.1089506.

13 U. Hasson, E. Yang, I. Vallines, D. J. Heeger, and N. Rubin. A hierarchy of temporal receptive windows in human cortex. *J Neurosci*, 28(10):2539–2550, 2008. doi:10.1523/JNEUROSCI.5487-07.2008.

14 Christopher J. Honey, Thomas Thesen, Tobias H. Donner, Lauren J. Silbert, Chad E. Carlson, Orrin Devinsky, Werner K. Doyle, Nava Rubin, David J. Heeger, and Uri Hasson. Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron*, 76(2):423–434, 2012. doi:10.1016/j.neuron.2012.08.011.

15 W. D. Hoyer and S. P. Brown. Effects of Brand Awareness on Choice for a Common, Repeat Purchase Product. *Journal of Consumer Research*, 17:141–148, 1990.

16 Laurent Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10):1304–1318, 2004. doi:10.1109/TIP.2004.834657.

17 D. Jerrick. The Effectiveness of Film Trailers: Evidence from the College Student Market. *UW-L, Journal of Undergraduate Research*, XVI, 2013.

18 J. Kato, H. Ide, I. Kabashima, H. Kadota, K. Takano, and K. Kansaku. Neural correlates of attitude change following positive and negative advertisements. *Front Behav Neurosci*, 3:6, 2009. doi:10.3389/neuro.08.006.2009.

19 D. R. J. Laming. *Information theory of choice-reaction times.* Academic Press, 1968. doi:10.1002/bs.3830140408.

20 Baojuan Li, Xiang Wang, Shuqiao Yao, Dewen Hu, and Karl Friston. Task-dependent modulation of effective connectivity within the default mode network. *Frontiers in Psychology*, 3(JUN), 2012. doi:10.3389/fpsyg.2012.00206.

21 Weisi Lin and C. C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011. doi:10.1016/j.jvcir.2011.01.005.

22 E. K. Macdonald and B. Sharp. Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product:: A Replication. *Journal of Business Research*, 2000. doi:10.1016/S0148-2963(98)00070-8.

23 M.Vranjes, S. Rimac-Drlje, and O. Nemcic. Influence of foveated vision on video quality perception. *2009 International Symposium ELMAR*, pages 28–30, 2009.

24 Bill Nichols. *Introduction to documentary.* Indiana University Press, 2001. arXiv:arXiv:1011.1669v3, doi:10.1017/CBO9781107415324.004.

25 S. Nichter. Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot. *American Political Science Review*, 102(1):19–31, 2008.

26 E. Qualman. *Socialnomics: How Social Media Transforms the Way We Live and Do Business.* Wiley, 2012.

27 Marcus E. Raichle. The Brain's Default Mode Network. *Annual review of neuroscience*, 38:433–447, 2015. doi:10.1146/annurev-neuro-071013-014030.

28 Mor Regev, Christopher J. Honey, Erez Simony, and Uri Hasson. Selective and invariant neural responses to spoken and written narratives. *The Journal of Neuroscience*, 33(40):15978–88, 2013. doi:10.1523/JNEUROSCI.1580-13.2013.

29 Lior Rokach and Oded Maimom. *Data mining with decision trees: theory and applications.* Springer, 2007. doi:10.1007/978-0-387-09823-4.

30 G. J. Stephens, L. J. Silbert, and U. Hasson. Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci U S A*, 107(32):14425–14430, 2010. doi:10.1073/pnas.1008662107.

31 HI storypilot. https://storypilot.org/. Accessed: 2016-05-02.

32 Zhou Wang and Qiang Li. Video quality assessment using a statistical model of human visual speed perception. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 24:B61–B69, 2007. doi:10.1364/JOSAA.24.000B61.

33 Stefan Winkler, Animesh Sharma, and D. McNally. Perceptual video quality and blockiness metrics for multimedia streaming applications. ... *Wireless Personal Multimedia* ..., 2001. URL: `http://stefan.winkler.net/Publications/wpmc2001.pdf`.

34 J. R. Zhang, J. Dmochowski, P. Sajda, J. R. Kender, and J. Sherwin. Correlating speaker gestures in political debates with audience engagement measured via EEG. In *ACM Multimedia 2014*, Orlando, FL, USA, 2014.

Animation Motion in NarrativeML

Inderjeet Mani

Atelier Milford Paradise, Prachuap Kiri Khan, Thailand
inderjeet.mani@gmail.com

Abstract

This paper describes qualitative spatial representations relevant to cartoon motion incorporated into NarrativeML, an annotation scheme intended to capture some of the core aspects of narrative. These representations are motivated by linguistic distinctions drawn from cross-linguistic studies. Motion is modeled in terms of transitions in spatial configurations, using an expressive dynamic logic with the manner and path of motion being derived from a few basic primitives. The manner is elaborated to represent properties of motion that bear on character affect. Such representations can potentially be used to support cartoon narrative summarization and question-answering. The paper discusses annotation challenges, and the use of computer vision to help in annotation. Work is underway on annotating a cartoon corpus in terms of this scheme.

1998 ACM Subject Classification H.5.1 Multimedia Information Systems, I.2.4 Knowledge Representation Formalisms and Methods, I.2.7 Natural Language Processing, I.4.8 Scene Analysis: Motion

Keywords and phrases Cinematography, Motion, Qualitative Reasoning, Narrative, NarrativeML

Digital Object Identifier 10.4230/OASIcs.CMN.2016.3

1 Introduction

Motion is the essence of animated cartoons. Animators go to great lengths to create gestures and sequences of poses that create a vivid and appealing illusion of many different varieties of motion. What would the Road Runner cartoons be without the thrill of the chase, the characters' prolonged braking motions and sudden propulsions? Why are we so entertained by Wile E. Coyote's fantastic object-penetrating collisions and varieties of chasm plunges? One would expect that qualitative representations of the characters' spatiotemporal dynamics would be more relevant to narrative than their precise geometries or the equations describing their highly constrained, cartoon-physics trajectories. Ideally, these qualitative representations should reflect the narratologically-relevant cognitive abstractions used by the audience in describing movies, and at the same time, be computable. This paper describes qualitative spatial representations relevant to cartoon motion incorporated into NarrativeML [28], an annotation scheme intended to capture certain core aspects of narrative.

As the film theorist David Bordwell [3] explains, films offer the same rich stimuli for inferring motion that are presented in the real world. He quotes Paul Messaris [31]: "What distinguishes images (including motion pictures) from language and from other modes of communication is the fact that images reproduce many of the informational cues that people make use of in their perception of physical and social reality." These inferences about motion involve, as is well-known, optical flow [13], which tracks the changing positions of points in sequences of images impinging on the retina (see Section 4). Building on Bordwell's account, I suggest that language-mediated inferences about static and dynamic spatial relations are crucial for narrative. In such an analysis, the spatial concepts are best represented qualitatively, a proposal which may be novel to humanities (including film) narratologists.

© Inderjeet Mani;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 3; pp. 3:1–3:16
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Figure 1 (a) Bugs entering the painting. (b) Woozy motion.

In any medium, manner of movement can be relevant for inferring character properties, including affect and traits. However, cartoon motion seems very different from the motions characters undergo in narrative texts (e.g., the ghost stories studied by [16]). In addition to expressing a parody of motion in the real world, cartoon physics allows for all sorts of creative manners of motion. Consider, as an example, the manner of motion in the 2003 cartoon movie *Looney Tunes: Back in Action*(LTBIA)[1]. In Figure 1(a), Bugs Bunny is about to leap into the painting *The Persistence of Memory* to escape Elmer Fudd, but in Figure 1(b), his motion within the landscape of the painting has become sluggish, as has that of Daffy who has joined him, as their shapes experience Daliesque distortions in this embedded storyworld. Such woozy movements are narratologically significant, as they convey struggle as well as exhaustion.

A large corpus of cartoon movies annotated with systematic characterizations of motion and other narratologically-relevant information would be useful in examining similarities and differences across medium and genre. Such an effort could have implications for humanities narratology [29] by way of providing a precise conceptual framework to enhance narratological theories for media such as cartoons. Corpora annotated with such representations can potentially also be of practical use for training algorithms aimed at movie scene search and summarization.

In 'silent' cartoon movies like the Road Runner ones, the fact that we are speakers of natural language influences our narratological inferences, even when these are drawn from non-textual media. This suggests that annotating the narrative content of a movie should start with an ekphrasis consisting of brief descriptions in natural language. Using natural language descriptions as an additional input for the annotator, beyond the video, not only leverages information that is not directly present in the video, but in addition allows one to harness the rich conceptual resources that natural language provides. Earlier work by [27] has described how qualitative spatial representations can be used to formally represent and reason about well-known aspects of the semantics of spatial prepositions and motion verbs. The contribution of this paper is twofold: extending the representations in NarrativeML to incorporate motion, and the application to the narratives in animated cartoons.

NarrativeML is based on multiple layers of annotation, relying on tagging predicates and arguments in the sentences of the text using PropBank [33]. Events and their temporal relations are represented using TimeML [34], which in turn leverages the interval calculus [1]. The automatic application of TimeML to classical narratological analyses of text is discussed in [25]. NarrativeML also includes a partial temporal ordering of narrative events that share a common protagonist, called a Narrative Event Chain (NEC) [4]. Once incidental events are pruned away, the NEC answers the question as to what the protagonist did in the story. References to places and simple static relations between them are modeled using SpatialML

[1] `https://www.youtube.com/watch?v=97PLr9FK0sw`.

[26] and ISO-Space [36]. All these concepts form part of the fabula (or story). NarrativeML also represents the mapping to *sjuzhet* (or discourse), including the seven varieties of ordering described by [11] as well as narrative tempo and subordinated discourse[2].

In contrast to film narratology such as [2], NarrativeML takes the position that there is always a narrator, but she may or may not be present in the film. For focalization, the annotator of a movie will have to record from whose point of view the scene is being displayed, deciding whether it is the narrator, the 'camera', the audience, or a particular character. Here film presents a challenge. Genette's three-way characterization of focalization into omniscient, internal, and external as in [11], [12] is text-based and involves overlapping categories, as [8] among others has argued. In film, as [21] points out, there may be many different shades of focalization, based on camera angles, deep focus, shot length and scale, etc. A related question is what sort of theory of mind the narrator has with respect to the characters; in the case of a silent movie, gazing into minds may be realized by thought balloons or the focus of attention of the 'camera'. NarrativeML sidesteps the complexities here by allowing for focalization a fourth mixed category, called OTHER, while requiring that the annotator record the position of the viewer relative to figure and ground objects. Thus, above and beyond its role in motion, spatial representation is key to capturing narrative information related to focalization.

2 Spatial Representation

The spatial representations discussed here are motivated by linguistic analyses of prepositions and motion verbs across languages. Being qualitative and linguistically motivated, they are at an entirely different level of abstraction from the fine-grained ones used in animation systems. However, as I will argue, they are useful in representing static and dynamic spatial aspects of narrative.

2.1 Static Spatial Relations

The analysis of spatial prepositions and adpositions in language has come from a variety of theoretical frameworks, including AI and psycholinguistics, e.g., [30], descriptive linguistics, [17], formal semantics, [46], and cognitive linguistics, [23], [18]. Much of the analysis has focused on representing phrases like "the book on the table" and "the fruit in the bowl" in terms of *topological relations* between objects involving notions of coincidence, contact and containment. To formally represent such relations, ISO-Space, and thus NarrativeML, uses the Region Connection Calculus (RCC-8) [37]. In RCC-8, objects are conceived as non-empty, equi-dimensional regions. Based on a single primitive relation of connection between regions, RCC-8 defines the set of eight base relations shown in Figure 2(a). Thus, "the book on the table" may be represented by `EC(book, table)` and "the fruit in the bowl" by `IN(fruit, bowl)`, where `IN` is the disjunction of the base relations `TPP` and `NTPP`.

In addition to topological relations, languages distinguish spatial relations that reflect *orientations* of objects. Studies across languages [24] reveal that they use a basic inventory of three varieties of coordinate systems to describe orientation, that are unevenly distributed across languages. In the *intrinsic* frame, used in examples like "in the front of the picture" and "by the side of the boat", the linguistic relation R between a figure object (F) and

[2] Other aspects of NarrativeML, involving characters, their goals, plot structure, and audience responses are not discussed here.

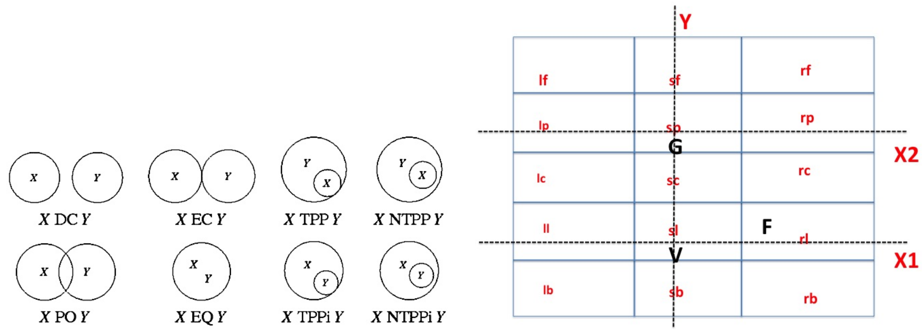


Figure 2 (a) *RCC-8* relations. (b) *Double Cross Calculus* example.

a ground object (G) is characterized in terms of particular facets of the ground object G, e.g., "front", "nose", "sides", etc., which are dependent on the object's affordances and are highly culture-dependent. In this frame, F lies "in a search domain extending from G on the basis of an angle or line projected from the center of G, through an anchor point A (usually the named facet 'R')" [24] (p. 42-3). The *absolute frame* of reference, e.g., "due north of St. Croix", involves a coordinate system where F is described in terms of fixed bearings (related to compass points and/or landscape markers) with respect to an origin on G. The *relative frame* involves a ternary relation, between F, G, and a third object, the viewer V, as in examples like "to the right of Bugs". Here a coordinate system is centered on V, with possibly another coordinate system centered on G arising from a geometric projection from V's coordinate system to G's, in turn providing intrinsic facets to G via V. Languages that have a relative frame always have an intrinsic frame as well, introducing ambiguity.

The ISO-Space representation, and as a result, NarrativeML, is neutral with respect to which qualitative representations should be used to capture orientation relations. Here we introduce three representations that will be used in the example under discussion.

A representation relevant to the *intrinsic frame* is the Dipole Calculus of [32], [7], which represents spatial relations based on oriented line segments called dipoles. Each dipole divides the plane into a left and right half, and the calculus accordingly specifies orientation relations between the start and end points of each dipole and the other. A start or end point on dipole B can be relatively to the left (l) or the right (r) of, or else start (s) or end (e) of, dipole A. Thus, in Figure 1(b), `llrr(Bugs, Daffy)`, meaning that the start and end of Daffy are to the left of Bugs and the start and end of Bugs are to the right of Daffy. This representation is compatible with "Daffy is to the left of Bugs" and "Bugs is on Daffy's right". When augmented with additional orientations: back (b), interior (i), and front (f), one gets a calculus with 69 base relations [32], which we will refer to as DC-69.

The *absolute frame* can be represented in the Cardinal Direction Calculus of [15], [39]. Here the minimum bounding rectangle of the ground region is made the central tile of a 9-element grid, and is labeled 'B', for bounding box. The figure region is then positioned on the grid, and the tiles it falls into are used to describe its orientation with respect to that central tile, yielding nine regions in all: B, S, SW, W, NW, N, NE, E, and SE. Thus, in Figure 1(b), with B over Bugs, we have `ESE(Daffy, Bugs)`. Given that the calculus has a base set of 511 relations, we will refer to it as CDC-511.

For the *relative frame* of reference, the Double Cross Calculus (DCC) of [10], [38], is relevant. Here we have a ternary relation between figure, ground, and viewer. As shown in Figure 2(b), the figure object F, viewer V, and ground G are construed as points, and a line Y from V to G is extended to create a pair of half-planes, left (l) and right (r). A pair of lines, one (X1) perpendicular to the line Y and through V, and the other (X2) parallel through it and through G, creates three regions, forward (f), back (b), with a central region (c) in between. Consider applying it to F=Bugs in Figure 1(a). He is in the plane between the viewer and the ground G, the painting, so we have `rf(Bugs, PersistenceofMemory, Viewer)`. This is compatible with "Bugs is in front of the painting" and "Bugs is on the right in front of the painting". Likewise, in Figure 1(b), Daffy is to the right of Bugs from the viewer's point-of-view, so we have `rc(Daffy, Bugs, Viewer)`. Adding the relations of equality and inequality, we get a base set of 17 relations (DCC-17).

2.2 Motion

Having represented aspects of time and space, one needs to incorporate motion into NarrativeML. A fundamental cross-linguistic insight regarding motion comes from Leonard Talmy [40], [41], who points out that languages have two distinct strategies for expressing concepts of motion. In *manner-type languages* (English and other Germanic languages, also Slavic languages), the main verb expresses the manner or cause of motion, while path information is expressed elsewhere in the form of 'satellite' constituents[3]. In contrast, in *path-type languages* (Romance, Turkish, Semitic, and other languages), the verb expresses the path, whereas the manner is optionally expressed by adjuncts.

Adopting this classification, which has been extensively studied cross-linguistically along with its exceptions, [27] introduce a procedural semantics for motion in natural language, where motion is viewed in terms of transitions in spatial configurations. A distinction is made between *action-based predicates* (for manner-of-motion verbs like "bike", "drive", "fly", etc.) and *location-based predicates* (e.g., for path verbs like "arrive", "depart", etc.). Action-based predicates do not make reference to distinguished locations, but rather to the 'assignment' and 'reassignment' of locations of the object, through the action. The location-based predicates focus on points on a path, and thus they reference a distinguished location, and the location of the moving object is 'tested' to check its relation to this distinguished value.

The semantics for these predicates is expressed in Dynamic Interval Temporal Logic (DITL) from [35], a first-order dynamic logic (introduced by James Pustejovsky) where events are modeled as programs, and states refer to preconditions or post-conditions of these programs. This approach to modeling the semantics of motion, is explained in detail in [27]. The following programs, from [27] (p. 95-107), describe the basic constructs of motion needed.

Definition 1 shows how directed movement away from a source is represented in DITL[4]:

▶ **Definition 1** (Moving away).

```
DITLmove_away(c, src) ≡ y:=src;
    (loc(c):=z, z ≠ y, dist(y, src) < dist(z, src); y:=z)+ /*
1.      Assign y to object location.
2.      Then reassign its location to z, which is further away
        from source than y.
3.      Iterate steps 1−2 one or more times.   */
```

[3] A satellite is "any constituent other than a noun-phrase or prepositional-phrase complement that is in a sister relation to the verb root" [40] (p. 102).

[4] In DITL, semicolon is a program sequencing operator and comma is a (higher-precedence) predicate conjunction operator.

One can now define non-primitive programs corresponding to motions that are lexicalized by motion verbs. Arriving is shown in Definition 2.

▶ **Definition 2** (Arriving as making contact at end of path).

```
reach(c, dest) ≡ (y:= loc(c); RCC-8DC(y, dest)?; move_toward(c, dest))+;
     (y:= loc(c); RCC-8EC(y, dest)?) /*
1.      Test if object is disconnected from the destination.
2.      If so, move towards the destination
3.      Iterate steps 1-2 one or more times
4.      Test if object touches the destination.  */
```

Manner of motion is not treated as a primitive, but arises as an elaboration of the components of the motion, namely figure, ground, event, path, and medium. This allows one to distinguish various manners of motion; for example, one can define sliding (Definition 3), which involves maintaining an extended connection with a surface, as well as bouncing (Definition 4), which involves alternating between an extended connection and disconnection.

▶ **Definition 3** (Sliding).

```
slide(c, surf) ≡ y:=loc(c),
    (loc(c):=z, z ≠ y, RCC-8EC(z, surf); y:=z)+
```

▶ **Definition 4** (Bouncing).

```
bounce(c, surf) ≡ y:=loc(c),
     (loc(c):=z, z ≠ y, RCC-8EC(z, surf); y:=z;
      loc(c):=z, z ≠ y, RCC-8DC(z, surf); y:=z)+
```

For representing affect associated with manners of motion, one has to introduce additional features into the framework. Here I build on the approach of [5], [45], who use natural language in input specifications to drive the motion of animated characters. I focus here on *Effort*, a concept taken from analysis of dance [22]. Effort is characterized (Table 1) in terms of four factors: Space, Weight, Time and Flow, with the left and right columns labeling the low and high ends respectively of a scale.

Thus, the woozy movement in Figure 1(b) is represented in NarrativeML as the event e, where `effort(e, f1)` is associated with the four factors, each on a five-point scale: space and weight as `space(f1, very_low)` & `weight(f1, very_high)`, with time and weight as `time(f1, very_low)` & `flow(f1, very_high)`. Bugs' and Daffy's flight across the landscape of the painting is increasingly tortured and slow, so in previous frames the flow value would have been freer.

Prolonged braking, a device essential to Road Runner and other cartoons, may be viewed as sliding with decreasing speed, as seen in Definition 5. A frazzled variant can be expressed via its Effort.

▶ **Definition 5** (Prolonged Braking).

```
slow-brake(c, surf) ≡ y:=loc(c);
    (loc(c):=z, z ≠ y, RCC-8EC(z, surf), speed(c, z) < speed(c, y);
     y:=z)+
```

Table 1 Effort in Laban's system, from [5].

Space: attention to the surroundings	
Indirect: flexible, meandering, wandering, multi-focus *waving away bugs, slashing through plant growth surveying a crowd of people, scanning a room for misplaced keys*	**Direct**: single focus, channeled, undeviating *pointing to a particular spot, threading a needle, describing the exact outline of an object*
Weight: attitude towards the impact of one's movement	
Light: buoyant, delicate, easily overcoming gravity, marked by decreasing pressure *dabbing paint on a canvas, pulling out a splinter, describing the movement of a feather*	**Strong**: powerful, having an impact, increasing pressure into the movement *punching, pushing a heavy object, wringing a towel, expressing a firmly held opinion*
Time: lack or sense of urgency	
Sustained: lingering, leisurely, indulging in time *stretching to yawn, stroking a pet*	**Sudden**: hurried, urgent *swatting a fly, lunging to catch a ball, grabbing a child from the path of danger, making a snap decision*
Flow: amount of control and bodily tension	
Free: uncontrolled, abandoned, unable to stop in the course of the movement *waving wildly, shaking off water, flinging a rock into a pond*	**Bound**: controlled, restrained *moving in slow motion, tai chi, fighting back tears, carefully carrying a cup of hot liquid*

3 Annotation Example

Sheep in the Island is a 2007 'silent' cartoon film from Korea that features a sheep stranded on a tropical island with a dragon duck[5]. It is a shipwreck narrative, with typical themes of dominance over nature and survival on a deserted island. Inspired by K-Pop culture, the film aims for universal appeal by limiting the presence of text and restricting the audio to non-linguistic verbal sounds and instrumental background music. It thus provides a simple test case for ekphrasis-based narrative annotation. A few sample frames relevant to the discussion below are shown in Figure 3.

The narrative is pre-segmented into sets of time intervals in the video, suggestive segment labels indicated with line comments (//). The time intervals are ordered chronologically, but are not contiguous. The input given to the annotator is shown here highlighted in yellow in Annotation 1. Its ekphrasis is shown alongside, along with the indices of events, entities, and times in NarrativeML[6].

▶ **Annotation 1** (SHEEP IN THE ISLAND).

1. // SETTING
2. 0:02-0:07$_{t1}$ island$_{x1}$ with rock$_{x2}$ and sand$_{x4}$ seen across sea$_{x3}$
3. // BOATS IN MOTION
4. 0:08-0:12$_{t2}$ gunboat$_{m1}$ approaches$_{e1}$ from right partly in front of island$_{x1}$
5. 0:13-0:17$_{t3}$ gunboat$_{m1}$ approaches$_{e2}$ seen from front looming large
6. 0:18-0:20$_{t4}$ gunboat$_{m1}$ approaches$_{e3}$ from left , seen from island$_{x1}$
7. 0:21-0:26$_{t5}$ larger boat$_{m2}$ approaches$_{e4}$ from right as gunboat$_{m1}$ approaches$_{e5}$ from left

[5] `https://mayhemandmuse.com/sheep-in-the-island-part-1/` and `https://www.youtube.com/watch?v=YvR8LGOUpNA`.

[6] This paper and the annotation environment use logical expressions rather than the underlying XML to which it is mapped. XML DTDs for NarrativeML are at `http://tinyurl.com/inderjeetmani/home/NarrativeML`.

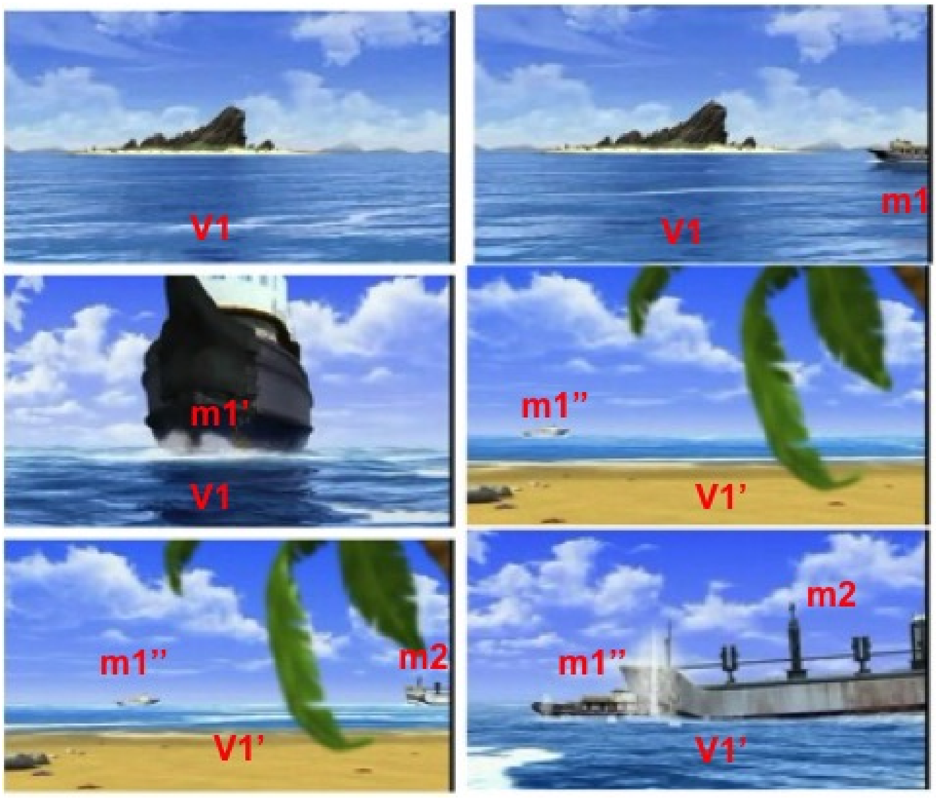


Figure 3 *Sheep in the Island* at 7, 11, 14, 19, 25, and 34 seconds.

8. // A SHIPWRECK
9. 0:33-0:38$_{t6}$ boats$_{m1,m2}$ crash$_{e6}$
10. 0:38-0:41$_{t7}$ boats$_{m1,m2}$ sink$_{e7}$ as three boxes$_{m3,m4,m5}$
    float$_{e8}$ towards island$_{x1}$
11. 0:42-O:46$_{t8}$ one box$_{m3}$ arrives$_{e9}$ on island$_{x1}$
12. 0.47-1:05$_{t9}$ box$_{m3}$ bounces$_{e10}$ on sand$_{x4}$
13. // ENTER THE SHEEP
14. 1:10-1:14$_{t10}$ sheep$_{c1}$ emerges$_{e11}$ from box$_{m3}$, seen from above
15. 1:15-1:17$_{t11}$ sheep$_{c1}$ jumps$_{e12}$ and lands$_{e13}$ on sand$_{x4}$
16. 1:20-1:22$_{t12}$ sheep$_{c1}$ approaches$_{e14}$
17. 1:23-1:26$_{t13}$ sheep$_{c1}$ turns$_{e15}$ and walks$_{e16}$ away
18. 1:27-1:28$_{t14}$ sheep$_{c1}$ turns$_{e17}$ facing forward in head shot
19. // A HUNT INTERRUPTED
20. 1:30-1:32$_{t15}$ sheep$_{c1}$ observes$_{e18}$ frog$_{c2}$ hopping$_{e19}$ on sand$_{x4}$ in front
21. 1:33-1:35$_{t16}$ sheep$_{c1}$ pursues$_{e20}$ frog$_{c2}$
22. 1:36-1:41$_{t17}$ sheep$_{c1}$ catches$_{e21}$ and holds$_{e22}$ frog$_{c2}$
23. 1:41-1:42$_{t18}$ sheep$_{c1}$ gets ready$_{e23}$ to devour$_{e24}$ frog$_{c2}$
24. :43-1:48$_{t19}$ sheep$_{c1}$ notices$_{e25}$ a large box$_{m4}$ to its left
25. 1:49-1:56$_{t20}$ sheep$_{c1}$ slams$_{e26}$ frog$_{c2}$ about
26. 1:58-2:03$_{t21}$ sheep$_{c1}$ strolls$_{e27}$ around box$_{m4}$ to right edge$_{z1}$,
    with stamp$_{z2}$ 'DANGER' on front face$_{z3}$
27. 2:05-2:06$_{t22}$ box$_{m4}$ shakes$_{e28}$
28. // ENTER THE DRAGON
29. 2:07-2:11$_{t23}$ dragon claw$_{y4}$ emerges$_{e29}$ from box$_{m4}$,
    seen from above along with sheep$_{c1}$

The individuation of events is based on the text, annotated in TimeML along with the time intervals[7]. The crucial thing in the BOATS IN MOTION segment is that in line 4, the gunboat m1 (with a front-protruding gun in the video) is seen in profile heading to the left parallel to the viewer V1 who is away from the island. This can also be seen visually in Figure 3 at 11 seconds.

Then, in line 5, the scene switches to a front shot of the same boat (Figure 3 at 14 seconds), the inference being that the viewer has changed orientation, not the boat[8]. In line 6, the gunboat is now seen from the island where the viewer now is, instead of from the sea, and it is now to the left of and parallel to the viewer (Figure 3 at 19 seconds). In line 7, the larger boat m2 (not the gunboat) approaches from the right, with the viewer still in the same position on the island (Figure 3 at 25 seconds), gearing up for a collision in the next segment A SHIPWRECK (line 9 ff., and Figure 3 at 34 seconds).

▶ **Annotation 2** (SETTING).

1. 0:02-0:07$_{t1}$ island$_{x1}$ with rock$_{x2}$ and sand$_{x4}$ seen across sea$_{x3}$
2. narrative(i1) & medium(i1, cartoon_animation) & narrative(i2)
   & medium(i2, text_annotation) & narrative_segment(i1, i3)
   & title(i3, 'SETTING')
3. & narrator(i1, N0) & narrator_type(N0, absent) & narrator(i2, N1)
4. & narrative_time(N0, =) & narrative_time(N1, =)
5. & narrative_order(N0, CHRONICLE) & narrative_order(N1, CHRONICLE)
6. & $_{RCC-8}$EC (x2, x4) // rock$_{x2}$ is connected to sand$_{x4}$
7. & $_{RCC-8}$NTPP (x4, x1) // sand$_{x4}$ is part of island$_{x1}$
8. & $_{RCC-8}$NTPP (x2, x1) // rock$_{x2}$ is part of island$_{x1}$
9. & $_{RCC-8}$EC (V1, x3) // Viewer$_{V1}$ is on sea$_{x3}$
10. & $_{DCC-17}$sf(x1, x3, V1)
    // island$_{x1}$ is in far background with respect to Viewer$_{V1}$

Annotation 2 shows the NarrativeML annotation of the SETTING segment. Line 2 distinguishes the filmic narrative from the textual description. Line 3 indicates that the narrator of the description is in fact the annotator N1, differentiated from the filmic narrator N0, who is absent. Line 4 states that N1 narrates the scene descriptions as in a running commentary, so that the narrative time is simultaneous. The filmic narrator is also not using any devices to suggest retrospective or other temporal distance. Line 5 indicates that the events are narrated by the film as well as by the annotator in (i.e., CHRONICLE) order of occurrence. The RCC-8 relations in lines 6-9 capture coarse-grained topological relations in the SETTING, and the Double Cross Calculus (DCC-17) in line 10 is used to convey point of view, namely the relative frame where the viewer 'camera' is shooting across the sea to the island.

▶ **Annotation 3** (BOATS IN MOTION).

1. 0:08-0:12$_{t2}$ gunboat$_{m1}$ approaches$_{e1}$ from right
   partly in front of island$_{x1}$
2. $_{IC-13}$EQUAL(e1, t2) & @($_{RCC-8}$DC(m1, x1), e1)
   // gunboat$_{m1}$ is disconnected from island$_{x1}$
3. & narrative_segment(i1, i4) & title(i4, 'BOATS IN MOTION')

[7] The BEFORE temporal relations indicating the chronological ordering of events in the fabula are left out for reasons of space.

[8] I use prime notation (V1', m1', etc.) in Figure 3 to remind the reader of an object's changed viewpoint.

```
4. & @(_{RCC-8}EC(m1, x3), e1)          // gunboat_{m1} floats on sea_{x3}
5. & face(m1, y1) & @(_{DC-69}rrrl(y1, m1), e1) // left face_{y1} of gunboat_{m1}
6. & @(_{DCC-17}rc(y1, x1, V1), e1)
     // left face_{y1} is between island_{x1} and viewer
7. & @(_{DITL}move_{away}(y1, RB), e1)
    // RB = right boundary of viewing frame
8. 0:13-0:17_{t3} gunboat_{m1} approaches_{e2} seen from front looming large
9. _{IC-13}EQUAL(e2, t3) & @(_{RCC-8}DC(m1, x1), e2) & effort(e2, f1)
10. & space(f1, very_high) & weight(f1, very_high)
     & time(f1, high) & flow(f1, high)
11. & @(_{RCC-8}EC(m1, x3), e2) & edge(m1, y2) & @(_{DC-69}sbsi(y2, m1), e2)
     // front edge_{y2} of gunboat_{m1}
12. & @(_{DITL}move_{toward}(y2, V1), e2)
13. 0:18-0:20_{t4} gunboat_{m1} approaches_{e3} from left, seen from island_{x1}
14. _{RCC-8}EC(V1, x1) & _{IC-13}EQUAL(e3, t4) // viewer is on island_{x1}
15. & @(_{RCC-8}EC(m1, x3), e3)
16. & face(m1, y3) & @(_{DC-69}lllr (y3, m1), e3)
      // right face_{y3} of gunboat_{m1}
17. & @(_{DCC-17}lf( y3, x1, V1), e3)
        // right face_{y3} is to the left of viewer
18. & @(_{DITL}move_{toward}(y3, RB), e3)
19. 0:21-0:26_{t5} larger boat_{m2} approaches_{e4} from right as gunboat_{m1}
             approaches_{e5} from left
20. _{RCC-8}EC(V1, x1) & _{IC-13}EQUAL(e4, t5) & effort(e4, f2)
21. & space(f2, very_high) & weight(f2, high) & time(f2, neutral)
            & flow(f2, high)
22. & _{IC-13}EQUAL(e5, t5) & effort(e5, f3)
23. & space(f3, high) & weight(f3, high)
            & time(f3, neutral) & flow(f3, neutral)
24. & @(_{RCC-8}EC(m2, x3), e4) & @(_{RCC-8}EC(m1, x3), e5)
            // boats float on sea_{x3}
25. & @(_{RCC-8}DC(m2, x1), e4) & @(_{RCC-8}DC(m1, x1), e5)
            // boats disconnected from island_{x1}
26. & face(m2, y4) & @(_{DC-69}rrrl(y4, m2), e4)    // left face_{y4} of boat_{m2}
27. & face(m1, y3) & @(_{DC-69}lllr(y3, m1), e5)
            // right face_{y3} of gunboat_{m1}
28. & @(_{DCC-17}lf( y3, x1, V1), e5)
            // right face_{y3} is to the left of viewer
29. & @(_{DCC-17}rf( y4, x1, V1), e4)
            // left face_{y4} is to the right of viewer
30. & @(_{DITL}move_{away}(y4, RB), e4) & @(_{DITL}move_{toward}(y3, RB), e5)
```

Annotation 3 turns to motion, which has until now not been represented in NarrativeML. In line 2, the @ predicate indicates that the separation of the gunboat from the island holds throughout e1. In line 5, the intrinsic left face y1 of the gunboat is characterized with an additional primitive spatial relation called `face`, using the Dipole Calculus (DC-69) to represent the left one, i.e., the gunboat dipole m1 is viewed as to the right and orthogonal to the left face dipole, i.e., `y1`$\uparrow$`m1`$\rightarrow$, yielding the relation `rrrl(y1, m1)`. This left face is moving away from the right boundary, as indicated by the $\texttt{move}_{away}$ predicate in line 7. In line 8, the scene changes to the front view of the gunboat, with its increased Effort, impelled as if by a sinister force, indicated in line 10. The gunboat's intrinsic front `edge` (another primitive) y2 is identified in line 11 using DC-69, where the two dipoles are represented as being on the same line. The DC-69 relation `sbsi(y2, m1)` expresses the fact that the start of the gunboat m1 is at the start of its front edge and its end is behind its front edge, and the start of its front edge is at the start of the gunboat and its end is in the interior of the gunboat. The gunboat's front edge y2 is moving towards the viewer as indicated in line 12. Capturing the fact that the gunboat is speeding towards the viewer V1 while looming steadily larger is narratologically important, as actions with the viewer as target have the potential to increase suspense.

The movement to the right of the other boat is captured in the remaining lines. The Effort of the boats approaching each other is indicated in lines 21 (larger boat) and 23 (gunboat), with the larger boat with its greater apparent momentum indicated by increased Effort.

▶ **Annotation 4** (A SHIPWRECK).

1. 0:33-0:38$_{t6}$ boats$_{m1,m2}$ crash$_{e6}$
2. $_{RCC-8}$EC(V1, x1) & $_{IC-13}$EQUAL(e6, t6)
   & @(reach(m1, m2), e6) & narrative_segment(i1, i5)
   & title(i5, 'A SHIPWRECK')
3. & @($_{RCC-8}$EC(m1, x3), e6) & @($_{RCC-8}$EC(m2, x3), e6)
4. & @($_{RCC-8}$EC(m2, x3), e6) & @($_{RCC-8}$EC(m1, x3), e6)
5. & @($_{RCC-8}$DC(m2, x1), e6) & @($_{RCC-8}$DC(m1, x1), e6)
6. & @($_{RCC-8}$PO(y3, y4), e6)
   // right face of gunboat$_{m1}$ telescopes into left face of boat$_{m2}$
7. & effort(e6, f4) & space(f4, very_high) & weight(f4, very_high)
   & time(f4, very_high) & flow(f4, very_high)
8. 0:38-0:41$_{t7}$ boats$_{m1,m2}$ sink$_{e7}$ as three boxes$_{m3,m4,m5}$
   float$_{e8}$ towards island$_{x1}$
9. & face(x3, y5) & $_{DC-69}$sfsi(y5, x3) // bottom of sea$_{x3}$
10. & $_{RCC-8}$EC(V1, x1) & @($_{DITL}move_{toward}$(m1, y5), e7)
    & @($_{DITL}move_{toward}$(m2, y5), e7)
    & $_{IC-13}$EQUAL(e7, t7) & $_{IC-13}$EQUAL(e8, t7)
11. & @($_{RCC-8}$NTPP(m1, x3), e7) & @($_{RCC-8}$NTPP(m2, x3), e7)
    // boats submerged
12. & @($_{RCC-8}$EC(m3, x3), e8) & @($_{RCC-8}$EC(m4, x3), e8)
    & @($_{RCC-8}$EC(m5, x3), e8)
13. & @($_{DITL}move_{toward}$(m3, x1), e8) & @($_{DITL}move_{toward}$(m4, x1), e8)
    & @($_{DITL}move_{toward}$(m5, x1), e8) // towards island$_{x1}$
14. 0:42-O:46$_{t8}$ one box$_{m3}$ arrives$_{e9}$ on island$_{x1}$
15. $_{RCC-8}$EC(V1, x1) & $_{IC-13}$EQUAL(e9, t8) & @(reach(m3, x1), e9)
16. 0.47-1:05$_{t9}$ box$_{m3}$ bounces$_{e10}$ on sand$_{x4}$
17. $_{RCC-8}$EC(V1, x1) & $_{IC-13}$EQUAL(e9, t9) & @(bounce(m3, x4), e10)
18. & NEC(m1, <e1,..,e7> & NEC(m2, <e4, e6, e7>)
19. & NEC(m3, <e8 ,..., e10)>) & NEC(m4, <e8>) & NEC(m4, <e8>)
20. & effort(e10, f5) & space(f5, low) & weight(f5, very_low)
    & time(f5, low) & flow(f5, low)

Annotation 4 begins with the boats crashing, which is seen as the right face of the gunboat telescoping into the left face of the larger boat (Figure 3 at 34 seconds). Line 7 indicates that the Effort is at the maximum for all its factors. In line 11, the boats are submerged below the sea, expressed in RCC-8. Line 12 has the three boxes floating on the sea, and in line 14 they move towards the island. The boxes emerge as by-products born of the crash, which is an early inflexion-point in the plot. In line 15, one box reaches the island, and in line 17, it bounces on the sand. Lines 18-19 indicate the NECs for the boats and the boxes. Line 20 characterize the effort involved in the bouncing of box m3, which is

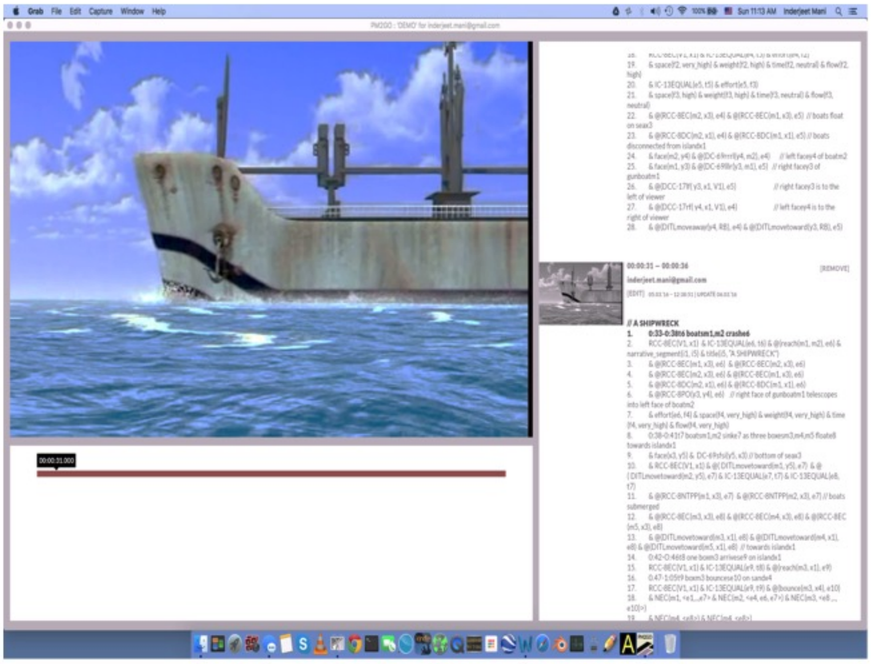


Figure 4 Annotating A SHIPWRECK at 0:33–0:38.

relatively unconstrained, propelled as the box is by the energy of the creature trapped inside. The self-propelled bouncing of the box foreshadows the emergence of new characters. Thus, although the entities in motion in the first three annotated segments (boats and boxes) do not involve the lifelike characters of the sheep and dragon duck, annotating their specific motions is relevant for plot structure as well as foreshadowing the arrival of those characters.

4 Annotation Challenges

Figure 4 shows the video annotation tool PM2GO being used to annotate segments of *Sheep in the Island*[9]. The video is shown on the left, with the player and interval selection below, and the annotations on the right: BOATS IN MOTION, above, and A SHIPWRECK below, using Annotations 3 and 4, respectively.

While individual movie ekphrases might be generated by crowd-sourcing, the annotations are too dense to be efficiently executed for large corpora without some level of automatic preprocessing. The good news here is that progress has been made on automatic labeling of semantic roles for PropBank, e.g., [14], SpatialML tagging [26] and Semantic Role Labeling (in the SemEval tasks) for figure-ground spatial relations [20]. For automatic TimeML tagging, there has been progress as well, though approaches seemed to have hit a ceiling of 70% F-measure on event-ordering across languages and tasks, e.g., TempEval [43], in part due to the paucity of annotated data[10]. Unfortunately, the annotation using PM2GO

[9] See `http://motionbank.org/sites/motionbank.org/files/pm2go_handbook_07_14.pdf`.

[10] Narrative texts auto-tagged with TimeML are available at `http://tinyurl.com/inderjeetmani/home/NarrativeML`.

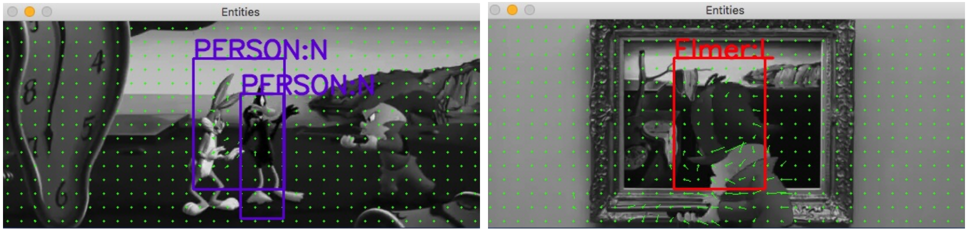


Figure 5 (a) Person detection. (b) Character and motion labeling.

does not use any automatic pre-processing. Integrating the TARSQI toolkit for TimeML tagging[11] and the SpatialML tagger[12] into an annotation pipeline is nontrivial since they are legacy software systems. Longer-term plans include re-implementing such capabilities on top of the far more modern Stanford CoreNLP toolkit[13] as well as migrating to a more narrative-friendly annotation workbench for video.

So far, the annotation of motion itself has not been automated. One possibility here is to leverage the field of computer vision, which has been advancing rapidly. It seems reasonable to populate some of the ekphrases and their annotations with suggestions from video processing. Figure 5 shows some results from applying computer vision tools from OpenCV[14] to *Looney Tunes: Back in Action*. In Figure 5(a), Bugs and Daffy have been classified as people using a Histogram of Oriented Gradients (HOG) [6] pre-trained on images of people; note that Elmer has been missed. Figure 5(b) shows that Elmer has been detected as an object and labeled correctly, using the Haar classifier cascade of [44], trained on labeled images from a corpus of Bugs Bunny cartoons. The system has also correctly identified Elmer's direction of movement (left) using an optical flow detector [9]. In addition to improving the accuracy of such computer vision methods with more training data, it should be possible to extend them to automatically label the type of motion, as in [42].

While NarrativeML has been used to annotate numerous examples, it has not as such been applied to text corpora in the large, let alone to ekphrases for movies, so important questions of annotation reliability and efficiency remain open. These latter questions are the focus of current research, applied to a corpus of cartoon movies. To simplify the task, the pre-selected set of frames to be annotated is restricted to relatively short time intervals, with the guidelines focused on creation of the ekphrasis and its NarrativeML for that set.

5 Conclusion

In terms of expressiveness, these additions to NarrativeML (constituting version 0.2) allow for the annotation of relevant narrative information in cartoon movies, at a level of abstraction guided by natural language and representing key semantic distinctions related to space and motion. The annotation scheme is thus attractive for representing spatial relations, focalization and motion in cartoons, and could potentially be used for humanities narratology and practical applications as described in Section 1. The scheme might also be embedded in authoring environments for animation.

[11] http://www.timeml.org/tarsqi/toolkit/download.html
[12] http://www.timeml.org/tarsqi/toolkit/download.html
[13] http://nlp.stanford.edu/software/corenlp.shtml
[14] http://opencv.org

Of course, there is still much that is missing that would shed light on narrative. For the intrinsic frame, where object shape is important, the dipole calculus is not that suitable. For focalization, there needs to be a characterization of relevant shot types, as discussed in [19], as well as the varieties of shot transition or cut. The varying distance, focus, orientation, and area of interest of the 'camera' are also crucial for film narrative. In addition, for the cartoon genre, character shape, as well as more elaborate motion manners and their velocities may be revealing of character affect. Recording this sort of information in narrative corpora could be very valuable. Nevertheless, reasoning with such qualitative representations is not always tractable, and maximal tractable subsets of calculus relations, when found, often require discarding key relations. Combining representations and adding dimensions only add to the complexity. Finally, there are numerous annotation challenges discussed in Section 4, some of which can be addressed by computer vision.

References

1. James Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843, 1983.
2. David Bordwell. *Narrative in the Fiction Film*.Madison: University of Wisconsin Press, 1985.
3. David Bordwell. *Common Sense + Film Theory = Common-Sense Film Theory?* `http://www.davidbordwell.net/essays/commonsense.php`.
4. Nathanael Chambers. *Inducing Event Schemas and their Participants from Unlabeled Text*. Ph.D. Dissertation, Department of Computer Science, Stanford University, 2011.
5. Diane Chi. *A Motion Control Scheme for Animating Expressive Figure Arm Movements*. PhD Thesis. University of Pennsylvania, 1999.
6. Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, 2005, pages 886–893.
7. Frank Dylla and Reinhard Moratz. Empirical complexity issues of practical spatial reasoning about relative position. In *Workshop on Spatial and Temporal Reasoning at ECAI 2004*, Valencia, Spain, August 2004.
8. William F. Edmiston. *Hindsight and Insight: Focalization in Four Eighteenth-Century French Novels*. University Park, PA: Penn State University Press, 1991.
9. Gunnar Farnebäck. Two-frame Motion Estimation Based on Polynomial Expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, pages 363–370, 2003.
10. Christian Freksa. Using orientation information for qualitative spatial reasoning. In A. U. Frank, I. Campari, and U. Formentini (eds.), *Theories and methods of spatiotemporal reasoning in geographic space*, Springer, Berlin, pages 162–178, 1992.
11. Gerard Genette. *Narrative Discourse* (trans. Jane Lewin). Ithaca: Cornell University Press, 1980.
12. Gerard Genette. *Narrative Discourse Revisited* (trans. Jane Lewin). Ithaca: Cornell University Press, 1988.
13. James J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.
14. Daniel Gildea and Daniel Jurafsky. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3), pages 245–288, 2002.
15. R. Goyal and M. J. Egenhofer. Consistent queries over cardinal directions across different levels of detail. In *Proceedings of the 11th International Workshop on Database and Expert Systems Applications*, 2000.

16 David Herman. Spatial Cognition in Natural-Language Narratives. In M. Mateas and P. Sengers (eds.), *Working notes of the Narrative Intelligence Symposium*, pages 21–25. AAAI Fall Symposium Series. Menlo Park, CA: AAAI Press, 1999.

17 A. Herskovits. *Language and Spatial Cognition: an interdisciplinary study of the prepositions in English*. Cambridge University Press, 1986.

18 Ray Jackendoff. *Semantic Structures*. Cambridge, MA: MIT Press, 1990.

19 Manfred Jahn. *Narratology: A Guide to the Theory of Narrative*. English Department, University of Cologne, 2003. `http://www.uni-koeln.de/~ame02/ppp.htm`.

20 Parisa Kordjamshidi, M. Van Otterlo, and M. F. Moens. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing (TSLP)*, 8 (3), 4, 2011.

21 Markus Kuhn and Johann N. Schmidt. Narration in Film. In Peter Huhn et al. (eds.), *The Living Handbook of Narratology*, paragraph 28. Hamburg: Hamburg University Press, 2014. `http://www.lhn.uni-hamburg.de/article/narration-film-revised-version-uploaded-22-april-2014`.

22 Rudolf Laban and F. C. Lawrence. *Effort: Economy in Body Movement*. Plays, Inc., 1974.

23 G. Lakoff. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press, 1987.

24 S. C. Levinson. *Space in Language and Cognition*. Cambridge University Press, 2003.

25 Inderjeet Mani. *The Imagined Moment*. Lincoln: University of Nebraska Press, 2010.

26 Inderjeet Mani, Christine Doran, David Harris, Justin Hitzeman, Robert Quimby, Justin Richer, Ben Wellner, Scott Mardis, and Seamus Clancy. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44(3):263–280, 2010.

27 Inderjeet Mani and James Pustejovsky. *Interpreting Motion: Grounded Representations for Spatial Language*. New York: Oxford University Press, 2012.

28 Inderjeet Mani. *Computational Modeling of Narrative*. Synthesis Lectures on Language Technologies, Morgan & Claypool, 2013.

29 Jan Christoph Meister. *Computing Action. A Narratological Approach*. Berlin: de Gruyter, 2003.

30 George A. Miller and Philip N. Johnson-Laird. *Language and Perception*. Belknap Press of Harvard University Press, 1976.

31 Paul Messaris. *Visual Literacy: Image, Mind, and Reality*. Boulder: Westview Press, page 165, 1994.

32 Reinhard Moratz, Jochen Renz, and DiedrichWolter. Qualitative spatial reasoning about line segments. In W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*. Berlin, Germany, IOS Press 2000.

33 Martha Palmer, Dan Gildea, and Paul Kingsbury. The Proposition Bank: a corpus annotated with semantic roles. *Computational Linguistics*, 31(1):71–105, 2005.

34 James Pustejovsky, Bob Ingria, Roser Sauri, Jose Castano, Jessica Littman, Rob Gaizauskas, Andrea Setzer, Graham Katz and Inderjeet Mani. The specification language TimeML. In Inderjeet Mani, James Pustejovsky and Robert Gaizauskas (eds.), *The Language of Time: A Reader*, New York: Oxford University Press, pages 49–562, 2005.

35 James Pustejovsky and Jessica L. Moszkowicz. The Qualitative Spatial Dynamics of Motion in Language. In M. Bhatt, H. Guesgen, S. Woelfl, and S. Hazarika (eds.), *Qualitative Spatial and Temporal Reasoning: Emerging Applications, Trends and Future Directions*. Journal of Spatial Cognition and Computation, 11(1): 15–44, 2011.

36 James Pustejovsky, Jessica L. Moszkowicz, and Marc Verhagen. The current status of ISO-Space. Joint ISA-7 Workshop on Interoperable Semantic Annotation SRSL-3, *Workshop on Semantic Representation for Spoken Language*, I2MRT Workshop on Multimodal Resources and Tools, 2012.

37 D. A. Randell, Z. Cui. and A. G. Cohn. A Spatial Logic on Regions and Connection. In *Proceedings of 3rd Int. Conf. on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, pages 165–176, 1992.

38 Alexander Scivos and Bernhard Nebel. Double-Crossing: Decidability and Computational Complexity of a Qualitative Calculus for Navigation. In *Proceedings COSIT-2001*, Springer-Verlag, 2001.

39 Spiros Skiadopoulos and Manolis Koubarakis. On the consistency of cardinal direction constraints. *Artificial Intelligence* 163, pages 91–135, 2005.

40 Leonard Talmy. *Toward a Cognitive Semantics*. MIT Press, 2000.

41 Leonard Talmy. Main Verb Properties and Equipollent Framing. In Guo JianSheng et al. (eds.), *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*. Lawrence Erlbaum Associates, 2009.

42 Subhashini Venugopalan, Huijuan Xu, Jeff Donahue, Marcus Rohrbach, Raymond Mooney and Kate Saenko. Translating Videos to Natural Language Using Deep Recurrent Neural Networks. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, Denver, Colorado, June 2015, pages 149–1504.

43 Marc Verhagen, Roser Sauri, Tommaso Caselli and James Pustejovsky. SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2)*, Uppsala, 2010, pages 57–62.

44 Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, pages 511–518, 2001.

45 Liwei Zhao, Monica Costa, and Norman I. Badler. Interpreting Movement Manner. In *Computer Animation 2000 (CA'00)*, Philadelphia, Pennsylvania, 2000, pages 98–103.

46 Joost Zwarts and Yoad Winter. Vector space semantics: A model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9(2):171–213, 2000.

Steps Towards a Formal Ontology of Narratives Based on Narratology*†

Valentina Bartalesi[1], Carlo Meghini[2], and Daniele Metilli[3]

1 **Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy**
valentina.bartalesi@isti.cnr.it

2 **Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy**
carlo.meghini@isti.cnr.it

3 **Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy**
daniele.metilli@isti.cnr.it

Abstract

Narrative is emerging as a notion that may enable overcoming the limitations of the discovery functionality (only ranked lists of objects) offered by information systems to their users. We present preliminary results on modelling narratives by means of formal ontology, by introducing a conceptualization of narratives and a mathematical expression of it. Our conceptualization tries to capture fundamental notions of narratives as defined in narratology, such as fabula, narration and plot. A validation of the conceptualization and of its mathematical specification is ongoing, based on the Semantic Web standards and on the CIDOC CRM ISO standard ontology.

1998 ACM Subject Classification H.3.7 Digital Libraries

Keywords and phrases Ontology, Narrative

Digital Object Identifier 10.4230/OASIcs.CMN.2016.4

1 Introduction

Information systems are object repositories of many different kinds, ranging from digital libraries (DLs) to institutional repositories, archives, and more. A common trait of these systems is their discovery functionality, based on the production of ranked lists of objects in response to queries in a natural language. This discovery functionality has been serving the users of these systems since a few decades now and there is common agreement that information systems, and DLs in particular, should move beyond it, offering a more sophisticated service. A way of doing so is to introduce narratives as first class objects in these systems. Narratives are natural candidates for advancing the performance of DL services for two fundamental reasons: (i) a narrative brings more information to the user than a simple list of unrelated objects; (ii) the introduction of the entities required for modelling narratives, *i.e.*, events and their contextualization properties, will enrich the information space of DLs, thereby producing beneficial effects on the functionality of DL systems. For

* This work was partially supported by the Italian national project (PRIN) *Towards a Digital Dante Encyclopedia*.

† This work has been partly supported by the European Commission under the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (Ariadne) project, grant agreement no. 313193.

© Valentina Bartalesi, Carlo Meghini, and Daniele Metilli;
licensed under Creative Commons License CC-BY

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 4; pp. 4:1–4:10
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

instance, it will be possible to obtain the events that happened in a defined range of time, as well as the events linked with a particular relation (*e.g.*, all the sub-events related to a main event through a hierarchical relation), or to compare different narratives on the same topic in order to identify similarities and differences.

As a necessary step towards the introduction of narratives in DLs, in this study we present a formal model of narratives. The model is derived from an analysis of the literature in classical narratology, and it is independent of any specific functionality one might desire to implement on narratives. In this sense, it aims at being an ontology of narratives. The model is also independent of any specific way of populating it: narratives conforming to the model may be constructed in a purely manual way, or with the support of automatic methods for extracting the involved knowledge from texts or other media.

In compliance with the theory of narratology, we view a narrative as a complex object spanning three main dimensions. The first one is the *fabula*, the network of events that the narrative purports to narrate. The second dimension is the textual *narration* of these events. The last one is the *association* between the narrated events and the narrating text.

Being rich in structure, narratives include many relationships between the entities composing them: relationships between events (*e.g.*, temporal and mereological), and between events and the objects that contextualize them, such as people, places, things, topics, and more. In this study we provide a first modelling of narratives, focussing on representing the factual aspects and the corresponding relations that characterize an event: *Where* an event happens, *Who* (persons) and *What* (things) are involved in it [23]. Our approach is to give a mathematical expression to these entities, to be used as a conceptualization for an ontology.

The paper is organized as follows: Section 2 presents a brief overview of the notions that we formalized from the (very vast) literature on narratives. The model is subsequently developed in two steps: a conceptualization step (Section 3) laying down the structure of the model, and a formalization step (Section 4) in which the structure is expressed in mathematical terms. Section 5 discusses further developments, while Section 6 reports our final remarks.

2 Background

In literary theory, narratology is a discipline devoted to the study of the narrative structure and the logic, principles, and practices of its representation [19]. The earliest antecedent to modern narratology can be found in the classical Aristotle's theory of aesthetics. Indeed in *Poetics*, Aristotle defines a narrative as the imitation of real actions (*praxis*) that forms an argument (*logos*) whose fundamental units, or events, can be arranged in a plot (*mythos*) [2].

For Russian formalism the narratology is based on the idea of a common literary language, or a universal pattern of codes that operates within the content of a work. A narrative can thus be conveyed through several different means of communication and a wide range of media, including speech, writing, gestures, music, etc. In particular, Vladimir Propp's *Morphology of the Folktale* (1928) [22] proposed a model to represent folktales as combinations of basic building blocks, including thirty-one "narrative functions" and seven roles, or "spheres of action", of the characters.

The theory of narratology was further developed by mid-20th Century structuralism. Claude Lévi-Strauss, in *Structural Anthropology* (1958) outlined a grammar of mythology; in *Structural Semantics* (1966) [13] A.J. Greimas proposed a system of six basic structural elements of narratives called *actants*; Tzvetan Todorov was the first to coin the term *narratologie* [25]. Later on, Gérard Genette [12] codified a system of analysis that studied both the narration and the act of narrating, considering them separately from the story and content of the text.

Since 1980, post-structuralist perspectives of narratology have been developed. In particular, Cognitive Narratology [14], which considers narratology a psychological phenomenon, and proposes a study of narrative aspects from a cognitive perspective. Empirical results from cognitive psychology highlight that most common-sense concepts cannot be characterised in terms of necessary/sufficient conditions. Monotonic description logics capture the aspects of compositional conceptual knowledge, but are insufficient in representing prototypical knowledge. However, a general description logic to represent concepts in prototypical terms does not exist yet [11, 16].

2.1 Fabula and Syuzhet

Russian formalism distinguishes between a *fabula*, defined as a series of events taking place at a certain time at a specific location, and a *syuzhet*, which is the particular way the story is narrated. Contrary to the order of the fabula, that is strictly chronological, the order of the syuzhet corresponds to the way the events are presented in the narrative by the author [22] [24]. A similar distinction is drawn in structuralism by Chatman [5], who identifies the opposing concepts of *story*, *i.e.*, the content that is transmitted, and *discourse*, *i.e.*, the particular organization of that content.

Currently, there is no universally accepted definition of the narrative structure. For instance, Crawford [6] posits that a narrative is a high-level structure based on causality, not on temporal or spatial relations. Genette [12] identifies five concepts that characterize the syntax of narratives: order, frequency, duration, voice and mood. In addition to the fabula and the syuzhet, Bal [3] defines a third level that constitutes the concrete representation of the content that is conveyed to the audience (e.g., the text in a novel).

2.2 Computational Narratology

The computational narratology studies narratives from a computational perspective. In particular, it focuses on "the algorithmic processes involved in creating and interpreting narratives, modelling narrative structure in terms of formal computable representations" [17].

The computational narratology is based on engineering disciplines aiming at developing Artificial Intelligence (AI) systems for reproducing human-like narrative behaviour and intelligent interfaces and game environments for interacting with narratives [18].

Computational narratology can assume different meanings according to different research contexts. In the AI perspective [4] we are interested in, computational narratology refers to the story generation systems, *i.e.*, any computer application that creates a written, spoken, or visual presentation of a narrative. Indeed, one of our aims is to develop a semi-automatic tool that allows users to construct a narrative, on top of the formal model we developed.

3 Conceptualization

In this Section, we present our formal computable representation of narrative, as derived from the above background, in an informal way. In particular, we envisage a *narrative* as consisting of three main elements:

1. the *fabula*, directly representing the fabula as defined by the Russian formalism, *i.e.*, the sequence of the events that composes the story in chronological order;

2. one or more texts that narrate the fabula, that we call *narrations* and that correspond to the Bal's definition of *presentation*;
3. a *reference* function that connects the narrations to the fabula and allows us to derive the *syuzhet* (or plot) as defined by the Russian formalism.

Fabula. The fabula is built on top of events, an event being an action or occurrence taking place at a certain time at a specific location. This definition of event is at the basis of the Event Calculus (EC) [15, 20, 21], a logic language developed in Artificial Intelligence for reasoning on the actions of a robot. In EC the terms Actions and Events are interchangeable and represent changes performed over time, whereas Davidson [8] defines actions as a particular subclass of events, that is the events endowed with intentionality. We subscribe to this view and consider actions as a special kind of events.

The narratives we are interested in are those found in the digital humanities. Therefore the events in one of our fabulae may be (a) real, such as those witnessed by a scholar using an information system, and recorded for communication purposes, or (b) hypothetical, such as those recorded by a historian in the process of reconstructing a particular piece of history, or (c) fictional, such as those created by writers in the literature.

In a fabula, events are connected to each other by three kinds of relations:

- a *mereological* relation, relating events to other events that include them as parts, *e.g.*, the birth of Dante Alighieri, the major Italian poet of the late Middle Ages, is part of the life of Dante;
- a *temporal occurrence* relation, associating each event with a time interval during which the event occurs. An event occurs before (or during, or after) another if and only if the period of occurrence of the former event is before (or during, or after) the period of occurrence of the latter. We formalize this relations between events using the Allen's temporal logic [1];
- a *causal dependency* relation, relating events that in normal discourse are predicated to have a *cause-effect* relation in the narrator's opinion, *e.g.*, the eruption of the Vesuvius caused the destruction of Pompeii. It is important to notice that in the Digital Humanities we are not interested in modelling the mechanical causal relationships that connect, for instance, events in a physical or chemical process. We are rather interested in a more generic notion of causality, whereby the connected events may be years apart in time (or centuries, like in history) and the causal connection may be indirect, *i.e.*, established through other events, which may be unknown or not represented as relevant. For this reason we prefer to speak about causal dependency as opposed to causality *tout court.* Technically, causal dependency can be thought as a generalization of scientific causality, produced by the transitive closure of the atomic relationships that constitute scientific causality.

Narrations. Each narration of a fabula consists of one or more narrators and a text, which is *authored by* (another relation in the conceptualization) the narrator(s) and constitutes the narration proper. Although the modelling of text is an active field of investigation at the crossroads of many disciplines, and there are many models of literary text that can be used in the present context, at this stage we focus on the only aspect that is functional to our model of narrative, namely *textual content,* that is the language expression that constitutes the content of a piece of text. We will therefore use textual content as identity of text, thereby adopting a purely extensional view. Notice that in this view the structure of a text, which is the decomposition of a text in *textual units* as established by the author(s), can be derived as a containment relation between individual texts.

Reference. The reference function connects each portion of text that narrates an event to the narrated event. In order to model reference we need to identify textual units, which we call *narrative fragments* (or simply fragments), each of which narrates a single event. The underlying assumption is that it is always possible to partition a text of a narrative into disjoint fragments. Based on our experience so far, this stands as a reasonable assumption, nevertheless it can be removed by modelling reference as a relation, whereby a fragment can be associated to more than one event.

Notice that the reference function allows deriving the plot of the narrative. Indeed, by visiting the text of the narration in its natural order, it is possible to access the *narrative fragments* and, via these, the events in the fabula, *in the order established by the narrator*, which may be different from the chronological ordering of the events in the fabula.

4 A Mathematical Specification of the Conceptualization

In this Section we provide a specification of the above conceptualization in mathematical terms. This will allow us to concentrate on the proper capturing of the notions highlighted above, postponing any language consideration to a later stage, once the mathematical specification will have brought forward the required machinery. As it will be shown, the elementary notions of set theory (see for instance [7]) will suffice for our purposes.

We start from three disjoint countable sets:

- events, denoted as E, members e, e_1, e_2, ...
- time points, denoted as T, members t, t_1, t_2, ..., totally ordered by a time precedence relation $<$
- texts, given by the strings of finite length over an alphabet S, $S^\star$, members s, s_1, s_2, ...

A fabula f is a 5-tuple $\mathsf{f} = \langle E_\mathsf{f}, p_\mathsf{f}, b_\mathsf{f}, d_\mathsf{f}, c_\mathsf{f} \rangle$ consisting of:

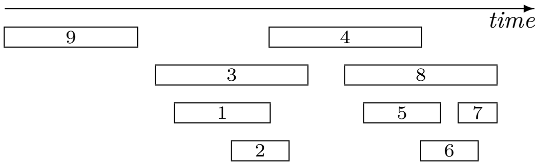
- A finite set of events, $E_\mathsf{f} \subset \mathsf{E}$
- The *event composition* function $p_\mathsf{f} : E_\mathsf{f} \to E_\mathsf{f}$ associating *some* event e_1 in E_f with a different event e_2 in E_f, such that e_1 is a part of e_2. In this case, we say that e_1 is a sub-event of e_2 or that e_2 is a super-event of e_1.
- The *event beginning* function $b_\mathsf{f} : E_\mathsf{f} \to T$, associating each event e in E_f with a time-point $t = b_\mathsf{f}(e)$ in T, such that event e starts at time $b_\mathsf{f}(e)$.
- The *event ending* function $d_\mathsf{f} : E_\mathsf{f} \to T$, associating each event e in E_f with a time-point $t = d_\mathsf{f}(e)$ in T, such that event e ends at time $d_\mathsf{f}(e)$.
- The *causal dependence* relation $c_\mathsf{f} \subseteq E_\mathsf{f} \times E_\mathsf{f}$, such that $e_1, e_2 \in c_\mathsf{f}$ if and only if event e_2 causally depends on event e_1.

For simplicity, we will omit subscripts from fabula components, when there is no ambiguity.

For each event $e \in E$, the pair $(b(e), d(e))$ is said to be the *period of occurrence of* e.

A well-formed fabula is a fabula satisfying the following conditions:

1. The event composition function p is acyclic, so that no event can be, at the same time, a sub-event and a super-event of some other event. Technically, acyclicity can be expressed as the condition that the transitive closure of p, $p^\star$, be an irreflexive relation.
2. No event finishes earlier than its beginning: for each event e in E, $b(e) \leq d(e)$.
3. The period of occurrence of a sub-event is always included in the period of occurrence of its super-event: for each event e in the domain of p, $b(p(e)) \leq b(e)$ and $d(e) \leq d(p(e))$.
4. Causal dependency is a reflexive and transitive relation.



level i	Units in $\sigma(i)$
1	[1,2000]
2	[1,1500] [1501,2000]
3	[1201,1500] [1901,2000]

Figure 1 Events in a fabula (left) and the structure of a narration (right).

From now on, we will tacitly consider only well-formed fabulae.

Notice that we allow events in the same fabula to overlap in time in an arbitrary way, enabling even the sub-events of the same event to do so. Also, we do not place any other condition on causal dependency other than the obvious reflexivity and transitivity.

Figure 1 left gives a pictorial representation of a fabula consisting of nine events, identified with the first nine positive integers, each represented by a rectangle whose horizontal extension gives the temporal extension of the event on the time scale depicted at the top of the figure. The event composition function is depicted by placing sub-events immediately below their super-events. As it can be seen, events 4 and 9 do not have any sub- or super-events; event 3 has 1 and 2 as sub-events, overlapping with each other; event 8 has 5, 6 and 7 as sub-events, also partially overlapping.

A narration n is a triple $\mathsf{n} = \langle s, k, \sigma \rangle$ consisting of:

1. A text $s \in S^\star$ giving the *content* of the narration, of length $|s|$.
2. A positive integer k giving the *depth* of the narration, that is the maximum number of *levels* in which the narration is structured. For instance, a narration structured in books and chapters has depth 3: level 1 is the level of the entire narration, level 2 is the level of books, and level 3 is the level of chapters. A narration that has no structure has depth 1. Note that depth is defined as a maximum, in order to capture the idea that not all levels need to be populated, *e.g.*, not all chapters need to have sections: it is sufficient that one chapter has a section to have depth 3.
3. A function σ giving the *structure* of the narration. σ has the first k positive integers $\{1, 2, \ldots, k\}$ as domain and sets of intervals in $[1, |s|]$ as range. Each interval $[i, j]$ in the range of σ is called a *structural unit,* or simply unit, and its *content* is the sub-string of s from the i-th to the j-th character. $\sigma(1)$ is always the set containing only the unit $[1, |s|]$, since the first level is the level of the entire narration. For $2 \leq j \leq k$, $\sigma(j)$ is a set of pairwise disjoint intervals, each one contained in one interval i of the previous level $\sigma(j-1)$ and giving the subdivision of i at the level j. Figure 1 right gives an example of the structure of a narration consisting of a text s of two thousand characters ($|s| = 2000$), divided in two chapters, one of fifteen hundred characters, the other of five hundred characters. Each chapter has one section partially covering its content.

This model of narration is kept simple to illustrate the concept for narrations with an acyclic structure, such as books. The model is not adequate to deal with narrations with possibly cyclic structures such as hypertexts. However, this is no real limitation, as it is always possible to capture arbitrary structures using more sophisticated models. Indeed, all the structures that can be used in a narration are expression of some grammar, therefore they can always be captured by a formal structure defined in set-theoretic terms.

Finally, we model the reference function. Given a fabula f and a narration n, a *reference function between* f and n, $\mathsf{ref}(\mathsf{f}, \mathsf{n})$, is a pair (F_n, r) where:

- F_n, the *fragmentation* of ref, is a set of intervals called *fragments,* each of which is contained in a unit of n, called the *source* of the fragment. Each fragment identifies the

Fragment f	$r(f)$
$[1, 500]$	4
$[600, 900]$	3
$[600, 800]$	1
$[820, 890]$	2
$[1250, 1850]$	8
$[1250, 1440]$	5
$[1501, 1700]$	6
$[1710, 1870]$	7
$[1950, 2000]$	9

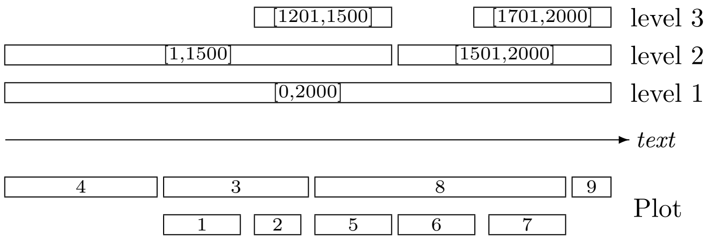


Figure 2 A reference function (left) and the resulting correspondence between narration structure (top right) and plot (bottom right).

portion of the narration that narrates an event of the fabula, and has as content the sub-string of the source's content delimited by the fragment.

- r, the *event association* of ref, is an injective function assigning to each fragment f in F_{n} an event $e = r(f)$ that is one of the events of the fabula f, that is $r(f) \in E$.

The above definition is meant to leave maximum freedom in constructing the plot of the narration. In particular:

- Fragments can be derived from any unit of the narration, not only from those that belong to the highest level.
- Fragments can be freely chosen, allowing them to arbitrarily overlap. Therefore the injectivity of the event association, which imposes that two fragments may not narrate the same event, does not represent a limitation to the creativity of narrators: a piece of text may narrate two or more events simultaneously.
- Similarly, we do not impose the event association to be surjective, so that each event in the fabula is associated to some fragment of the narration, leaving to narrators the possibility of omitting the narration of some events.
- Finally, we do not impose the narration of a sub-event to be a part (technically, a sub-string) of the narration of the super-event. This condition may well apply to history texts, in which, *e.g.*, the narration of the battle of Ludford is part of the narration of the War of the Roses. But it does not necessarily apply to other kinds of narrations, therefore it is not included in our model.

In this way, the plot of the narration can be displayed on a line, similarly to the fabula, except that in the fabula the line represents the flow of time, while in the narration it represents the sequence of characters that constitutes the content of the narration. Figure 2 illustrates this similarity between fabula and plot. The left-hand side of the Figure gives the reference function between the fabula and the narration presented in the previous examples (see Figure 1). The right-hand side shows the narration content against which both the plot (bottom) and the structure of the narration (top) are displayed.

Tying things up, we define a narrative N as a $(k+1)$-tuple, $k \geq 1$, $N = \langle \mathsf{f}, (\mathsf{n}_1, \mathsf{ref}_1), \ldots, (\mathsf{n}_k, \mathsf{ref}_k)\rangle$ where f is a fabula, and each pair $(\mathsf{n}_i, \mathsf{ref}_i)$, $1 \leq i \leq k$, consists of a narration n_i and a reference function between the fabula f and the narration n_i. This definition directly reflects the concept of narrative as spelled out in the conceptualization, that is as a fabula endowed with one or more narrations, each related to the fabula by a reference function.

5 Future Developments

In order to create an ontology for narratives, the above mathematical specification must be expressed in a logical language, providing (a) names for the notions in the specification, and (b) more importantly, axioms for encoding the various conditions on these notions.

We use the Semantic Web language OWL (Web Ontology Language)[1] for encoding narratives. We focus on this language for technical interoperability reasons, looking at Linked Data and at the Web as the ideal medium and the ideal infrastructure, respectively, for producing and consuming narratives.

In addition, we aim at semantic interoperability, based on the sharing of ontologies. Semantic interoperability is a two-way concept: on the one hand, we aim at widening the usage of our ontology for narratives by making it re-usable; on the other, we aim at re-using existing ontologies as much as possible in developing our own. A natural candidate of this latter category is the CIDOC CRM ontology [9], an ISO standard largely employed in the digital library domain. The CRM aims to be monotonic in the sense of Domain Theory. That is, the existing CRM constructs and the deductions made from them must always remain valid and well-formed, even as new constructs are added by extensions to the CRM. The CRM includes temporal entities for capturing time-dependent concepts such as events; moreover, its harmonization with the FRBR ontology, known as FRBRoo [10] provides fundamental notions for the modelling of text, such as expressions and expression fragments. Because the CRM's primary role is the meaningful integration of information in an Open World, it seems natural to embed our narrative ontology in the CRM, by introducing the required extensions into the current expression of this ontology. Furthermore, we have already started the validation of our model by partially expressing it in the CRM and by using it to formally represent the biography of Dante Alighieri as case study. Our representation of Dante's life is derived from a biography of the poet written by an authoritative Italian biographer of Dante, who collaborated with us constructing a narrative.

In order to support the construction of this narrative, we implemented a semi-automatic tool that allowed the biographer/narrator to define the events of the biography of Dante, and to connect these events to each other based on their temporal, mereological or causal relations, through a simple GUI. The objects that contextualize the events, *e.g.*, people, places, times, things, are automatically extracted from the Wikidata knowledge base[2], as well as manually added by the narrator. The resulting knowledge base is expressed in OWL. In order to allow the biographer to evaluate the created narrative, we included in our tool a visualization component that allows visualizing the narrative on a timeline. We then asked the biographer to evaluate the ability of the ontology to capture in a formal way the main aspects of the narrative. After the analysis of the events and their components shown on the timeline, the scholar expressed a positive evaluation confirming that the ontology was able to represent the events of the narrative, their relations, and their components, as described in his text.

Due to the encouraging results of this first experiment, we plan to make the tool available to a community of scholars in the context of an Italian national research project[3], in order to perform a larger scale evaluation.

---

1 https://www.w3.org/TR/owl-features/
2 http://wikidata.org
3 http://perunaenciclopediadantescadigitale.eu

6 Conclusions

We have presented a conceptualization of narratives based on fundamental notions in narratology, and a first mathematical expression of it, to be used as a basis for the development of an ontology of narratives, encoded in OWL. Our model of narratives includes three dimensions: the fabula, the narration and the connection between them via a reference function, through which it is possible to derive the plot of the narrative. A validation of the model is ongoing. Indeed, using a CRM expression of the model, we have formally represented a narrative of the biography of Dante Alighieri. The fabula of this narrative is given by the main events in Dante's life reconstructed by an authoritative scholar from various primary sources. The narration of the fabula consists of the text written by the scholar and the reference function connects each portion of text that narrates an event to the narrated event. The validation has given positive results so far. We plan to conclude our study developing an ontology of narratives with an associated tool for building, visualizing, managing and sharing narratives.

Acknowledgements. We wish to thank Mirko Tavoni for his patience in introducing the computer scientist to the Humanities, and Giuseppe Indizio for evaluating our model.

References

1 James F. Allen. Towards a general theory of action and time. *Artificial intelligence*, 23(2):123–154, 1984.
2 Aristotele. *Poetica*. Laterza, 1998.
3 Mieke Bal. *Narratology: Introduction to the theory of narrative*. University of Toronto Press, 1997.
4 Marc Cavazza and David Pizzi. Narratology for interactive storytelling: A critical introduction. In *Technologies for Interactive Digital Storytelling and Entertainment*, pages 72–83. Springer, 2006.
5 Seymour Chatman. Characters and narrators: Filter, center, slant, and interest-focus. *Poetics Today*, 7(2):189–204, 1986.
6 Chris Crawford. *Chris Crawford on interactive storytelling*. New Riders, 2012.
7 B. A. Davey and H. A. Priestley. *Introduction to lattices and order*, chapter 3. Cambridge University Press, second edition, 2002.
8 Donald Davidson. *Essays on actions and events: Philosophical essays*, volume 1. Oxford University Press, 2001.
9 Martin Doerr. The cidoc conceptual reference module: An ontological approach to semantic interoperability of metadata. *AI Mag.*, 24(3):75–92, September 2003.
10 Martin Doerr, Chryssoula Bekiari, Patrick LeBoeuf, and Bibliothèque nationale de France. Frbroo, a conceptual model for performing arts. In *2008 Annual Conference of CIDOC*, pages 6–18, 2008.
11 Marcello Frixione and Antonio Lieto. Representing concepts in formal ontologies. compositionality vs. typicality effects. *Logic and Logical Philosophy*, 21(4):391–414, 2012.
12 Gérard E. Genette and Jane E. Lewin. *Narrative discourse: An essay in method*. Cornell University Press, 1983.
13 Algirdas Julien Greimas, Daniele McDowell, and Alan R. Velie. *Structural semantics: An attempt at a method*. University of Nebraska Press, 1983.
14 David Herman. Narratology as a cognitive science. *Image and Narrative*, 1(1), 2000.
15 Robert Kowalski and Marek Sergot. A logic-based calculus of events. In *Foundations of knowledge base management*, pages 23–55. Springer, 1989.

16 Antonio Lieto and Rossana Damiano. A hybrid representational proposal for narrative concepts: A case study on character roles. In *OASIcs – OpenAccess Series in Informatics*, volume 41. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2014.

17 Inderjeet Mani. Computational modeling of narrative. *Synthesis Lectures on Human Language Technologies*, 5(3):1–142, 2012.

18 Inderjeet Mani. Computational narratology. *Handbook of narratology*, pages 84–92, 2014.

19 Jan Christoph Meister. Narratology, 2012. Synthesis. American Museum of Natural History. Available at `http://ncep.amnh.org`.

20 Rob Miller and Murray Shanahan. Some alternative formulations of the event calculus. In *Computational logic: logic programming and beyond*, pages 452–490. Springer, 2002.

21 Erik T. Mueller. *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann, 2014.

22 Vladimir Propp. *Morphology of the Folktale*, volume 9. University of Texas Press, 1973.

23 Ryan Shaw, Raphael Troncy, and Lynda Hardman. Lode: Linking open descriptions of events. In *The Semantic Web*, pages 153–167. Springer, 2009.

24 Victor Shklovsky. Art as technique. *Russian formalist criticism: Four essays*, 3, 1965.

25 Tzvetan Todorov. *Grammaire du décaméron*. Mouton, The Hague, 1969.

Exploring “Letters from the Future” by Visualizing Narrative Structure

Sytske Wiegersma[1], Anneke M. Sools[2], and Bernard P. Veldkamp[3]

1 **Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, The Netherlands**
`s.wiegersma@utwente.nl`
2 **Department of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands**
`a.m.sools@utwente.nl`
3 **Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, The Netherlands**
`b.p.veldkamp@utwente.nl`

Abstract

The growing supply of online mental health tools, platforms and treatments results in an enormous quantity of digital narrative data to be structured, analysed and interpreted. Natural Language Processing is very suitable to automatically extract textual and structural features from narratives. Visualizing these features can help to explore patterns and shifts in text content and structure. In this study, streamgraphs are developed for different types of “Letters from the Future”, an online mental health promotion instrument. The visualizations show differences between as well as within the different letter types, providing directions for future research in both the visualization of narrative structure and in the field of narrative psychology. The method presented here is not limited to “Letters from the Future”, the current object of study, but can in fact be used to explore any digital or digitalized textual source, like books, speech transcripts or email conversations.

1998 ACM Subject Classification I.2.7 Natural Language Processing

Keywords and phrases Text visualization, NLP, Narrative structure

Digital Object Identifier 10.4230/OASIcs.CMN.2016.5

1 Introduction

In this paper “Letters from the Future”, a narrative-based instrument used by [29] to study the human capacity to imagine the future, is studied using a combination of quantitative analysis, Natural Language Processing and text visualization methods. Traditionally, in narrative psychology, qualitative methods for analysing narrative content and structure are predominantly based on hand-coded data. The underlying structure of a narrative is represented for example by defining clusters or counting word frequencies. A widely used approach in narrative studies is the componential analysis, which focusses on identifying and examining the structural elements that narratives consist of. The narrative framework of [18], who originally divided narratives into five structural units (orientation, complication, resolution, evaluation, and coda), is a prime example of the componential approach.

The many features and feature combinations that can potentially be extracted from narratives can quickly result in an overwhelming quantity of data to be processed and interpreted. In addition, as a consequence of the growing popularity of e-mental health

© Sytske Wiegersma, Anneke M. Sools, and Bernard P. Veldkamp;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 5; pp. 5:1–5:18
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

interventions, more and more digital narrative data becomes available for analysis. Processing and interpreting all this data by hand is a tremendous, if not impossible, task. However, the growing availability of digital narrative data also generates new opportunities. There now is sufficient narrative data available to scale up the study of narratives by applying Natural Language Processing (NLP) methods. In NLP computers are used to process and manipulate natural language [9], [19], “natural language” being any spoken or written language used by humans in everyday life [2]. The main benefits of NLP by computers over the manual processing of narratives is that it is far less time consuming and less error prone than human coders. Moreover NLP enables researchers to process and compare large datasets or very detailed textual data.

A recent systematic literature review on text mining applications in psychiatry [1] showed that the use of NLP and text mining methods is still in its infancy in the fields of psychology and psychiatry. NLP applications have also only recently found their ways in the field of humanities. From the humanities perspective, Computational Narratology [21] can be described as a methodological instrument to develop narratological theories, enabling researchers to extend and test their models on larger text corpora and to specify and apply concepts and models automatically and thus more consistently [24]. As described by [21], in computational narratology narratives and narrative structures are explored using computation and information processing methods.

[5] state that an efficient approach to explore the underlying mathematical structure of narratives is text visualization. The mathematical structure is generally captured using first-, second- and third-order statistics like word frequencies, clustering and natural language algorithms [37]. Visually representing this structure enables researchers to reveal and interpret differences and relationships within and between text documents that would have been difficult, or even impossible, to identify solely from the texts or from tables of numerical data extracted from these texts [5], [37], [33]. Contrary to graphs, visualizations are generally used as an exploratory tool to explore and analyse the data and not to present study results to the public [33].

The current study is a response to the suggestion of [29] that more in-depth insight into how and why narrative futuring works can be gained by combining traditional qualitative with quantitative methods. The principal aim is to gain more detailed understanding of the differences in letter content, specifically the distribution (sequential order) and proportion of narrative processes and grammatical elements, both within and between the letter types. The new insights from this study can be used not only to confirm the previous findings of [30] but also to develop new theories and hypotheses regarding the human capacity to imagine the future.

This paper is organized as follows. First the object of study, the instrument “Letters from the Future”, is described in detail, followed by a description of the dataset and data-processing steps in the Methods section. In this section, the two essential topics in the development of text visualizations are addressed as well: first an existing NLP package used to extract the mathematical structure of the narratives is described. Second it is investigated how these structures can be visualized in such a way they can be used to study differences both within and between the different types of letters. In the Results section, the developed text visualizations are compared to the previous findings based on qualitative methods by [30] and linked to the existing narrative framework of [18]. Finally, conclusions and implications for future work are described in the Discussion section.

Table 1 Letter structure and characteristics.

	Imagining/experiencing a future situation Extended core with imaginative component, information on events, places, persons, experience	**Generic letter** No/limited imaginative components. Possibly global descriptions of future situations at end of letter
Retrospective evaluation Look back from future or present to past	*Type 1* *Imagining and evaluating the futured past* Structure: - Narrative imagination of desired future situation (present tense) - Anticipated reminiscence of the future past (past tense) - Conclusion/insight from evaluated experiences and/or - Worldly wisdom (self-praising remarks) - Comments on implications for the future (moral advice/future promises)	*Type 4* *Reminiscing and evaluating the past without imagination of the future* Structure: Equal to structure of type 1. Recounted/ evaluated period in past instead of futured past, presented as current concern taking place before moment of writing.
Prospective orientation Look forward from present to future/ from future even further ahead	*Type 2* *Imagining and orienting to the futured present and futured past* Structure: - Statement about present position in life (present/past perfect tense) - Imaginary goals/purposes - Description of how to realize these objectives	*Type 5* *Intentional orientation with expression of emotions* No clear structure: No clear action orientation or path from present to future. Sometimes written from future instead of present. Much use of intentional time (hope/wish), future tense (shall/will) and hesitation.
Present-oriented Focus on moment in time (present/ future present) instead of period	*Type 3* *Expressive imagining of the futured present* No clear structure: No orientation/evaluation or path from present to future. Sometimes conclusions are drawn. Contains sensory details (hopes, wishes, gratitude and self-appraising). Imagined future described mainly in present-tense.	*Type 6* *Advisory letters about current practical and moral concerns* No clear structure: Consists mainly of general insights/ conclusions, generic (existential/moral) advices or worldly wisdoms. No path to origination of conclusions or insights.

1.1 Letters from the Future

In this study, computational narratology is used to explore "Letters from the Future", an online narrative-based mental health promotion instrument developed by [29]. The instrument is adapted from an earlier exercise by [4], in which storytelling groups are used to enhance mental health. Using a web-based tool, participants are asked to write a letter from a particular situation and moment in the future to someone in the present. [30] studied the human capacity to imagine the future by hand-coding narrative processes within each individual letter on sentence-level. They clustered these narrative processes into five

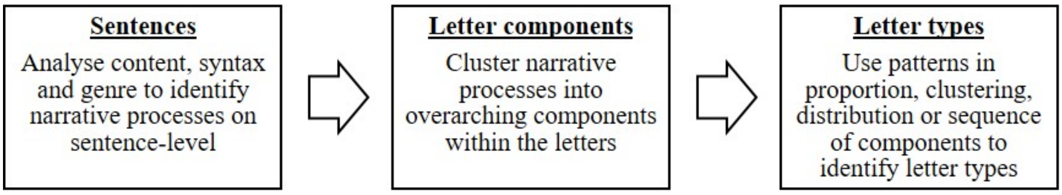


Figure 1 Schematic overview of procedure.

overarching components which were then used to identify six different letter types (see overview in Figure 1).

The letter types were defined based on a comparative analysis of the following elements: 1) the dominant narrative process (imagining, evaluating, orienting, expressing emotions or engaging in dialogue); 2) the use of certain grammatical elements like past, past imperfect, present and future tense, modals ("would", "could", "should"), intentional time ("hope", "wish", "want"), or the imperative ("go!", "remember!"); 3) the presence and clearness of a path between present and future; and 4) the level of detail of the imagination. Table 1 gives an overview of the six letter types and the corresponding structures found by [30].

As shown in Table 1, [30] found a clear distribution and sequence of narrative processes and grammatical elements for half of the letters (letter types one, two and four). However, these structures are not always uniformly applicable to all letters of the corresponding type. For example, type one letters generally consist of five elements, but the order of these elements can differ; letters can start either with narrative imagination of the desired future, anticipated reminiscence of the future past or an evaluative part preceding narrative imagination. The same goes for letters of type two, about which [30] write: "The orienting function could be prominent from the first sentence, in letters starting with goal-setting or value orienting phrases rather than with a situation (but the order could be reversed as well)." (p. 19). Another remark on type two letters was the finding that hope, a prominent feature in those letters, could occur either at the beginning or end of a letter.

In the following section the dataset is described in more detail, followed by a description of the methods used to pre-process the dataset in order to capture and visualize the narrative structure and content of the different letter types.

2 Methods

2.1 Dataset

An existing dataset of 492 letters collected for a previous study by the Storylab, the Dutch expert centre for narrative psychology and mental health promotion at the University of Twente, was used (see [29], [30] for more information on the data collection process). Informed consent to re-use these letters for on-going research by the Storylab was obtained. The letters were written by a relatively diverse, mainly Dutch (70%) and German (27%) participants. The letters were manually categorized into six categories by three independent raters (interrater reliability score = 0.672). Table 2 shows an overview of the number of letters and mean text length per category. In the current study only Dutch letters that were clearly categorized in one of the six letter types are used, resulting in a dataset of 351 letters.

Table 2 Dataset characteristics.

	Imagination letter	**Generic letter**
Retrospective evaluation	Type 1: N(letters) = 137 Mean N(words/text) = 324	Type 4: N(letters) = 19 Mean N(words) = 292
Prospective orientation	Type 2: N(letters) = 47 Mean N(words) = 303	Type 5: N(letters) = 9 Mean N(words) = 196
Present-oriented	Type 3: N(letters) = 94 Mean N(words) = 289	Type 6: N(letters) = 45 Mean N(words) = 270

2.2 Pre-processing

Salutations, recipient and sender names, location and dates at the beginning and end of the letters are removed. This is done because these elements are considered non-informative and may cause difficulties when splitting and concatenating the letters into segments, distorting the results of subsequent analyses and visualizations. After that the narratives are split into equally sized segments, for which word frequencies can be plotted along the horizontal axis. In a previous study by [10], document streamgraphs were created for the book "Tom Sawyer" by splitting the text into ten segments. Although using ten segments is suitable for long text documents like books, the narratives used in the current study are much shorter (see Table 2 for mean number of words per letter type). Therefore a smaller number of segments may be more appropriate. To decide on the number of segments to use, three different splits were made and the resulting visualizations were compared.

First, following [10], the narratives were split into ten segments, which resulted in very dynamic and detailed visualizations. However, these results were too fine-grained, making it difficult to use the visualizations for their initial purpose; to confirm previous findings and develop new hypotheses. Second, the narratives were split into three segments (representing the beginning, middle and end of the story, a structure often used in the formation and analysis of narratives, [14]). It was expected that the three segments would result in more interpretable visualizations revealing major trends. The resulting visualizations were however very global and flat, making it difficult to draw conclusions or gain new insights. Therefore third, based on the framework of [18], widely used to represent narrative information and analyse personal narratives, the narratives were split into five segments: orientation, complication, resolution, evaluation, and coda.

Although five segments may still seem too fine-grained for short narratives like the letters used in the current study, the 'narrative clause' used by [18]) as the basic unit of narrative can be as short as one sentence. This framework therefore is very suitable (and widely used) for analysing short narratives like daily life stories or therapeutic interviews [17]. In addition, splitting the narratives into five segments is in line with the five narrative processes used by [30] to identify the different letter types and letter structures. The five segments resulted into well-interpretable visualizations, showing the same trends as the visualizations for ten segments but then for larger-grained sections more inherent in personal letters.

Since the aim is to develop visualizations per letter type, for each type the letters are split into five equal segments and concatenated in one new text file per segment. This results in five new text files for each letter type, as shown in Figure 2. The five segments are analysed and visualized for each letter type separately.

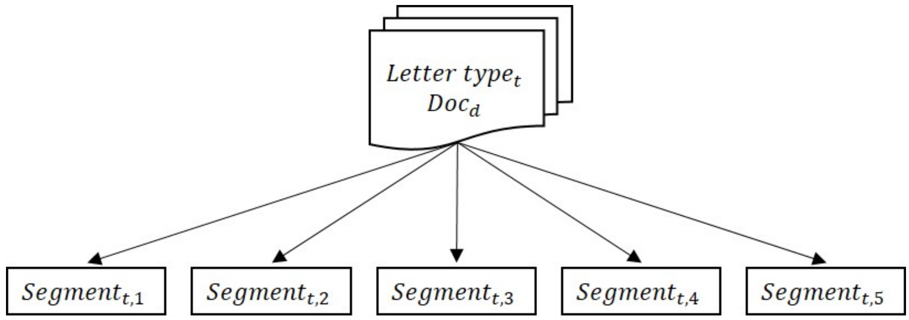


Figure 2 Splitting text documents into segments for each letter type.

2.3 Mathematical structure

To explore the differences in letter content and structure, plotting word-frequencies within each text segment for each letter type seems appropriate. However, since plotting frequencies for all used words will probably not lead to legible and interpretable visualizations, generally a sub selection of the occurring words is included in the visualizations. [10] for example only used words starting with capital letters or only the most prominent words as series in his graph. Another way to reduce the number of series is by categorizing them into word classes, as [36] did. In the current study words are categorized hierarchically using the text analysis program Linguistic Inquiry and Word Count (LIWC, [25]). LIWC is a structured, knowledge-rich method, relying on tight structures from existing software and dictionaries. LIWC processes texts on word level, comparing each word to a dictionary files for each category. It is a validated, ready-to use efficient and effective method to study a range of cognitive, emotional and structural components in spoken and written narratives [26].

In order to process Dutch texts, the Dutch LIWC dictionary developed by [38] was used. Contrary to the more complete English dictionary, the Dutch dictionary contains variables for the grammatical tenses past, present and future, but not for modals, intentional time or the imperative. The Dutch dictionary is based on the English LIWC dictionary (2001 version) and, as shown in Table 3, consists of 66 word categories divided over five dimensions. The words can be assigned to one or more categories, scoring the occurrences as percentages. The 66 LIWC categories are organized into a hierarchy of eleven main categories and 55 subcategories, which, when applied to the range of letter segments, results in a set of hierarchical additive time series.

2.4 Analytical procedure

As stated earlier, in this paper visualizations are used to explore differences in letter content both within the letters of the same type as between different types of letters. Two separate analyses were used to find the most informative categories. First, to find which categories best visualize the differences in category occurrence *between* the letter types a one-way analysis of variance (ANOVA) is used. The ANOVA is used to determine if there are significant differences between the means of multiple groups [22]. The mean category occurrence is calculated for each letter type by summing up the category scores for all segments and then dividing the sum by five (the total number of segments). The mean occurrences were compared using Welch's statistic [31].

Second, to find which LIWC categories fluctuate the most *within* each letter type, the spread in category occurrence values for the segments was evaluated. The most commonly

Table 3 Categories Dutch LIWC dictionary (translated from [38]).

I. Linguistic processes
1. Pronouns (I, them, our)
2. 1st person singular (I, me, mine)
3. 1st person plural (we, our, us)
4. Total 1st person (I, we, me)
5. Total 2nd person (you, your, thou)
6. Total 3rd person (they, their, she)
7. Negations (no, not, never)
8. Assent (agree, ok, yes)
9. Articles (a, an, the)
10. Prepositions (to, with, above)
11. Numbers (second, thousand)
II. Psychological processes
12. Emotional (happy, sad, down)
13. Positive emotions (happy, pleased)
14. Positive feelings (fun, love, smile)
15. Optimism (proud, passionate)
16. Negative emotions (hurt, hostile)
17. Anxiety (nervous, fearful, worried)
18. Anger (hate, annoyed, threat)
19. Sadness (grief, disappointment)
20. Cognitive (cause, know, ought)
21. Causation (because, effect, hence)
22. Insight (think, know, consider)
23. Discrepancy (should, would, could)
24. Inhibition (block, constrain, stop)
25. Tentative (maybe, perhaps, guess)
26. Certainty (always, never)
27. Perceptual (observe, heard, feeling)
28. See (view, saw, seen)
29. Hear (listen, hearing)
30. Feel (feel, touch)
31. Social (share, talk, help)
32. Communication (interview, rumour)
33. Other references (we, them, they)
34. Friends (buddy, friend, neighbour)
35. Family (daughter, husband)
36. Humans (adult, baby, boy)
III. Relativity
37. Time (end, until, season)
38. Verbs in past tense (went, ran)
39. Verbs in present tense (is, does)
40. Verbs in future tense (will, going)
41. Space (nearby, place, North)
42. Up (above, higher, top)
43. Down (deeper, lower, bottom)
44. Including (and, inclusive, too)
45. Excluding (unless, except, out)
46. Motion (approach, walk, climb)
IV. Personal concerns
47. Occupation (achieve, promote)
48. School (student, exam)
49. Work (job, career, colleague)
50. Achievement (earn, hero, win)
51. Leisure (cook, bike, movie)
52. Home (kitchen, home, garden)
53. Sports (game, fitness, work-out)
54. Television (film, video, tv)
55. Music (sing, song, guitar)
56. Money (profit, cash, owe)
57. Metaphysical (altar, church)
58. Religion (pray, honour, bless)
59. Death (bury, mourn, kill)
60. Physical (ill, faint, appetite)
61. Body (vital, thirsty, cramp)
62. Sexual (flirt, love, kiss)
63. Ingestion (drink, hungry, dish)
64. Sleep (dream, wake, sleepy)
65. Groom (shower, make-up)
V. Experimental dimensions
66. Swear words

used measure of spread in a set of values is the standard deviation (SD). As low SD values indicate that all data points are close to the mean, LIWC categories with low SD values can be presumed to show little to no fluctuation in occurrence within the letter. LIWC categories with high SD values can be presumed to be highly fluctuating and thus showing more differences in occurrence within the concerning letter type. Since there are big differences in the means of the occurrence categories, to be able to compare the variation each SD is normalized with respect to its mean by: SD/Mean. The resulting value is known as the coefficient of variation (CV, also known as relative standard deviation), which shows the amount of variability in relation to the mean [20]. A major limitation of the CV is that when the mean is very small, a small variation in the dataset will already result in a large CV value [6]. Therefore the LIWC categories with Mean < 1 were excluded. Then for each letter type the ten most fluctuating LIWC categories (thus the ten categories with the highest CV scores) were selected and included in the letter type specific visualizations.

2.5 Text visualization design

Time series analysis, the study of changes in variables over time, can focus on one given variable, or the change of a specific variable compared to others over a certain time period. The time series consists of a sequence of measurements over a continuous, equal distanced time interval [28]. Time series are often visualized using simple line graphs, which works well when comparing a small number of series since the line graph shows direct values for each series at each time point. Another way to visualize time series are stacked graphs, where the series, represented by coloured layers, are stacked on top of each other, showing not only the individual values for each layer but also the total value at certain time points on the horizontal axis [7]. Stacked graphs are very useful to visualize hierarchical time series. However, both line and stacked graphs become illegible when using a large number of series [7], [10]. To overcome this problem [13] created ThemeRiver, a smooth, continuous graph stacked symmetrically around the x-axis, which is situated at the centre of the graph instead of at the bottom. ThemeRivers later became known as Streamgraphs, thanks to a popular visualization in the New York Times by [11]. Streamgraphs differ mainly from ThemeRivers in the design and layout decisions (like colour, interaction or geometry) made to make the graph visually attractive and more organic. Although originally applied to music, movies [3] and (baby) names [34], streamgraphs have also been applied to text documents and seem very suitable for the visualization of narratives.

The smooth, continuous lines that distinguish between the layers are the main advantage of a streamgraph, since this visualizes the data in an intuitive and easily interpretable way [13], [7]. Continuous data is required to generate such smooth, curving lines. However, splitting the texts into five separate segments results in a discrete dataset with different values for y at the data points x_1, x_2, $\ldots$, x_5. This problem is solved by interpolating between the discrete data points as suggested by [13]. By using interpolation, intermediate values between data points are estimated from the neighbouring data points [12]. This results in smooth, continuous lines connecting the discrete data points [23]. There are different interpolation methods, of which the Cubic Splines model based on third degree polynomials results in the smoothest curve fits [12] and is therefore used in the current study.

There are two final notes with regard to the graph design. First, for all visualizations counts that since the layers are stacked symmetrically at the centre of the graph, the values on the y-axis are of no added value and are therefore not included in the plots, as in [13] and [7]. Second, sequential colour palettes were used to visualize the hierarchical structure of the time series, as was done by [35]. Each main category is assigned its own colour with a range colours with slightly different shades to reflect the corresponding subcategories.

3 Results

In this section first the results of the quantitative analysis are described, followed by the resulting visualizations.

3.1 Selected LIWC categories

To determine for which LIWC categories there are significant differences in mean occurrence between the different letter types, a one-way ANOVA was used. For 31 of the 66 LIWC categories (indicated by an asterisk (*) in Table 4) significant differences between the means ($p \leq 0.05$) were found. The coefficient of variation (CV) was used to measure the amount of variability in category occurrences throughout the letters. Table 4 contains the mean proportion (in %) and standard deviation of each LIWC category over all segments. For each letter type, the values for the most fluctuating categories are printed in bold. These categories are selected for the visualizations in Figures 4, 5 and 6.

Table 4 Means and standard deviations for each letter type

LIWC category	Type 1 Mean (SD)	Type 2 Mean (SD)	Type 3 Mean (SD)	Type 4 Mean (SD)	Type 5 Mean (SD)	Type 6 Mean (SD)
1. Pronoun*	12.75 (0.66)	11.63 (1.14)	10.56 (0.47)	13.41 (1.11)	13.67 (2.44)	13.19 (0.82)
2. I*	5.07 (0.80)	**3.42 (1.14)**	4.84 (0.62)	6.60 (0.36)	6.95 (1.93)	**3.55 (0.69)**
3. We*	0.46 (0.07)	0.56 (0.15)	0.66 (0.18)	0.43 (0.29)	0.45 (0.73)	0.12 (0.11)
4. Self*	5.54 (0.81)	**3.99 (1.00)**	5.50 (0.71)	7.04 (0.58)	7.40 (1.32)	**3.67 (0.74)**
5. You*	5.50 (0.75)	5.93 (0.76)	**3.11 (0.84)**	4.69 (0.84)	4.01 (1.16)	8.05 (0.30)
6. Other	0.67 (0.19)	0.78 (0.26)	**1.02 (0.29)**	0.65 (0.32)	0.96 (0.51)	0.46 (0.15)
7. Negation	**1.42 (0.41)**	**1.27 (0.35)**	**1.52 (0.38)**	1.51 (0.28)	**1.42 (0.53)**	2.09 (0.31)
8. Assent	0.17 (0.05)	0.14 (0.09)	0.15 (0.06)	0.13 (0.10)	0.00 (0.00)	0.26 (0.20)
9. Article	7.37 (0.76)	7.68 (0.74)	8.21 (0.91)	7.27 (0.94)	7.18 (0.92)	7.16 (0.51)
10. Prepos.*	11.03 (0.72)	11.40 (0.76)	11.13 (0.08)	10.54 (0.95)	10.96 (2.10)	10.24 (0.44)
11. Number	**1.24 (0.60)**	**1.15 (0.63)**	**1.21 (0.58)**	**1.26 (0.64)**	**1.25 (0.82)**	0.83 (0.51)
12. Affect	4.20 (0.66)	3.72 (0.34)	3.90 (0.63)	4.02 (0.94)	4.64 (1.43)	**4.18 (0.76)**
13. Pos. emo.	**2.94 (0.62)**	**2.62 (0.49)**	**2.89 (0.50)**	**2.54 (0.64)**	3.84 (1.18)	**2.67 (0.81)**
14. Pos. feel.	0.77 (0.27)	0.55 (0.17)	0.83 (0.23)	0.52 (0.21)	0.73 (0.47)	0.56 (0.14)
15. Optimism*	0.55 (0.17)	0.56 (0.17)	0.48 (0.17)	0.50 (0.21)	**1.64 (0.54)**	0.51 (0.23)
16. Neg. emo.*	1.15 (0.09)	0.98 (0.16)	0.92 (0.19)	**1.44 (0.34)**	0.73 (0.33)	1.45 (0.14)
17. Anxiety	0.21 (0.07)	0.22 (0.11)	0.11 (0.06)	0.25 (0.12)	0.00 (0.00)	0.27 (0.10)
18. Anger	0.11 (0.03)	0.13 (0.07)	0.10 (0.04)	0.05 (0.08)	0.06 (0.13)	0.21 (0.12)
19. Sadness*	0.28 (0.07)	0.15 (0.04)	0.34 (0.11)	0.45 (0.16)	0.28 (0.20)	0.33 (0.09)
20. Cognitive*	5.72 (0.86)	5.53 (0.28)	5.11 (0.71)	5.99 (1.25)	8.14 (1.39)	7.60 (0.34)
21. Causation	0.57 (0.13)	0.63 (0.07)	0.56 (0.14)	0.52 (0.21)	0.56 (0.34)	0.72 (0.10)
22. Insight	2.10 (0.29)	1.83 (0.18)	1.71 (0.28)	2.35 (0.41)	**2.04 (0.99)**	2.90 (0.24)
23. Discrep.*	**2.33 (0.47)**	2.46 (0.24)	**2.19 (0.37)**	**2.35 (0.65)**	5.26 (0.75)	2.99 (0.24)
24. Inhibition	0.06 (0.03)	0.03 (0.04)	0.06 (0.03)	0.07 (0.10)	0.00 (0.00)	0.08 (0.04)
25. Tentative	**1.50 (0.28)**	1.57 (0.15)	1.49 (0.17)	**1.71 (0.65)**	**2.65 (1.14)**	1.72 (0.18)
26. Certainty	1.58 (0.23)	1.20 (0.14)	**1.32 (0.23)**	1.62 (0.37)	**1.53 (0.62)**	**1.60 (0.46)**
27. Senses	1.27 (0.13)	1.23 (0.16)	1.27 (0.14)	**1.55 (0.39)**	0.68 (0.51)	**1.46 (0.35)**
28. See	0.48 (0.05)	0.39 (0.11)	0.52 (0.17)	0.41 (0.15)	0.28 (0.28)	0.53 (0.12)
29. Hear*	0.44 (0.06)	0.51 (0.15)	0.46 (0.09)	0.76 (0.15)	0.23 (0.37)	0.57 (0.25)
30. Feel	0.34 (0.07)	0.34 (0.09)	0.26 (0.10)	0.38 (0.20)	0.17 (0.25)	0.35 (0.08)
31. Social*	9.61 (1.12)	10.24 (0.65)	8.04 (0.57)	8.99 (0.95)	8.47 (1.52)	11.29 (0.12)
32. Comm.*	0.81 (0.13)	0.80 (0.11)	0.77 (0.12)	1.03 (0.21)	0.23 (0.24)	1.00 (0.11)
33. Others*	6.71 (0.88)	7.33 (0.61)	4.94 (0.42)	5.97 (0.88)	5.48 (1.50)	8.74 (0.19)
34. Friends*	0.24 (0.05)	0.24 (0.05)	0.20 (0.10)	0.14 (0.10)	0.39 (0.16)	0.24 (0.10)
35. Family*	0.89 (0.05)	0.80 (0.19)	0.80 (0.09)	0.88 (0.24)	1.02 (0.33)	0.46 (0.09)
36. Humans*	0.65 (0.16)	0.56 (0.18)	0.86 (0.10)	0.54 (0.25)	0.96 (0.43)	0.69 (0.17)
37. Time	**7.10 (1.23)**	**6.61 (1.97)**	6.82 (0.81)	7.13 (1.16)	**6.04 (2.25)**	**6.73 (1.30)**
38. Past*	**4.46 (0.83)**	**3.87 (0.97)**	**2.91 (0.52)**	5.83 (1.26)	0.96 (0.42)	**2.61 (0.93)**
39. Present*	12.61 (1.23)	12.57 (0.85)	13.44 (0.92)	12.18 (1.35)	14.12 (1.11)	15.05 (0.83)
40. Future*	0.93 (0.19)	**1.21 (0.20)**	0.83 (0.17)	0.90 (0.49)	2.71 (0.82)	1.42 (0.13)

Table 4 Means and standard deviations for each letter type (Continued)

LIWC category	**Type 1** Mean (SD)	**Type 2** Mean (SD)	**Type 3** Mean (SD)	**Type 4** Mean (SD)	**Type 5** Mean (SD)	**Type 6** Mean (SD)
41. Space	1.91 (0.28)	**1.95 (0.44)**	1.93 (0.28)	**1.62 (0.45)**	**2.03 (0.94)**	1.40 (0.20)
42. Up	1.11 (0.16)	0.96 (0.22)	1.21 (0.17)	**1.06 (0.35)**	0.73 (0.51)	0.97 (0.19)
43. Down	0.04 (0.03)	0.02 (0.02)	0.05 (0.05)	0.05 (0.05)	0.06 (0.13)	0.02 (0.02)
44. Incl.*	8.66 (0.21)	9.13 (0.84)	8.52 (0.19)	8.12 (0.54)	10.00 (0.63)	7.71 (0.77)
45. Excl.*	3.92 (0.54)	3.16 (0.24)	**3.64 (0.66)**	**4.15 (1.00)**	**4.01 (1.54)**	4.71 (0.51)
46. Motion	**1.87 (0.33)**	**2.15 (0.39)**	1.91 (0.19)	**2.04 (0.53)**	**1.75 (0.83)**	**1.98 (0.40)**
47. Occup.*	**2.04 (0.38)**	1.84 (0.27)	**1.33 (0.39)**	0.90 (0.22)	0.96 (0.71)	**1.60 (0.28)**
48. School*	0.76 (0.23)	0.72 (0.23)	0.38 (0.10)	0.40 (0.08)	0.40 (0.47)	0.72 (0.11)
49. Job*	**1.01 (0.26)**	0.91 (0.25)	0.75 (0.30)	0.45 (0.17)	0.51 (0.36)	0.49 (0.14)
50. Achieve*	0.32 (0.09)	0.25 (0.13)	0.22 (0.11)	0.11 (0.08)	0.11 (0.15)	0.41 (0.14)
51. Leisure*	0.59 (0.15)	0.91 (0.37)	0.87 (0.21)	0.95 (0.38)	0.73 (0.74)	0.29 (0.12)
52. Home*	0.49 (0.18)	0.73 (0.30)	0.68 (0.20)	0.45 (0.25)	0.45 (0.37)	0.22 (0.13)
53. Sports*	0.05 (0.02)	0.17 (0.07)	0.14 (0.07)	0.27 (0.14)	0.23 (0.37)	0.05 (0.04)
54. TV	0.03 (0.03)	0.01 (0.02)	0.01 (0.01)	0.07 (0.08)	0.06 (0.13)	0.00 (0.00)
55. Music	0.02 (0.02)	0.01 (0.03)	0.04 (0.03)	0.22 (0.14)	0.06 (0.13)	0.02 (0.02)
56. Money*	0.25 (0.12)	0.32 (0.12)	0.39 (0.12)	0.07 (0.08)	0.34 (0.12)	0.27 (0.05)
57. Metaphys.	0.07 (0.01)	0.06 (0.06)	0.08 (0.03)	0.13 (0.12)	0.00 (0.00)	0.04 (0.04)
58. Religion	0.04 (0.02)	0.04 (0.04)	0.07 (0.03)	0.07 (0.08)	0.00 (0.00)	0.04 (0.04)
59. Death	0.03 (0.02)	0.03 (0.03)	0.02 (0.01)	0.05 (0.08)	0.00 (0.00)	0.00 (0.00)
60. Physical	0.66 (0.10)	0.54 (0.14)	0.83 (0.09)	0.74 (0.56)	0.73 (0.32)	0.66 (0.14)
61. Body	0.30 (0.03)	0.24 (0.06)	0.33 (0.05)	0.34 (0.39)	0.34 (0.12)	0.40 (0.09)
62. Sexual	0.06 (0.03)	0.06 (0.03)	0.07 (0.03)	0.02 (0.04)	0.06 (0.13)	0.10 (0.08)
63. Eating	0.07 (0.02)	0.07 (0.02)	0.20 (0.11)	0.11 (0.12)	0.17 (0.16)	0.05 (0.03)
64. Sleep	0.24 (0.08)	0.19 (0.12)	0.23 (0.07)	0.32 (0.19)	0.23 (0.13)	0.14 (0.07)
65. Groom	0.00 (0.00)	0.00 (0.00)	0.03 (0.03)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
66. Swear**	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.02 (0.02)

Note:
Bold values: ten most fluctuating LIWC categories for each letter type.
* Significant differences ($p \leq 0.05$) between means of different letter types
** Only occurred in one letter type so means could not be compared

3.2 Visualizations

The figures below contain the streamgraphs for each letter type. The mean proportion of each LIWC category (over all the letters of the concerning letter types) is plotted per segment (s_1, s_2, ..., s_5) on the x-axis. The panel in Figure 3 contains six streamgraphs, one for each letter type. These visualizations show differences in occurrence proportions of LIWC categories throughout each letter type. All 66 LIWC categories are included in these graphs. The darkest shades of every colour show the main (overarching) categories, followed by the corresponding sub categories. The categories are plotted in the same order for each graph. These graphs can be used to find central themes within the letters and overall differences *between* the letters. In the legend, the asterisk (*) behind LIWC categories indicates that there are significant differences between the mean occurrences of the letter types for these categories. The visualizations in Figures 4, 5 and 6 show the ten most fluctuating LIWC categories for each letter type. These graphs can be used to find specific patterns and shifts in the occurrence proportions of LIWC categories *within* the letters.

The streamgraphs in Figure 3 show some clear similarities and differences between the six letter types. An interesting finding is that the visualizations for types 1–3 do not seem to differ as much as was expected based on the previous findings of [30]. Overall, the imagination

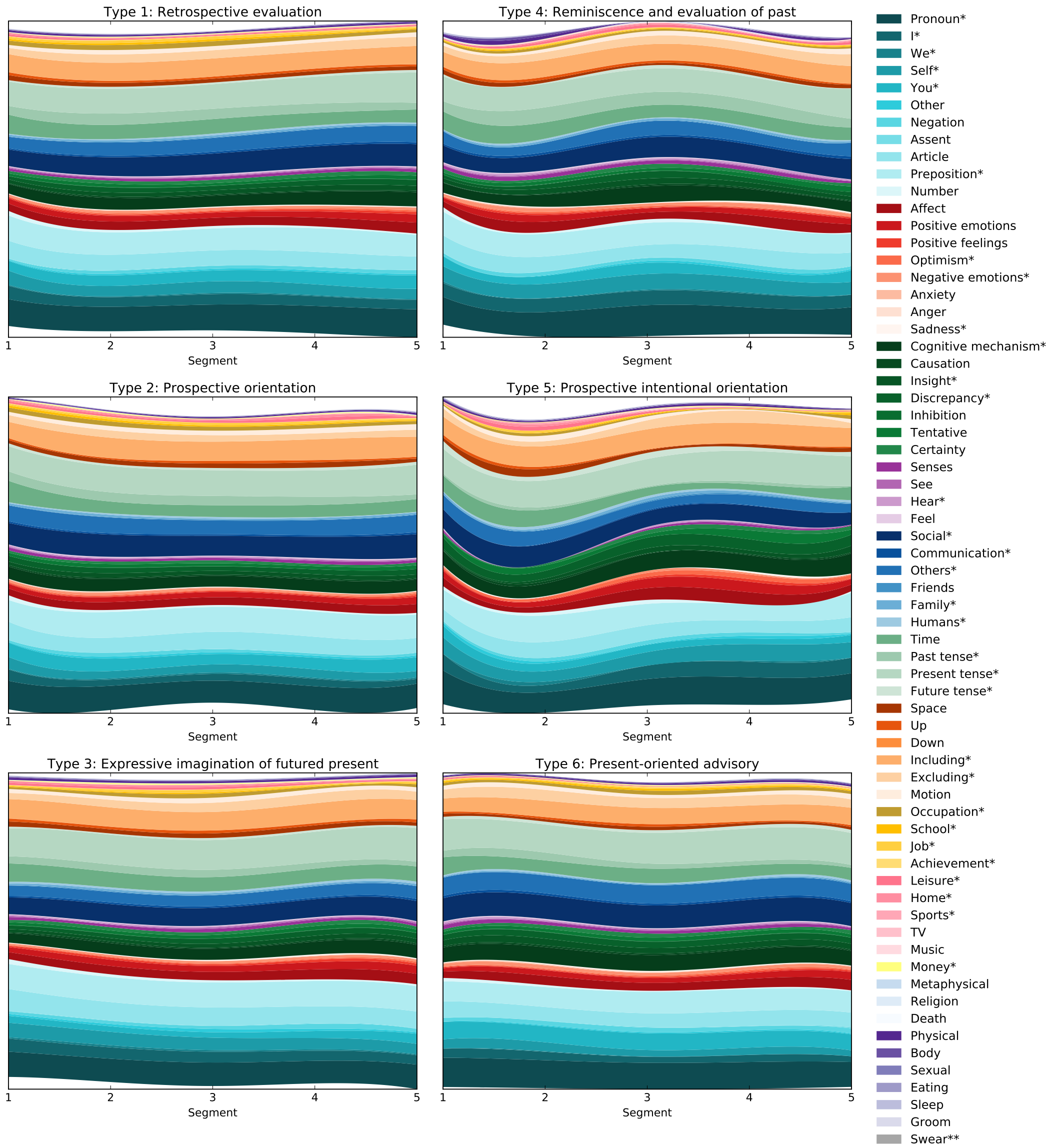


Figure 3 Overview LIWC categories per letter type.

letters (type 1–3) seem to have a calmer flow than the general letters (type 4–6), which show bigger differences in the proportions of the LIWC categories over the five segments. The differences between and within the streamgraphs will now be described in more detail and compared pairwise for the retrospective letters (types 1 and 4), prospective letters (types 2 and 5) and present-oriented letters (types 3 and 6).

3.2.1 Retrospective letters

[30] found that imagination and general retrospective letters generally have the same structure. This is also reflected by the streamgraphs in Figure 3, which show that for the majority of the LIWC categories the distribution of the category proportions over the segments is quite similar for both retrospective letters.

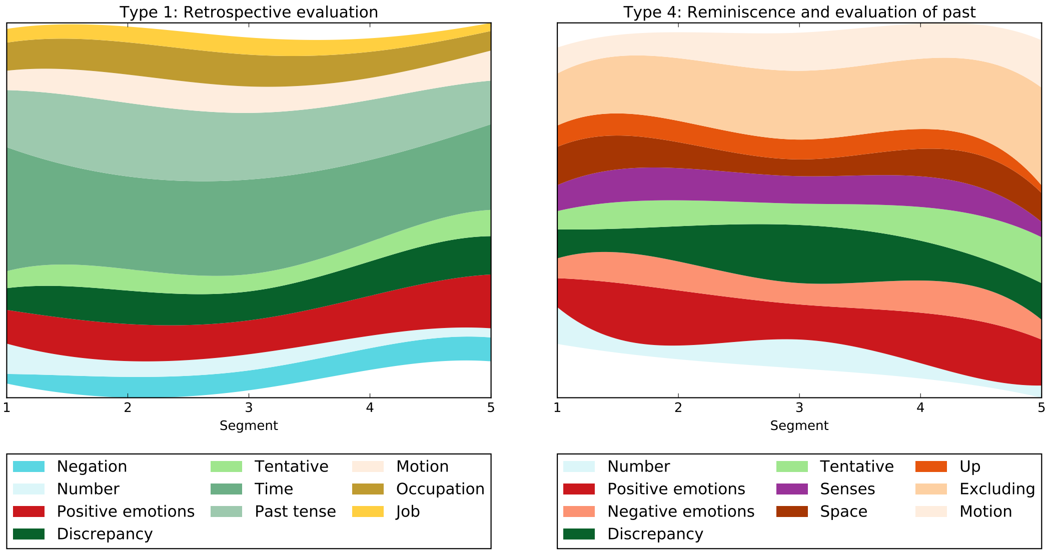


Figure 4 Ten most fluctuating categories retrospective letters.

According to [30], the main difference between both retrospective letters would be the verb tenses and the sequence in which these are used. The type 1 letters start with an imaginative future situation in the present tense followed by reminiscence of the future past in the past tense, whereas in the type 4 letters the recounted period actually lies in the past instead of the futured past and is described as a present concern. Based on these findings, one would expect to observe differences in the proportions of used verb tenses both between (Figure 3) and within (Figure 4) the letters. However, Figure 3 shows no observable differences in the proportions of the LIWC categories that regard verb tenses (past, present and future) between letters 1 and 4. Figure 4 does show "past tense" as one of the ten most fluctuating categories for letter type 1: the use of past tense slightly increases towards the middle of the letter and then decreases towards the end. The use of present tense does not seem to differ much throughout the type 1 letter as it is not amongst the ten most fluctuating categories included in the graph. None of the used tenses fluctuates much throughout the type 4 letters, as they are not amongst the ten categories included in the graph in Figure 4.

Overall it can be observed from both Figure 3 and Figure 4 that the imagination letters (type 1) contain more words regarding occupation and job, combined with motion words and positive emotions. The words related to occupation and job could be linked to the narrative element "orientation", the first narrative element distinguished by [18]. The motion and positive emotion words could be used to describe the (path towards) the desired future situation or a period of personal growth ("complicated action", [18]). The graphs further show an increasing use of discrepancy words (e.g. should, could, would) from the middle to the end. This supports the findings of [30], who state that towards the end of the letters conclusions or insights are drawn (pointing towards the narrative elements "evaluation" and "resolution", [18]), followed by statements of worldly wisdom self-praising remarks (which could be defined as the "coda", [18]).

For type 4 letters, Figure 3 and Figure 4 show an increase in the use of words from the categories "physical" and "body" combined with both positive and negative emotion words at the beginning and end of the letter. This could indicate that the element "orientation" from the framework of [18], contains mainly physical characteristics in letter type 4, as opposed to

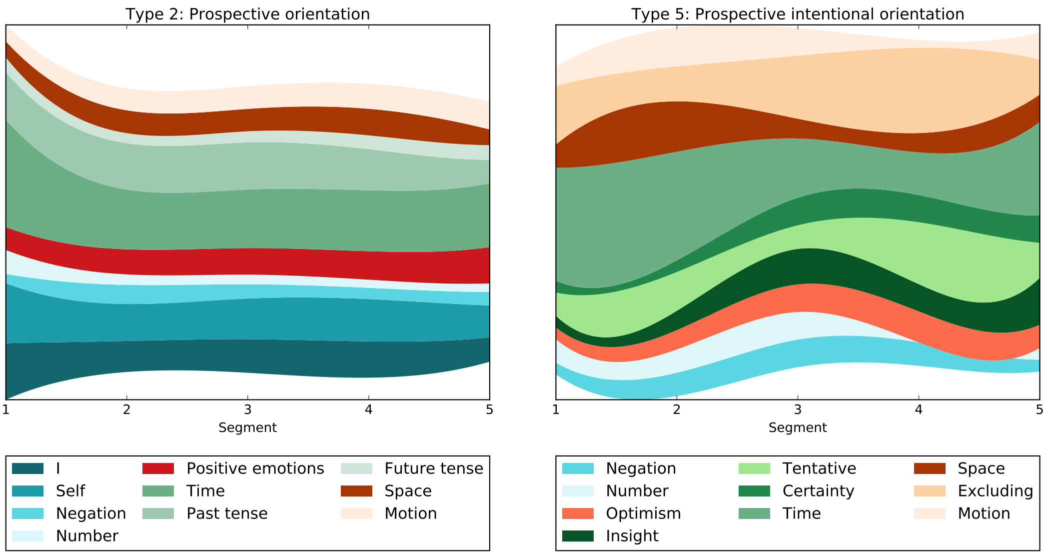


Figure 5 Ten most fluctuating categories prospective letters.

the professional characteristics used in letter type 1. Cognitive mechanisms are used more from the middle (insight and discrepancy) to the end (tentative) of the letters. This could be because these writers are still in the process of reminiscing and evaluating past events (pointing towards the elements "evaluation" and "resolution" of [18]). It could be that these letters start with a description of physical or emotional complaints or by a recollection of a happier past, which is then processed and evaluated, followed by moral advice or a tentative promise for a better future (distinguished by [30] as the "coda"). Finally the type 4 letters also contain more words related to senses and leisure. Overall, the general letters seem to be more sensitive, expressive and detailed than the type 1 letters.

3.2.2 Prospective letters

[30] found a clear structure for the imaginative letters (type 2), but not for the general letters (type 5). The imaginative letters were expected to start with a statement about one's present position in life (in present or past tense). This is reflected in the high occurrence of words in the categories "I" and "Self" and words related to "Time" (e.g. end, until) and numbers at the beginning of the letters (see Figure 5), which could be used to describe one's present position in life (narrative element "orientation" [18]). The increase in the use of words regarding space (e.g. nearby, places, directions), in the middle can reflect concrete imaginary goals and purposes. The path towards the futured situation (possibly the "complicated action", [18]) could be indicated by the increasing use of motion words and positive emotions.

With regard to the used tense, Figure 3 further shows that, in addition to the present tense, more past tense is used in the type 2 letters, whereas more future tense is used in the type 5 letters. This is in line with the findings of [30]. Overall the general letters contain more affect and emotions, and more cognitive mechanisms towards the end, which could point towards encouraging oneself to realize their goals, as described by [30].

The intentional element, the major characteristic of the type 5 letters, is clearly reflected in Figure 3 by the use of future tense and the high occurrence of tentative words (like "hope", "believe", "try", "possible") and discrepancy words ("must", "wish","want"). Figure 5 further

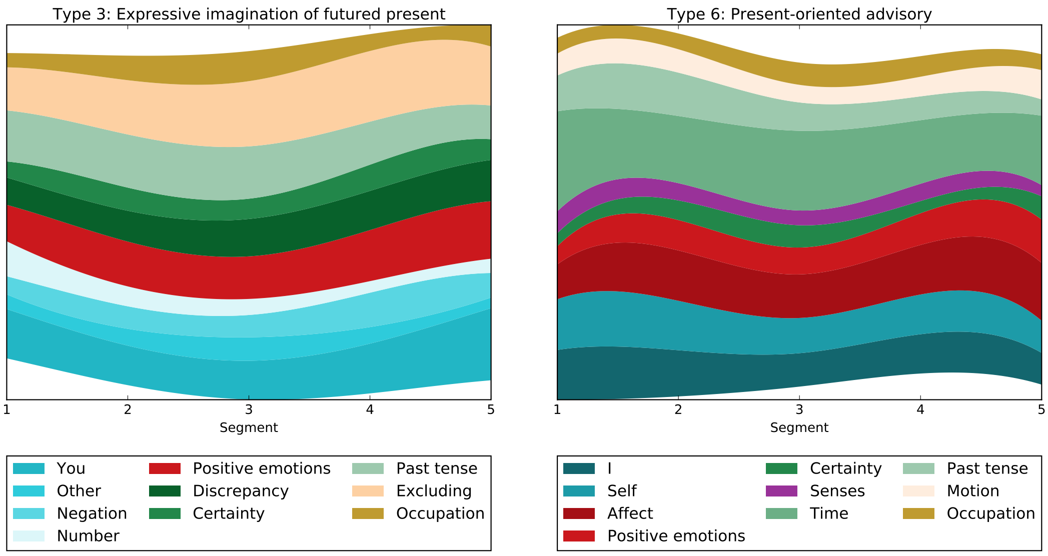


Figure 6 Ten most fluctuating categories present-oriented letters.

shows that the type 5 letters start and end more tentative, alternated with insight in the middle and end (pointing towards "evaluation", [18]). This letter is increasingly optimistic and certain, combined with an increasing use of excluding words. This might point towards an increasing insight in desired versus non-desired situations or future aspects, which may lead to more a more positive and concrete vision for the future in the "coda" [18]. However, the increasing use of excluding words combined with the high use of tentative and hesitative words could also reflect the doubt and uncertainty regarding the future related to prospective intentional orientation.

3.2.3 Present-oriented letters

[30] found no specific sequential order in narrative processes for the present-oriented letters. Figure 3 shows that the present-oriented letters are quite similar for both categories. However, the imagination letters (type 3) do contain more words regarding family, leisure, more superlatives (category "up") and slightly more positive emotions and feelings. This is in line with the findings of [30], who found that type 3 letters are positive, content, and joyful letters.

The letters generally end with hopes and wishes (shown by the increase in discrepancy words) and contain a lot of self-praising remarks (shown by the high increase in the use of "you" in the middle and end). This could point to the narrative elements "resolution" and "coda" [18]. The low use of cognitive mechanism and insight words supports the findings of [30] that the letter contains almost no orientation or evaluation, two of the five narrative elements distinguished by [18]. The high use of excluding words could point towards a breach with the past, without describing the current situation or the path from past to future (no "complicated action" [18]). The additional increase in the use of certainty towards the end indicates that the letters become more stimulating and convincing at the end (indicating "result/resolution" or "coda" [18]). It seems that the confidence of the writer increases by imagining the future situation. Finally, regarding the used tense, the type 3 letters are written mainly in the present tense, although Figure 6 shows that in both letters the past tense is used more in the beginning than in the middle and end of the letters.

In the general letters (type 6), more insight and discrepancy words are used. These letters also contain more negative emotions and feelings and slightly more sensory words. This supports the findings of [30], who state that the function of these letters is mainly to provide insight in and guidance for current problems or concerns, followed by statements of worldly wisdom. The finding of [30] that these letters do not contain a clear path or clarification of how and where certain knowledge or insights have been gained is supported by the fact that these letters contain almost no causation words. The high use of certainty words in the middle of the letter may be explained by the statements of wisdom and moral advice, combined with the fact that these letters do not contain evaluative aspects, which introduce more uncertainty. Apart from the elements "resolution" and "coda" it is difficult to link the letter characteristics from the visualizations to the narrative elements of [18].

4 Discussion

In this paper, a combination of Natural Language Processing, quantitative analysis and visualization techniques was used to explore differences in letter content, specifically the distribution (sequential order) and proportion of narrative processes and grammatical elements, both within and between the different types of "Letters from the Future". The visualizations could be used for two purposes; to confirm findings of previous studies on the content of the letters and to explore the letters in a broader sense to come to new insights or theories. Two essential topics in the development of text visualizations – capturing the underlying mathematical narrative structure and choosing a suitable format to visualize changes in letter content throughout the letter – were addressed. In general, the use of text visualizations proved to be a good method to globally explore and compare the underlying structures and differences in contents within and between the letter types. Thanks to the shape of the streamgraphs and the use of sequential colour palettes, the hierarchical time series plots of the letters were easily interpretable and comparable. By combining the visualizations with quantitative analysis of variance and the coefficient of variation, more specific insights in the distribution and proportion of narrative processes and grammatical elements throughout the letters was gained.

All in all, the visualizations were found to be very usable to at least partially confirm the previous findings of [30]. Finding strong additional characteristics or differences between and within the letters turned out to be more challenging. An interesting finding is that the proportional distributions of the LIWC categories, especially those of letter types one, two and three do not differ as much as expected based on the previous findings of [30]. The visualizations for those types look very similar, as opposed to the visualizations for letter types four, five and six. An explanation for this may be that the LIWC categories used as underlying structure are too global or do not directly apply to the current dataset. A more specific categorization system developed especially for the "Letters from the Future" dataset might perform better. A possibility is to develop a new LIWC dictionary based on the previous findings of [30] and the visualizations generated in this study, and apply this to a new dataset. Potential features to include in this dictionary could be the most informative features that discriminate between the six letter types. These most informative features have been extracted from the current dataset for a different study by the authors in which supervised text classification algorithms are used to automatically categorize the letters to their corresponding classes. It would be interesting to visualize the occurrence of these features within the letters.

It could also be that the way the letters are split into five segments influences the proportional distributions. For example, when a certain narrative process starts at the end of the first segment and finishes at the beginning of the second segment, the characteristics for this process are evened out between the first to segments. This may cause a blur in the resulting visualization. It would be interesting to see if splitting the letters manually into five segments, based either on the narrative elements of [18] or the five narrative processes distinguished by [30] would lead to more distinctive variations both between the letter segments and the letters as a whole.

Splitting the narratives into the structural elements distinguished by [18] also opens up to a new avenue for future research, namely to investigate variations in the narratives that depend on the characteristics of the writer. The framework of [18] has already been used to investigate differences in narrative content between classes [17], [15], gender [16], [8], age [27], [32] and geography [16]. Visualizing the narrative structures for groups with different characteristics may lead to new insights or hypotheses for further research on these topics.

As a final note, although the current focus is on visualizing the content of “Letters from the Future”, the resulting method can in fact be used to explore any available digital text document or corpus. The methods and results described in this paper can be seen as a first step in an ongoing study by the authors and the Storylab to study therapy-related textual features in e-mental health interventions. By using methods like NLP and text visualization to analyse patterns in therapy-related textual features, extracted for example from written narratives or the linguistic interaction between counsellor and client, more insight can be gained in what happens within therapy, when progress is made, or for which persons a certain type of therapy is more effective. This could greatly improve e-mental health interventions and advance therapy change process research. Future research will therefore include expanding the time series to include more letters written by the same person, studying changes between subsequent narratives and analysing counsellor-client interaction.

References

1. Adeline Abbe, Cyril Grouin, Pierre Zweigenbaum, and Bruno Falissard. Text mining applications in psychiatry: a systematic literature review. *International Journal of Methods in Psychiatric Research*, 2015. doi:10.1002/mpr.1481.
2. Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python*. O'reilly Media, Inc, Sebastopol, 2009.
3. M. Bloch, L. Byron, S. Carter, and A. Cox. The ebb and flow of movies: Box office receipts 1986-2007. *New York Times*, 2008. URL: http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html?_r=0.
4. Ernst Bohlmeijer. *De Verhalen die we leven. Narratieve psychologie als methode.* Boom, Amsterdam, 2007.
5. J. Bradley and G. Rockwell. What scientific visualization teaches us about text analysis. In *ALLC/ACH Conference*, Paris, 1994.
6. Charles E. Brown. Coefficient of Variation. In *Applied Multivariate Statistics in Geohydrology and Related Sciences*, pages 155–157. Springer Science & Business Media, 2012.
7. Lee Byron and Martin Wattenberg. Stacked graphs – Geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1245–1252, 2008. doi:10.1109/TVCG.2008.166.
8. Jenny Cheshire. The telling or the tale? narratives and gender in adolescent friendship networks. *Journal of Sociolinguistics*, 4(2):234–262, 2000.
9. Gobinda G. Chowdhury. Natural language processing. *Annual Review of Information Science and Technology*, 37(1):51–89, 2005. doi:10.1002/aris.1440370103.
10. Jeff Clark. Tom Sawyer Character StreamGraph, 2008. URL: http://www.neoformix.com/2008/TomSawyer.html.

11 A. Cox and L Byron. The Ebb and Flow of Box Office Sales, 1986-2007,. *The New York Times, February 23, 2008*, 2008.

12 O. A. de Carvalho, Renato Fontes Guimarães, Roberto Arnaldo Trancoso Gomes, and Nilton Correia da Silva. Time series interpolation. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 1959–1961. IEEE, IEEE, 2007.

13 Susan Havre, Beth Hetzler, and Lucy Nowell. Themerivertm: In search of trends, patterns, and relationships. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

14 P. C. Hogan. Continuity and change in narrative study. Observations on componential andfunctional analysis. *Narrative Inquiry*, 16(1):66–74, 2006.

15 Barbara Horvath. Text on conversation: Variability in storytelling texts. In K. Denning, S. Inkelas, F. McNair-Knox, and Rickford. J., editors, *Variation in Language*. Department of Linguestics, Stanford, 1987.

16 Barbara Johnstone. Variation in discourse: Midwestern narrative style. *American Speech*, 65(3):195–214, 1990.

17 William Labov. Some further steps in narrative analysis. *Journal of Narrative and Life History*, 7:395–415, 1997.

18 William Labov and Joshua Waletzky. Narrative analysis: oral versions of personal experience. In J. Helm, editor, *Essays on the verbal and visual arts*, chapter Narrative, pages 12–44. Washington University Press, Seattle, 1967.

19 E. D. Liddy. Natural language processing. In M.A. Drake, editor, *Encyclopedia of library and information science*. Marcel Decker, New York, 2nd edition, 2001.

20 Pat Lovie. Coefficient of Variation. In *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd:, Oxford, UK, 2005.

21 Inderjeet Mani. Computational Narratology. In Peter Hühn, Christoph Meister, Jan, John Pier, and Wolf Schmid, editors, *Handbook of Narratology*, pages 84–92. De Gruyter, 2014.

22 Scott E. Maxwell, Harold D. Delaney, and Ken Kelley. *Designing Experiments and Analyzing Data*. Taylor & Francis Group, New York, 2003.

23 Duncan M. McGreggor. *Mastering Matplotlib*. Packt Publishing, 2015.

24 Jan C. Meister and Alastair Matthews. *Computing Action*. De Gruyter, Berlin, 2003.

25 J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic Inquiry and Word Count (LIWC): LIWC2001, 2001.

26 James W. Pennebaker, Cindy K. Chung, Molly Ireland, Amy Gonzales, and Roger J. Booth. The development and psychometric properties of LIWC2007, 2007.

27 C. Peterson and A. McCabe. *Developmental psycholinguistics: Three ways of looking at a child's narrative*. Plenum Press, New York, 1983.

28 Robert H. Shumway and David S Stoffer. *Time series analysis and its applications. With R examples*. Springer Science & Business Media, New York, 2006. `doi:10.1016/j.peva.2007.06.006`.

29 Anneke Sools and Jan Hein Mooren. Towards Narrative Futuring in Psychology: Becoming Resilient by Imagining the Future. *Graduate Journal of Social Science*, 9(2):203–226, 2012.

30 Anneke M. Sools, Thijs Tromp, and Jan H. Mooren. Mapping letters from the future: Exploring narrative processes of imagining the future. *Journal of Health Psychology*, 20(3):350–364, 2015. `doi:10.1177/1359105314566607`.

31 Andrew J. Tomarken and Ronald C. Serlin. Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, 99(1):90–99, 1986. `doi:10.1037/0033-2909.99.1.90`.

32 M. J. Toolan. *Narrative: A critical linguistic introduction*. Routledge, London, 1988.

33 P. Valéry, G. Rockwell, and J. Bradley. Printing in Sand; Scientific Visualization and the Analysis of Texts. *Geoffreyrockwell.Com*, pages 1–16, 1999.

34 Martin Wattenberg. Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization. INFOVIS 2005.*, pages 1–7. IEEE, 2005.

35 Martin Wattenberg and Jesse Kriss. Designing for social data analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):549–557, 2006.

36 Wibke Weber. Text visualization-what colors tell about a text. In *Information Visualization, 2007. IV'07. 11th International Conference*, pages 354–362. IEEE, 2007.

37 James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual. Spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Information Visualization Symposium 1995*, pages 51–58, Atlanta, Georgia, 1995. IEEE.

38 Hanna Zijlstra, Tanja Van Meerveld, Henriët Van Middendorp, James W. Pennebaker, and Rinie Geenen. De Nederlandse versie van de 'Linguistic and Word Count' (LIWC). Een gecomputeriseerd tekstanalyseprogramma. *Gedrag & Gezondheid*, 32(4):271–281, 2004.

Comparing Extant Story Classifiers: Results & New Directions*

Joshua D. Eisenberg[1], W. Victor H. Yarlott[2], and Mark A. Finlayson[3]

1 **Florida International University, School of Computer and Information Sciences, Miami, FL, USA**
jeise003@fiu.edu
2 **Florida International University, School of Computer and Information Sciences, Miami, FL, USA**
wvyar@cs.fiu.edu
3 **Florida International University, School of Computer and Information Sciences, Miami, FL, USA**
markaf@fiu.edu

Abstract

Having access to a large set of stories is a necessary first step for robust and wide-ranging computational narrative modeling; happily, language data–including stories–are increasingly available in electronic form. Unhappily, the process of automatically separating stories from other forms of written discourse is not straightforward, and has resulted in a data collection bottleneck. Therefore researchers have sought to develop reliable, robust automatic algorithms for identifying story text mixed with other non-story text. In this paper we report on the reimplementation and experimental comparison of the two approaches to this task: Gordon's unigram classifier, and Corman's semantic triplet classifier. We cross-analyze their performance on both Gordon's and Corman's corpora, and discuss similarities, differences, and gaps in the performance of these classifiers, and point the way forward to improving their approaches.

1998 ACM Subject Classification I.2.7 Natural Language Processing: Language Models

Keywords and phrases Story Detection, Machine Learning, Natural Language Processing, Perceptron Learning

Digital Object Identifier 10.4230/OASIcs.CMN.2016.6

1 Motivation

Computational narrative researchers often rely on textual story data to inform and support their analyses. There is a wealth of textual story content that can be harvested from electronic sources, including online newspapers, journals, blogs, e-books, emails, electronic records, and so forth. Nevertheless, what text does or does not constitute a story is not *prima facie* obvious: text is rarely labeled explicitly as story. Thus, to take advantage of the vast potential sources of narrative data available in electronic form, we need to be able to automatically find stories in a sea of otherwise non-story text.

* This work was partially supported by National Institutes of Health (NIH) grant number 5R01GM105033-02. Implementation work on this project was supported by Nicholas Zessoules, who was funded under Research Experience for Undergraduates (REU) grant number 1263124 from the National Science Foundation (NSF) and Department of Defense (DoD).

© Joshua D. Eisenberg, W. Victor H. Yarlott, and Mark A. Finlayson;
licensed under Creative Commons License CC-BY

There has been prior work on automatic story identification, in particular, Gordon's unigram-based story classifier [9] and Corman's semantic triplet story classifier [2, 3] each tested on their own data sets. In this work we re-implement and cross-evaluate both these classifiers on both datasets, and thereby gain insight into the strengths and weaknesses of both approaches. This comparison also points the way forward to more accurate and effective story classifiers.

1.1 What is a Story?

In this study we are trying to classify the linguistic artifact known as the story, but what is a story? Let's start with the definitions of story that Gordon and Corman employ. Gordon is searching for *personal stories*: "textual discourse that describes a specific series of causally related events in the past, spanning a period of time of minutes, hours, or days, where the author or a close associate is among the participants" [9]. This definition implies knowledge of who the narrator is in relation to the *dramatis personae*, and the diegesis or narrative distance of the story teller. Corman says a story has three components: actors, the actions they perform, and the resolution which they "must result in resolution" [2]. But, not all stories have resolution. We propose a more succinct definition of story, a definition from author E. M. Forster, "A story is a narrative of events arranged in their time sequence" [8]. Typically these events are related for the story to be intelligble, but not all stories are composed this way. We feel that this definition is a beacon towards constructing a more accurate story classifier.

1.2 Outline of the Paper

To evaluate the classifiers on both sets of data, we needed access to implementations of both classifiers as well as to their training and testing corpora. Gordon's paper [9] provided a high level description of his classifier, which we used as a blueprint for implementing it in Java. He also provided a Python implementation of a story classifier[1], but it is not the same algorithm described in the paper. Our re-implementation of the Gordon classifier is close to the original design, which is discussed in §2.1. During reimplementation we noted that replacing the original classifier with a Perceptron implementation improved the results somewhat, and the details of this mildly improved version of Gordon's classifier are given in §2.2.

Corman et al. did not provide an open-source implementation of their code, and so we reimplemented that classifier from descriptions in their papers[2]. They provided a high level description of their preprocessing and feature extraction pipeline in [2, 3], described in §2.3. We used this as a guide for building our reimplementation, but there were aspects that performed significantly differently due to specific, difficult-to-reimplement design choices.

To train and test the classifiers we obtained the datasets used in the original papers: the ICSWM Spinn3r 2009 Weblog dataset [1], which Gordon used to test his classifier, and the CSC Islamic Extremist corpus, which Corman used for his study. We report the results of our reproductions of the original experiments in §3.1, and report on our new cross-test experiments in §3.2. We discuss our observations of the results of the experiments in §4, and outline some potential future work in §4.1. We conclude with a list of our contributions in §5.

[1] `https://github.com/asgordon/StoryNonstory`
[2] We intend to release all of our code as supplementary material associated with a future article.

2 Extant Story classifiers

The computational identification of stories is relatively a new problem, and there are only two extant approaches: one developed by Gordon et al. at USC [9], the other by Corman et al. at ASU [2, 3]. Both the Gordon and Corman classifiers leverage supervised machine learning algorithms trained on large annotated datasets, and both use linguistic features to separate story from non-story text. Gordon's classifier uses unigram frequencies to classify stories. This classifier was originally tested on the ICWSM 2009 Spinn3r blog dataset, which contains personal stories that were posted to blogs in 2009 [1]. Corman's classifier, on the other hand, focuses on verbs and their patient and agent arguments (semantic triplets). It also considers the unigram frequencies and density of various features such as part of speech tags, named entities, and stative verbs. Corman's classifier was originally tested on the CSC corpus of Islamic Extremist texts, in which each paragraph was annotated as either *story*, *exposition*, *supplication*, *religious verse*, or *other*.

2.1 Reimplemented Gordon Story Classifier

The philosophy of Gordon's classifier is stories are made up of words, and so to find stories one should look for story-relevant words. Gordon's story classifier uses unigram features [9] and so makes associations based on what words appear in stories vs. non-stories (a "bag of words" approach). The features (words) extracted from each text in the training set are used to train a confidence weighted linear classifier [5]. This is similar to a perceptron [13, 16], but it has augmentations which can improve how it learns NLP features. Although the confidence weighted linear classifier can be better suited than the perceptron for classifying certain NLP phenomena, we did not find that to be true in our story classification experiments. To train the confidence weighted linear classifier each word is run through the classifier one time (one epoch); it is uncertain how many epochs of training Gordon used, but we found that performance did not improve significantly by training this classifier for more epochs. After training, the confidence weighted linear classifier has assigned each individual word a weight that represents how relevant it is in classifying text as a story. Weights closer to zero imply that the word occurs in both stories and non-stories with similar frequency, while weights far from zero imply that the relevant word appears correlates with one of the two classes (stories or non-stories). To classify a document, its words are extracted in the same manner as they are in the training set. Then the feature counts are capped: all counts above 7 are brought down to 7. Finally, the feature values are normalized to values between 0 and 1.

There are two parts to our reimplementation: the feature extractor and the confidence weighted linear classifier. The feature extraction pipeline is built in Java, and the classifier is written in the Go programming language, in order to make use of the gonline library[3], a library of online machine learning algorithms written in Go (we could not find a usable version of the confidence weighted linear classifier in Java). We modified the gonline library in order to produce more fine-grained error statistics, and to suppress false parser errors. For the feature extractor, we use some of the same text preprocessing that Gordon provided[4]. We use the same regular expressions from his Python code to break up clitics, punctuation, and irregular characters. Then we feed the filtered data through the Stanford Tokenizer [12] to turn each document into a stream of tokens. This stream of tokens is used for the unigram counting.

[3] `https://github.com/tma15/gonline`
[4] `https://github.com/asgordon/StoryNonstory`

2.2 Improved Gordon Story Classifier

During reimplementation we experimented with replacing the confidence-weighted linear classifier with traditional single layered biased perceptron network [13, 16]. This “Perceptron” version of the Gordon classifier scored an F_1 measure 5 points higher than our implementation of the confidence weighted Gordon classifier. This is interesting because the confidence weighted linear classifier has been reported as better at classifying data from the NLP domain [5]. In this case, not only was the Perceptron classifier easier to develop than the confidence weighted one, it also had better performance. These findings are expanded upon in §3.2. In our final implementation of this improved classifier, we trained the Perceptron for 10 epochs and used a learning rate of 0.005. We settled on these hyperparameters via tuning and experimentation. Additionally, for each epoch, we randomized the order that the training vectors are shown to the Perceptron, because Perceptrons are known to be quite sensitive to the order that the training examples are learned.

The final difference between the original implementation and our improved implementation is that our encoding scheme for the frequencies is slightly different. Our augmentation is a smoothing of all the feature values: 0.07 (roughly 1/14) is added to each feature value. We did this to guarantee that the weights for each feature will be updated during the Perceptron learning, as features with value 0 do not contribute to the learning.

2.3 Reimplemented Corman Classifier

Corman’s semantic triplet based classifier is trained on a wide variety of features “of varying semantic complexity” so it requires a robust linguistic pipeline to facilitate feature extraction [2, 3]. Some of the features are based on lexical properties: the densities of the 30 Penn Tree Bank part of speech tags [18], stative verbs, and person, location, and organization named entities. The term frequency inverse document frequency (tf-idf) [17] measure is calculated for each word in each document of the training corpus and, in this implementation, the words with the highest 20,000 tf-idf measures are features. The final feature is the semantic triplet. These are triplets of each verb with their respective patient and agent arguments. Sometimes this feature is actually a four-tuple where the fourth term is a locational preposition. The lexical features and triplets are extracted from each document in the training set and used to train a support vector machine (SVM) with an RBF function [10].

Corman’s 2012 paper gives a high-level description of how these features are extracted from each document. Our classifier differs in a few non-trivial ways:

- We do not use the Illinois Semantic Role Labeler [14]. We use a semantic role labeler built from scratch in our lab, which is included in the Story Workbench linguistic annotation tool [7]. We used our own tool because the Illinois SRL is quite heavyweight: it requires installation of the Illinois Curator Server [4] and MongoDB[5].
- We do not do coreference resolution to replace the pronouns with their corresponding referent entity.
- We do no alias standardization. It was unclear how this should be carried out, since the named entity tagging is run on each token, and there was no explanation for how multi-word named entity boundaries were determined.
- We performed no spell checking on our named entities for the same reason as the alias standardization.

---

[5] `https://www.mongodb.org/`

- We only use the Stanford named entity recognizer [6] and the Illinois named entity tagger [15]. We do not use the Open Calais named entity recognizer[6] because we wanted the classifier to run without needing to query a limited resource on the internet.

We removed three portions of the text preprocessing, used a different SRL, and one fewer NER than Corman's original implementation. Even though we built a less complicated version of Corman's classifier, it performed approximately the same as (or even slightly better than) the original.

To extract semantic triplets, we first extract the parse tree for a sentence. We pass this parse tree to the SRL, which extracts all the predicates and their arguments. We take the lemmatized predicates and search for a VerbNet [19] category based on the number of arguments in Propbank [11]; we take an exact match if there is one, but take the closest match otherwise. Failing this, we return the lemma of the word from Wordnet. We follow the following rules, set forth by Corman et al. in their paper:

- If our object has multiple verbs, it is complex. We create new triplets for the object recursively and assign a pointer to the new triplets as the object for this triplet.
- If the SRL doesn't return Arg1, we substitute Arg2 for the object.
- If Arg2 is a location preposition, we include it as a fourth element.
- If Arg2 is a preposition, we use Arg1 as the subject and Arg2 as the object.

Due to some shortcomings in our SRL we may find that some or all of the arguments for a given predicate are null; in the event that some of our predicates are null, we simply tag the remaining slot (either subject or object) with a "-1" indicating a lack of an argument. If all of our arguments are null we attempt to find the closest noun behind the predicate, which we assign as the subject, and the closet noun ahead of the predicate, which we assign as the object. As above if we result in a null argument from finding the closet nouns we tag those slots with "-1".

To use the triplets as a feature, we extract all the possible subjects, objects, and location prepositions for a given verb or verb category from across the entire corpus. Then we assign each verb and verb category a specific index. These indices are determined by a simple alphabetical sort: We check every document's triplets and assign it a "1" at a given index if it has that verb or that argument for the verb and a "0" otherwise. These features are used to train a SVM that uses a radial bias function (RBF) kernel function and a soft margin C of 10,000, which is a relatively standard setting (we are not sure what soft margin parameter was used by Corman). This produces a model that can classify whether a text contains a story. To test text on the model, the same types of features extracted in training, are extracted from the test document. The feature values are used with the model to obtain a classification value.

3 Experimental Results

3.1 Reproduction of Original Experiments

To show that the reimplemented classifiers behave the same way as the originals, they were trained and tested on the same data sets as in the original studies. The Gordon confidence weighted (CW) linear classifier was trained and tested on the Spinn3r Weblog Corpus. We used the same texts that Gordon used in his study. As can be seen in Table 1, in terms of

[6] `http://www.opencalais.com/`

Table 1 Results for all three Gordon classifiers tested and trained on the Weblog corpus.

System	Training	Testing	Precision	Recall	F_1
Gordon Reported	**Web**	**Web**	0.66	0.48	0.55
Gordon CW	**Web**	**Web**	0.475	0.5	0.471
Gordon Perceptron	**Web**	**Web**	0.648	0.457	0.522

Table 2 Results for Corman's classifier- tested and trained on the Extremist corpus.

System	Training	Testing	Precision	Recall	F_1
Corman Reported	**Ext**	**Ext**	0.731	0.559	0.634
Corman	**Ext**	**Ext**	0.773	0.573	0.658

F_1 our Gordon CW reimplementation performs almost 8 points worse than what Gordon reported. Our Gordon CW classifier has a precision of 0.475, recall of 0.5, and an F_1 of 0.471. This could be because Gordon used a different version of the CW algorithm than we did, but it is unclear why our implementation performs so differently. Although our modified Gordon Perceptron is arguably simpler than our Gordon CW, it has a higher F_1 by almost 5 points. The Gordon Perceptron's F_1 is 0.522, which is almost 3 points less than the performance Gordon reported. We cannot say that the Gordon CW is a good reimplementation of the original Gordon classifier since the F_1 measures are significantly different. On the other hand, the performance of our Perceptron and Gordon's original classifier are quite similar, which is encouraging.

Corman et al. trained and tested their classifier on the CSC Islamic Extremist Corpus. We used the same texts during this experiment on the reimplementation. As can be seen in Table 2, our reimplementation of Corman's classifier performs similarly to the original version. The reimplementation scored a precision of 0.773, recall of 0.573, and F_1 of 0.658, which compares favorably with the originally reported results.

In terms of F_1 our performance scored slightly higher than that of the original system. This may be due to the differences in the preprocessing pipeline and the triplet extraction, as discussed in §2.3. Nevertheless, we take the similarity of the results as evidence that our reimplementation is roughly faithful to the original.

3.2 Cross-Testing the Classifiers

Because we have both classifiers and both datasets, we performed experiments to compare how each classifier performs on the other data, as well as on both datasets simultaneously. We trained and then tested both our reimplemented and improved classifiers on all combinations of the three corpora. For the cross-tests of the Gordon Classifier we use the Gordon Perceptron. As shown in the previous section, the Gordon Perceptron performs more close to the original Gordon Classifier than our reimplementation of the Gordon CW. Using the Gordon Perceptron allows for a more accurate comparison of the classifier than our Gordon CW implementation. For each of the cross-tests, the corpora go through 10-fold cross validation to generate the training and testing sets, as follows:

- If the training and testing corpora are the same, divide up the stories into ten subsets of equal size, and the not stories into ten sets of equal size. For each fold of cross validation a different story set and the not story set (of the same index) are used as the testing set and the remaining 9 are used for training.
- If the training is done on the combined corpus, and the test corpus is either the weblog or extremist corpus, which we will refer to as the single corpus, first divide the stories

Table 3 Results for Gordon Classifiers across all experiments.

System	Training	Testing	Precision	Recall	F_1
Gordon Reported	Web	Web	0.66	0.48	**0.55**
Gordon CW	Web	Web	0.475	0.5	0.471
Gordon Perceptron	Web	Web	0.648	0.457	**0.522**
Gordon Perceptron	Web	Ext	0.238	0.354	0.285
Gordon Perceptron	Web	Comb	0.594	0.008	0.014
Gordon Perceptron	Ext	Web	0.25	0.413	0.311
Gordon Perceptron	Ext	Ext	0.65	0.545	0.472
Gordon Perceptron	Ext	Comb	0.433	0.407	0.322
Gordon Perceptron	Comb	Web	0.363	0.488	0.299
Gordon Perceptron	Comb	Ext	0.62	0.508	0.426
Gordon Perceptron	Comb	Comb	0.639	0.471	0.464

Table 4 Results for Corman classifier across all experiments.

System	Training	Testing	Precision	Recall	F_1
Corman Reported	Ext	Ext	0.731	0.559	**0.634**
Corman	Ext	Ext	0.773	0.573	**0.658**
Corman	Ext	Web	0.229	0.372	0.283
Corman	Ext	Comb	0.46	0.495	0.477
Corman	Web	Ext	0.591	0.003	0.007
Corman	Web	Web	0.656	0.314	0.425
Corman	Web	Comb	0.612	0.005	0.01
Corman	Comb	Ext	0.779	0.554	**0.647**
Corman	Comb	Web	0.412	0.34	0.365
Corman	Comb	Comb	0.756	0.543	**0.632**

into ten equal sized sets, and then divide up that corpus' not stories into ten equal sets. First divide the stories in the single corpus into ten equal sets, and the not stories of the same corpus into ten equal sets. These can be split into the training and testing sets the same as in the previous situation. Additionally, the whole other corpus, the one that is not the single corpus, is added to the training set.

- If training is done on a single corpus, and the test corpus is the combined corpus, first break up the stories and not stories of the single corpus each into ten equal subsets. Assign them to the training and testing set as in the first situation. Then add the whole other corpus, the one that is not the single corpus, to the testing set.

The results can be seen in Tables 3 and 4. Macro averages for precision, recall and F_1 are reported for each experiment. We chose to report macro averaging since it was less sensitive to outliers and atypical models. Unless stated otherwise, the F_1 measures reported in this study are relative to the story class. Using story as the label to classify correctly is a good way to measure performance for this task, since the overall goal is to produce a system that can accurately identify stories.

4 Discussion

The Corman classifier has it's best performance when tested and trained on the extremist corpus, while the Gordon Perceptron does best on the weblog corpus. Both classifiers perform best when tested and trained on the corpus that it was tested on in their original studies. The Corman classifier scores best across all the experiments: 0.658 F_1 for when it is trained and tested on the extremist corpus. Yet, the Corman classifier F_1 is 10 points worse than the Gordon Classifier when tested and trained on the weblog corpus. This suggests that the Corman classifier has a harder time learning from the weblog corpus than the extremist corpus.

The Gordon classifier performed best when trained and tested on the Webblog corpus. The F_1 measure for this experiment was 0.522, which is about 3 percentage points lower than the best result Gordon reported, and we hypothesize that this is because differences in our CW classifier implementations.

When the Corman classifier, with the Extremist model, is tested on the Weblog corpus it only has an F_1 of 0.283. This poor performance is because the model has not been trained to recognize the type of stories in the Weblog corpus. The stories in the Extremist corpus are typically 3rd person, second hand accounts, not 1st person personal accounts. Even though the Weblog corpus has significant syntactic irregularity, it contains a different type of story than that present in the Extremist corpus, so it is still useful for model training.

None of the cross-tests perform as well as the tests on the original corpus. The only cross test that comes close to the original performance is when the Corman classifier is trained on the combined corpus and tested on either the extremist or combined corpora, respectively their F_1 measures are 0.647 and 0.632. This makes sense, because the training set mostly contains examples from the extremist corpus, thus the model is more heavily influenced by that corpus. The extremist corpus has about five times more texts than the weblog corpus, hence the training set will have five times more texts from the extremist corpus. In effect, the combined model is quite similar to the extremist model.

The two weakest of all the cross tests are when the models are trained using the Weblog data and tested on the combined corpus. When trained on the Weblog corpus and tested on the combined corpus the F_1 measure Gordon's F_1 is 0.014, and Corman's F_1 is 0.007. Another particular weak experiment was when the Corman classifier is tested on the webblog corpus but tested on the extremist corpus. The F_1 for this experiment is 0.007. This suggests that the Weblog data does not generalize well to the Extremist set. So although the Corman classifier has experiments with the highest F_1 measures, it also produces the weakest models when trained on only the Weblog corpus.

We can draw a few additional conclusions from these results. The Corman classifier has better performance when trained on the extremist corpus while the Gordon classifier has better performance with the weblog corpus. This is interesting because from the stories, the Extremist corpus mainly contains second hand accounts of events, often with a 3rd person narrator. On the other hand, the Weblog corpus mainly contains person stories and 1st person narrators. So it is possible that the Gordon classifier is better at finding personal stories while the Corman classifier is better at finding second hand accounts of events. It naturally follows that the Corman classifier has better performance when trained on the combined corpus than Gordon. This is due to the extremist corpus comprising 80% of the combined corpus.

4.1 Future Work

There are a number of features not considered by Corman or Gordon that we hypothesize could lead to a boost in performance.

First, the temporal nature of stories is not considered in either system. Forster said "A story is a narrative of events arranged in their time sequence" [8]. At a minimum a story narrates time. Both classifiers are aware of whether certain lexical and verb-based features occur but not when they occurred. Implementing features that reflect the passage of time may help improve performance.

Second, stories are collections of events, and thus it may help to consider detection of events rather than verbs *per se*. The semantic triplets reflect each verb and their arguments, but these verbs do not always refer to an event. As of now, Corman's classifier generates many, many triplets, and a large number of them have don't directly have to do with the progression of events in a story.

Third, referential cohesion of the characters carrying out the events could also help with the classification. Usually events in stories are carried about by a group of characters. Incorporating features reflecting whether certain characters are talked about more, or less, or across the timeline of the story, may also improve the classifier.

5 Contributions

The experiments in this paper move the study of story classification forward. First, we reimplemented both the Gordon and Corman story classifiers. Our reimplementations are more or less faithful to the behavior of the originals, as shown by the results in Tables 1 and 2. Additionally, both classifiers were tested across the Weblog, Extremist, and combined corpora, yielding 16 new experiments (Tables 3 and 4). This is the first time these different story classifiers have been tested on the same corpora, and so could be directly compared. We also detailed the process of constructing these classifiers: the nuances and difficulties associated with their construction and how to test them. Through cross-testing both classifiers we were able to find strengths and weakness in both methods.

Acknowledgements. We thank Reid Swanson especially, as well as Andrew Gordon, for working to retrieve, format, and providing their original annotations of the Weblog corpus. Thanks to Steve Corman for facilitating the transmission of the Extremist story data, which is covered by special U.S. government FOUO rules. Also thanks to Betul Ceran and Ravi Karad for providing the list of stative verbs and examples of the extracted semantic triplets.

References

1 Kevin Burton, Akshay Java, and Ian Soboroff. The icwsm 2009 spinn3r dataset. In *Proceedings of the Third Annual Conference on Weblogs and Social Media (ICWSM 2009), San Jose, CA*, 2009.
2 B. Ceran, R. Karad, S. Corman, and H. Davulcu. A hybrid model and memory based story classifier. In *The Third Workshop on Computational Models of Narrative (CMN). Istanbul, Turkey*, 2012.
3 Betul Ceran, Ravi Karad, Ajay Mandvekar, Steven R. Corman, and Hasan Davulcu. A semantic triplet based story classifier. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 573–580. IEEE, 2012.
4 James Clarke, Vivek Srikumar, Mark Sammons, and Dan Roth. An NLP curator (or: How I learned to stop worrying and love NLP pipelines). In *LREC*, pages 3276–3283, 2012.
5 Mark Dredze, Koby Crammer, and Fernando Pereira. Confidence-weighted linear classification. In *Proceedings of the 25th international conference on Machine learning*, pages 264–271. ACM, 2008.
6 Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 363–370. Association for Computational Linguistics, 2005.
7 Mark A. Finlayson. The story workbench: An extensible semi-automatic text annotation tool. In *Intelligent Narrative Technologies*, 2011.

8 Edward Morgan Forster. *Aspects of the Novel*. RosettaBooks, 2010.

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

10 S. Sathiya Keerthi and Chih-Jen Lin. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural computation*, 15(7):1667–1689, 2003.

11 Paul Kingsbury and Martha Palmer. Propbank: the next level of treebank. In *Proceedings of Treebanks and lexical Theories*, volume 3. Citeseer, 2003.

12 Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.

13 Hwee Tou Ng, Wei Boon Goh, and Kok Leong Low. Feature selection, perceptron learning, and a usability case study for text categorization. In *ACM SIGIR Forum*, volume 31, pages 67–73. ACM, 1997.

14 Vasin Punyakanok, Dan Roth, and Wen-tau Yih. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287, 2008.

15 Lev Ratinov and Dan Roth. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 147–155. Association for Computational Linguistics, 2009.

16 Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.

17 Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

18 Beatrice Santorini. Part-of-speech tagging guidelines for the Penn Treebank Project (3rd revision). Technical report, University of Pennsylvania, 1990.

19 Karin Kipper Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, University of Pennsylvania, 2005.

Learning a Better Motif Index: Toward Automated Motif Extraction*

W. Victor H. Yarlott[1] and Mark A. Finlayson[2]

1 **Florida International University, School of Computing & Information Sciences, Miami, FL, USA**
wvyar@cs.fiu.edu

2 **Florida International University, School of Computing & Information Sciences, Miami, FL, USA**
markaf@fiu.edu

Abstract

Motifs are distinctive recurring elements found in folklore, and are used by folklorists to categorize and find tales across cultures and track the genetic relationships of tales over time. Motifs have significance beyond folklore as communicative devices found in news, literature, press releases, and propaganda that concisely imply a large constellation of culturally-relevant information. Until now, folklorists have only extracted motifs from narratives manually, and the conceptual structure of motifs has not been formally laid out. In this short paper we propose that it is possible to automate the extraction of both existing and new motifs from narratives using supervised learning techniques and thereby possible to learn a computational model of how folklorists determine motifs. Automatic extraction would enable the construction of a truly comprehensive motif index, which does not yet exist, as well as the automatic detection of motifs in cultural materials, opening up a new world of narrative information for analysis by anyone interested in narrative and culture. We outline an experimental design, and report on our efforts to produce a structured form of Thompson's motif index, as well as a development annotation of motifs in a small collection of Russian folklore. We propose several initial computational, supervised approaches, and describe several possible metrics of success. We describe lessons learned and difficulties encountered so far, and outline our plan going forward.

1998 ACM Subject Classification I.2.7 Natural Language Processing, J.5 Arts and Humanities

Keywords and phrases Text analysis, automated feature extraction, folklore, narrative, Russian folktales

Digital Object Identifier 10.4230/OASIcs.CMN.2016.7

1 Motifs as a Source Cultural Information

Motifs are distinct, recurring narrative elements found in folklore and, more generally, cultural materials. Motifs are interesting because they provide a compact source of cultural information: many motifs concisely communicate a related constellation of cultural ideas, associations, and assumptions. For example, "troll under a bridge" is an example of a motif common in the west. To members of many western cultures, this combination entails a number of related ideas that are by no means directly communicated by the surface meaning

* This research was made possible by an FIU's Presidential Fellowship and FIU's SCIS's Director's Fellowship, both awarded to W. Victor H. Yarlott. This work was also partially supported by National Institutes of Health (NIH) grant number 5R01GM105033-02.

© W. Victor H. Yarlott and Mark A. Finlayson;
licensed under Creative Commons License CC-BY

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 7; pp. 7:1–7:10
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of the words: the bridge is along the critical path of the hero, and he must cross to achieve his goal; the troll often lives under the bridge, crawling out to waylay innocent passers-by; the troll charges a toll or exacts some other payment for crossing the bridge; the troll is a squatter, not the 'officially' sanctioned master of the bridge; the troll enforces his illegitimate claim through threat of physical violence; and the hero often ends up battling (and defeating) the troll instead of paying the toll.

Because of this density of information, motifs are often retained as a tale is passed between cultures and down generations, and folklorists have observed that a tale's specific composition of motifs can be used to trace the tale's lineage [35, Part 4, Chapter V]. This has led folklorists to construct motif indices that identify motifs and note their presence in specific tales (usually as represented in a particular folkloristic collection). The most well-known motif index is the Thompson motif index (TMI) [34], which references tales from over 614 collections, indexed to 46,248 motifs and sub-motifs, 41,796 of which have references to tales or tale types. In Thompson's index each motif is given a designating code; for example, "troll under a bridge" is referenced by the codes G304 and G475.2. In this case, "troll under a bridge" is represented by two motifs as Thompson generalizes trolls to ogres, a general class of monstrous beings; thus, the motifs are "troll as ogre" (G304) and "ogre attacks intruders on bridge" (G475.2).

While Thompson's index is the best known, there are many other motif indices targeting specific cultures and periods, for example, early Irish literature [8], traditional Polynesian narratives [27], or Japanese folk-literature [21]. In addition, the idea of motif was incorporated into another useful notion, the *tale type*, which seeks to classify whole tales based on their motifs. Antti Aarne constructed an index of tale types in 1910 [1], with translations and revisions by Thompson [34] and Uther [37] (the last being known as the *ATU* catalog).

Thompson informally defines a motif as items "worthy of note because of something out of the ordinary, something of sufficiently striking character to become a part of tradition, oral or literary. Commonplace experiences, such as eating and sleeping, are not traditional in this sense. But they may become so by having attached to them something remarkable or worthy of remembering" [34, p. 19]. He notes that motifs generally fall into one of three subcategories: an event, a character, or a prop [35, pp. 415–416]. Here we give an example of each (with their associated Thompson's motif code):

A **hero rescuing a princess** (B11.11.4) is perhaps one of the most well-known event motifs in western culture. Ask a westerner the following question: "A princess has been kidnapped: who kidnapped her, who rescues her, and what does the rescuer need to do to effect a rescue?", and common answers will be "a dragon kidnapped her, the knight must rescue her, and he must kill the dragon." This motif may be the climax of the story, with a "happily ever after" ending just after the hero defeats the dragon, or it may just happen in the course of a story: in *Ivan Dogson and the White Polyanin*, a Russian folktale [2, Tale #139], Ivan slays three dragons, each with more heads than the last, rescuing a princess each time. The motif is prolific, found across the tales, literature, and movies of many cultures.

Old Man Coyote (A177.1) is a character motif: known in some Native American Indian tribes as Coyote, he is one of the most recognizable gods. In Native American Crow folklore, Old Man Coyote creates the earth and all the creatures on earth. He travels the world, teaching the animals how they should behave. Old Man Coyote, however, is far from a noble and elegant creator. He creates ridiculous costumes and tries to trick the Crow tribe into wearing them, only to be run off. He purposefully bungles rituals to produce food, such as transforming skin from his back to meat, in order to guilt his guests into performing the ritual correctly to get free food, later performing it correctly to discredit his former guests when they tell others he erred. Anywhere Old Man Coyote is referenced, he calls to mind someone who has done great things, but is lazy and often far too clever for their own good, falling prey to their own cunning.

A **magic carpet** (D1155) is a prop that allows the hero to fly through the sky, and is familiar to anyone who has watched Disney's *Aladdin*. In *One Thousand and One Nights*, Prince Husain encounters a merchant selling a carpet for an outrageous price; the merchant says: "O my lord, thinkest thou I price this carpet at too high a value? ...Whoever sitteth on this carpet and willeth in thought to be taken up and set down upon other site will, in the twinkling of an eye, be borne thither, be that place nearhand or distant many a day's journey and difficult to reach" [6, p. 496]. Solomon, said to be the third king of Israel, was said to have a carpet 60 miles on each side that could transport him vast distances in a short amount of time. In Russian hero tales, magic carpets are common items that aid the hero in his quest.

While the above examples are drawn from folklore, motifs have importance beyond folktales: they occur in common tales, news stories, press releases, propaganda, novels, movies, plays, and anywhere that cultural materials are found. A powerful recent example is the use of the *Pharaoh* motif in modern middle eastern discourse. The Pharaoh appears in Qur'an, and comes into conflict with Moses and his attempts to free the Hebrews from Egyptian slavery. The Pharaoh is an arrogant and obstinate tyrant who defies the will of God and is punished for it. In modern Islamist extremist narratives, the Pharaoh is a symbol of struggles against anti-Islamic regimes and has been invoked against leaders such as Anwar Sadat of Egypt, Ariel Sharon of Israel, and George W. Bush, whom Osama bin Laden referred to as the "pharaoh of the century" [19].

Because of their prominence and ubiquity, the ability to automatically identify and extract motifs would open up a vast repository of important cultural information to computational analysis. Currently, motifs must be extracted by hand by trained cultural experts, and any indices manually constructed, with all the attendant delay, error, and incompleteness. Motifs are rarely identified explicitly in new stories, due to size of the indices and the amount of work involved: developing automated motif extraction would allow extensive, exhaustive identification of motifs across textual cultural materials, and allow us to apply all the power of statistical machine learning and related techniques on this new wealth of information.

In this work, we outline a supervised approach to solving this problem. We use supervised techniques in service of our goal of learning a model of how folklorists create and understand motifs: the ubiquity of motifs suggests that there may be some interesting cognitive processes at work and in modeling them we may get closer to understanding these underpinnings.

2 Motif Extraction: Problems

There are a number of barriers to overcome in automatically extracting motifs.

First, current motif indices are **incomplete** and **inconsistent**. Many interesting narrative elements are not listed in the existing motif indices, and the motifs that are listed are not identified at a uniform level of abstraction. This means that extracting motifs from even well-studied materials (such as folklore) is not just a matter of looking for motifs listed in the index. Rather, the problem is two-fold: motif-like elements must be identified within a text and it must be determined whether they represent an existing motif, a specialization or variant of an existing motif, an entirely new motif, or a spurious false positive.

Second, there is a **lack of data**. No one, to our knowledge, has undertaken even the most basic annotation of text with motifs. Therefore we have no data to which machine learning techniques can be immediately applied to quickly make progress. As is well known, generating manually annotated data is a time-consuming and labor-intensive process [20], making it difficult to learn what a motif is from the text of the narrative itself.

Third, and even more fundamentally, motifs are **ill-defined**. There is no formal definition of motifs and current definitions fall short of the specificity needed for computational work. Further, folklorists have not laid out the principles behind motif identification, nor do we understand the cognitive principles which would drive people to naturally identify motif-like information. We believe, however, that there are some underlying principles, as motifs are not only transmitted culture to culture, but often arise independently between cultures.

2.1 Defining Motifs

In this project we seek to overcome these barriers to demonstrate that automated motif identification, extraction, and annotation is feasible. The first step in this process is to tighten up our definition of motifs, ideally creating a formal model which describes exactly what a motif is. While we do not present a formal model here, we lay the groundwork by identifying the features of motifs such a model would need to address.

Thompson defined motifs as *something remarkable or out of the ordinary*: eating is not a motif, but eating from a magical table is. Even so, Thompson described his analysis as selecting anything he felt was of interest to future scholars, which suggests a somewhat less principled and more intuition-driven approach. From Thompson's discussion on motifs, a concise version of Thompson's definition might be: a motif is any remarkable or non-commonplace element in a story. We get a definition of "element" from *The Folktale*, where Thompson defines the classes most motifs fall into as actors, items, and single incidents [35, pp. 415–416]. Within this paper, we refer to these elements as characters, props, and events, respectively.

This simple first attempt at a definition has some problems. What does *remarkable or non-commonplace* mean? Practically, it means that the element that comprises a motif is not an unremarkable, everyday narrative element, such as eating or an ordinary table. In folklore, such commonplace elements are often excluded (or not consistently retained) as they are not interesting enough to be retained over generations of retellings. Motifs are maintained across many variations of the same tale because they carry culturally relevant, interesting information. Further, even if a commonplace element is inserted into a particular telling, these elements are not likely to show up consistently across tales with the same tale type, suggesting that it would be possible to smooth out remaining commonplace elements from a selection of tales.

Second, the definition does not address the appropriate level of abstraction for motifs. On the one hand, many motifs as listed in existing motif indices are highly specific: for example, a runner who keeps his leg tied up to prevent himself from running away (F681.1), as opposed to the individual presence of a *runner*, *tying a leg*, or *prevention of running away*. This suggests that motifs should tend toward more specific forms that are repeated across tales: for example, "eating oneself up" (F1035) would be preferred over "eating."

On the other hand, motifs often have closely-related variants that lead to the creation of more abstract entries. For example, in the Russian folktale *Bukhtan Bukhtanovich* [18, p. 168], a fox tricks the Tsar into thinking Bukhtan is very wealthy by pretending to have to measure Bukhtan's money using a large bucket. In Thompson's motif index, there is only an entry about a *cat* using such a trick (K1954.1). In this particular case, while we could imagine creating a new motif specific to the *fox*, other examples suggest a preference for generalizing the existing motif or grouping the motifs together under a category like *animal* *uses the measuring trick*. Examples of both methods can be found in Thompson's index: *kindness rewarded* (Q40) is a single motif entry with ten sources cited and no submotifs, but *conception from eating animal* (T511.5) has four motifs as children (T511.5.1–T511.5.4).

One important note is that motifs are not necessarily constitutive elements – that is, the presence of absence of a motif is not definitional for the identity of a particular tale. Motifs, rather, impose a "family resemblance" relationship between different version of the same tale. For example, in the well-known tale *Cinderella* [7], found across many different cultures, several motifs commonly recur across retellings: three evil step-sisters, a fairy godmother, a glass slipper, and so forth. But the story will continue to be recognizable as *Cinderella* if the pumpkin carriage (F861.4.3 – *Carriage from pumpkin.*) is replaced by another means of transportation or does not appear at all. A story having all the motifs of other tales of the same type is sufficient, but not necessary, for it to be recognized as a member of that tale type. Fisseni and Lawrence [17] have shown results where, in some cases, modifying the motifs involved may result in a story very similar to the original in what they refer to as a "simple solution to the problem of integrating the proposed change" (p. 103). Ignoring these non-constitutive motifs smooths over details that may potentially contain cultural information and, thus, is not in our interests.

Jason [23] makes an effort to more clearly define motifs, leveling similar complaints on the clarity of Thompson's definition of motifs to those in this paper. Jason provides a definition of motifs as narrative elements that meet the following criteria: they must be (1) the simplest unit of content that fill a primary formal slot of literary structure (a character or deed) and (2) context-free (not belonging to a certain plot). There are issues with this definition. Jason does not appear to define what simplest means beyond filling a slot of literary structure. Restricting motifs to characters or deeds ignores the importance of props within a story, such as magic carpets (D1155). And context-free motifs ignore the vast wealth of cultural knowledge that motifs contain: to encapsulate cultural knowledge, motifs necessarily arise from related tales (a tale type) within a culture.

To address the concerns raised by other definitions, we propose our own definition based on Thompson's original definition: "A motif is a set of closely-related variants of a non-commonplace, specific narrative element that is repeated across tales of the same type." In future work we intend to specify the components of this definition more formally; below, in discussions of our experimental procedure, we indicate how we might do this.

3 Experimental Design & Pilot Work

3.1 Goals

We have two general goals for our experimental work. While we do not achieve either of these goals in this paper, we do make substantive progress toward them, identifying key data, revealing hard problems, and sketching implementations of solutions.

The first goal is to develop a system to extract motifs from narrative text. Automatic extraction is the ability of a system to identify and extract motif-like elements from a raw-text document. We do not expect that the system be capable of assigning a proper, descriptive name for each motif, but it should be capable of grouping of individual occurrences of motifs together (that is, clustering tokens into types), as well as clustering motifs by topic.

The second, longer-term goal is to learn a model of what folklorists think a motif is: that is, how folklorists define motifs, how they determine what elements of a story comprise a motif, and how they extract motifs from narrative texts. We would expect our model to be capable of examining a narrative and identify the same motifs that folklorists identify.

Importantly, the motifs identified by our system in the first goal, and the motifs identified by folklorists in the second goal, will not necessarily be the same. As with all people, folklorists are fallible, prone to errors of commission, omission, and inconsistency. We will seek to expose the basic principles used by folklorists to uncover motifs, apply those principles uniformly, and show how the computational approach can add value to the manual approaches of folklorists by correcting errors, filling gaps, and enforcing consistency.

3.2 Experimental Procedure

We outline here an experimental procedure using supervised techniques that could be used to accomplish our first goal. While we have not implemented this whole procedure, we have made concrete progress on a number of steps as discussed in later subsections.

1. **Input of Raw Narrative Text.** Texts containing narratives (folklore or other cultural materials) are input to the system.
2. **Initial Processing via NLP Pipeline.** The texts are processed by a natural language processing (NLP) pipeline that performs common analyses such as tokenization, lemmatization, part of speech tagging, chunking, syntactic parsing, word sense disambiguation, latent semantic analysis, semantic role labeling, and event detection [15, 16, 24].
3. **Grouping by Term Distribution.** Using term frequency–inverse document frequency (tf–idf)[32, 33], the system will identify the most important terms in each narrative document. The system will sort the narratives into rough similarity groups based on these terms and annotate each document with the group to which it belongs. Another option for this step is to do topic modeling [4, 28], to cluster texts into groups by topic distribution. This step enables the system to smooth out commonplace events, as described in step 6.
4. **Candidate Identification.** For each text, the system will identify spans of text that could potentially be motifs. Spans of text will be identified that meet the general criteria of a motif, in that they are a narrative “element”, often indicated by an event or a nominal representing a character or prop. Another strategy would be to look for common, important terms using tf-idf and then attempting to expand a window around these terms. This would allow the system to prefer *three-headed dragon* over *three-headed* or *dragon* if *three-headed dragon* appeared in multiple tales.
5. **Candidate Classification.** Then, within each group identified above, the system will assess the commonality of each identified span, classifying them as motif or not using either a rule-based system or a machine classifier trained on annotated data. A rule based approach, for example, could look for cut-off points in the tf–idf score distribution for a group. A machine learning approach could use features learned from texts manually annotated with the presence of motifs.
6. **Commonplace Event Elimination.** At this stage, candidate motifs can be compared across groups. When a candidate motif appears across multiple tale groups, it is biased against, as these are more likely to be commonplace events, such as eating or sleeping. Candidate motifs that fall below a threshold (either hard-coded or learned) will be eliminated. The remaining motifs candidates are graduated to identified motifs.
7. **Motif Alignment.** The system attempts to group motifs together into variant groups: groups of closely-related motifs that would be subtypes of a single motif in a motif index using semantic role labeling and the relations catalogued in WordNet [36]. For example, if *cat uses the measuring trick* and *fox uses the measuring trick* were both identified, they would be grouped together as a single variant. Narratives are annotated with both the specific motif and with the variant group that these motifs belong to.

8. **Evaluation.** We will evaluate the performance of the system by comparison with manually annotated motif data, using agreement statistics such as the F_1 measure or Fleiss' kappa. Throughout the project we will consult with folklorists to check our work. Specifically, we plan to consult with them after developing a model, when annotating stories, when determining unique motifs, and as we continue to develop the system. This is a crucial step in achieving our second goal of understanding how folklorists identify and extract motifs from narratives and enabling an automated system to do the same.

3.3 Structured Parsing of Thompson's Index

Several keys steps of the experimental design (e.g., steps 4, 5, and 8) require annotated data in the form of a motif index applied to actual text. Therefore we have been working on generating a structured, electronic version of the most comprehensive motif index available, Thompson's motif index [34]. An electronic form of such a resource would allow us to map motifs to the individual stories that contain them, and provide a starting point for manual annotation. Such a resource would allow us, for example, to identify a dense subnetwork of motifs and narratives that would be used as a pilot corpus for testing and training. Thompson's motif index is one of the most widely-used sources for motif information and having it in a structured, easily queryable form will be a great benefit to the community at large.

We have uncovered numerous challenges in this apparently simple task: first, there is no high-quality digitized version of Thompson's motif index. One commonly cited online source, hosted at `Ruthenia.ru` [31], a joint effort between Moscow-based publisher OGI and the Department of Russian Literature at Tartu University to provide sources for Russian language research, has inconsistent HTML and numerous OCR errors that makes parsing of the index difficult. The MOMFER effort to parse the motif index with the intention of creating a search engine [26], provides code for parsing the HTML motif index hosted at Ruthenia.ru, but is incomplete and does not accurately parse large parts of the index.

Even if we had a pristine digitization of Thompon's index, the text suffers from inconsistent formatting, abbreviations, and typographical conventions throughout. Delimiting bibliographic sources is inconsistent: semi-colons are usually a delimiter, but Thompson also uses commas and periods to delimit motifs. This is particularly troublesome, as commas are used to separate multiple references within a single collection ("Grimm Nos. 3, 35, 81 ...") and periods are used to abbreviate and end entries.

Through this effort, other issues with Thompson's motif index have come to light: many of his references to "tales" are simply cross-references to other collections (such as Cross' index of Irish literature [8] or Boberg's index of Icelandic literature [5], among many others). Thus the index does not provide in many cases a direct link between motifs and tales: many stories are cited for only a single motif, despite containing more. Many of the cited stories and collections are hard to find or may no longer be accessible. Due to these issues, the motif index will likely not provide a solution to the initial problem we identified: the need for a corpus with many related motifs.

Despite all the challenges and issues with Thompson's motif index, we are developing a sequence of regular expressions to handle parsing of the index into structured form.

3.4 Preliminary Analysis

Even with a motif index, comprehensive or not, to train and test a motif extraction system we would need the motifs annotated on actual text. To that end, we have completed an initial development annotation of the motifs in fourteen tales[1] using Thompson's motif index as a resource. Thompson's motif and tale type index contain information taken from Andreev's tale-type index for Russian tales [3, 22], making Thompson's index a suitable reference. The tales we annotated are English translations of Aleksandr Afanase'ev's collection of Russian folklore [2], chosen in part because of the large body of work already related to them due to their prominence in Vladimir Propp's Morphology of the Folktale [30].

This preliminary annotation serves to inform us as we develop an annotation scheme for motifs, a necessary step in developing a corpus of stories. In the near future, we plan to consult a folklorist regarding our motif analysis, formalize our annotation scheme, and fully annotate a substantive set of stories Story Workbench, a narrative annotation tool [15, 16]. These stories will form a gold standard pilot corpus for motif annotation.

4 Related Work

On the folklore side, there are many motif indices, with the Aarne-Thompson Motif-index of Folk-literature [34] being the primary resource. There are numerous other indices, most primarily focused on a specific culture. Thompson also has substantial discussion on motifs and the compilation of indices in his book *The Folktale* [35]. While Thompson's motif index is perhaps the primary source of motif information used today, it has been criticized because of overlapping motif subcategories, censorship (primarily of obscenity), and missing motifs [14].

Additionally, much work has been done identifying tale-types: recent work by Hans-Jörg Uther expands and improves upon the Aarne-Thompson tale classification system, resulting in the ATU classification system [37].

Darányi [9] has called attention to the need for research into the automation of extraction and annotation of motifs in folklore. Further work by Darányi and Forro [10] has determined that motifs may not be the highest level of abstraction in narrative, Darányi *et al.* [11] have made substantial headway towards using motifs as sequences of "narrative DNA", and Ofek, *et al.* [29] have demonstrated learning tale types based on these sequences. Declerck *et al.* [13] have also done work on converting electronic representations of TMI and ATU to a format that enables multilingual, content-level indexing of folktale texts, building upon past work [12]. Currently, this work appears to be focused on the descriptions of motifs and tale types, without reference to the stories.

With regard to analyzing motif annotation schemes, Karsdorp *et al.* [25] present an analysis of the degree of abstraction present in the ATU catalog and the methods used to note what motifs belong to a given tale type. They find the ATU annotation insufficient for analyzing recurring motifs across types, in that it the ATU scheme fails to capture commonalities across closely related types.

5 Contributions

The information content and ubiquity of motifs makes them an important consideration for anyone working with culturally-influenced narratives. In this paper we have motivated a deeper computational look at motifs, and outlined an experimental plan for developing

[1] The fourteen tales analyzed were: Bukhtan Bukhtanovich (#163); Kozma Quickrich (#164); Shabarsha the Laborer (#151); The Serpent and the Gypsy (#149); Burenushka, the Little Red Cow (#101); Wee Little Havroshechka (#100); Ivan Popyalov (#135); Ivan the Bull's Son (#137); Ivan the Peasant's Son and the Thumb-sized Man (#138); The Flying Ship (#144); Ivan the Cow's Son (#136); Nikita the Tanner (#148); The Magic Swan Geese(#113); The Crystal Mountain (#162).

a system that automatically extracts motifs from text. Importantly we have identified numerous barriers: the lack of annotated data, the incompleteness of existing motif indices, the need for a formal model of motifs, and the lack of clarity into the cognitive processes that lead to the generation and identification of motifs in narrative.

We have already made progress toward addressing these problems. We identified several features of motifs that will form the foundation of a formal, computational model, a necessary step towards motif extraction; we have made significant progress in understanding how to parse Thompson's motif index into a structured resource; we have analyzed fourteen stories for motifs, which has revealed several important questions with regard to how motifs should be appropriately reliably annotated.

Motifs are key features in culturally-relevant narratives, and we seek to enable access to this vast resource and open up a new dimension in computational narrative analysis.

References

1 Antti Amatus Aarne. *Verzeichnis der Märchentypen*. Suomalainen tiedeakatemia, 1910.

2 Aleksandr Nikolaevich Afanas'ev. *Narodnye Russkie Skazki*. Moscow: Gos. Izd-vo Khudozh Lit-ry., 1957.

3 Nikolaĭ Petrovich Andreev, Antti Aarne, and Heda Jason. *Index of Tale-plots According to the System of Aarne*. Rand Corporation, 1968.

4 David M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

5 Inger Margrethe Boberg. *Motif-index of early Icelandic literature*. Munksgaard, 1966.

6 Richard Francis Burton. *The Arabian nights*. Barnes & Noble, 2009.

7 Marian Roalfe Cox. *Cinderella: Three hundred and Forty-Five Variants of Cinderella, Catskin, and Cap o'Rushes*, volume 31. Folklore Society, 1893.

8 Tom Peete Cross. *Motif-index of early Irish literature*. Indiana University, 1952.

9 Sándor Darányi. Examples of Formulaity in Narratives and Scientific Communication. In *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, pages 29–35, 2010. URL: `http://ilk.uvt.nl/amicus/amicus_ws2010_proceedings.html`.

10 Sándor Darányi and László Forró. Detecting Multiple Motif Co-occurrences in the Aarne-Thompson-Uther Tale Type Catalog: A Preliminary Survey. *Anales de Documentación*, 15(1), 2012. URL: `http://revistas.um.es/analesdoc/article/view/analesdoc.15.1.134691/131801`.

11 Sándor Darányi, Peter Wittek, and László Forró. Toward Sequencing "Narrative DNA": Tale Types, Motif Strings and Memetic Pathways. In Mark A. Finlayson, editor, *Third Workshop on Computational Models of Narrative (CMN)*, pages 2–10, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).

12 Thierry Declerck and Piroska Lendvai. Linguistic and semantic representation of the thompson's motif-index of folk-literature. In *Research and Advanced Technology for Digital Libraries*, pages 151–158. Springer, 2011.

13 Thierry Declerck, Piroska Lendvai, and Sándor Darányi. Multilingual and Semantic Extension of Folk Tale Categories. In *Proceedings of the 2012 Digital Humanities Conference (DH 2012)*, 2012. URL: `http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/multilingual-and-semantic-extension-of-folk-tale-catalogues/`.

14 Alan Dundes. The motif-index and the tale type index: A critique. *Journal of Folklore Research*, pages 195–202, 1997.

15 Mark A. Finlayson. The story workbench: An extensible semi-automatic text annotation tool. In *Intelligent Narrative Technologies*, 2011.

16 Mark Alan Finlayson. Collecting semantics in the wild: The story workbench. In *AAAI Fall Symposium: Naturally-Inspired Artificial Intelligence*, pages 46–53, 2008.

17 Bernhard Fisseni and Faith Lawrence. A Paradigm for Eliciting Story Variation. In *Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13)*, volume 32, pages 100–105. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013.

18 Norbert Guterman. *Russian Fairy Tales.* Pantheon Books, 1973.

19 Jeffry R. Halverson, Steven R. Corman, and H. L. Goodall Jr. *Master narratives of Islamist extremism.* Palgrave Macmillan, 2011.

20 Nancy Ide and James Pustejovsky, editors. *Handbook of Linguistic Annotation.* Springer, 2016. Forthcoming.

21 Hiroko Ikeda. *A type and motif index of Japanese folk-literature.* Orient Cultural Service, 1971.

22 Heda Jason. *NP Andreev,'Index of Tale-Plots According to the System of Aarne': A Partial Translation.* Rand Corporation, 1968.

23 Heda Jason. About 'motifs','motives','motuses','-etic/s','-emic/s', and 'allo/s-', and how they fit together. an experiment in definitions and in terminology. *Fabula*, 48(1-2):85–99, 2007.

24 Daniel Jurafsky and James H. Martin. *Speech and Language Processing.* Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

25 FB Karsdorp, P Kranenburg, Theo Meder, Dolf Trieschnigg, and A Bosch. In search of an appropriate abstraction level for motif annotations. In *Proceedings of the 2012 Workshop on Computational Models of Narrative*, 2012.

26 Folgert Karsdorp, Marten van der Meulen, Theo Meder, and Antal van den Bosch. Momfer: A search engine of thompson's motif-index of folk literature. *Folklore*, 126(1):37–52, 2015.

27 Bacil F. Kirtley. *A motif-index of traditional Polynesian narratives.* University of Hawai'i Press, 1971.

28 Jon D. Mcauliffe and David M. Blei. Supervised topic models. In *Advances in neural information processing systems*, pages 121–128, 2008.

29 Nir Ofek, Sándor Darányi, and Lior Rokach. Linking Motif Sequences with Tale Types by Machine Learning. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13)*, volume 32, pages 166–182, Hamburg, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.CMN.2013.166.

30 Vladimir Propp. *Morphology of the Folktale*, volume 9. University of Texas Press, 1968.

31 Ruthenia. S. Thompson. Motif-index of folk-literature. http://www.ruthenia.ru/folklore/thompson/. Accessed: 2016-03-09.

32 Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

33 Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

34 Stith Thompson. *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books and local legends*, volume 4. Indiana University Press, 1960.

35 Stith Thompson. *The folktale.* University of California Press, 1977.

36 Princeton University. About wordnet, 2010. Retrieved on May 9, 2016 from: http://wordnet.princeton.edu.

37 Hans-Jörg Uther. *The types of international folktales: a classification and bibliography, based on the system of Antti Aarne and Stith Thompson.* Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, 2004.

ProppML: A Complete Annotation Scheme for Proppian Morphologies

W. Victor H. Yarlott[1] and Mark A. Finlayson[2]

1 **School of Computing & Information Sciences, Florida International University, Miami, FL, USA**
`wvyar@cs.fiu.edu`
2 **School of Computing & Information Sciences, Florida International University, Miami, FL, USA**
`markaf@fiu.edu`

Abstract

We give a preliminary description of ProppML, an annotation scheme designed to capture all the components of a Proppian-style morphological analysis of narratives. This work represents the first fully complete annotation scheme for Proppian morphologies, going beyond previous annotation schemes such as *PftML*, ProppOnto, Bod *et al.*, and our own prior work. Using ProppML we have annotated Propp's morphology on fifteen tales (18,862 words) drawn from his original corpus of Russian folktales. This is a significantly larger set of data than annotated in previous studies. This pilot corpus was constructed via double annotation by two highly trained annotators, whose annotations were then combined after discussion with a third highly trained adjudicator, resulting in gold standard data which is appropriate for training machine learning algorithms. Agreement measures calculated between both annotators show very good agreement ($F_1 > 0.75$, $\kappa > 0.9$ for functions; $F_1 > 0.6$ for moves; and $F_1 > 0.8$, $\kappa > 0.6$ for dramatis personae). This is the first robust demonstration of reliable annotation of Propp's system.

1998 ACM Subject Classification I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Narrative structure, Computational folkloristics, Russian folktales

Digital Object Identifier 10.4230/OASIcs.CMN.2016.8

1 Introduction

Vladimir Propp's *Morphology of the Folktale* is Propp's attempt to "make an examination of the forms of the tale which will be as exact as the morphology of organic formations" [39, p. xxv]. While previous work on motifs and tale types by Roman Volkov [44] and Antti Aarne [1] made attempts at creating a system for expressing and describing components of folktale, Propp dismissed these as both unscientific and suggesting the incorrect notion that there is a clear-cut division into types. As the first example of its kind, Propp's work aimed to capture the formulaic repetition that is present in folktales in a precise and relatively formal way.

With today's more advanced mathematical and conceptual machinery, Propp's theory can be described as a *plot grammar*, which lists the elemental plot pieces and their possible orders that may occur in folktales (Propp called his plot pieces *functions*). Propp's theory also describes the high-level organization of tales (*moves*), the instantiated forms of functions (what we have called elsewhere *function subtypes*), long-distance dependencies between function subtypes, exceptions and other complications (*inversions* and *trebling*), and character archetypes (*dramatis personae*), which correlate with particular functions. To derive and

© W. Victor H. Yarlott and Mark A. Finlayson;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 8; pp. 8:1–8:19
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

provide evidence for his theory he analyzed approximately 100 Russian hero tales drawn from the collection of Aleksandr Afanas'ev [2], providing for each tale a list of functions that appear and the order of their combination. Propp noted that his particular list of functions and subtypes, and their orders, was applicable only to this specific set of Russian tales; others, such as Alan Dundes [8] and Benjamin Colby [6], have shown, however, that Proppian-style analyses may profitably be performed for other cultures.

Propp's morphology is a seminal work in the field of folklore and narratology, inspiring many counter-reactions [31, 32] and extensions [8, 25], and it continues to inspire generations of scholars. Within the field of computational narrative, Propp's morphology has a substantial number of potential applications for three reasons. First, Propp's morphology is one of the most formal narratological treatments developed so far, having a relatively clear method for determining and extracting theory components from text. Second, Propp's work separates content from form, allowing a description and analysis of a plot without requiring its instantiation directly into language. These two points combine to make Propp's morphology readily applicable to the creation of computational models. Third, properties of Propp's morphology—that functions always occur in the same sequence—makes them a powerful tool for story generation. Propp's work has been applied in systems as diverse as textual story generation [12, 23, 24], support and guidance for children during story creation activities [33], and as a means of varying sign language for virtual characters based on conflict to enhance computer generated sign language [41].

There have been a number of objections to Propp. Propp's work has been criticized for its separation of form from content, treating content as "less important" than form [32, p. 179], making the mistake of trying to "characterize a tale without mentioning the motifs" [5, p. 194]. His work has been considered inconsistent, because even as he attempts to separate content from form, morphological criteria "reintroduce some aspects of content" [32, p. 179]. Additional criticisms target the reproducibility of Propp's work [4] and that Propp's work is not sufficient to account for the diversity of plots in fairy tales [5].

Surprisingly, despite the deep interest in Propp's work in computational narrative circles, one of the most fundamental questions regarding Propp's morphology has not been satisfactorily answered: taking Propp's list of functions as a given, is the morphology specified precisely enough that an independent person will find, in the same tales, the same functions that Propp did?

In this work we address this question. We designed a complete annotation scheme for Propp's morphology, which goes beyond all previous annotation schemes to capture every piece of Propp's morphology in a linguistically sophisticated way: past work (e.g., [35]) does not address implicit functions, dramatis personae, or moves, nor does it provide a fully annotated corpus. We ran a double-annotation experiment where we extensively trained three people over several months in the workings of Propp's system (especially the list of functions and subtypes), and had them annotate 15 single-move tales drawn from Propp's original corpus. It should be noted that while these tales are all technically single-move, in accordance with Propp's original annotations, preparatory functions are not a part of the main move and thus comprise, in effect, their own 'preparatory' move. Agreement measures calculated over this set of data shows that Propp's system can be reliably applied by human annotators to single-move tales. While prior work has addressed this question (e.g., [4]), prior work has failed to adequately train annotators, our work addresses this shortcoming and is the first to robustly demonstrate the reliability of applying Propp's system. This result shows that human annotation of Propp's morphology on single-move tales has a high agreement ($F_1 > 0.75$ and $\kappa > 0.9$ for functions, $F_1 > 0.6$ for moves, $F_1 > 0.8$ and $\kappa > 0.6$

for dramatis personae). In general, these results are very good, with the κ for functions being "near perfect." Annotation and analysis of multi-move tales is ongoing, but show the same trends and will be reported in future articles.

Our work lays the foundation for the machine learning of Proppian functions. Preliminary versions of the data presented here [13] has already been used to demonstrate that learning Propp's function from text is possible [16].

1.1 Outline of the Paper

In §2, we describe how the experimental design of double-annotation addresses the question at hand. Then, in §3, we explain how our annotation scheme improves on prior work to design of our annotation scheme, and give a high-level description of the scheme with an example annotation. We give a full formal BNF specification of the scheme in Appendix A. In §4, we describe the selection of data, the annotation procedure (including training of the annotators), and the tools used during annotation. In §5 we show the results and discuss the agreement measures. In §6 we cover related work, including prior attempts at answering the experimental question.

2 Experimental Design

The question at hand is whether Propp's theory is *reliable*. By *reliable*, we mean something quite specific: will independent people agree with each other when applying Propp's theory? There are several ways to define "agree with each other," in particular:

Q1. Taking as given Propp's general approach, list of functions, and identified functions in specific tales, will independent people agree with each other as to where and whether those functions appear in those tales?

Q2. Taking as given Propp's general approach and list of functions, will independent people agree with each other, and also agree with Propp (as appropriate), when asked to find his functions in tales?

Q3. Taking as given Propp's general approach, will independent people agree with each other, and with Propp (as appropriate), as to the set of functions that are indicated by a particular set of tales?

As can be seen the level of generality of the questions progresses from quite specific in Q1 to fairly general in Q3. In this work we address Q1 with regard to single-move tales. We will address Q1 for multi-move tales and Q2 generally in future work. As for Q3, others have addressed this manually [9, 6], and we have begun to address it computationally [16].

What experimental design is appropriate to answer Q1? As pointed out by Bod *et al.* [4], a double-annotation paradigm is one appropriate approach to addressing the reliability of a textual marking scheme, and is the approach we follow here. In a double-annotation experiment, two people are trained in the operation of the scheme (these are called the *annotators*), and are asked to independently mark up texts with the scheme. The agreement between the two sets of markings created by the annotators is measured using appropriate statistical agreement measures such as the F-measure [20, 43] or Fleiss' kappa [19]. High agreements indicate a positive answer to the question. Further, a gold standard marking of the texts (suitable for machine learning) can be generated by having the two annotators confer to resolve disagreements, sometimes assisted by a third party (called the *adjudicator*).

Conducting such a double-annotation experiment entails the following steps:

1. Define an appropriate and complete annotation scheme. (§3)
2. Select the texts to be annotated. (§4.1)
3. Assemble or build the tools required to do the annotation. (§4.2)
4. Train the annotators and adjudicator in both the scheme and the tools. (§4.3)
5. The annotators perform the annotation. (§4.4)
6. Measure agreement between the annotators. (§5)
7. *Optional:* The adjudicator eliminates disagreements to generate gold standard data. (§4.4)

In reality, there is often a loop between steps 6 and 1, as noted elsewhere [40, 17], because analysis of the data reveals flaws in the scheme, which requires revision and a repeat of the experiment. We had already progressed through such a loop several times during previous attempts at annotation of Propp's scheme, and we discuss the lessons we learned in those loops, and how they were integrated into this final scheme, in the next section.

3 Design of ProppML

The point of an annotation scheme is to capture, in a precise way, all the different "moving parts" of the phenomena to be annotated. There should be a way of notating every important distinction provided by the theory that backs the scheme, and, for text annotation schemes, associating those notations to the relevant spans of text.

In our approach to annotating Propp's morphology, we split the task into three separate schemes: functions, moves, and dramatis personae. Collectively we refer to these as *ProppML*. Although the schemes are separate, they do cross-reference each other in specific places (i.e., the move scheme refers to the function scheme), as well as reference other related annotation schemes as described below.

All three schemes allow association of Propp's theoretical constructs of functions, moves, and dramatis personae with the text under consideration. This association is implemented by reference to character offsets, anchored by identified token boundaries. As described in [13], the texts are first run through a tokenizer (in this case, from the Stanford CoreNLP suite [36]) that splits the text into single tokens. These tokens are indexed to character offsets in the text as described in Appendix A. The ProppML schemes then refer to the tokens by unique id numbers.

We sketch out the important parts of ProppML in each of the three following subsections. The full details for each of these parts are found in Appendix A.

3.1 ProppML: Function Scheme

The first of the three schemes allows an annotator to notate the presence of functions in the text. Functions are split into two portions: the **function tag** and one or more function **instances**. The function tag allows the annotator to mark three pieces of information: the **type** of the function, whether or not the function is **inverted**, and the **symbol** for the function.

The function type marks the function as a *Normal*, a *Prepatory* function, or the *Initial* situation. Separating these out is important because Propp does not specifically note where the Preparatory and Initial functions occur, and preparatory and initial functions do not participate in the normal Move structure of the tale [39, pp. 108–109].

The primary symbol of the function is one of the most memorable parts of Propp's theory. He identified, for example, the function *Villainy*, and gave it the symbol *A*. Propp gave the function *Reward* the symbol W. Importantly, Propp also identified various function subtypes, indicated by a combination of superscripts and subscripts on the primary symbol. Our scheme allows the symbol and its super- and sub-scripts to be marked by alphanumeric strings in the appropriate positions.

One of the important innovations of our marking scheme is the notion of a **function instance**. In our early approaches to marking Propp [13], when Propp indicated a function was present, we noted that there often was not one unambiguous occurrence of that function. Propp himself admited this when he identified the phenomenon of *trebling*, which is where a particular function is repeated three times. We found that repetition of this sort occurred in many places where Propp did not indicate it. Our scheme allows these occurrences to be marked as separate "instances" of the same function, which is important because instances of one function can interleave with instances of other functions when the functions are repeated in sequence. For example, in The Magic Swan Geese, the little girl encounters three potential helpers (the stove, the tree, the river) at three separate times both when chasing the swan geese and fleeing them[1]. In each case the *DEF* function sequence happens three times, and in each repetition an instance of D precedes an instance of E which precedes an instance of F. Therefore, without the ability to separate these instances, the scheme looses fidelity to Propp's theory.

Annotators may also mark a function instance as **inverted**, which was a part of Propp's theory which indicated when a plot piece did not fulfill its intended plot function, or was fulfilled by a semantically opposite form [39, p. 116, note 4].

For each function instance the annotator is required to mark the **type**. This is an important innovation of our scheme, because often Propp indicates that a function is present in the text, but there is no span of text directly corresponding to the function. For example, in Tale 148, Nikita the Tanner, the function B (dispatch of the hero on the quest) appears explicitly when the Tzar asks Nikita to fight the dragon and rescue his daughter. In these cases the annotator marks the instance with the type *Actual*, which indicates that the instance actually occurs in the text of the tale. In the next sentence we see Nikita preparing to fight, and so clearly the function C (the decision to go on the quest), has already occurred. In these cases the annotators are instructed to find the closest logically connected event, mark that as the instance, then mark the instance as either *Antecedent* or *Subsequent*, depending on whether the connected event is before or after the presumed occurrence of the function in the timeline.

Finally, annotators are also able to mark text spans corresponding to the *signals* for the function and any inversion, and the full extent of the function. The separation of spans into a *signal* portion and an *extent* is important. The identification of a signal span allows the annotators to mark the key verbs (or other words) that most strongly indicate the presence of the function or the inversion. The extent, on the other hand, allows the annotators to indicate the full portion of the text that represents the function. The details of these spans are covered in the Appendix.

3.2 ProppML: Move Scheme

The second of the three schemes allows annotators to mark the Move structure of the tale. Moves are defined as the development from villainy or lack (functions A or a) to a terminal

[1] Propp calls this "trebling", although in practice this sort of repetition occurs in our corpus anywhere from two to four times

function; each lack or act of villainy creates a new move [39, p. 92]. The representation of the move in our scheme is rather simple. Annotators first mark the *move number*, which is a non-negative integer. The move numbered zero is conventionally deemed the "preparatory" move, meaning all the preparatory functions, as well as the initial situation, are contained within it. Thus even "single-move" tales in our scheme can have two moves: the preparatory "pseudo-move", and the actual move.

Each move is then represented as a sequence of function *instances*. Importantly, moves are sets of function instances rather than functions themselves, as different instances of a repeated function can be spread across different moves (see, for example, Tale 93 in [39, pp. 136–137]). Within a move, function instances are ordered by their appearance in the text.

3.3 ProppML: Dramatis Personae Scheme

The final annotation scheme in ProppML is the dramatis personae scheme, which marks the role of various characters. As Propp describes, there are seven different character roles: Hero, Villain, Donor, Helper, Princess, Dispatcher, and False Hero [39, pp. 79–80]. In the data presented below, all the tales were previously marked with "coreference groups" corresponding to bundles of co-refering referential expressions [13]. For example, in the text, a single character might be referred to at different times as "Nikita", "Nikita the Tanner", "he", "him", or "a tanner in the city of Kiev". All these referential expressions referring to Nikita are marked and bundled together into a single coreference group that is assigned a unique identifying number. Annotators are then asked to assign any number of the seven labels (including none) to each co-reference group in the text. Allowing multiple labels to be assigned to a single character is important, as Propp notes that a single character can fulfill multiple different roles, or a single role may be spread among different characters [39, pp. 80–81].

3.4 Example Annotations

Figure 1 shows excerpts from actual annotated files, with the function, move, and dramatis personae annotations. These files are in Story Workbench annotation format, which is described elsewhere [14, 15]. A detailed BNF for each of the three annotation schemes is found in Appendix A.

4 Data Production

4.1 Selection of Texts

Because answering Q1 requires us to know Propp's list of functions for a tale, our raw text is necessarily drawn from Propp's original corpus, which he selected from Aleksandr Afanas'ev's collection of Russian folktales [2]. Propp analyzed 100 of these tales, publishing a subset of his analyses (45 tales) in a table at the end of his monograph [39].

For this pilot study we restricted our analysis to single-move tales. The primary reason for this is that the annotation is time-consuming, and the single move tales are among the shortest in the collection. Selecting the single-move tales facilitated the production of a pilot study, as well as provided initial evidence that further effort applied to the multi-move tales (a much larger and longer set of text), would not be a waste. Further, we have already deeply annotated a set of fifteen single-move tales (18,862 words) for our other work [13], and it seemed appropriate to begin with those as the translations were already available.

```
<?xml version="1.0" encoding="UTF-8"?>
<story>
  <rep id="edu.mit.story.char">
    <desc id="0" len="4064" off="0">
    ...
    An old man lived with his old wife; they had a daughter and a little son. "Daughter, daughter," said the
    mother, "we are going to work; we shall bring you back a bun, sew you a dress, and buy you a kerchief. Be
    careful, watch over your little brother, do not leave the house." The parents went away and the daughter what
    they had told her; she put her brother on the grass beneath the window, ran out into the street, and became
    absorbed in games. Some magic swan geese came, seized the little boy, and carried him off on their.
    ...
    </desc>
  </rep>
  <rep id="edu.mit.parsing.token">
    <desc id="3" len="2" off="353">An</desc>
    <desc id="4" len="3" off="356">old</desc>
    <desc id="5" len="3" off="360">man</desc>
    ...
    <desc id="866" len="6" off="4048">father</desc>
    <desc id="867" len="7" off="4055">arrived</desc>
    <desc id="868" len="1" off="4062">.</desc>
  </rep>
  ...
  <rep id="edu.mit.semantics.rep.function" ver="0.5.0">
    <desc id="2555" len="73" off="353">
      INITIAL::alpha:false::|ACTUAL:::3~4~5~6~7~8~9~10~11~12~13~14~15~16~17~18~19~20
    </desc>
    <desc id="2558" len="182" off="450">
      PREPARATORY::gamma:false:1:2|ACTUAL:27::27~28~29~30~31,57~58~59~60~61~62~63~64~65~66~67~68~69~70~71~72
    </desc>
    <desc id="2560" len="186" off="468">
      PREPARATORY::beta:false::1|ACTUAL:34,75::32~33~34~35~36,73~74~75~76
    </desc>
    <desc id="2559" len="107" off="704">
      PREPARATORY::delta:false::|ACTUAL:88,98,105::87~88~89~90~91~92~93~94~95~96~97~98~99~100~101~102~103~104~
      105~106~107~108
    </desc>
    <desc id="2514" len="87" off="814">
      NORMAL::A:false::1|ACTUAL:116,122::110~111~112~113~114~115~116~117~118~119~120~121~122~123~124~125~126~
      127~128
    </desc>
  ...
  </rep>
  <rep id="edu.mit.discourse.rep.coref">
    <desc id="2456" len="3701" off="353">father|2158,2159,...</desc>
    <desc id="2457" len="3668" off="375">mother|2160,2167,...</desc>
    <desc id="2458" len="3665" off="389">parents|2161,2169,2170,2179,2182,...</desc>
    <desc id="2459" len="3538" off="398">the children|2163,...</desc>
    <desc id="2460" len="3638" off="398">daughter|2162,2165,2166,2171,2174,2176,2180,2183,2184,2185,...</desc>
    <desc id="2461" len="3211" off="413">son|2164,2177,2186,2192,2193,...</desc>
    <desc id="2462" len="235" off="467">what they had told her|2168,2181</desc>
    <desc id="2463" len="5" off="515">a bun|2172</desc>
    <desc id="2464" len="7" off="530">a dress|2173</desc>
    <desc id="2465" len="10" off="551">a kerchief|2175</desc>
    <desc id="2466" len="3338" off="621">the house|2178,...</desc>
    <desc id="2467" len="29" off="727">the grass beneath the window|2187</desc>
    <desc id="2468" len="10" off="746">the window|2188</desc>
    <desc id="2469" len="10" off="771">the street|2189</desc>
    <desc id="2470" len="5" off="806">games|2190</desc>
    <desc id="2471" len="3117" off="814">Swan Geese|2191,2194,...</desc>
    <desc id="2472" len="2691" off="889">Geese wings|2195,...</desc>
    ...
  </rep>
  <rep id="edu.mit.semantics.rep.move" ver="0.1.0">
    <desc id="2556" len="458" off="353">0|2555,2558,2560,2559</desc>
    <desc id="2549" len="3015" off="814">1|2514,...</desc>
  </rep>
  <rep id="edu.mit.semantics.rep.archetype" ver="0.1.0">
    <desc id="2529" len="3638" off="398">HERO|2460</desc>
    <desc id="2544" len="3211" off="413">PRINCESS|2461</desc>
    <desc id="2530" len="3117" off="814">VILLAIN|2471</desc>
    ...
  </rep>
</story>
```


Figure 1 Selection of the annotation of The Magic Swan Geese [26, p. 349], provided for illustrative purposes. Ellipses indicate removal of data to improve readability. We have included **six** annotation layers, enclosed in **<rep>** tags: the text itself, tokens, Proppian function annotation, coreference annotation, move annotation, and archetype annotation. For the function, coref, move, and archetype layer, we only include the data applicable to the selection of text in the text layer. An in-depth description of the annotation scheme can be found in Appendix A.

While Propp's work was performed on the original Russian text, all of our work is performed on English translations. This is generally considered acceptable for first-order structural analysis [18]. Even given these restrictions, this pilot corpus is still a significantly larger set of data than annotated in previous studies.

4.2 The Story Workbench

Linguistic annotation usually requires some sort of computer tool support. In our case, we had already developed the Story Workbench, a robust tool for annotating text in multiple representations [14, 15]. The Story Workbench is a cross-platform and highly extensible generic text annotation tool that was used to actually apply the ProppML annotation schemes to the texts. The Story Workbench builds on top of popular, well-supported tools, uses open-source libraries, and adopts widely-used and well-documented standards..

Story Workbench distinguishes between the programmatic format of annotations (*representations*) and the actual annotations in that format (*descriptions*). Representations in Story Workbench are designed to be reused and applied to multiple texts, rather than existing only in annotations. There are currently more than 17 representations implemented in Story Workbench, including the Proppian functions, archetypes, and moves we describe in this paper.

Story Workbench supports single annotation, double annotation, and annotation development. It supports the automatic creation of annotations, annotation problem identification and migration, inspection of annotations, manual creation and correction of annotations, comparing and contrasting annotations, annotation adjudication, and annotation scheme implementation, among other features [13, The Story Workbench].

To perform the experiment described here, we implemented all three ProppML annotation schemes in the Story Workbench codebase. The Story Workbench provides a user interface for manually creating all the annotations described here, and enforces syntactic consistency and well-formedness. The annotation tool, along with documentation and references fully explaining the capabilities and usage of Story Workbench, may be downloaded online[2] and used for other annotation projects involving ProppML.

4.3 Annotator Training

As discussed previously, annotation was done in a double-blind manner by two highly-trained annotators. Both annotators were students at Harvard University in Cambridge, MA[3] To begin we trained the annotators for three weeks where the annotators were first asked to read Propp's book from start to finish, and then asked to annotate the Magic Swan Geese tale, the analysis of which is explained in detail in Propp's book [39, pp. 96–99]. Reading the book and the initial annotation of the Magic Swan Geese took approximately 20 hours total. The annotators were then brought together then with the adjudicator for a three-hour meeting to discuss any questions and compare their annotations. The adjudicator was already highly trained in Propp's system: he was one of the original annotators who helped produce the first set of Propp annotations [13], and was also a Ph.D. student in English literature at Harvard University. After the annotators received feedback on their annotations, they re-did their annotations on the Magic Swan Geese and had another meeting (another 10 hours total). At this point agreement was very good, and annotation of the remainder of the data began.

Annotators were also trained separately in the operation of the Story Workbench, which took approximately one hour.

[2] `http://projects.csail.mit.edu/workbench`
[3] The study was begun while Dr. Finlayson was a researcher at MIT, also in Cambridge.

Table 1 Annotation agreement measures for all stories. F_1^s is the strict F_1 measure, F_1^i is the identification F_1 measure, and F_1^g is the grouping F_1 measure. κ_{sym} is the Fleiss Kappa score for primary function symbols and κ_{dp} is the Fleiss Kappa score for dramatis personae markings. The cumulative measures are the microaverages of the scores in the rest of the column.

		Functions			Moves		Dramatis Personae		
#	English Title	F_1^s	F_1^i	κ_{sym}	F_1^s	F_1^g	F_1^s	F_1^i	κ_{dp}
148	Nikita the Tanner	0.80	1.00	0.89	0.00	0.90	1.00	1.00	1.00
113	The Magic Swan Geese	0.33	0.83	0.93	0.00	0.70	0.74	0.93	0.48
163	Bukhtan Bukhtanovich	0.00	0.70	0.66	0.00	0.57	0.67	0.67	0.37
162	The Crystal Mountain	0.41	0.69	1.00	0.50	0.69	0.73	0.73	0.47
151	Sharbarsha the Laborer	0.22	1.00	1.00	0.50	0.72	0.67	0.67	0.35
152	Ivanko the Bear's Son	0.24	0.82	0.67	0.00	0.55	0.86	0.86	0.62
131	Frolka Stay-at-Home	0.40	0.80	1.00	0.50	0.85	0.90	0.90	0.77
108	Ivashko and The Witch	0.29	0.79	1.00	0.00	0.78	0.89	1.00	0.73
145	The Seven Simeons	0.12	0.71	0.80	0.00	0.63	0.95	0.95	0.76
135	Ivan Popyalov	0.10	0.76	1.00	0.00	0.64	0.67	0.71	0.38
149	Serpent & Gypsy	0.10	0.57	0.57	0.00	0.39	0.67	1.00	0.36
114	Prince Danila Govorila	0.21	0.72	0.92	0.00	0.70	0.75	1.00	0.46
127	Merchant's Daughter	0.21	0.71	1.00	0.00	0.63	0.73	1.00	0.50
140	Dawn, Evening, & Midnight	0.33	0.73	0.90	0.00	0.63	0.95	0.95	0.88
154	Runaway Soldier & the Devil	0.27	0.73	1.00	0.00	0.71	0.89	0.86	0.76
Cumulative		0.27	**0.76**	**0.92**	0.11	**0.67**	**0.82**	**0.88**	**0.63**

4.4 Annotation Procedure

Annotation of the texts after the Magic Swan Geese was performed at a rate of approximately 2,000 words/week, with annotators spending approximately 7 hours a week annotating, and 3 hours/week in an adjudication meeting. Therefore, after the initial training period of 30 hours, the annotators spent approximately 9 weeks annotating, and each annotator spent a total of approximately 120 hours on the project. The adjudicator spent approximately 24 hours on the project, not counting the annotation that he performed in previous years (which constituted, at a minimum, approximately 100 hours of work). During each adjudication meeting disagreements between the annotators were discussed, and additional discussion of subtleties of Propp's system was held as needed. Further, a gold standard marking was produced by the team. Thus, the project produced three sets of marked texts: one marked by annotator 1, another marked by annotator 2, and a gold-standard set corrected by the adjudicator. If the team had a disagreement that could not be resolved, they consulted the project manager[4].

5 Agreement Results

The agreement between the annotators was assessed for each of functions, moves, and dramatis personae in several ways.

For functions, we report two F_1 measures: a 'strict' F_1 which requires the annotator markings to be *exactly* the same in every aspect; and an 'identification' F_1 that marks

[4] Dr. Finlayson managed the project.

agreement if there is any overlap at all between two annotator's marked instances. Agreements are reported for each text individually, as well as microaveraged over the whole corpus. As would be expected, the strict F_1 result is low (0.27), as any disagreement at all (including disagreement over the exact extent of a function, which might include several hundred tokens) gives a penalty. However, the identification measure measures agreement by performing a "best effort" alignment, where function markings that agree exactly are paired off, and then function markings are then paired off in descending order of degree of overlap. This measure of agreement is very good (0.76), especially considering the complexity of Propp's system.

We also report the Fleiss kappa for identification of the primary function symbols. Although annotators were given the list of functions to mark on the texts, they were allowed to change the symbol identity or subtype if they felt necessary. However, this was not necessary in most cases, as the microaveraged kappa was a "near perfect" 0.92.

For moves, we report a strict and a 'grouping' F_1 measure, where the latter measures an F_1 measure calculated between the function instances involved in moves (rather than the moves as a whole, as for the strict measure). Because these tales were single move, these measures assess only how well annotators marked functions as either "preparatory" or "normal"; as described in our annotation scheme (§3.2), preparatory functions are contained within their own move. 'Grouping' F_1 agreement was good at 0.67.

For dramatis personae, we report a strict and a identification F_1 measures, as well as Fleiss kappa for assignment of dramatis personae labels. The identification F_1 measure marks whether the annotators agreed on whether a character was a dramatis personae, not necessarily on the label(s). Both cumulative F_1 measures are quite high, at 0.82 and 0.88 for the strict and identification, respectively. The cumulative Fleiss Kappa agreement was good at 0.63.

Given these high measures, especially the cumulative F_1 measures, it is evident that Propp's morphology can reliably be applied to narrative texts by human annotators.

6 Related Work

One of the most notable attempts to annotate Proppian morphologies is the Proppian Fairy Tale Markup Language (PftML) [35]. Malec discusses the method and difficulties encountered while creating his annotation scheme and applying it parts of 20 Russian magic tales, but the work is hard to assess as it does not include examples of annotation nor a description of the annotation scheme in the version available online [35]. PftML, as described, does not appear to handle the annotation of implicit functions nor the annotation of dramatis personae and Proppian moves. Further work on PftML brings in additional linguistic information [30, 7] and includes brief examples of annotation, but does not appear to support signals nor discontinuous regions representing functions. Recent work on PftML looks towards the possibility of automatically classifying and annotating Russian folktales [34].

Work in story generation has resulted in ProppOnto, an OWL ontology based on Propp's morphology [38]. Peinado *et al.* explicitly state that their system is not intended to be complete (p. 6) and appear to use the system solely for story generation rather than annotation.

Recent work by Lendvai *et al.* [29, 30] attempts to integrate PftML and ProppOnto together with linguistic information, demonstrating an approach to enrich both schemes. Currently, this work appears to be a proposal, with it being unclear whether or not work has proceeded in this integration. Further, this integration does not alleviate some of the issues that PftML and ProppOnto suffer from as annotation schemes.

Work by Bod *et al.* [4] is the work most suited to direct comparison with our results: they directly studied the reproducibility of Proppian narrative annotations. Bod's study was split into two trials. The first study consisted of nine students who were briefed for 45 minutes, given a small handout describing the functions and dramatis personae, and asked to annotated four single-move stories with functions, subfunctions, and dramatis personae. The second study was similar to the first, but omitted subfunctions, had dramatis personae already assigned, and only had six participants [4, pp. 18–20]. Bod *et al.* concluded that the dramatis personae had an important effect on the assignment of functions. Further, they conclude from the first study that dramatis personae cannot be reliably annotated and from the second study that even given the dramatis personae, human annotators cannot reliably annotate some functions (p. 20).

While Bod *et al.* state that their previous suggestion of a large-scale study to determine how reliably humans can apply Proppian morphology to narratives is "not worthwhile" (p. 21), they themselves admit that making Propp's "vague descriptions" (p. 20) understandable to annotators may require more time and training than they were given in the study. Bod *et al.* suggest that a property necessary for a formal framework to be the basis of an automated system is that sufficiently trained human annotators will annotate a narrative in the same way (p. 17). We believe that our study has sufficiently addressed the training shortcoming identified by Bod *et al.*, by giving annotators more than 30 hours training, as well as another 90 hours of time to produce annotations.

6.1 Use of Propp in Story Generation

A substantial amount of work has been built on top of Propp's morphology in the field of story generation. Some of the first work using Propp to generate stories was done by Klein *et al.* [27, 28], creating a system capable of generating folktale text based on Propp's morphology.

Grasbon and Braun [24] describe an implementation of an interactive story-telling system that used the dependencies and sequences inherent in Propp's functions to generate narratives from pre-written scenarios for each function.

Arinbjarnar [3] describes the creation of a murder mystery game engine based on a Bayesian network designed with a morphology similar to Propp's morphology. Also in games, Fairclough and Cunningham's work [10, 11, 12] integrates Proppian characters and character functions into a game as part of a case-based planning and constraint satisfaction system designed to make agents react appropriately to player actions while following a plot.

Thomas [42] describes two methods for generating folktales using Propp's morphology. The first method expands existing analyses of tales (called *schemes* by Thomas) and filling in roles with random character. The second method attempts to generate fully-formed moves by considering the sequential order of Propp's functions and filling in subtypes for the functions in the move.

Early work by Gervás uses ProppOnto [38] as part of a case-based reasoning system to measure the semantic distance between situations and maintain "an independent story structure from the simulated world" [22, p. 4]. Peinado and Gervás [37] also raise questions about the creativity of narratives generated by Proppian morphology, determining that it is possible to produce relatively novel narratives through generation, but admitting that more experiments are needed. Newer work by Gervás describes the partial implementation of a story generation prototype based on Propp's morphology [21] and work by Gervás, León, and Méndez [23] attempts to reconcile existing schemes with Proppian character functions and extend a Propp-based generation system (Propper) to support schema-driven generation.

7 Contributions

Propp's morphology has had a deep impact in narratology, narrative understanding, and computational approaches to narrative, especially in story generation. Propp's work has inspired and influenced a substantial body of work on narrative structure since its translation into English. ProppML, the scheme we present here, is a complete annotation scheme for Proppian morphologies, succeeding previous work on annotation schemes targeting Propp's work. We have produced a pilot corpus of annotated data larger than any previous study.

We have also shown that Propp's morphology can be reliably applied by human annotators to single-move tales with a cumulative agreement for functions of $F_1^i = 0.76$ and $\kappa = 0.92$; for moves, $F_1^g = 0.67$; and for dramatis personae $F_1^s = 0.82$, $F_1^i = 0.88$, and $\kappa = 0.6$. The definitions for how F_1^i, F_1^g, and F_1^s are calculated is given in §5.

Our work provides a powerful tool for the machine learning of Proppian morphologies. Our lab has already used a prior version of the annotation scheme to demonstrate the machine learning of Propp's morphology [16]; our current version improves on this past scheme and paves the way toward even more accurate systems in the future.

Acknowledgments. This research was made possible by an FIU's Presidential Fellowship and FIU's SCIS's Director's Fellowship, both awarded to W. Victor H. Yarlott. This work was also partially supported by National Institutes of Health (NIH) grant number 5R01GM105033-02. We would like to thank our two annotators, David Roberson and Aeron Commins, and our adjudicator, Jacob Stulberg, for their hard work and attention to detail.

References

1 Antti Amatus Aarne. *Verzeichnis der märchentypen*. Suomalainen tiedeakatemia, 1910.
2 Aleksandr Nikolaevich Afanas'ev. *Narodnye Russkie Skazki*. Moscow: Gos. Izd-vo Khudozh Lit-ry., 1957.
3 María Arinbjarnar. *Murder She Programmed: Dynamic Plot Generating Engine for Murder Mystery Games*. Thesis, Reykavik University, 2005. URL: `http://www-users.cs.york.ac.uk/~maria/greinar/BSc.pdf`.
4 Rens Bod, Bernhard Fisseni, Aadil Kurji, and Benedikt Löwe. Objectivity and Reproducibility of Proppian Narrative Annotations. In Mark Alan Finlayson, editor, *Third Workshop on Computational Models of Narrative (CMN)*, pages 17–21, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).
5 Claude Bremond, Jean Verrier, Thomas G. Pavel, and Marylin Randall. Afanasiev and propp. *Style*, pages 177–195, 1984.
6 Benjamin N. Colby. A Partial Grammar of Eskimo Folktales. *American Anthropologist*, 75:645–662, 1973.
7 Thierry Declerck, Kerstin Eckart, Piroska Lendvai, Laurent Romary, and Thomas Zastrow. Towards a standardized linguistic annotation of fairy tales. In *Workshop on Language Resource and Language Technology Standards*, pages 60–63, 2010.
8 Alan Dundes. From etic to emic units in the structural study of folktales. *The Journal of American Folklore*, 75(296):95–105, 1962.
9 Alan Dundes. *The Morphology of North American Indian Folktales*. Folklore Fellows Communications, 1964.
10 Chris R. Fairclough and Pádraig Cunningham. An Interactive Story Engine. In *Proceedings of the 13th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2002)*, pages 171–176, 2002.

11 Chris R. Fairclough and Padraig Cunningham. A Multiplayer Case Based Story Engine. In *Proceedings of the 4th International Conference on Intelligent Games and Simulation (GAME-ON 2003)*, pages 41–47. EUROSIS, 2003. URL: `https://www.scss.tcd.ie/publications/tech-reports/reports.03/TCD-CS-2003-43.pdf`.

12 Chris R. Fairclough and Pádraig Cunningham. AI Structuralist Storytelling In Computer Games. In *Proceedings of the 5th International Conference on Computer Games: Artificial Intelligence, Design and Education (CGAIDE 2004)*. University of Wolverhampton, 2004.

13 Mark A Finlayson. Propplearner: Deeply annotating a corpus of russian folktales to enable the machine learning of a russian formalist theory. *Digital Scholarship in the Humanities*, page fqv067, 2015.

14 Mark Alan Finlayson. Collecting Semantics in the Wild: The Story Workbench. In Jacob Beal, Paul Bello, Nick Cassimatis, Michael Coen, and Patrick Winston, editors, *Proceedings of the AAAI Fall Symposium on Naturally Inspired Artificial Intelligence (published as Technical Report FS-08-06, Papers from the AAAI Fall Symposium)*, volume 1, pages 46–53, Arlington, VA, 2008. AAAI Press, Menlo Park, CA. URL: `http://www.aaai.org/Papers/Symposia/Fall/2008/FS-08-06/FS08-06-008.pdf`.

15 Mark Alan Finlayson. The Story Workbench: An Extensible Semi-Automatic Text Annotation Tool. In Emmett Tomai, Jonathan P. Rowe, and David K. Elson, editors, *Proceedings of the 4th Workshop on Intelligent Narrative Technologies (INT4)*, pages 21–24, Stanford, CA, 2011. AAAI Press, Menlo Park, CA. URL: `http://aaai.org/ocs/index.php/AIIDE/AIIDE11WS/paper/view/4091`.

16 Mark Alan Finlayson. Inferring Propp's Functions from Semantically-Annotated Text. *Journal of American Folklore, Special Issue on Computational Folkloristics*, 129(511):53–57, 2016.

17 Mark Alan Finlayson and Tomaz Erjavec. Overview of Annotation Creation: Processes & Tools. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*. Springer, 2016. `arXiv:arXiv:1602.05753`.

18 J. L. L. B. Fischer. The Sociopsychological Analysis of Folktales. *Current Anthropology*, 4(3):235–295, 1963. URL: `http://www.jstor.org/stable/2739608`, `doi:10.1086/200639`.

19 Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.

20 William B. Frakes and Ricardo Baeza-Yates. *Information retrieval: data structures and algorithms*. Prentice Hall PTR, 1992.

21 Pablo Gervás. Propp's Morphology of the Folk Tale as a Grammar for Generation. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13)*, volume 32, pages 106–122. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013. `doi:10.4230/OASIcs.CMN.2013.106`.

22 Pablo Gervás, Belén Daíz-Agudo, Federico Peinado, and Raquel Hervás. Story plot generation based on CBR. *Knowledge-Based Systems*, 18(4-5):235–242, 2005. `doi:10.1016/j.knosys.2004.10.011`.

23 Pablo Gervás, Carlos León, and Gonzalo Méndez. Schemas for Narrative Generation Mined from Existing Descriptions of Plot. In *Proceedings of the 6th Workshop on Computational Models of Narrative (CMN'15)*, pages 54–70, 2015. URL: `http://narrative.csail.mit.edu/cmn15/schedule_cmn15.html`.

24 Dieter Grasbon and Norbert Braun. a morphological approach to interactive storytelling. In *Proceedings of cast01, Living in Mixed Realities*, pages 337–340. FhG – Institut Medienkommunikation (IMK), German Federal Ministry of Education and Research, 2001.

URL: http://netzspannung.org/version1/extensions/cast01-proceedings/pdf/by_name/Grasbon.pdf.

25 A. J. Greimas. *Structural Semantics: An Attempt at a Method.* University of Nebraska Press, Lincoln, Nebraska, 1983.

26 Norbert Guterman. *Russian Fairy Tales.* Pantheon Books, 1973.

27 Sheldon Klein. Meta-compiling text grammars as a model for human behavior. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing*, pages 84–88. Association for Computational Linguistics, 1975.

28 Sheldon Klein, John F. Aeschlimann, Matthew A. Appelbaum, D. F. Blasiger, Elizabeth J. Curtis, Mark Foster, S. D. Kalish, S. J. Kamin, Y. D. Lee, L. A. Price, et al. Modeling Propp and Lévi-Strauss in a metasymbolic simulation system. *Patterns in Oral Literature*, pages 141–222, 1977.

29 Piroska Lendvai, Thierry Declerck, Sándor Darányi, Pablo Gervás, Raquel Hervás, Scott Malec, and Frederico Peinado. 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 (LREC)*, pages 1996–2001. European Language Resources Association (ELRA), 2010.

30 Piroska Lendvai, Thierry Declerck, Sándor Darányi, and Scott Malec. Propp revisited: Integration of linguistic markup into structured content descriptors of tales. In *Proceedings of the Conference for Digital Humanities 2010*, 2010.

31 Claude Lévi-Strauss. L'analyse morphologique des contes russes. *International Journal of Slavic Linguistics and Poetics*, 3:122–149, 1960.

32 Claude Levi-Strauss. Structure and Form: Reflections on a Work by Vladimir Propp. In Vladimir Propp, editor, *Theory and History of Folklore*, chapter 11, pages 167–210. University of Minnesota Press, Minneapolis, MN, 1984.

33 Isabel Machado, Ana Paiva, and Paul Brna. Real characters in virtual stories. In *Virtual Storytelling Using Virtual Reality Technologies for Storytelling*, pages 127–134. Springer, 2001.

34 Scott Malec. Autopropp: Toward the automatic markup, classification, and annotation of russian magic tales. In *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, pages 112–115, 2010.

35 Scott A. Malec. Proppian structural analysis and xml modeling. *Proc. of Computers, Literature and Philology (CLiP 2001)*, 2001. URL: https://www.researchgate.net/publication/247286265_Proppian_Structural_Analysis_and_XML_Modeling.

36 Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014. URL: http://www.aclweb.org/anthology/P/P14/P14-5010.

37 Federico Peinado and Pablo Gervás. Creativity Issues in Plot Generation. In *Working Notes on Workshop on Computational Creativity, at 19th International Joint Conference on Artificial Intelligence (2nd IJWCC'05)*, pages 45–52. Departamento de Ingeniera del Software e Inteligencia Artificial, Universidad Complutense de Madrid, 2005. URL: http://www.fdi.ucm.es/profesor/fpeinado/publications/2005-peinado-creativity.pdf.

38 Federico Peinado, Pablo Gervás, and Belén Díaz-Agudo. A description logic ontology for fairy tale generation. In *Language Resources for Linguistic Creativity Workshop, 4th LREC Conference*, pages 56–61. Citeseer, 2004.

39 Vladimir Propp. *The Morphology of the Folktale (2nd ed.).* University of Texas Press, Austin, TX, 1968.

40 James Pustejovsky and Amber Stubbs. *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications.* O'Reilly, Sebastopol, CA, 2013.

41 Thomas Rieger and Norbert Braun. Narrative use of sign language by a virtual character for the hearing impaired. *Computer Graphics Forum*, 22(3):651–660, 2003.

42 Craig Michael Thomas. *The Algorithmic Expansion of Stories.* Thesis, Queen's University, 2010. URL: `http://hdl.handle.net/1974/6127`.

43 Cornelis Joost van Rijsbergen. *Information Retrieval.* London: Butterworths, 1979. URL: `http://www.dcs.gla.ac.uk/Keith/Preface.html`.

44 Roman M Vólkov. *Shazka. Rozyskanija po sjužetosloženiju narodnoj skazki.*, volume 1. Ukrainian State Publishing House, 1924.

A ProppML Detailed Specification

In this appendix we give a detailed description of the annotation scheme. Figure 1 shows an excerpt from an annotated file (The Magic Swan Geese, tale #113). Each annotated file is an XML document with the basic format shown in Figure 2

A Story Workbench annotated file represents a stand-off annotated text, with each annotation layer encoded within a `<rep>` tag, standing for *representation*, which collects together a set of annotations that all share an annotation scheme. Each representation has a unique string identifier in reverse-domain-namespace form, e.g., `edu.mit.story.char`, and contains multiple `<desc>` tags, each of which encodes a single annotation. As shown in Figure 3, each annotation has three XML attributes: a unique id number (`id`), an offset (`off`), and a length (`len`). The id number is unique among all the annotations in the document. The offset refers to the starting character in the characters of the text to which the annotation applies; the length attribute refers to the length of the annotation, in characters.

The first annotation layer of a Story Workbench file is always the *Character* layer, and contains the text to be annotated in a description with id zero. All offsets and lengths for annotations in the remainder of the document refer to the character content of this `<desc>` tag (minus the first and last linebreak characters). The format of the character body of each other `<desc>` tag in the document is determined by the BNF associated with its parent representation. For example, in Figure 1, the representation layer corresponding to Propp's functions is begun by `<rep id="edu mit.semantics.rep.function" ver="0.5.0">`. The first annotation in this layer begins with `<desc id="2555" len="73" off="353">`. The text that immediately follows that tag is formatted according to the BNF in §A.1.

All of the annotations that follow have fields that share a common structure relating to marking spans of characters in the text, as shown in Figure 4. The general mechanism for marking text spans is the *segment*, which is a sequence of consecutive tokens, with none skipped. Segments may be brought together in a *segment set* to mark discontinuous spans. Each segment is encoded as referring to a particular set of token annotation ids.

SegmentSet. A segment set contains zero, one, or more segments. Multiple segments in a set are encoded as separated with commas.

Segment. Segment describe the set of contiguous tokens that make up a span of text. Segments must contain at least one token, and each additional token is appended to the segment encoding using a tilde.

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT story (rep+)>
<!ELEMENT rep (desc+)>
<!ATTLIST rep id CDATA #REQUIRED>
<!ELEMENT desc (#CDATA)>
<!ATTLIST desc id ID #REQUIRED>
<!-- both len and off should be a non-negative integer -->
<!ATTLIST desc len CDATA #REQUIRED>
<!ATTLIST desc off CDATA #REQUIRED>
```

Figure 2 DTD that represents the high-level structure of a XML-formatted Story Workbench annotation file.

```
<desc id="#" len="#" off="#">Annotation</desc>
```

Figure 3 Structure of a `desc` XML tag.

```
SegmentSet        ::= "" | Segment | Segment "," SegmentSet
Segment           ::= TokenID | TokenID "~" Segment
TokenID           ::= <postive integer>
```

Figure 4 BNF for SegmentSets.

TokenID. Tokens are one of the lower levels of linguistic annotation that is performed automatically by Story Workbench. TokenIDs refer to the automatically generated id that is a part of the annotation for each token.

Note that in this BNF, the empty string (`""`) indicates that a segment set may be empty. This convention is followed in the other BNFs below.

A.1 BNF for Propp's Functions

Figure 5 shows the BNF for the annotation of functions. A function annotation is split at it's top level into a function tag and one or more instances.

FunctionTag. The function tag contains all the essential information about the function we're annotating: the function type, the function itself, the modifiers, and whether or not the function is inverted.

FunctionType. The three types of functions are defined as follows: `NORMAL` functions are those that are represented by roman letters in Propp's work, as well as $\uparrow$ and $\downarrow$. The `INITIAL` situation is the function α. Finally, `PREPARATORY` functions are those listed with other Greek letters. Within a narrative, there may only be a single `INITIAL` situation, which may not be inverted, has only one instance, has no signal, and has no inversion signal.

Symbol. Function symbols are listed in Propp's Morphology of the Folktale, and can be represented by any alphanumeric string containing the symbols `[a-zA-Z0-9]`. Those function symbols that do not have a corresponding symbol in the alphanumeric character set are written out (e.g. `alpha`, `beta`, `up`, `down`).

```
Annotation       ::= FunctionTag Instances
FunctionTag      ::= FunctionType ":" PreSuperscript ":" Symbol
                     ":" Inverted? ":" Subscript ":" Superscript
FunctionType     ::= "NORMAL" | "PREPARATORY" | "INITIAL"
Symbol           ::= <string>
PreSuperscript   ::= <string>
Subscript        ::= <string>
Superscript      ::= <string>
Inverted?        ::= "true" | "false"
Instances        ::= FuncInstance | FuncInstance Instances
FunctionInstance ::= InstanceType ":" Signal ":" InversionSignal ":" Extent
InstanceType     ::= "ACTUAL" | "ANTECEDANT" | "SUBSEQUENT"
Signal           ::= SegmentSet
InversionSignal  ::= SegmentSet
Extent           ::= SegmentSet
SegmentSet       ::= "" | Segment | Segment "," SegmentSet
Segment          ::= TokenID | TokenID "~" Segment
TokenID          ::= <postive integer>
```

Figure 5 BNF for Propp's functions. Unlisted non-terminals are explained below. <string> means an alphanumeric string, possibly empty.

Superscript, Subscript, Presuperscript. Function subtype modifiers are also listed in Propp's monograph, and can be represented by any alphanumeric string containing the symbols `[a-zA-Z0-9]`. Those modifiers symbols that do not have a corresponding key in the alphanumeric character set may written out.

Inverted? Functions may be marked as "inverted," meaning the function is fulfilled by something semantically opposite of its usual filler, but this is not the same as negation. For example, refusing to wed would be an inversion of W, but not getting caught is not an inversion of Rs.

Instances. Functions have one or more function instances.

FunctionInstance. Each instance of function has four field: (1) the type, (2) the signal, (3) the inversion signal, and (4) the extent.

InstanceType. Most function instances are explicit, in which case the instance type is "ACTUAL". However, for implicit instances, the closest logically related event is marked as the extent of the instance. If the event happens directly before the function, the instance type is "ANTECEDENT", and if it occurs afterwards, the instance type is "SUBSEQUENT". If there is a tie between events, "ANTECEDENT" is preferred.

Signal. Instance signals are the single word or phrase that most strongly indicates the presence of the function, usually a verb.

InversionSignal. Inversion signals are words that signal the inversion of a function, such as "refused" or "not," depending on the context.

```
Annotation       ::= MoveNum "|" Instances
MoveNum          ::= <non-negative integer>
Instances        ::= FunctionInstance | FunctionInstance "|" Instances
FunctionInstance ::= InstanceIndex "," FunctionID
InstanceIndex    ::= <non-negative integer>
FunctionID       ::= <positive integer>
```

Figure 6 BNF for Propp's moves.

Extent. The extent of an instance should be the smallest region that covers the function. The extent of an instance must cover, at the very least, all the signal words and the inversion signal, if there is one. Extents can be continuous or discontinuous. If a verb is marked as a signal, the core arguments to that verb should be included, but non-core arguments should be excluded. Sentence-ending punctuation should only be included when the extent includes the whole sentence or covers words on either side of the sentence boundary.

A.2 BNF for Propp's Moves

Figure 6 shows the BNF for our annotation of moves within a folktale. A move is that part of a tale characterized by starting with a lack or an act of villainy and ending with the resolution of the lack or the defeat of the villain. A move comprises a move number and one or more function instances.

MoveNum. Each move is given a non-negative integer. By convention, move 0 contains the preparatory functions, including the initial function.

Instances. Each move may contain one more function instances. As noted in the body of the paper, functions may have multiple instances that are part of different moves.

FunctionInstance. This refers to the specific instance of a function that is part of a move.

InstanceIndex. Each function instance is represented by two numbers: the id of the funciton annotation of which it is a part, and the index of that function instance in the list of instances for that function. Function instances are indexed starting at zero.

FunctionID. Functions are covered in Section A.1. The function id refers to the automatically generated id for each function annotation. There can be multiple functions in a single move, but there must always be at least one function in a move.

A.3 BNF for Propp's Dramatis Personae

Figure 7 shows the BNF for dramatis personae. Dramatis personae annotations comprise a label and a coreference bundle id.

DramatisPersonae. A string corresponding to one of the seven character archetypes defined by Propp.

```
Annotation        ::= DramatisPersonae "|" CorefID
DramatisPersonae ::= "HERO" | "VILLAIN" | "DONOR" | "HELPER" | "PRINCESS" |
                     "DISPATCHER" | "FALSE HERO"
CorefID           ::= <positive integer>
```

Figure 7 BNF for Propp's dramatis personae.

CorefID. The coreference bundle id refers to the automatically generated id for each coreference bundle annotation.

Summarizing and Comparing Story Plans*

Adam Amos-Binks[1], David L. Roberts[2], and R. Michael Young[3]

1 **North Carolina State University, Raleigh, NC, USA**
`aaamosbi@ncsu.edu`
2 **North Carolina State University, Raleigh, NC, USA**
`dlrober4@ncsu.edu`
3 **North Carolina State University, Raleigh, NC, USA**
`rmyoung@ncsu.edu`

Abstract

Branching story games have gained popularity for creating unique playing experiences by adapting story content in response to user actions. Research in interactive narrative (IN) uses automated planning to generate story plans for a given story problem. However, a story planner can generate multiple story plan solutions, all of which equally-satisfy the story problem definition but contain different story content. These differences in story content are key to understanding the story branches in a story problem's solution space, however we lack narrative-theoretic metrics to compare story plans. We address this gap by first defining a story plan summarization model to capture the important story semantics from a story plan. Secondly, we define a story plan comparison metric that compares story plans based on the summarization model. Using the Glaive narrative planner and a simple story problem, we demonstrate the usefulness of using the summarization model and distance metric to characterize the different story branches in a story problem's solution space.

1998 ACM Subject Classification I.2.8 [Problem Solving, Control Methods, and Search] Plan execution, formation and generation

Keywords and phrases artifical intelligence, planning, narrative, comparison, story

Digital Object Identifier 10.4230/OASIcs.CMN.2016.9

1 Introduction

Branching story games, such as Mass Effect and The Walking Dead, have gained popularity for creating unique playing experiences by adapting story content in response to user actions. Research in interactive narrative (IN) uses *automated planning* to generate story plans as they offer an action-oriented and causally-related representation consistent with narrative [3, 19, 34]. For a given story problem, a story planner can generate multiple story plan solutions, all of which equally-satisfy the story problem definition but contain different story content. Characterizing the qualitatively different story plans in a story problem's solution-space would capture the story branches experienced by a user and enable plan-based INs to adapt in a story-theoretic manner to user actions.

A first step to characterizing the different story branches in a story plan problem is to develop the capability to compare two story plans based on their narrative-theoretic

* This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

© Adam Amos-Binks, David L. Roberts, and R. Michael Young;
licensed under Creative Commons License CC-BY

properties. Conflict Partial Order Causal Link (CPOCL) plans represents story plans with additional narrative-theoretic structures, namely character subplans and conflict, derived from classical plan structures. While conforming to the formal structures of classical planning limits the narrative representation, it allows us to directly leverage the *automated planning* community of research for topics such as plan comparison. Thus, in this paper we limit the scope of narrative-theoretic plan comparisons to those with a CPOCL representation.

Historically, plan comparison research has focused on metrics to assess minimal length or optimal plans. More recently however, mixed-initiative planning systems have employed additional advanced distance metrics to capture the differences in plan syntax. Still, these distance metrics have been restricted to domain-independent properties and do not account for domain-specific semantics of story plans. We address this limitation by differentiating between the necessary *syntactical* structure required by a story plan to be considered a valid plan and the additional *semantic* structures derived by CPOCL. This difference allows us to define a story plan *summary*. By first summarizing a story plan's narrative-theoretic content and then making comparisons allows us to emphasize the narrative-theoretic representation in story plans to more accurately capture a users experience.

The first contribution we make is the definition of a computational model to summarize the story semantics of CPOCL story plans. The model captures key story semantics of the logical progression of plot and character believability, both factors associated with narrative comprehension. Our second contribution is the definition of a CPOCL story plan distance metric. This metric builds upon previous work on plan distance metrics to calculate differences in story semantics between plans. We bring these two models into operation together on a story domain-problem pair to demonstrate their sensitivity to changes in story plan semantics and contrast them to existing domain-independent plan distance metrics.

2 Previous Work

In this section, we review key areas of research that support developing a model of CPOCL story plan summarization and distance metric. First, we outline previous work in story planning to highlight the specific story plan structures that are applicable to plan summarization research. This is followed by assessing research on plan comparison for concepts that translate to story plan comparison. Finally, we discuss story comparison from a wide variety of disciplines to identify principles to operationalize in story plan comparison.

2.1 Story Plan Summarization

The origins of automated story generation can be traced to the Tale-Spin system [17]. Tale-Spin makes use of an inference engine to direct character action when an initial state and environment are provided. Automated planning was later used to capture author goals in the Universe system [15], which focused on causal coherence of actions in the stories it generated and ensured specific author outcomes in a goal state. Following the advances made by Universe, the Fabulist system [26] developed the Intent-driven Partial Order Causal-Link (IPOCL) planner which determines the intentions each character could have and motivates these intentions through story actions. This supports character believability which is linked to narrative comprehension. Conflicting character goals are viewed as an essential literary element of stories [1, 12]. Work to operationalize a notion of conflict by Ware *et al.* [33] resulted in the Conflict Partial Order Causal Link (CPOCL) algorithm, which uses non-executed steps to model foiled character goals. While other story generation mechanisms are less formal and explore less quantifiable aspects of narrative (e.g MEXICA [25]), we limit the scope of this paper to story plans with more formal representation as it allows the direct application of previous work in classical planning to plan summarization and comparison.

The work of Clement *et al.* [5] uses abstraction in a Hierarchical Task Network (HTN) planner to dramatically reduce the run time of planning algorithms. Myers *et al.* [20] investigated more appropriate ways to interact with a user with an approach that seeks an explanation of a plan from the knowledge stored in a meta-theory, rather than from the syntax of a plan. Another attempt to provide a human-centric interface to plan summaries was the research by Mellish *et al.* [18] that used Natural Language Generation (NLG) to elucidate the steps of a generated plan. Additional work by Myers [21] uses a temporal based domain-theory to summarize key temporal regularities and exceptions in support of domain-independent summarization techniques in temporal plans.

The use of Grice's maxim of quantity [11] motivated the work of Young [35] to find an equilibrium between the level of detail and abstraction communicated from plan structures. Plan summaries using the Local Brevity Algorithm (LBA) are computed by weighting the importance of individual plan steps based on the causal inferences of story events as identified by Trabasso and Sperry [31]. Specifically, Trabasso and Sperry [31] define a causal chain to contain an opening, a closing, and to continue the chain of events. The causal chains of six folk-tales were used to compute both a story event's membership on the causal chain and causal connectedness. The experimental results showed that these two factors account for a significant portion of the variance in agreement of story event recall among study participants and did not have a statistically significant interaction. This leads the authors to conclude that causal inferences are foundational to the process of story representation in memory and a source of how we privilege information. This finding will prove to be valuable research, as it will form the basis for our arguments in Section 3.2.

2.2 Story Plan Comparison

Like plan summarization, many plan distance metrics are domain independent. In classical planning, plan distance is well-served by assessing a candidate plan's syntactic structure to determine whether it is *minimal* or *redundant* with respect to an optimal plan. The application of planning algorithms to new domains where human decision-making is augmented, termed mixed-initiative planning, has increased interest in developing more robust plan distance metrics to capture differences in plan structure.

Research by Srivastava *et al.* [30] and subsequently by Nguyen *et al.* [23] into domain-independent plan distance metrics founded on Jaccard distance resulted in the action, causal-link and state-space distance metrics. The authors define action distance as

$$\delta_A(p, p') = 1 - \frac{|A(p) \cap A(p')|}{|A(p) \cup A(p')|}, \tag{1}$$

where A is the actions of a plan and p, p' are complete plans. Similarly, the causal link distance is

$$\delta_C(p, p') = 1 - \frac{|C(p) \cap C(p')|}{|C(p) \cup C(p')|}, \tag{2}$$

where C is the set of causal links between plan actions of a plan and p, p' are complete plans. These two metrics are of particular interest, as they capture the differences in actions and causal links, which enable plans to model story properties.

There have also been several other domain-independent distance metrics developed. Goldman and Kuter [10] designed *normalized compression distance* as a way to measure the conditional information from one plan to another. Plan landmarks, on the other hand, capture the essential conditions all solution plans must contain and are used by Bryce [4] to assess the differences between plans in the *landmark distance* metric. Lastly, Roberts *et al.* [28] use the differences in plan length to compute *parsimony*. Each of these metrics capture syntactic differences between plans; however, none can be immediately applied to story plan semantics.

To date, computational models of story comparison have tended to apply comparison methods from other disciplines. For example, representing story actions as a sequence of string characters enables the use of sequence-alignment tools commonly used in textual analysis (spell-checking) and bioinformatics (DNA sequence alignment). Porteous *et al.* [24] manipulate characters' social relationships to explore the qualitative differences in interactive narratives as reflected in Levenshtein distance [16]. Another sequence-alignment algorithm is evaluated by Fay *et al.* [8] who make use of the Needleman-Wunsch algorithm [22] to compare linear story sequences from Genesis and show its ability to greatly reduce the compute time for matching and comparison of stories. Finally, the use of an intelligent Drama Manager (DM) to negotiate the balance between authorial intent and player autonomy in interactive systems motivated the work of Jones and Isbell [13], who empirically evaluated story similarity metrics when gameplay is represented as a Targeted Trajectory Distribution-Markov Decision Process (TTD-MDP) [27]. While methods from other disciplines have shown to be helpful in computing the differences in stories, they do not make use of human-centered models of story comparison.

Human-centered research on story comparison has also been a recent focus, such as the work by Fisseni and Lowe [9] on non-structural dimensions of narratives for the purpose of story equivalence. Of particular relevance to the work presented in this paper, Kypridemou and Michael [14] validate the *common summary* of two stories as a model of their similarity. Through human subject validation, the authors confirm that the more appropriate a common summary of two stories, the more similar the two stories are judged to be. While they did not employ a computational model for generating summaries of the stories, their work offers a valuable human-centered principle upon which to define a computational model of a story summary for use in story comparisons.

2.3 Summary of Previous Work

Story generation models that adhere to classic planning principles offer a rich causal representation of events, which has lead to the capability to represent character intentions [26] and conflict [32] as story plan semantics. Unfortunately, plan comparison metrics have focused on syntactic properties of plans [4, 10, 23, 28, 30] not domain-specific semantics related to story plan semantics. Reasonably, computational models of story comparison have leveraged methods from other disciplines [8, 13, 24]; however, human-centered models of story comparison have shown that a common summary of two stories is important to similarity assessments [14]. A plan summarization model based on a story event's causal degree and membership on the causal chain [35] offers a principled plan-based representation to begin comparing CPOCL story plans.

3 Story Planning

As discussed in Section 2.3, a limitation of domain-independent comparison metrics is their focus on measuring syntactic differences between plans. This results in the difference between the story semantics of two plans not being represented in current plan comparison metrics.

We address this limitation by first defining the story plan semantics that support a story plan summarization model. This narrative-theoretic summarization model captures character intentions and causally significant actions from the syntax required to conform to planning formalisms. This is followed by defining a story plan distance metric which computes the differences between two story plan summaries.

3.1 Story Plan Semantics

Story planning algorithms generate story plans to solve a story planning problem, of which there can be many equally-satisfying solutions. The story semantics between these story plan solutions can vary and are not captured in current plan distance metrics. We provide precise definitions of story plans and their related parts to both leverage previous work in classical planning and to ground work from other disciplines in this context. In the tradition of being precise, the following definitions are formalized based on the established body of work in story planning.

A planning domain theory models the way in which the world can change through the application of actions to a world state. Story planners implement algorithms which instantiate and combine actions from a domain theory to reach a goal state from some initial state.

▶ **Definition 1** (Action)**.** An action is defined as the preconditions that must be satisfied before its execution and the effects that result. A precondition is a function-free positive literal in a state space and the conjunction of an action's preconditions must evaluate to true before it can execute. An action's effects are function-free positive literals whose conjunction is the result of the change in state space when an action is executed.

Together with an action's name and parameter list, the precondition and effects describe an *action schema*. An action schema can be instantiated into various forms, dependent on the literals to which the variables unify. Note that the use of the terms *action, step,* and *operator* are used interchangeably.

▶ **Definition 2** (Story planning problem)**.** A story planning problem Φ is a four-tuple $\langle \mathcal{I}, A, \mathcal{G}, \Lambda \rangle$ where $\mathcal{I}$ is a conjunction of function-free ground literals which are true in the initial state, A the set of symbols referring to character agents, $\mathcal{G}$ a conjunction of function-free ground literals which are true in the goal state, and Λ a set of action schemata. This definition is in the way of Riedl & Young [26].

An action's preconditions are satisfied through causal links to an earlier step's effects and, in turn, its effects can satisfy preconditions to later steps.

▶ **Definition 3** (Causal Links)**.** Causal links are denoted $s \overset{p,q}{\rightarrow} u$, where s, u are steps in S of story plan P, with an effect p and precondition q respectively. In this case, q is satisfied for u because s had p as an effect and there exists a literal r that unifies with p and q. The step s is a causal parent of u, while u is the causal child of s. The causal parents of s are causal ancestors of u and the causal children of u are the causal descendants of s. These transitive relations extend until an action with no parents or children is reached.

Steps which establish causal links for subsequent steps are at risk of some other step's effect undoing a causal parent's effect, thereby transforming the state of the world into one which the causal child action cannot execute.

▶ **Definition 4** (Causal link threat)**.** A causal link threat occurs when a causal link is established $s \overset{p,q}{\rightarrow} u$, and some other step w has the effect $\neg p$ and could be executed after s but before u. Executing w in this interval means the precondition q of u is no longer satisfied by s and u will not execute.

A causal link threat to $s \overset{p,q}{\rightarrow} u$ can be resolved by introducing an ordering constraint such that w is executed before s or after u. Resolving all causal link threats to this causal link guarantees that p remains true after s is executed until u is executed, preserving the causal link with u.

► **Definition 5** (Ordering constraint). An ordering constraint is of the form $w \prec s$ and is interpreted as w must be executed at some time before s.

Establishing causal links between steps requires maintaining which precondition and effect variable pairs must unify. Additionally, it may be desired that the free variables in action parameters are not equivalent. Binding constraints are used to capture both these conditions.

► **Definition 6** (Binding constraint). A binding constraint is a pair of variables (u, v) or negated pair $\neg(u, v)$ where the pair must unify or not be allowed to unify, respectively.

Characters are an important element of stories. In support of this, character intention frames were implemented in story plans to justify individual character actions in service of their goals.

► **Definition 7** (Intention Frame). An intention frame in a plan P is a tuple of five elements $\langle c, g, m, \sigma, T\rangle$ where c is a character, a goal that c intends to make true is represented by g. The motivating step m is in $S(P)$ with the effect $intends(c, g)$, the satisfying step σ is also in $S(P)$ and has g as an effect. The set of steps T is a subset of steps from $S(P)$ taken by c to achieve the goal effect g. T is called the character's subplan to achieve the goal effect g. All steps in T must occur after the motivating step m and before σ.

The POCL (Partial Order Causal Link) family of planners searches through a story plan space to produce story plans of the form defined below.

► **Definition 8** (Story plan). A story plan P is a tuple of five elements $\langle S, B, O, L, I\rangle$ where the set of steps is S (with executed steps denoted S_e and non-executed S_{ne}), the set of binding constraints on the free variables of S is defined as B, the partial ordering of the steps in S defined as O, L the set of causal links joining steps from S, and finally I, the set of intention frames which define character subplans in S. This definition is consistent with the definition of a CPOCL plan in Ware *et al.* [33].

The differentiation of executed and non-executed steps is made in support of representing character conflict. Non-executed steps are part of foiled sub-plans that a character had the intent of completing, but could not due to an unresolved causal link threat with another step.

► **Definition 9** (Story plan solution). A story plan P is a *solution* to the story plan problem Φ if its actions, which has $\mathcal{I}(\Phi)$ as effects of the first action and $\mathcal{G}(\Phi)$ as preconditions to the goal step, have no open preconditions and is *consistent*. A story plan is *consistent* if there are no cycles in the ordering constraints $O(P)$ and no causal link threats remain between two *executed* steps in $S_e(P)$.

While these definitions may appear verbose, they are necessary to formally characterize the representation of a story plan solution. This representation leads to a story planning problem to, in fact, have many story plan solutions and they can differ in both their syntax (*e.g.* actions) and in more human centric ways. The next sections focus on operationalizing the ability of humans to recall certain events as a model of summarization and comparison.

3.2 Story Plan Summarization

The story semantics detailed in Section 3.1 lay the principles for a computational model of a story plan summary. The importance of causal degree and the causal chain in the recall of story events [31] suggests a mechanism to summarize the logical progression of plot. Additionally, intention frame summaries play an important role in the comprehension of a story. These two properties will form the basis of a story plan summary, which is incrementally defined below.

We capture a story plan's causal links in the $n \times n$ causal matrix $\mathcal{S}_P$,

$$\mathcal{S}_P = \begin{bmatrix} 0 & s_{12} & s_{13} & \dots & s_{1n} \\ 0 & 0 & s_{23} & \dots & s_{2n} \\ 0 & s_{32} & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & s_{n-1n} \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix}. \tag{3}$$

where $n = |S(P)|$. The s_{ij} entry contains the number of effects of step i used as preconditions by step j. Note the zeroes in the initial state's column (s_1) and of the goal state's row (s_n), which denote the lack of preconditions and effects, respectively. It is necessary to compute $\mathcal{S}_P$ for each story plan solution, and not for the action schemata, as $\mathcal{S}_P$ captures the number of times the effects of a step are used in a story solution plan, which can be equal to, greater than, or less than the effects in the action schema.

To support the identification of important events in a story plan, the causal degree of the step i, s_i, is computed directly from $\mathcal{S}_P$ by adding both the ith row sum,

$$deg_c^-(s_i) \quad = \quad \sum_{j=1}^{n} s_{ji}, \tag{4}$$

a step's used effects, and column sum,

$$deg_c^+(s_i) \quad = \quad \sum_{j=1}^{n} s_{ij}, \tag{5}$$

a step's satisfied preconditions.

▶ **Definition 10** (Causal degree)**.** The causal degree of a step s_i in a plan P is the sum of the step's preconditions $deg_c^-(s_i)$ and the sum of the effects used by each causal child $deg_c^+(s_i)$, defined in Equation6:

$$deg_c(s_i) \quad = \quad deg_c^-(s_i) + deg_c^+(s_i). \tag{6}$$

The criteria for a causal chain having a motivation, beginning, and end emerged from Trabasso and Sperry's work [31] and is directly applicable to story plans as character goals, the initial state, and the goal state, respectively.

▶ **Definition 11** (Causal chain)**.** A causal chain C of a story plan solution P, $C(P)$, is a subset of $S(P)$, which consists of all the steps that are causal ancestors of the goal step, plus those steps' causal descendants. The causal chain excludes both the initial and goal step, s_1 and s_n, respectively.

The causal chain excludes the initial and goal steps. While they are both elements of $S(P)$, they are never executed by the planner. Rather they are only states to delineate the start and end of a plan. Including these steps in the causal chain would only serve to complicate deriving story plan summaries.

▶ **Definition 12** (Important steps)**.** The important steps E of a story plan solution P is the set of *executed* steps in the causal chain $C(P)$ with the *highest* causal degree computed from the matrix $\mathcal{S}_P$.

While non-executed steps provide a construct to identify conflicts between characters, they are excluded from important step calculations as their effects are not realized in the story world state. Additionally, we use highest causal degree to mean an exact number, and it is expected that the number of important steps in story plan will be small.

The set of steps T of an intention frame $I = \langle c, g, m, \sigma, T \rangle$ (Definition 7) captures the means by which a character achieves a goal and it has been used to reflect character traits [2]; however, this could include steps without any story plan semantics, such as movement actions. While such actions are necessary for the progression of the story plan, we do not have a character-centric model to capture which steps are meaningful in story plans. In order to avoid these steps influencing our comparisons, we simply remove T to form an intention frame summary, until such a time that plan steps can be validated as a robust character trait model.

▶ **Definition 13** (Intention frame summary)**.** An intention frame summary j of some intention frame in I from the story plan P is a four-tuple $\langle c, g, m, \sigma \rangle$ where c,g,m and σ are preserved from Definition 7. The set of intention frame summaries of P is denoted as $J(P)$, where each intention frame in $I(P)$ has a corresponding intention frame summary in $J(P)$.

We combine the causally important steps (Definition 12) and intention frames summaries (Definition 13) to define the *Important-Step Intention-Frame* story plan summary as

$$\psi^{ISIF}(P) \quad = \quad \langle E, J \rangle, \tag{7}$$

where P is a story plan solution, important steps E are computed from the steps $S(P)$ and intention frame summaries $J(P)$ derived from the intention frames $I(P)$.

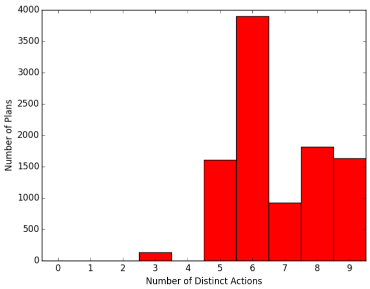
3.3 Story Plan Distance Metric

Previous plan comparison metrics compare, on a limited basis, the semantics between story plans as they apply uniform importance to all plan syntax. We address this limitation by using the formalization of story plan summaries in Section 3.2, which extracts the story plan semantics to make them available for direct comparison. We first define the notion of the distance between two story plan summaries formally.

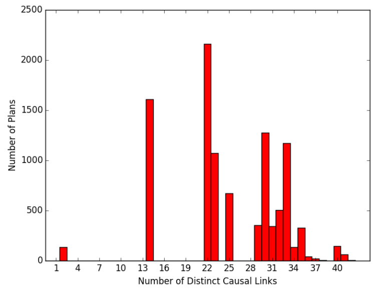
▶ **Definition 14** (Story plan distance metric)**.** A distance metric between two story plan summaries, ψ_1 and ψ_2, is a function Δ where $\Delta(\psi_1, \psi_2) \longrightarrow [0, 1]$. The value of zero represents perfect similarity between the two summaries, where the value of one denotes perfect dissimilarity.

In order for any Δ to be mathematically consistent as a distance metric, it needs to satisfy the identity, symmetry, and triangle inequality properties. As a result, proper subsets of plans must not be made by the metric itself, as it violates the identity property of a distance metric. As an example, a metric that only compared the penultimate steps of two plans would assess them as being identical when their penultimate step is equal, when in fact the rest of the steps of the plan could be all different. To maintain mathematical consistency, the distance metric defined in this section is based on *story plan summaries.*

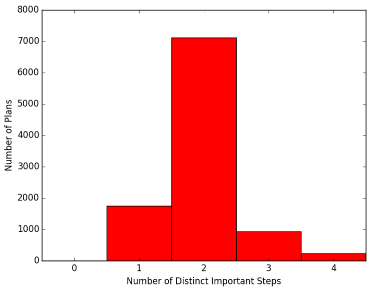
Previous work in plan comparison has made use of the Jaccard distance metric to compare plans. While not theoretically motivated or explicitly denoted by the authors in [23, 30], Jaccard distance is often used as a simple distance metric for assessing non-normal data, with no underlying distribution and no linear relationships. A key feature of the Jaccard distance is the use of the *intersect* set operator in the numerator as it accounts for the



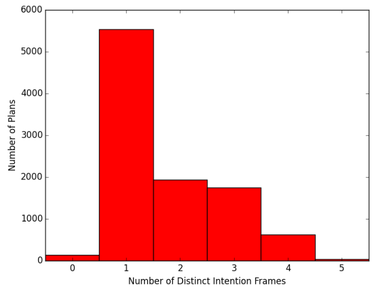
(a) Ground actions (S) per plan



(b) Ground causal links (L) per plan



(c) Ground important steps (E) per plan



(d) Unique intention frames (I) per plan

Figure 1 Solution space distributions of Π.

common elements between two story plan summaries, a factor for human judgment of story similarity [14].

We define the Jaccard distance metric between two story plan summaries as

$$\delta_{ISIF}(\psi_1^{ISIF}, \psi_2^{ISIF}) = 1 - \frac{1}{2}\left(\frac{\left|E(\psi_1^{ISIF}) \cap E(\psi_2^{ISIF})\right|}{\left|E(\psi_1^{ISIF}) \cup E(\psi_2^{ISIF})\right|} + \frac{\left|J(\psi_1^{ISIF}) \cap J(\psi_2^{ISIF})\right|}{\left|J(\psi_1^{ISIF}) \cup J(\psi_2^{ISIF})\right|}\right), \qquad (8)$$

where ψ_1^{ISIF} and ψ_2^{ISIF} are ISIF story plan summaries (Equation 7), E is the important events of a story plan summary, and J is the intention frame summaries of a story plan summary. We introduce a factor of $\frac{1}{2}$ over the Jaccard similarity of E and J to ensure the metric remains between 0 and 1. This also has the consequence of equally weighting both E and J. We can interpret the δ_{ISIF} distance metric as story plan summaries which have little in common in terms of intention frames and causally important steps will have a score close to 1, whereas similar summaries will have a δ_{ISIF} close to 0.

3.4 Summary of Definitions

Past research in plan comparison has classified distance metrics into domain-independent and domain-specific. While the distance metric described in Section 3.3 can be classified as a domain-specific metric, the use of narrative-theoretic constructs in the story plan summary affords the metric to generalize to all story domains and would classify as a *Narrative Metric* under the StoryEval framework [29]. In practice, story plan distance metrics are constrained to the representational capability of story planners. In Section 4, we turn our focus to determining whether the δ_{ISIF} distance metric can capture differences in story semantics in the solution space of a simple story planning problem.

4 Experimental Results

We desire to characterize the story branches in an IN that uses automated planning for story generation. This requires differentiating between story plans which equally-satisfy a single story planning problem. With this in mind, the evaluation uses a single solution plan-set generated from the Glaive narrative planner [32], which has been previously used for IN story generation, and does not consider comparing plans across story planning problems or story planning domains.

Our commitment to the CPOCL story plan representation introduces another consideration in evaluating the δ_{ISIF} story plan distance metric. We must consider what alternative methods exist to compare two CPOCL story plans. Previous story comparison efforts fall into two classes. The first are human judgments of story comparisons [9, 14], making them difficult to formalize to a CPOCL representation. A second class uses existing domain-independent measures that accept the author's representation as input, such as the use of string comparison [8, 24] and subgraph isomorphism [7]. In short, previous story comparison methods either require capturing the stories in CPOCL story plans or adapting CPOCL story plans to the method's required input. Rather than undertake the task of further formalization, we use CPOCL plans' adherence to automated planning formalisms to find appropriate plan-based domain-independent distance metrics to compare the δ_{ISIF} story plan distance metric. Conveniently, established plan-based distance metrics which capture the action-oriented and causally-linked representation overlaps between narrative and automated planning already exist, namely the action (δ_A) and causal-link (δ_C) distance metrics.

The goal in the preliminary evaluation in the following section is two-fold. Firstly, to investigate the ability of δ_{ISIF} story plan distance metric and existing distance metrics to capture the story branches in an IN story problem's solution plan-set. Secondly, demonstrate that when comparing a pair of plans the $\delta_{ISIF} \bullet$ story plan distance metric is more sensitive to small syntactical differences that result in significant story semantic differences than existing automated-planning distance metrics.

4.1 Story Domain and Problem

To demonstrate the differences between the different distance metrics, we use the *space* story domain, which was previously defined to evaluate the Glaive planner and highlight the CPOCL representation [32]. The space story domain was the simplest of the domains used in the evaluation and consists of two variable types, eleven predicates, and ten actions. This domain was used as it presents enough complexity to highlight the differences between distance measures, but not so much as to lead the analysis into extreme cases.

The story problem used is one in which an astronaut ("Zoe") must resolve conflicting intentions of exploration and self-preservation as a volatile planet becomes inhospitable. This story problem is referred to as the *exploration* problem. The initial state of the *exploration* problem consists of fourteen predicates and five character intentions, three belonging to Zoe and two to an alien. A key feature of this problem is the goal step consisting of a single predicate; $\neg$(*habitable surface*), which is only possible by one action in the domain; *begin-erupt*.

4.2 Solution Plan-Set Analysis

The Glaive heuristic search planner (HSP) [32] was able to generate a solution plan-set of $10,000$ CPOCL solution plans, Π, to the exploration problem. While each story plan is

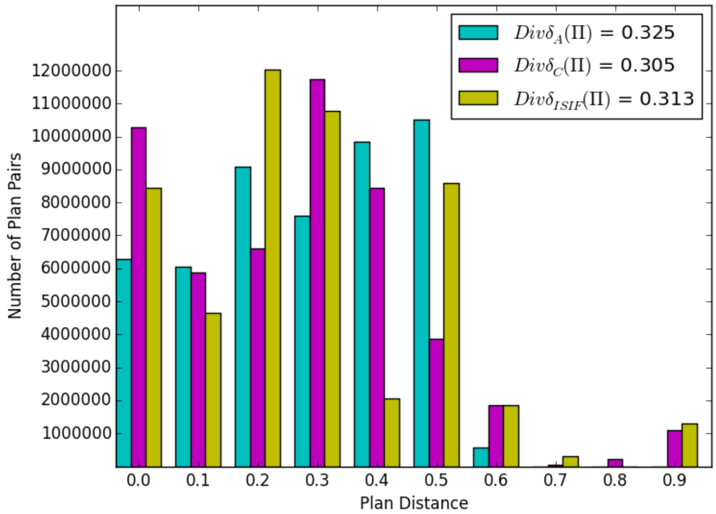


Figure 2 Solution diversity using different distance metrics.

equally-satisfying in their ability to solve the story planning problem, significant differences exist between their story semantics. Of the plan comparison metrics reviewed, δ_A (Equation 1) and δ_C (Equation 2) are the most relevant to compare with δ_{ISIF} as they capture first principle overlaps between planning and stories: namely, action-oriented nature and causally-related events. Additionally, the metrics do not require total orderings, a necessary criteria when comparing partial order plans.

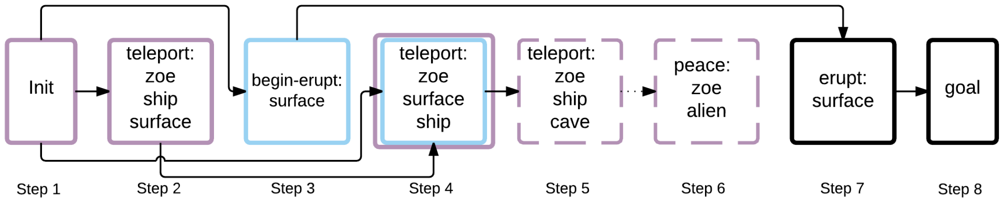
We can obtain some insight of the pertinent elements of Π to each distance metric in Figures 1a–1d. Specifically, both the distribution of steps (Ground actions, Figure 1a) and causal links (Figure 1b) show that a non-trivial number of elements are used in the calculations made by the δ_A and δ_C distance metrics. While the number of important steps (Figure 1c) and intention frames (Figure 1d) are smaller relative to ground actions and causal links, they are combined together in the δ_{ISIF} metric to achieve greater size. To sum, we can see from Figures 1a–1d that despite the simplicity of the space exploration problem, the distribution of the relevant properties in plans of Π support substantive syntactic and story semantic comparisons.

We use a standard calculation from the automated planning community to capture the differences in the solution space. The *plan-set diversity* [6] is a pairwise comparison made between every solution plan in Π,

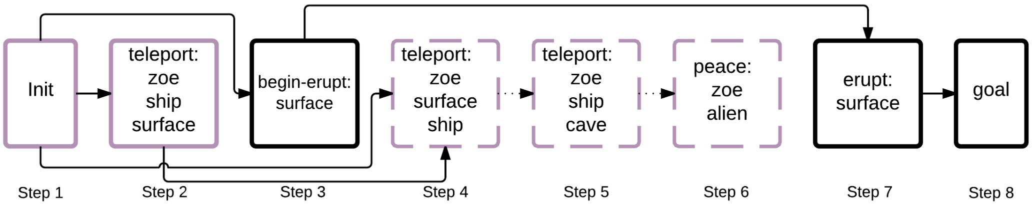
$$Div(\Pi) \quad = \quad \frac{\sum\limits_{\pi,\pi' \in \Pi} D(\pi,\pi')}{\frac{|\Pi| \times (|\Pi|-1)}{2}}, \tag{9}$$

where π and π' are plans in a plan-set, Π, and $D(\pi,\pi')$ is a distance metric. The distribution of plan-pair comparisons using the three distance metrics and their associated diversity measures are shown in Figure 2.

We can observe the agreement in plan-set diversity using the three metrics (0.325, 0.305, 0.313); yet, this does not capture the differences in distribution. Take for instance the lack of scores near 1.0 for the δ_A metric but the presence of such scores for δ_{ISIF}. This is an artifact of the *exploration* problem's goal state containing the $\neg$(*habitable surface*) predicate, which requires all solutions to have the erupting volcano action as it is the only action with this



(a) Plan π_1



(b) Plan π_2

Figure 3 Story plans with intention frames in color.

Table 1 Story plan properties.

π_i	$\lvert S(\pi_i)\rvert$	$\lvert L(\pi_i)\rvert$	$\lvert I(\pi_i)\rvert$	$\lvert E(\pi_i)\rvert$	$C(\pi_i)$	$E(\pi_1)$
π_1	8 (6 executed, 2 non-executed)	35	2	2	$\{s_2, s_3, s_4, s_7\}$	$\{s_2, s_4\}$
π_2	8 (5 executed, 3 non-executed)	35	1	1	$\{s_2, s_3, s_7\}$	$\{s_2\}$

predicate as an effect. The presence of scores at 1.0 when using the δ_{ISIF} metric indicates that there is more than a single narrative experience in the solution plan-set. While we can glean some insights from the plan-set diversity measure when using the δ_{ISIF} metric, plan-set diversity does not capture, for instance, the exact number of branches in a story problem's solution space. We leave the development and application of other plan-set measures that capture important IN characteristics for future work.

4.3 Solution Plan-Pair Analysis

The solution space distance metric results presented in Section 4.2 illustrate qualities of the solution space coupled to properties of the story domain and story problem; however, specific examples of the story plan distance measure are needed to demonstrate the effectiveness of the story summary distance metric. The following example uses two story plans, π_1 and π_2, from Π to illustrate when small syntactic differences have large semantic differences between story plans.

We observe the difference in structure of the two plans in Figure 3, where executed actions are solid rectangles, non-executed actions dashed rectangles, and the links between them represent the existence of one or more causal links. The plans have a similar syntactic structure, containing the same number of steps and causal links (8 and 35 respectively) with a single difference in s_4 (executed vs non); however, this single syntactic difference changes the story semantics in a significant manner by affecting the number of important events and intention frames.

We capture the common intention frame between the two plans in purple and the additional intention frame in π_1 with blue. The common intention frame in each plan is Zoe's goal to make peace with the alien, which is motivated in s_1, and her first action towards

$$\mathcal{S}_{\pi_1} \begin{bmatrix} & & & & & & & & & deg_c^+ \\ & 0 & 6 & 1 & 5 & 5 & 6 & & & 23 \\ & 0 & & & 1 & & & & & \mathbf{1} \\ & 0 & & & & & & 1 & & 1 \\ & 0 & & & & 1 & & & & \mathbf{1} \\ & 0 & & & & & & & & x \\ & 0 & & & & & & & & x \\ & 0 & & & & & & & 1 & 1 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ deg_c^- & 0 & \mathbf{6} & 1 & \mathbf{6} & x & x & 1 & 1 & \end{bmatrix} \quad \mathcal{S}_{\pi_2} \begin{bmatrix} & & & & & & & & deg_c^+ \\ 0 & 6 & 1 & 5 & 5 & 6 & & & 23 \\ 0 & & & 1 & & & & & \mathbf{1} \\ 0 & & & & & & 1 & & 1 \\ 0 & & & & & & & & x \\ 0 & & & & & & & & x \\ 0 & & & & & & & & x \\ 0 & & & & & & & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{6} & 1 & x & x & x & 1 & 1 & \end{bmatrix}$$

Figure 4 Causal Matrices.

Table 2 Example story plan-pair properties.

π_i, π_k	$\delta_A(\pi_i, \pi_k)$	$\delta_C(\pi_i, \pi_k)$	$\delta_{ISIF}(\pi_i, \pi_k)$
π_1, π_2	0.11 (1 - $\frac{8}{9}$)	0.00 $(1 - \frac{35}{35})$	0.50 $(1 - \frac{1}{2}(\frac{1}{2} + \frac{1}{2}))$

completing this goal is s_2; however, the surface begins to erupt in s_3 making Zoe $\neg safe$. Zoe now has conflicting goals and must choose between continuing her existing goal while $\neg safe$ or adopting a new goal of ensuring her life is preserved by teleporting back to the ship and becoming $safe$. We can observe that s_4 is executed in π_1, which constitutes Zoe's first and satisfying step of her goal to be $safe$ instead of continuing her intention to make peace with the alien. This additional intention frame contrasts with π_2 where s_4 is a non-executed step, indicating Zoe does not adopt the goal of being $safe$, thus only a single intention frame of her search for the alien.

The change of s_4 from non-executed to executed step also affects the number of causally-important steps. Both plans have s_2 as an important step, since it is both on the causal chain and has the highest causal degree, seven as $deg_c^-(s_2) = 6$ and $deg_c^+(s_2) = 1$ (see Figure 4). Identically, we observe that the causal degree of s_4 is also seven, $deg_c^-(s_4) = 6$ and $deg_c^+(s_4) = 1$; however since the definition of a causal chain requires steps to be executed, s_4 is excluded from $E(\pi_2)$. This additional member of $E(\pi_1)$ demonstrates how important steps can vary from equally-satisfying solutions and have a semantic impact on story plan content.

The semantic properties, as well as the syntactic properties, which are relevant for computing the three distance metrics between π_1 and π_2, are shown in Table 1. We observe in Table 2 that the syntactic change in s_4 between the plans results in a difference of 0.11 when calculated using the δ_A metric. The same change results in the δ_C metric calculated as 0.00, as the causal structure is identical from π_1 to π_2. In contrast, the δ_{ISIF} calculation results in a much more significant 0.50 as it captures the semantic importance of the small syntactic change. This relatively large difference between the domain-specific (δ_{ISIF}) and domain-independent (δ_A, δ_C) metrics demonstrates the ability of a metric based on story plan semantics to more accurately capture subtle differences in story plan syntax.

4.4 Summary of Experimental Results

In this section, we described the *space* domain's *exploration* problem then evaluated three plan distance metrics on their ability to compare two equally-satisfying story plans. The δ_A metric provided a useful insight into the composition of the problem definition, namely all solution plans will contain the *begin-erupt* step. It further suggests that meaningful insights could be ascertained by describing the solution space using the δ_{ISIF} metric. Lastly, we used two equally-satisfying story plans with a single syntactic difference to illustrate the difference in story plan semantics that δ_{ISIF} captures. These results support semantic summaries as a promising approach to story plan comparison.

5 Limitations

We have presented and demonstrated formalized models of story plan summarization and comparison for a CPOCL story plan representation. It is worth emphasizing that while CPOCL represents validated forms of character intentions and conflict, its adherence to a strict automated planning formalism limits its capability to represent narrative. Story comparisons extend far beyond those stories which CPOCL and the models presented in the paper capture. As more advances are made in deriving story semantics from classic plan representations into story plans, summarization models and comparison metrics will be able to build on the contributions of this paper.

6 Conclusions

Branching stories have enjoyed success in recent games and quantifying the story branches in a narrative planner requires a story-plan comparison metric. In support of addressing this need, this paper draws together research from both computer science and narrative theory to synthesize two primary contributions to advance CPOCL story plan comparison.

The first is the definition of a computational model for a story plan summary supported by cognitive psychology and narratology. A story plan summary uses story plan steps' causal degrees, in addition to character intention frame summaries to capture story plan semantics while ignoring syntactic structure of less importance. The resulting story plan summary is a concise semantic representation of a story plan and provides a foundation upon which to compare other story plans.

A second contribution is the definition of the δ_{ISIF} story plan distance metric. The distance metric leverages previous work in plan comparison, namely the use of Jaccard similarity, to compare domain-specific properties. While the formulation of the distance metric weights the score equally over the important step and intention frame components, it avoids learning weighted parameters and violating the definition of a distance metric.

We demonstrated the above contributions using a CPOCL solution plan-set to the *exploration* problem. At both the solution plan-set level and in specific story plan-pair examples, we calculate the differences between two existing syntactic plan comparison metrics and the δ_{ISIF} story plan comparison metric. We exhibit that when using the δ_{ISIF} distance metric in plan-set comparisons it can confirm the existence of mutually-exclusive story plan summaries, while at the story plan-pair level it captures small syntactic differences that have larger semantic differences. These results are significant, not only do the story plan summarization model and distance metric enable more robust comparisons of story plans but we expect them to generalize to all story plan domains due to their cognitive psychology and narratology foundations. We expect to validate these models in a human subject evaluation in the near future.

References

1 H. Porter Abbott. *The Cambridge Introduction to Narrative*. Cambridge University Press, second edition, 2008. `doi:10.1017/CBO9780511816932`.

2 Julio César Bahamón and R. Michael Young. Toward a computational model for the automatic generation of character personality in interactive narrative. In *Intelligent Virtual Agents*, pages 520–522. Springer, 2012.

3 Mieke Bal. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, 1997.

4 Daniel Bryce. Landmark-based plan distance measures for diverse planning. In *ICAPS*, pages 56–64, 2014. URL: `http://www.aaai.org/ocs/index.php/ICAPS/ICAPS14/paper/view/7903/8011`.

5 Bradley J. Clement, Edmund H. Durfee, and Anthony C. Barrett. Abstract Reasoning for Planning and Coordination. *Journal of Artificial Intelligence Research*, 28:453–515, 2007. `doi:10.1613/jair.2158`.

6 Alexandra Coman and Héctor Muñoz avila. Generating Diverse Plans Using Quantitative and Qualitative Plan Distance Metrics. In *AAAI Conference on Artificial Intelligence*, pages 946–951, 2011.

7 David K. Elson. Detecting Story Analogies from Annotations of Time, Action and Agency. In *Proceedings of the Third Workshop on Computational Models of Narrative*, pages 91–99, 2012.

8 Matthew P. Fay. Story Comparison via Simultaneous Matching and Alignment. *Computational Models of Narrative Workshop*, pages 100–104, 2012.

9 Bernhard Fisseni and Benedikt Löwe. Which dimensions of narrative are relevant for human judgments of story equivalence? *Computational Models of Narrative Workshop*, pages 114–118, 2012. URL: `http://dare.uva.nl/document/362429`.

10 Robert P. Goldman and Ugur Kuter. Measuring Plan Diversity : Pathologies in Existing Approaches and A New Plan Distance Metric Normalized Compression Distance for Plan. In *AAAI Conference on Artificial Intelligence*, 2015.

11 H. Paul Grice, Peter Cole, and Jerry L. Morgan. Syntax and semantics. *Logic and conversation*, 3:41–58, 1975.

12 David Herman. *Narrative theory and the cognitive sciences*. Center for the Study of Language and Information, 2003.

13 Joshua Jones and Charles L. Isbell. Story Similarity Measures for Drama Management with TTD-MDP. In *International Conference on Autonomous Agents and Multi-agent Systems*, pages 77–84, 2014. URL: `http://dl.acm.org/citation.cfm?id=2615747`.

14 Elektra Kypridemou and Loizos Michael. Narrative Similarity as Common Summary. In *Workshop on Computational Models of Narrative*, pages 129–146, 2013.

15 Michael Lebowitz. Story-telling as planning and learning. *Poetics*, 14(6):483–502, 1985. `doi:10.1016/0304-422X(85)90015-4`.

16 Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

17 James R. Meehan. TALE-SPIN: An Interactive Program that Writes Stories. In *International Joint Conference on Artificial Intelligence*, volume 77, pages 91–98, 1977. URL: `http://www.ijcai.org/PastProceedings/IJCAI-77-VOL1/PDF/013.pdf`.

18 Chris Mellish and Roger Evans. Natural language generation from plans. *Computational Linguistics*, 15(4):233–249, 1989.

19 David S. Miall. Experiencing narrative worlds: On the psychological activities of reading. *Journal of Pragmatics*, 32(3):377–382, Feb 2000. `doi:10.1016/S0378-2166(99)00017-X`.

20 Karen L. Myers. Metatheoretic Plan Summarization and Comparison. *Artificial Intelligence*, pages 182–191, 2002.

21 Karen L. Myers. Temporal Summarization of Plans. In *International Conference on Automated Planning and Scheduling*, 2007.

22 Saul B. Needleman and Christian D. Wunsch. 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, 1970. `doi:10.1016/0022-2836(70)90057-4`.

23 Tuan Anh Nguyen, Minh Do, Alfonso Emilio Gerevini, Ivan Serina, Biplav Srivastava, and Subbarao Kambhampati. Generating Diverse Plans to Handle Unknown and Partially

Known User Preferences. *Artificial Intelligence*, 190:1–31, 2012. doi:10.1016/j.artint.2012.05.005.

24 Julie Porteous, Fred Charles, and Marc Cavazza. NetworkING : Using Character Relationships for Interactive Narrative Generation. In *International Conference on Autonomous Agents and Multi-agent Systems*, pages 595–602, 2013.

25 Rafael Pérez ý Pérez and Mike Sharples. MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(2):119–139, 2001. doi:10.1080/09528130118867.

26 Mark O. Riedl and R. Michael Young. Narrative Planning : Balancing Plot and Character. *Journal of Artificial Intelligence Research*, 39:217–267, 2010.

27 David L. Roberts, Andrew S. Cantino, and Charles L. Isbell. Player Autonomy versus Designer Intent: A Case Study of Interactive Tour Guides. *Proceedings of the Third Artificial Intelligence and Interactive Digital Entertainment Conference*, pages 95–97, 2007. URL: http://www.aaai.org/Papers/AIIDE/2007/AIIDE07-020.pdf.

28 Mark Roberts, Adele E. Howe, and Indrajit Ray. Evaluating Diversity In Classical Planning. In *ICAPS*, pages 253–261, 2014.

29 Jonathan P. Rowe, Scott W. McQuiggan, Jennifer L. Robison, Derrick R. Marcey, and James C. Lester. STORYEVAL: An Empirical Evaluation Framework for Narrative Generation. In *AAAI Spring Symposium: Intelligent Narrative Technologies*, pages 103–110, 2009.

30 Biplav Srivastava, T. A. Nguyen, and A. Gerevini. Domain Independent Approaches for Finding Diverse Plans. In *International Joint Conference on Artificial Intelligence*, pages 2016–2022, 2007. URL: http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-325.pdf.

31 Tom Trabasso and Linda L. Sperry. Causal Relatedness and Importance of Story Events. *Journal of Memory and Language*, 24:595–611, 1985. doi:10.1016/0749-596X(85)90048-8.

32 Stephen G. Ware and R. Michael Young. Glaive: A State-Space Narrative Planner Supporting Intentionality and Conflict. In *International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2014. URL: http://stephengware.com/publications/ware2014glaive.pdf.

33 Stephen G. Ware, R. Michael Young, Brent Harrison, and David L. Roberts. A Computational Model of Plan-based Narrative Conflict at the Fabula Level. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(3):271–288, 2014. doi:10.1109/TCIAIG.2013.2277051.

34 R. Michael Young. Notes on the Use of Plan Structures in the Creation of Interactive Plot. In *AAAI Fall Symposium on Narrative Intelligence*, pages 164–167, 1999. URL: http://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-01/FS99-01-028.pdf.

35 R. Michael Young. Using Grice's Maxim of Quantity to Select the Content of Plan Descriptions. *Artificial Intelligence*, 115(2):215–256, 1999. doi:10.1016/S0004-3702(99)00082-X.

Leveraging a Narrative Ontology to Query a Literary Text

Anas Fahad Khan[1], Andrea Bellandi[2], Giulia Benotto[2], Francesca Frontini[2], Emiliano Giovannetti[2], and Marianne Reboul[6]

1 **Dipartimento di Studi Umanistici – Università di Ca' Foscari, Venezia, Italy and Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa, Italy**
`fahad.khan@unive.it`

2 **Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa, Italy**
`andrea.bellandi@ilc.cnr.it`

3 **Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa, Italy**
`giulia.benotto@ilc.cnr.it`

4 **Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa, Italy**
`francesca.frontini@ilc.cnr.it`

5 **Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa, Italy**
`emiliano.giovannetti@ilc.cnr.it`

6 **Labex OBVIL, Sorbonne Paris-IV, Paris, France**
`marianne.reboul@free.fr`

Abstract

In this work we propose a model for the representation of the narrative of a literary text. The model is structured in an ontology and a lexicon constituting a knowledge base that can be queried by a system. This *narrative* ontology, as well as describing the actors, locations, situations found in the text, provides an explicit formal representation of the timeline of the story. We will focus on a specific case study, that of the representation of a selected portion of Homer's Odyssey, in particular of the knowledge required to answer a selection of salient queries, formulated by a literary scholar. This work is being carried out within the framework of the Semantic Web by adopting models and standards such as RDF, OWL, SPARQL, and lemon among others.

1998 ACM Subject Classification I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Ontology, Lexicon, Computational Narratology, Time, Semantic Web

Digital Object Identifier 10.4230/OASIcs.CMN.2016.10

1 Introduction

In this article we present a model for the computational representation of narrative using ontologies encoded in the Web Ontology Language (OWL) with the aim of querying literary texts on a semantic basis. We have chosen to focus on Homer's Odyssey as a test case both because of its importance as one of the foundational texts of Western literature but also because of the great deal of research that has been already carried out on the identification of possible narrative structures in Homer's epic poem (for example see the well known study: [11]). In general most of the published literature on the Odyssey makes at best limited use of computational techniques and there seems to be a lack of research into how one might make such data more accessible and more useful from a computational viewpoint. We have attempted to fill this gap by bringing together methods, techniques, and tools from fields such as ontology engineering and linked data in addition to the more traditional scholarly

© Anas Fahad Khan, Andrea Bellandi, Giulia Benotto, Francesca Frontini, Emiliano Giovannetti, and Marianne Reboul;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 10; pp. 10:1–10:10
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

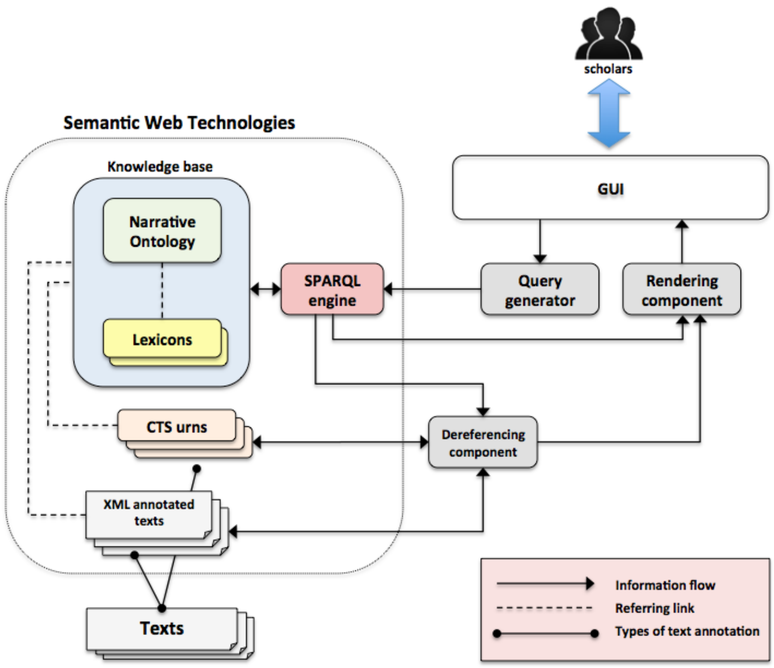


Figure 1 A sketch of the overall architecture of the system.

methodologies applied to the Homeric text, to create a resource that will assist and enable scholarly research on the text of the Odyssey.

Within the wider context of the Digital Humanities, machine-readable representations of literary texts are already regarded as essential for a number of purposes, amongst which those of literary and narratological research [23, 22] as well as automatic storytelling [3, 8, 20, 18]. In order to pursue these aims, many aspects of narrative have been formally modelled such as the descriptions of full narratives as logically-and temporally-ordered streams of formalized elementary events [21, 17], the notions of characters and narrative world [23, 1], stories and actions [3].

In Section 2, we will detail the construction of our Homeric "narrative" ontology and describe how it models the content of the Odyssey by including formal descriptions of events, characters, places, etc, as well as the timeline (or more accurately timelines) of the story. This ontology is only one component in a more general Semantic Web based system for studying literary texts, in this case one specific text and its translations. We chose to work with Semantic Web based technologies because they make it far easier to publish and share such scholarly work, as well as to link different datasets and resources together. The overall system can be seen in schematic form in Figure 1, and is essentially a Web application (the use of which, thus, just requires a browser) implementing an advanced querying system. In our system the ontology is linked to one or more RDF lexicons in Greek and in a number of other languages. These datasets are themselves linked to a corpus of TEI-XML files representing different editions and/or translations of the same literary work; references to the corpus are also provided as CTS-URNs[1] [16] addressing specific parts of the texts. The ontology and

[1] CTS stands for Canonical Text Services, the protocol of requests and responses that allow one to work with texts over networks.

the lexicons together constitute a knowledge base to be queried using the SPARQL Query Language. We use a graphical interface that makes use of a controlled natural language to allow users unfamiliar with the SPARQL query language to interact with the knowledge base and retrieve facts about both the narrative structure and the linguistic features of the text.

One of the core motivations behind this work was to enable the study of a literary text in its different versions and translations from several different points of view, including, but not limited to, the (inter) textual, the linguistic, as well as the semantic and narratological; as mentioned above the use of RDF-based technologies in particular facilitates the linking and comparison of different versions/translations of the same text. The system will be evaluated based on: i) how efficiently a user is able to set the kinds of queries that will help him/her answer questions relating to the different descriptions of the storyline in the text, ii) how different aspects of the story or of the protagonists are represented at various points in the text, and iii) whether the system responds correctly to the given queries. Since an evaluation of the functions of a system like this (usually including precision and recall) can be particularly challenging, we will initially ask to a group of users (students and scholars) to fill in a survey to state their opinion about system's performance.

2 Case Study: Homer's Odyssey

For this case study we will only focus on a specific part of the Odyssey: that part called the Telemachy (or Telemacheia), that is to say books I to IV, and books XVI to XXIV. At the beginning of the Odyssey, Telemachus is portrayed as a young prince who is not yet capable of ruling Ithaca; and yet at the end of the poem not only has he acquired "regal" characteristics, but he has began to prove himself by acting like a true ruler.

2.1 The ODY-ONT Knowledge Base

The technical challenges of representing dynamically evolving information in the Web Ontology Language (OWL), which being based on the RDF model permits only unary and binary relations and thus makes it difficult to add temporal arguments, led us to adopt the perdurant based approach developed in [19] in order to model the situations and changes described in the narrative (we rejected other approaches to add temporal information, such as ontology versioning [7], reification, n-ary relations [12] as being less intuitive and harder to work with). In our perdurant based approach, it is a timeslice of each protagonist that participates in an event or process unfolding over time; these timeslices can be seen as bundles of properties that are stable over a given period of time (the temporal extent of each timeslice), at least in our model of the narrative. Using this approach we are able to represent protagonists in the plot of the Odyssey as possessing a given attribute at a certain interval of time and not at other times. We used the vocabularies OWL-TIME and TIME-PLUS as a basis for the temporal parts of ODY-ONT ontology; we used the Proton ontology (`http://ontotext.com/products/proton/`) as an upper level for ODY-ONT. Proton is a basic upper-level ontology providing coverage of the general concepts necessary for a wide range of tasks, including semantic annotation, indexing, and retrieval of documents. It contains about 300 classes and 100 properties and it is formalised in OWL-Lite.

The diagram in Figure 2 represents the event of Telemachus' passing through the corridor ("striding forth") purposefully. Notice that it is a temporal part of Telemachus that participates in the event. ODY-ONT captures all the events (up to a given granularity) described in the Telemachy as well as the main protagonists in the text and some necessary background knowledge not explicitly stated in the text (e.g., Ithaca is a place). The first stage in the

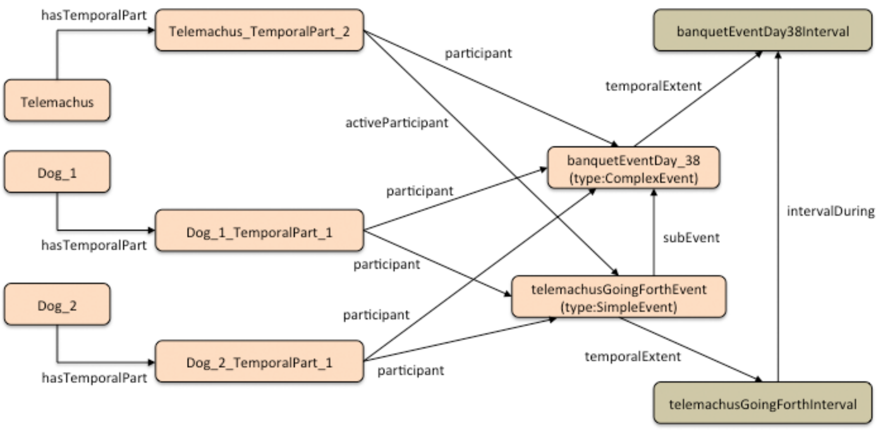


Figure 2 The perdurantist approach. Example from Book XVII of the Odyssey.

creation of ODY-ONT was the identification of the pertinent events in the text. Afterwards, when we had created an ontology individual for each event in the narrative we were able to tag an XML copy of the original Greek text and add the ontology ID for each event referred to in the text, we were also able to do the same for the French translations too. The granularity of the events represented was determined in large part by the kinds of queries that we determined would be interesting and useful both for researchers from a literary or classics background as well as for students and educators. We represented a simple event as an event which is identified by means of exactly one verb or verbal phrase; a complex event is a sequence of many simple events. It is important to point out that events in a story may be ordered in terms of the story timeline, i.e., where they appear in the story described the narration, as well as the narrative timeline, i.e., where they appear in the narration; for the Telemachy these two tend to coincide on the whole.

Figure 3 presents an excerpt from the annotated texts, first in Greek, then in French. The passage in question, from Book I of the Odyssey, describes a first meeting between Telemachus and Athena and is essential in the Odyssey's narrative as it is the first time in the Odyssey a god actually meets a mortal, who in this case will be there explicitly to guide him throughout his adventures. Athena is traditionally the goddess who kept a close watch on Odysseus in the Iliad, and is now about to do the same for his son: the audience immediately understands that Telemachus is about to "grow up" and become the real ruler that he was destined to become. The use of the verb **ἴδε** shows how Telemachus is first and foremost a resigned character, confined to passive expectations, and that he needs to gain control, and learn how to master his own house and country. We link entities such as characters and places to both DBpedia and Wikidata entries (when available). Telemachus of Figure 2, for example, is linked to URIs http://dbpedia.org/page/Telemachus and https://www.wikidata.org/wiki/Q192482. Furthermore, we link items in the ontology with the fragments of text mentioning those items. Currently we use both CTS URN's to address specific portions of the Greek text of the Odyssey as well as X-Pointer to refer to TEI-XML versions of some selected French translations and the original Greek annotated with elements from the ontology [2]. Note that unlike some other similar works in textual narratology [5] we have chosen to (temporarily) put aside all the issues related to discourse analysis and, instead, we chose to annotate the events and predicates in the text and to focus on the modeling of a much richer ontological

```
<ody-Event>
    τὴν δὲ πολὺ πρῶτος
    <ody-pred>ἴδε </ody-pred>
    Τηλέμαχος
    <ody-attribute>θεοειδής</ody-attribute>
</ody-Event>
    ,
<ody-Event>
    <ody-pred>ἧστο</ody-pred>
    γὰρ ἐν μνηστῆρσι
    <ody-Event>
        φίλον
        <ody-pred>τετιημένος</ody-pred>
        ἦτορ
    </ody-Event>
...
```

```
<ody-Event>
    Et , le premier de tous , le
    <ody-attribute>divin</ody-attribute>
    Tèlémakhos
    <ody-pred>vit</ody-pred>
    Athènè
</ody-Event>
    .
<ody-Event>
    Et il
    <ody-pred>était assis</ody-pred>
    parmi les Prétendants ,
    <ody-Event>
        le cœur triste
    </ody-Event>
...
```

Figure 3 Custom XML tags used for annotating the text.

level that includes information on the structure of the narrative and the individual events described in the text. For instance, we codified the temporal sequence of the events in the ontology instead of annotating the text with, for example, TimeML tags. Similarly, in the present version of the work we were not interested in techniques for the automatic or semi-automatic annotation of narrative, as discussed in [6, 14].

In order to add linguistic information to our system we have also included a lexicon incorporating information from an abridged version of the well known Liddell-Scott Greek lexicon in the dataset: we have converted this lexicon into RDF using the lemon model (http://lemon-model.net/). Lemon is a model for representing lexico-semantic data on the Semantic Web. It is heavily influenced by previous computational lexicographic models such as LMF [10]. It has already been used to model a number of other important lexical resources such as the Princeton Wordnet [9], Framenet and Verbnet. We link Greek words tagged in the texts to the relevant senses in the lexicon. We also plan to include RDF French lexicons in the near future. The overall representation model is depicted in Figure 4.

2.2 Querying the ODY-ONT

We designed our system in collaboration with a scholar in the field to ensure that it could answer at least three pertinent types of queries, each differing with respect to the type of knowledge involved and each corresponding to a different typology of research question. The first set of questions involve the events in the Telemachy without reference to their temporal sequence. A possible example relating to our test topic would be to search for all interactions between humans and gods, this for example, would enable us discover the character who interacts the most with divine entities, apart from Odysseus is Telemachus himself.

The corresponding query deals both with the ontological layer as well as with the text. In particular all instances of characters that take part in a dialogue (formalised as a type of event) with a divinity are picked out, together with the instances of the dialogue itself: for each one the system should return the relevant sections of text using the CTS-URN scheme. Indeed, it should be possible to reach the relevant part of the text from each event through the CTS-URN process. For example, in order to access the global section where the action defined by the verb **ἴδε** (cf. Figure 3) is described, the following CTS-URN can be used: `http://data.perseus.org/citations/urn:cts:greekLit:tlg0012.tlg002.perseus-grc1:1.80-1.124`.

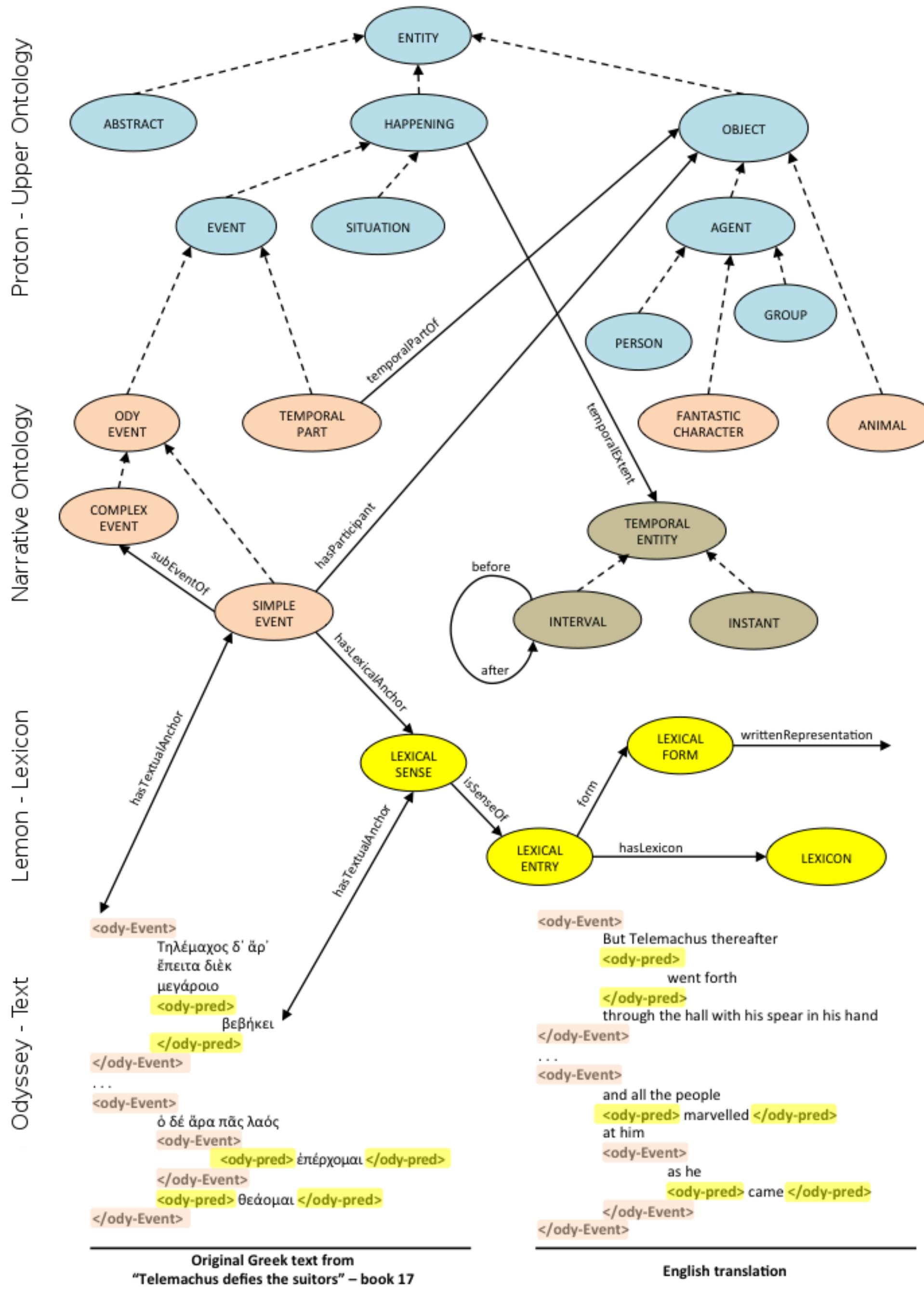


Figure 4 The proposed representation model of ODY-ONT. An example of representation of a simple event. Only the main classes and properties are depicted.

The second set of questions pertain to the temporal aspects of the story and the evolution of the plot and characters. One such query might look into Telemachus' role in the Odyssey, his status as an actor in the story, and in particular which verbs are used to refer to his actions. Our system allows researchers to easily search out a variety of verbs, and to therefore learn about Telemachus' actions. Users can then track the changes, within the narrative

Listing 1 SPARQL code

```
SELECT ?event ?strippedForm ?writtenRepresentation ?cts_urn
WHERE { ?event ody-ont:hasActiveParticipant ?p .
        ?p ody-ont:temporalPartOf ody-ont:Telemachus .
        ?event ody-ont:hasLexicalAnchor ?sense .
        ?event ody_ont:hasTextualAnchor ?cts_urn .
        ?sense praclex:strippedForm ?strippedForm .
        ?sense lemon:isSenseOf ?le .
        ?le lemon:form ?lf .
        ?lf lemon:writtenRep ?writtenRepresentation }
```

structure, in the use of verbs, and the evolution in Telemachus' characteristics throughout the story.

Finally, a specialist in traductology may be interested in seeing the evolution of translation motifs throughout the centuries as well as for instance the evolution of the characters. He/She may therefore want to examine the differences between translations of the verbs in the above query group for our Telemachus case study, and if the translators of the text were able to identify and reproduce the changes in Telemachus' behaviour. The scholar can then measure both an internal evolution, within one translator's text, and an external evolution, within a wider chronological scale, comparing different translators with each other. This kind of query would exploit all the resources of our system, by involving the ontology, the Greek lexicon, the French lexicons, as well as the source text and its translations. As a matter of fact, our system is equipped with an interface to help scholars produce complex queries to answer to complex queries (as, for example, the latter) in a more efficient and effective way.

2.3 Query Interface

Accessing structured data in the form of ontologies requires training and learning formal query languages which poses significant difficulties for non-expert users. In Listing 1, we show a SPARQL query related to the second question presented above; for reasons of lack of space we omit the other queries. The query foresees the interrogation of events that deal with actions performed by Telemachus, i.e., those actions for which Telemachus is the agent (or active participant). This type of query may involve both the Greek text and the Greek lexicon, in order to retrieve the verbs used to describe the actions and to visualize them as a list, which is ordered on the basis of the story or the narration.

One way of lowering the learning overhead and making ontology queries more straightforward to formulate is through a Natural Language Interface (NLI) [4]. In order to make our resource easily accessible to scholars, we are developing a controlled natural language interface [15] for querying the ontology. Research on the text can be performed by taking into account the lexical level, the ontological level or both. The solution we provide in this work refers to the creation of query templates [13]. By means of these, different types of queries can be carried out, using the menus inside each template. The end user can then put together a natural language question, guided by the chosen model. Figure 5 shows an example of our interface related to the query above presented. As illustrated, each template is made of a fixed part that typifies a specific querying model and a variable part that allows the user to chose a specific element of the ODY-ONT knowledge base from the drop-down list. Question templates are processed by the software and converted into SPARQL queries aimed at interrogating the ontology, providing the user with the desired answers. Concerning

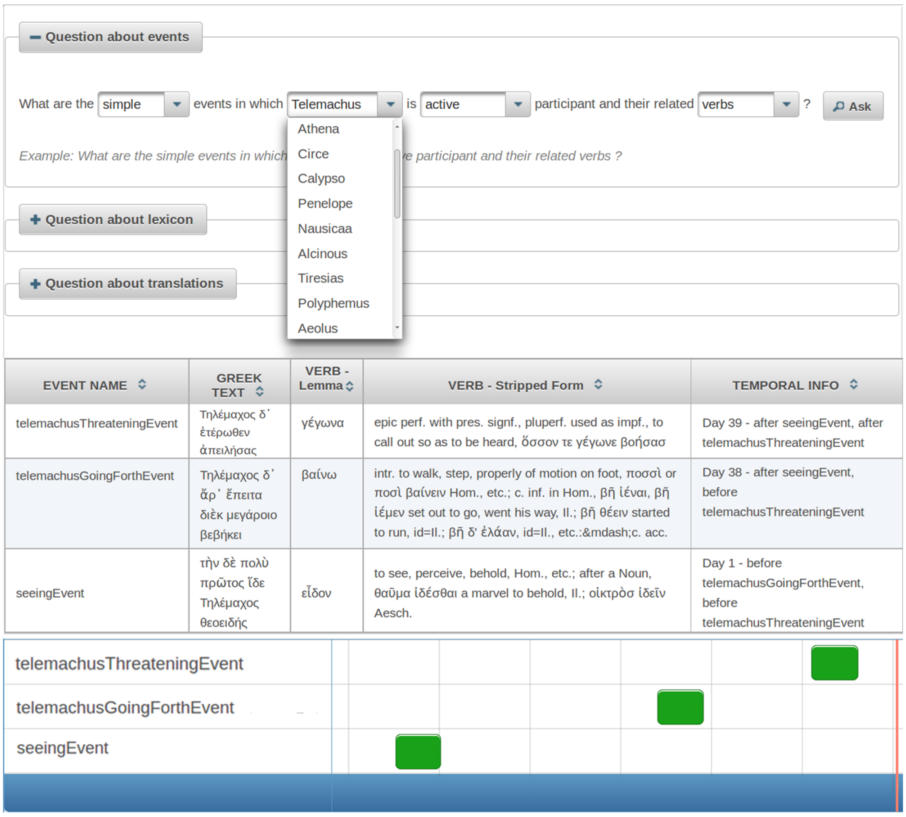


Figure 5 Final user GUI. Running example of the query "What are the simple events in which Telemachus is agent and their related verbs"?

the query of Figure 5, the system retrieves three simple events and for each event it returns the lemma of the verb representing the action, the related text snippet accessed by means of the appropriate CTS-URN, the stripped form of the verb, and the available temporal information. The inference engine uses this information to compute the temporal relations among the retrieved events, and the system draws the appropriate timeline.

3 Conclusions

We feel that the computational representation of the narrative of the text has not yet been adequately addressed in the context of the Semantic Web, despite the fact that linked data technologies are now mature and well-established. In this paper we have proposed a system for the querying of a literary text that integrates existing semantic web technologies. The open issues are still numerous and we intend to address them (in future publications) by carrying on the construction of the ODY-ONT resource (currently being implemented using Stanford's Protégé application), developing the relative lexicons and by carrying out the implementation of the user interface.

References

1 Fabio Ciotti. Toward a formal ontology for narrative. *Matlit: Revista do Programa de Doutoramento em Materialidades da Literatura*, 4:29–44, 2016. doi:10.14195/2182-8830_4-1_2.

2 Fabio Ciotti, Maurizio Lana, and Francesca Tomasi. TEI, ontologies, linked open data: Geolat and beyond. *Journal of the Text Encoding Initiative*, 8, 2014. doi:10.4000/jtei.1365.

3 Rossana Damiano and Antonio Lieto. Ontological Representations of Narratives: a Case Study on Stories and Actions. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 76–93, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.CMN.2013.76.

4 Danica Damljanovic, Valentin Tablan, and Kalina Bontcheva. A text-based query interface to OWL ontologies. In *The 6th International Conference on Language Resources and Evaluation (LREC)*, pages 205–212, 2008.

5 David K. Elson. *Modeling Narrative Discourse*. PhD thesis, Columbia University, 2012.

6 Mark A. Finlayson. *Learning narrative structure from annotated folktales*. PhD thesis, Massachusetts Institute of Technology, 2011.

7 Fabio Grandi and Maria R. Scalas. The valid ontology: A simple OWL temporal versioning framework. In *Advances in Semantic Processing, 2009. SEMAPRO'09. Third International Conference on*, pages 98–102, 2009. doi:10.1109/SEMAPRO.2009.12.

8 Michael O. Jewell, K. Faith Lawrence, Mischa M. Tuffield, Adam Prugel-Bennett, David E. Millard, Mark S. Nixon, Nigel R. Shadbolt, et al. Ontomedia: An ontology for the representation of heterogeneous media. In *In Proceedings of SIGIR workshop on Mutlimedia Information Retrieval*, 2005.

9 John P. McCrae, Christiane Fellbaum, and Philipp Cimiano. Publishing and Linking WordNet using lemon and RDF. In *Proceedings of the 3rd Workshop on Linked Data in Linguistics*, 2014. URL: https://github.com/jmccrae/wn-rdf-paper/raw/master/rdf-wordnet.pdf.

10 John P. McCrae, Dennis Spohr, and Philipp Cimiano. Linking Lexical Resources and Ontologies on the Semantic Web with lemon. In *Proceedings of the 8th Extended Semantic Web Conference*, pages 245–249, 2011.

11 Glenn W. Most. The structure and function of Odysseus' Apologoi. *Transactions of the American Philological Association (1974-)*, 119:15–30, 1989.

12 Natalya Fridman Noy and Alan Rector. Defining N-ary relations on the semantic web, 2006.

13 Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, and Angel Kirilov. KIM–a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10(3-4):375–392, 2004. doi:10.1017/S135132490400347X.

14 Elahe Rahimtoroghi, Thomas Corcoran, Reid Swanson, Marilyn A. Walker, Kenji Sagae, and Andrew Gordon. Minimal narrative annotation schemes and their applications. In *Seventh Intelligent Narrative Technologies Workshop*, 2014.

15 Rolf Schwitter. Controlled natural languages for knowledge representation. In *23rd International Conference on Computational Linguistics*, pages 1113–1121. Association for Computational Linguistics, 2010.

16 Neel D. Smith and Christopher W. Blackwell. Four URLs, limitless apps: separation of concerns in the Homer multitext architecture. *Donum natalicium digitaliter confectum Gregorio Nagy septuagenario a discipulis collegis familiaribus oblatum: A Virtual Birthday Gift Presented to Gregory Nagy on Turning Seventy by His Students, Colleagues, and Friends*, 2012.

17 Mischa M. Tuffield, Dave E. Millard, and Nigel R. Shadbolt. Ontological approaches to modelling narrative, 2006. URL: http://eprints.soton.ac.uk/261962/.

18 Jasper R. R. Uijlings. Designing a virtual environment for story generation. Master's thesis, University of Amsterdam, 2006.

19 Chris Welty and Richard Fikes. A reusable ontology for fluents in OWL. In *Proceedings of the 2006 Conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, pages 226–236, Amsterdam, The Netherlands, 2006. IOS Press.

20 Dov Winer. *Language, Culture, Computation. Computing of the Humanities, Law, and Narratives: Essays Dedicated to Yaacov Choueka on the Occasion of His 75th Birthday, Part II*, chapter Review of Ontology Based Storytelling Devices, pages 394–405. Springer Berlin Heidelberg, 2014. `doi:10.1007/978-3-642-45324-3_12`.

21 Gian Piero Zarri. Representation and management of complex 'narrative' information. In *Part II of Essays Dedicated to Yaacov Choueka on Language, Culture, Computation. Computing of the Humanities, Law, and Narratives – Volume 8002*, pages 118–137. Springer-Verlag New York, Inc., 2014. `doi:10.1007/978-3-642-45324-3_8`.

22 Amélie Zöllner-Weber. *Noctua literaria : a computer-aided approach for the formal description of literary characters using an ontology.* PhD thesis, Bielefeld University, 2008.

23 Amélie Zöllner-Weber and Andreas Witt. Ontology for a formal description of literary characters. In *Proceedings of Digital Humanities 2006*, 2006.

Annotating Musical Theatre Plots on Narrative Structure and Emotional Content*†

Pablo Gervás[1], Raquel Hervás[2], Carlos León[3], and Catherine V. Gale[4]

1 **Universidad Complutense de Madrid, Spain**
`pgervas@ucm.es`
2 **Universidad Complutense de Madrid, Spain**
`raquelhb@ucm.es`
3 **Universidad Complutense de Madrid, Spain**
`cleon@ucm.es`
4 **Wingspan Productions, London, UK**
`cat@wingspanproductions.co.uk`

Abstract

Although theoretical models of the structure of narrative arising from systematic analysis of corpora are available for domains such as Russian folk tales, there are no such sources for the plot lines of musical theatre. The present paper reports an effort of knowledge elicitation for features that characterise the narrative structure of plot in the particular domain of musical theatre. The following aspects are covered: identification of a valid vocabulary of abstract units to use in annotating musical theatre plots, development of a procedure for annotation–including a spread-sheet format for annotators to use, and a corresponding set of instructions to guide them through the process – selection of a corpus of musical theatre pieces that would constitute the corpus to be annotated, the annotation process itself and the results of post-processing the annotated corpus in search for insights on the narrative structure of musical theatre plots.

1998 ACM Subject Classification D.2.1 Requirements/Specifications Elicitation methods, I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Narrative annotation, conceptual representation of narrative, character functions, narrative schemas, musical theatre

Digital Object Identifier 10.4230/OASIcs.CMN.2016.11

1 Introduction

Storytelling systems require the representation and manipulation of large amounts of knowledge. This involves not only the product itself – stories represented at various levels of detail – but also the knowledge resources that are required to inform the construction processes. There are a number of possible sources for extracting such knowledge from existing literature on narrative. However, there is little to be found if plot generation for musical theatre is required.

* This work was partially supported by the project WHIM (611560), funded by the European Commission under FP7, the ICT theme, and the Future Emerging Technologies (FET) program.
† The authors want to acknowledge the fundamental contribution made to this annotation initiative by the following people: Julian Woolford (GSA, University of Surrey), Kat Mace (Researcher, Wingspan Productions), James Sandy (Assistant Producer, Wingspan Productions), Archie Baron (Creative Director, Wingspan Productions). For more information on the associated television and theatrical production see: `http://www.wingspanproductions.co.uk/`.

© Pablo Gervás, Raquel Hervás, Carlos León, and Catherine V. Gale;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 11; pp. 11:1–11:16
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The present paper reports an effort of knowledge elicitation for features that characterise the narrative structure of plot in the particular domain of musical theatre.The need for such knowledge arose when researchers at Universidad Complutense de Madrid (UCM) were asked by Wingspan Productions to explore the possibility of producing a software tool that could automatically generate plot lines for musical theatre. This took place as part of a broader effort to explore the extent to which it would be possible to generate the premise, setting, structure and music/lyric content for a new piece of musical theatre, using computational methods. The scientific and theatrical production processes invovled in such an effort were to be documented for Sky Arts (UK) in 2x 1hr documentaries. The researchers at UCM had been working on tools for narrative generation within the domain of Russian folk tales, based on the Vladimir Propp's "Morphology of the Folk Tale" [16]. However, the background research carried out to inform those tools had also identified the limited applicability of that particular formalism beyond the domain of Russian folk tales, which presents significant differences with the domain of musical theatre. For instance, there is no mention of love affairs between characters in Russian folk tales, whereas this is a very important feature in musical theatre. As a result, if the available story generation functionalities were to be applied to the domain of musical theatre, it became imperative to undertake an analysis of the regularities of musical theatre plots much in the line of the work Propp had done for Russian folk tales. Although the present effort cannot be compared to Propp's original work in terms of thoroughness and academic rigour, it constitutes a valuable empirical source for any further analyses that may be attempted.

2 Previous Work

There are multiple dimensions when considering knowledge representation for story generation. Gervás and León [10] provided a list of the most relevant classifications, and proposed their own list of suitable dimensions obtained from the different aspects of a narrative. The dimensions proposed were: discourse, simulation, causality, character intention, theme, emotion, authorial intention, and narrative structure.

The annotation effort carried out to build the case base described in the present paper concerned only two of these aspects: the narrative structure aspect, and the emotional aspect.

2.1 Existing Schemas for Modelling Narrative

Narratives are created and told according to a certain structure (or set of structures). The aristotelian narrative arc, for instance, is one of the first instances of this kind of structure. Vladimir Propp's analysis of Russian folk tales [16] is known to have produced a semi-formal description of the structure of these tales that has acted as inspiration for several story generation systems, both sequential and interactive. Its exhaustive description of the constituent elements of tales of this kind, and the enumeration of the patterns they follow provided a very useful starting point for researchers looking for computational implementations of story generators.

Propp identified a set of regularities across a corpus of Russian folk tales in terms of *character functions*, understood as acts of the character, defined from the point of view of their significance for the course of the action. These character functions have been used as building blocks for story generators in many instances [7, 8].

Gervás et al. [11] describe an effort to mine existing literature on narrative to identify descriptions of the structure of narrative that might serve as instances of the schemata that might be used as skeletons over which to build more elaborate instantiations of narrative during narrative generation processes. This approach led to the identification of a particular way of abstracting units of description that carry meaning with respect to the overall plot (following the work of Vladimir Propp on character functions [16]) and a number of descriptions of plot lines (based on Booker [1], Tobias [18], and Polti [15]).

2.2 Existing Approaches to the Elicitation of Knowledge Pertinent to Narrative Modelling

Storytelling systems are known to require large amounts of explicit knowledge to operate successfully. Generated stories are only as good as the knowledge they have been derived from. Recent attempts have been made to address this problem via crowdsourcing [13]. In this work, a number of human authored narratives are mined to construct a *plot graph*, which models the author-intended logical flow of events in the virtual world as a set of precedence constraints between plot events [19]. When required to develop stories on a given topic that it has no knowledge of, the system is designed to send a query to Amazon Mechanical Turk (AMT) soliciting typical narratives in natural language on that topic. Such narratives are parsed and merged into a combined representation in terms of plot graph for the domain being explored, which is later used to inform the process of constructing narratives.

To inform the development of the Dramatis system for modelling suspense [14], O'Neill carried out an effort of knowledge engineering driven by methods adapted from qualitative research. The goal was to collect typical reader genre knowledge while simultaneously limiting engineer bias. The process was to acquire a corpus of natural language text and the conversion of that corpus into the knowledge structures required by Dramatis.

The Story WorkBench [6] is a free, open-source, cross-platform framework for text annotation that provides a number of common text annotation operations, including representations (e.g., tokens, sentences, parts of speech), functions (e.g., generation of initial annotations by algorithm, checking annotation validity by rule, fully manual manipulation of annotations) and tools (e.g., distributing texts to annotators via version control, merging doubly-annotated texts into a single file). It has been used extensively to build resources that capture particular aspects of narrative, such as narrative structure [5] or use of referring expressions [12], or that focused on particular topics such as Islamic terrorism [4].

3 Development of an Annotated Case-Base of Musical Theatre Plots

Very little analysis of the structure of musical theatre in terms of narrative has been carried out in the past. Talor and Simmonds [17] provides a brief description of the narrative structure of musicals in relation to the classic view of Aristotle and Vladimir Propp [16] and mentioning concepts from Booker's work [1]. However the description is sketchy and a very limited number of musicals is analysed in a manner more didactic than systematic. Woolford [20] provides a more methodical approach, based on Joseph Campbell [2], and includes detailed analysis of a small number of musicals. In both cases the size of the analised samples and the level of detail where insufficient to support the type of computational learning effort envisaged for the current project. In view of this – after initial consultation with the UCM researchers – the Wingspan Productions team organised a mass narrative annotation of musical theatre plots, the intention being that the data resulting from this could be used by the UCM researchers to further explore musical theatre narratives. Successful annotation of 42 musicals provided material for the UCM researchers to develop their narrative generation system, specifically with musicals in mind.

The research effort carried out on this initiative involved several tasks:

1. identification of a valid vocabulary of abstract units to use in annotating musical theatre plots
2. development of a spreadsheet format for annotators to use, and a corresponding set of instructions to guide them through the process
3. selection of a corpus of musical theatre pieces that would constitute the corpus to be annotated
4. carrying out the annotation process itself
5. post-processing the results of the annotation

3.1 Annotation Vocabulary

The work carried out by the UCM researchers on narrative generation [7, 8] had identified a number of valuable abstractions for the representation of narrative that were then employed in the case base of schemas described in [11] and the case base reasoning procedure described in [9]. These abstractions were based on Propp's concept of a character function as an abstract contribution by a particular character to the overall plot. However, the background research carried out to inform those processes had also identified the need for a wider range of actual character functions to consider. The set of character functions described by Propp was originally designed to capture the abstract descriptions of character's contributions to plot in the context of Russian folk tales. This is a limited domain and the set of abstractions that resulted is constrained to the basic actions that are considered valid for that genre. This presented significant differences with the domain of musical theatre. For instance, there is no mention of love affairs between characters in Russian folk tales, whereas this is a very important feature in musical theatre.

To account for these differences, a new specific set of abstractions in the spirit of character functions was defined. This set was constructed from a number of sources:

- the original set of Propp's character functions [16] was mined for abstractions general enough to be applicable across different domains
- instances of Propp's character functions that were specific to the domain of Russian folk tales were generalised to produce a set of options (for instance, Propp only considered the alternative of the hero succeeding in his struggles with the villain, whereas a character function for defeat of the hero was considered necessary to achieve the required generality)
- additional abstractions that had been identified during the exploration of the set of schemas reported in [11] were introduced as additional terms for the vocabulary (including functions like **Repentance** and **Repentance Rewarded** and a large set of abstractions to describe plot-relevant contributions related with love between characters)
- a number of abstractions specific to the description of the plot of musical theatre, such as abstractions for scenes where the plot so far is summarised (**Summary**), or where characters describe themselves and/or their point of view (**I Am What I Am**)

A more detailed description of how this process of identifying abstractions was carried out may be found in [11].

An initial set of abstractions was proposed by the UCM researchers. This initial set was reviewed, revised and extended by a group of experts in musical theatre convened by Wingspan Productions. The revision and extension was carried out keeping in mind that the set was to be used to annotate plot relevant events in pieces of musical theatre. The final set of abstractions is described in Appendix A.

For ease of reference, these labels were grouped into subsets that fell under one of 6 generic labels. This classification is described in Appendix B.

Additionally, it was decided that the term "character function" would be difficult for untrained annotators to relate to, and the term "plot element" was chosen instead to describe these labelled abstractions of character's contributions to plot.

3.2 Annotation Tool and Annotation Instructions

The set of abstractions (from now on referred to as *plot elements*) so obtained was used to build an annotational tool in the form of a spreadsheet that annotators could use to produce a standard register of an annotated musical.

This involved making available the set of plot elements paired with brief descriptions of their meaning. It also allowed for inclusion of the synopsis of the plot of the musical that was to be annotated, so that annotations could be made with respect to a canonical reference (specific abstractions are to be assigned to spans of text in the synopsis).

The tool provided a specific spreadsheet that allowed for the following operations:

- identifying the span of the synopsis covered by a particular plot element (by indicating the start and end of the text span in question)
- selecting which label out of the vocabulary to assign to that plot element (this was provided as two drop down lists, one to select the subset of the labels and to select which label out of that subset to assign)
- annotating whether that plot element corresponded to a musical number, and, if so, what was its title
- declaring an additional label for plot elements not covered by the vocabulary provided
- the possibility of indicating, for each member of the cast, which emotions that character experienced in the context of the annotated plot element, and whether the character sang or danced
- indicating which emotions where prevalent in the context of the annotated plot element for the set of characters involved overall

The set of emotions that were chosen to annotate where: Love, Hate, Danger, Violence, Happiness, Sadness, Humour, Surprise, and Fear.

Additionally, a plot element called OTHER was introduced for annotators to use whenever they reckoned that a new label was needed, not covered by the ones in the list, but fundamental for annotating a particular plot.

A screenshot of the basic spreadsheet for annotation is shown in Figure 1.

3.2.1 Selection of the Corpus to Annotate

The corpus to annotate was compiled by the team at Wingspan Productions based on a list produced by an earlier analysis carried out at the University of Cambridge (categorising musicals by length of run and numbers of awards won).

The musicals listed in Table 1 were annotated.

3.3 Annotation Process

The annotation process was carried out at the University of Surrey, supervised by Julian Woolford, Programme Leader for the MA in Musical Theatre and the team from Wingspan Productions.

The 35 annotators were predominantly Musical Theatre MA (Masters) students, ranging in age from 20–30; male/female 40/60%. All participants will have completed secondary

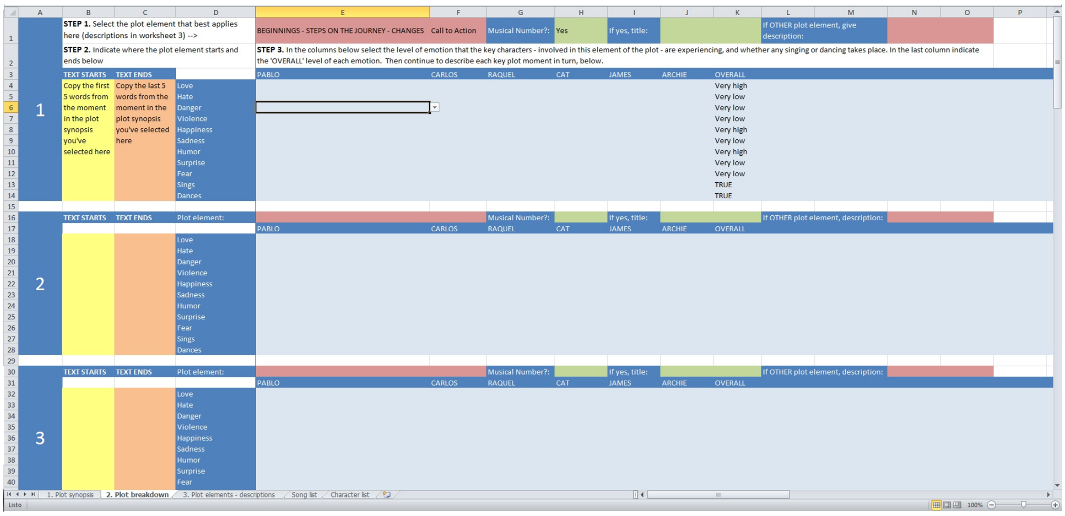


Figure 1 Screenshot of the annotation spreadsheet.

education and the majority also a first degree in a related subject. Two members of academic staff from University of Surrey also participated.

Annotators were offered a long list of musicals to choose from, prior to the session, to ensure they were familiar with the musical they annotated. If they enjoyed the first session they were then invited to choose a second musical to annotate (that they were familiar with) which was completed at home.

Annotators were provided with the instruction sheet prepared initially by Raquel Hervás, modified slightly for the group annotation scenario. They were emailed this the day before, to read as preparation, and were then provided with a hard copy of the instructions on the day. All participants were briefed verbally at the start of the session, and three people were on hand to answer questions and provide support as necessary.

An additional small set of annotations were completed by members of the Wingspan Productions team, completed during development of the methodology.

4 Discussion of Results

Results are discussed from two different points of view: observations on how the annotation process took place, and observations arising from the postprocessing of the annotated data.

4.1 Discussion of the Annotation Process

The majority of people took around 90 minutes to complete the annotation of a single musical (ranging from ca. 45 mins to 2 hours). Those that completed the process quickly did not turn in incomplete annotations/use fewer plot elements than others – they simply seemed to have genuine aptitude for the task, and were very comfortable thinking about the shows they annotated in the terms the process required.

From the annotations session, it was clear that familiarity with the musical being annotated is an enormous help. Having a keen knowledge of the structure of a musical (particularly how it is performed, in addition to how it appears on the page) makes it much more straightforward to break it down into coherent units, expressed as plot elements.

Table 1 Musicals annotated.

A Little Night Music	Oliver
Amour	Miss Saigon
Annie	Oklahoma!
Avenue Q	Parade
Barnum	Passion
Beauty and the Beast	Rent
Billy Elliot	Pippin
Brass	South Pacific
Cabaret	Spring Awakening
Carousel	State Fair
Chicago	Sunday In The Park With George
The Fields of Ambrosia	Sweeney Todd
Grease	The Book of Mormon
Gypsy	The Lion King
Jekyll and Hyde	The Pirate Queen
Jesus Christ Superstar	The Rocky Horror Show
Joseph and the Amazing Technicolor Dreamcoat	The Sound Of Music
La Cage Aux Folles	Thoroughly Modern Millie
Mary Poppins	Urinetown
Matilda	West Side Story
Memphis	Wicked

The process most people adopted was to read the synopsis first (as advised), then proceed with the annotation alternating between the main spreadsheet and the list of plot element definitions. Most annotators found the given instructions to be clear, with only a small number requiring significant help. Help tended to be required where people were unfamiliar with using Microsoft Excel.

The chosen platform used in the annotation raised some minor technical problems. For instance, there were some problems with compatibility of the spreadsheet across different operating systems, although these could be resolved on the day. Having predicted that all user would have Microsoft Office installed in their devices turned out to be false, and several ad-hoc solutions had to be taken in order to have every subject complete their annotation(s). With this in mind, an online form with cross-browser compatibility could be much more straightforward both for annotation and data collection.

Annotators frequently missed fields – a check was done as each person finished, but some gaps still went unnoticed. From questions asked by the annotators it is apparent that there was some doubt as to whether the emotion to be annotated was the emotion attributed to the character or the emotion felt by the spectator. Annotators needed frequent reminding to add in information about whether a song happened during particular plot elements.

An interesting observation made by one of the volunteers helping with the annotation process concerns the fact that different people seemed to interpret the plot of a musical in different ways, so that what is three plot elements to one person, is a single plot element to another.

4.2 Discussion of the Annotated Data

The volume of data obtained is substantially large, and only preliminary analyses of it have been carried out. Nevertheless, these reveal some interesting conclusions. Exhaustive listing of all the data would exceed the available space in this kind of paper so an attempt has been made to summarise features that have been considered significant.

Regarding the set of plot elements included in the vocabulary, there was generic concern among the researchers that developed the vocabulary that either some elements may not be employed at all or that some elements crucial for capturing the essence of the plots in question might be missing.

Of the whole set of 116 plot elements in the vocabulary only 6 were never used by evaluators in the annotation of the 42 plots that were annotated. The plot elements in question were: **Rescue from pursuit**, **Erroneous Judgement**, **Involuntary Crimes of Love**, **The Enigma**, **Unfounded Claims**, **Riches**. Of these, **Rescue from pursuit** – originating from Propp's original set of character functions and retained because it more explicitly linked up with the **Pursuit** plot element – may have been displaced by the more intuitive **Rescue** – which was introduced as a generalization and is actually used twice in the annotated plots. The **Involuntary Crimes of Love** plot element originates in Polti's dramatic situations, and has significant overlap with more explicit plot elements such as **Adultery** or **Murderous Adultery**. Plot element **The Enigma** arises from one of Tobias' possible plots, and it is very abstract. Its full description concerns plots that have an underlying mystery that must be solved in parallel by the protagonist and the reader – as in thrillers, mystery novels, and whodunnits. It probably had no role to play as part of a vocabulary for annotating musicals, and it was in any case poorly described in the material provided to annotators. The **Unfounded Claims** plot element is also inherited from Propp's character functions, but may also have been poorly described in the material, so that annotators would be unlikely to choose it.

The remaining 110 plot elements provided were used to label 992 segments of musical theatre plots, giving an average of 9 appearances per plot element. However, observed use differed significantly from this average.

Table 2 shows the most frequently used plot elements in each category, with their associated frequency of use. The most frequently used plot element was **Decision to Take Action** which appears 50 times in the annotated musicals. Five of the plot elements appear more than 20 times: **Aspiration**, **Boy Meets Girl**, **Tested**, **Character's Reaction** and **Bond Strengthened**. In the range of between 10 to 20 appearances we find 22 plot elements. Of these, 8 relate to love (**Class differences**, **Love shift**, **Love Triangle**, **Obstacles to Love**, **Couple Wants to Marry**, **Forbidden Love**, **Reconciliation** and **One-sided Love**); 2 to initiating events (**Call to Action** and **Departure**); 5 to conflicts that require resolution (**Difficult task**, **Imprisoned**, **Struggle**, **Loss of Loved Ones**, **Revenge**); 5 to possible developments from earlier situations (**Reward**, **Solution**, **Exposure**, **Discovery**, **Epiphany**); and 2 cases of help provided to a character (**Guidance** and **Assistance**). Overall, these details constitute an interesting sample of the ingredients of musical theatre as a genre.

There were three plot elements in the vocabulary provided that were deemed important to make it possible to annotate musical plots even though they had not been suggested by any of the sources studied on plot. They refer to either references to the plot itself or references to the annotation vocabulary. Two plot elements were intended to help mark parts of the plot designed to help the audience either by introducing an initial situation (**Initial situation** which was used 45 times in the annotation) or by summarising prior parts of the

Table 2 Extract of the most frequently used plot elements in each category, with their associated frequency of use.

BEGINNINGS – STEPS ON THE JOURNEY – CHANGES	
Aspiration	20
Epiphany	18
Departure	15
Call to Action	13
Discovery	12
Guidance	11
BIG DECISIONS - CHOICES	
Decision to Take Action	50
Character's Reaction	23
TESTS, TRIAL AND ORDEALS	
Tested	21
Loss of Loved Ones	17
Difficult task	14
Imprisoned	14
Exposure	11

TRIALS AND TRIBULATIONS OF LOVE – FRIENDSHIP	
Boy Meets Girl	20
One-sided Love	15
Forbidden Love	13
Couple Wants to Marry	12
Obstacles to Love	11
Class differences	10
Love shift	10
Love Triangle	10
CONFLICT	
Bond Strengthened	30
Assistance	15
Reconciliation	13
Solution	10
Reward	10

plot at a given stage (**Summary** which was used 34 times in the annotation). The results support the impression that these plot elements are useful in describing musical theatre plots.

The OTHER plot element was introduced to cover the possibility that some elements crucial for capturing the essence of the plots in question might be missing. Annotators were advised to resort to this label when none of the plot elements provided covered the essence of the fragment they were trying to annotate. The OTHER label was used 19 times. A large variation in usage is observed:

- A small number of annotators used the OTHER field where they felt they could not find a fitting plot element. It is unclear if these choices genuinely do represent plot elements missing from the list provided, or whether the OTHER field was used for ease
- Some people used the OTHER field in addition to choosing a primary plot element, presumably because two plot elements together enabled them to express finer-grained detail about what was happening at that point in the story
- On occasion the OTHER field was used to provide additional non plot-related information (e.g. a diegetic song)

On a related note, at least in one occasion the generic label for a subgroup of plot elements was used to annotate instead of a particular member of the group. Again, it is unclear whether this is due to inadequacy of the lower level plot elements or for simplicity.

Working from insights obtained in the development of story generators based on Propp's character functions [7, 8], it was expected that some of the plot elements introduced in the vocabulary would show some evidence of cross-dependency. This would affect cases like **Aspiration** and **Aspiration Achieved** or **Misunderstanding Arises** and **Misunderstanding Cleared**. While a detailed analysis of this type of relation in individual annotated plots is still pending, a rough analysis of the statistical data collected shows that this expectation is likely to be unfulfilled. Relative frequency for a number of such pairings between plot elements are shown in Table 3.

Along similar lines, Table 4 compiles statistics on the relative frequency between pairs of plot elements that represent different alternatives for resolving a given initial situation. It might seem from these data that, overall, characters in the annotated musicals seem to

Table 3 Relative frequency of appearance of hypothetically related pairs of plot elements.

Misunderstanding Arises	4
Misunderstanding Cleared	1
Aspiration	20
Aspiration Achieved	7
Couple Wants to Marry	12
Wedding	2
Forbidding/Warning	3
Warning/Forbidding Disregarded	2
Lack	7
Lack Fulfilled	2

Table 4 Relative frequency of appearance of related plot elements to correspond to alternative solutions.

Succumbing to Temptation	9
Temptation Resisted	1
Moral Dilemma Triumph	6
Moral Dilemma Failure	4
Boy Meets Girl	20
Boy Loses Girl	7
Couple Wants to Marry	12
Wedding	2

succumb to temptation more often than resist it and triumph more often than fail when faced with a moral dilemma. Also that boys that meet girls do not often lose them, and that couples that want to marry do not always succeed. However, more detailed analysis of the individual annotated plots is required. First, to ensure that these pairings are actually present simultaneously in individual plots. Second, to double check that the expected resolution is not actually implied by some other plot element that may describe the same solution in different terms.

With respect to the annotations on emotions that were collected, it has only been possible to carry out a quick processing of the application of emotional labels to the plot elements as a whole. Labels on emotional content were also collected for each individual character involved in a plot element, but these data have not been processed yet.

The assignment of emotional labels to plot elements as used in the annotation in the context of given musical plots shows little indication that particular emotional labels may be consistently applied to specific plot elements. The annotations collected have been processed with the ultimate goal of identifying cases where similar values for the emotional features considered have been consistently applied to specific plot elements across all their appearances in the set of annotated plots. To achieve this, the standard deviation over the values assigned to the plot element for each emotional feature has been computed. Values of this standard deviation below 0.5 have been taken to be probable indication of consistency across appearances. This occurs only infrequently. Where values are assigned consistently, only cases where either low or high values of some of the emotional features are attributed to a plot element are considered. The resulting data are compiled in Table 5.

Overall, the data on emotion assignment to plot elements shows few surprises. The **Deception to Fit In** plot element shows low values of both happiness and sadness, and

Table 5 Plot elements consistently assigned significant emotional values across the annotated set of plots.

Emotional feature	Plot elements
Love+	Couple Wants to Marry Maturation Pursuit Recovery of a Lost One Rescue Wedding
Love-	Abduction Madness Warning/Forbidding Disregarded
Hate+	Cross-Rank Rivalry
Hate-	Cross-Dressing Deception to Fit In Deliverance Return Reward Victory Wedding
Danger+	Abduction Misfortune
Danger-	Adultery Boy Loses Girl Moral Dilemma Failure Rescue Victory Wedding
Sadness+	Abduction
Sadness-	Couple Wants to Marry Deception to Fit In Rescue Wedding
Humour-	Abduction Moral Dilemma Failure Shame of Loved One Warning/Forbidding Disregarded

Emotional feature	Plot elements
Violence+	Abduction
Violence-	Aspiration Achieved Couple Wants to Marry Disguise High Status Revealed Love shift Maturation Obstacles to Love High Status Revealed Pursuit Recovery of a Lost One Reward Victory Wedding
Fear+	Shame of Loved One
Fear-	Adultery Ambition Aspiration Achieved Deception Moral Dilemma Failure Reward Aspiration Achieved
Happiness+	Aspiration Achieved Couple Wants to Marry Lack Fulfilled Pursuit Wedding
Happiness-	Cross-Rank Rivalry Deception to Fit In Lack Shame of Loved One

the **Pursuit** plot element shows high values of happiness. More interesting is the relation between some emotions for certain plot elements. For example, the set of plot elements **Reward**, **Victory**, and **Wedding** presents low values of hate and violence and, in the case of **Victory** and **Wedding**, low values of danger. Other dependencies appear for **Couple Wants to Marry** and **Wedding**, both having low values of sadness and high values of happiness and love, and **Abduction** showing high values of danger and violence. More studies will be performed automatically in order to find more dependencies like these ones.

The low level of consistency in attribution of emotional features to specific plot elements may be the result of having insufficient data. Larger volumes of data would have to be collected to obtain significant results. The current annotation suffers from two disadvantages: each musical has been annotated by a different person, so a single subjective view is available for each musical, and there are insufficient occurrences of most of the plot elements to provide a significant set of values.

5 Conclusions

The annotation reported in this paper was carried out to inform the development of a generator of musical theatre plots. The ultimate aim of the annotation was to provide a reference corpus to test the validity of a Proppian approach to the generation of musical theatre plots, supported by a limited process of adaptation to the new domain. The generator for musical plots had to be ready to be used in the composition of a musical that had to open in the London West End on February 2016. It therefore had tight deadlines to meet. Although every effort was made to apply academic rigour wherever possible, constraints, both on time and resources, made it imperative to reduce both the number of musicals that could be annotated and the amount of annotations that could be collected for them. Nevertheless, the corpus was ready on time and served to inform the development of the PropperWryter software – a generator of plot lines for musical theatre pieces – which was used for the composition of "Beyond the Fence", the first ever experimental computer-generated musical [3]. "Beyond the Fence" opened successfully at the Arts Theatre in London on 22nd February 2016 and has by now enjoyed a successful two-week run.

The present paper reports on the knowledge resources that were developed to inform that process. The team that worked on the initial annotation is currently applying for funding to carry out a more refined annotation, over a larger corpus, with a view to improving the results obtained so far.

References

1 C. Booker. *The Seven Basic Plots: Why We Tell Stories*. The Seven Basic Plots: Why We Tell Stories. Continuum, 2004.

2 J. Campbell, P. Cousineau, and S. L. Brown. *The Hero's Journey: Joseph Campbell on His Life and Work*. Collected works of Joseph Campbell. New World Library, 1990.

3 Simon Colton, M. T. Llano, R. Hepworth, J. Charnley, C. V. Gale, A. Baron, F. Pachet, P. Roy, P. Gervas, N. Collins, B. Sturm, T. Weyde, D. Wolff, and J. Lloyd. The beyond the fence musical and computer says show documentary. In *Proceedings of the Seventh International Conference on Computational Creativity, ICCC 2016, Paris, 27 June – 1 July 2016.*, 2016.

4 M. A. Finlayson, J. R. Halverson, and S. R. Corman. The N2 corpus: A semantically annotated collection of islamist extremist stories. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, pages 896–902, 2014. URL: `http://www.lrec-conf.org/proceedings/lrec2014/summaries/48.html`.

5 M.A. Finlayson. *Learning Narrative Structure from Annotated Folktales*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2011.

6 M.A. Finlayson. The story workbench: An extensible semi-automatic text annotation tool. In *Intelligent Narrative Technologies IV, Papers from the 2011 AIIDE Workshop, Stanford, California, USA, October 10-11, 2011*, 2011. URL: `http://aaai.org/ocs/index.php/AIIDE/AIIDE11WS/paper/view/4091`.

7 P. Gervás. Reviewing Propp's Story Generation Procedure in the Light of Computational Creativity. In *AISB Symposium on Computational Creativity, AISB-2014, April 1-4 2014*, Goldsmiths, London, UK, 2014.

8 P. Gervás. Computational Drafting of Plot Structures for Russian Folk Tales. *Cognitive Computation*, pages 1–17, 2015. `doi:10.1007/s12559-015-9338-8`.

9 P. Gervás, R. Hervás, and C. León. Generating Plots for a Given Query Using a Case-Base of Narrative Schemas. In *Proceedings of the Experience and Creativity Workshop, ICCBR*, 2015.

10 P. Gervás and C. León. The need for multi-aspectual representation of narratives in modelling their creative process. In *5th Workshop on Computational Models of Narrative*, OASIcs-OpenAccess Series in Informatics, 2014.

11 P. Gervás, C. León, and G. Méndez. Schemas for Narrative Generation Mined from Existing Descriptions of Plot. In M.A. Finlayson, B. Miller, A. Lieto, and R. Ronfard, editors, *6th Workshop on Computational Models of Narrative (CMN 2015)*, pages 54–71, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. `doi:10.4230/OASIcs.CMN.2015.54`.

12 R. Hervás and M. Finlayson. The prevalence of descriptive referring expressions in news and narrative. In *48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, Uppsala, Sweden, 2010.

13 Boyang Li, S. Lee-Urban, G. Johnston, and M. O. Riedl. Story generation with crowd-sourced plot graphs. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, AAAI'13, 2013.

14 B. O'Neill. *A Computational Model of Suspense for the Augmentation of Intelligent Story Generation*. PhD thesis, Georgia Institute of Technology, Atlanta, Georgia, 2013.

15 G. Polti and L. Ray. *The Thirty-six Dramatic Situations*. Editor Company, 1916.

16 V. Propp. *Morphology of the Folktale*. University of Texas Press, 1968.

17 M. Taylor and D. Symonds. *Studying Musical Theatre: Theory and Practice*. Palgrave Macmillan, 2014.

18 R. B. Tobias. *20 Master Plots: And How to Build Them*. F+W Media, 2012.

19 P. Weyhrauch. *Guiding interactive drama*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 1997.

20 J. Woolford. *How Musicals Work: And how to Write Your Own*. A Nick Hern book. Nick Hern Books, 2012.

A Appendix A

This appendix presents the full list of abstractions proposed as plot elements, with their accompanying descriptions as presented to the annotators.

Abduction a character is abducted
Adultery two characters commit adultery, deceiving a partner
Ambition character tries to obtain/achieve the possession or status of another
An Enemy Loved one of two lovers is hated by a friend(s) of the other
Aspiration Achieved an aspiration held earlier the story is achieved
Aspiration a member of a group wants or aspires to have, or be, something better
Assistance a character assists another to achieve their goals
Bond Strengthened the bond between two characters is strengthened
Boy Loses Girl two characters that were emotionally attached drift apart
Boy Meets Girl two characters meet and become emotionally attached
Branding a character acquires a mark or a token by which they will later be recognised
Call to Action faced with a misfortune/lack/aspiration, a character is exhorted to take action
Character Flaw a character's misfortune is caused by their own character flaws
Character's Reaction a character reacts to the actions of another character testing them
Class differences relationship impossible due to class differences between partners

Complicity a character colaborates with an enemy, deceiving their friends
Conflict with a God a mortal struggles with a deity
Couple Wants to Marry two characters want to marry
Crimes of Love a lover and beloved incur in questionable acts
Cross-Dressing a character dresses up as one of the opposite sex
Cross-Rank Rivalry two masculine or feminine rivals with different rank
Daring Enterprise character attempts to recover an object or person from an adversary
Deception to Fit In character pretends to be something they're not, in order to fit in
Deception character is deceived, unwittingly helping an enemy
Decision to Take Action a character agrees to or decides to take action
Defeat focal character is defeated by another
Deliverance protector comes to the rescue of the distressed
Departure character leaves the place where they were, the story follows them
Difficult task a difficult task is proposed to a character
Disaster someone is defeated by an enemy or catastrophe
Disconnected from Reality a character becomes disconnected from the reality of their surroundings (either due to overwhelming love, grief, dreaming, flashbacks, etc.)
Discovery character discovers/realises/learns something about themselves
Disguise a character puts on a disguise
Epiphany a character has a significant moment of realization
Erroneous Judgement any kind of mistaken judgement
Escape character that was imprisoned or abducted escapes by their own means
Exposure a character operating under deception is exposed
Forbidden Love unconventional love relation
Forbidding/Warning a character is forbidden from or warned against a specific course of action
Guidance a character is guided towards a new possession or goal
Hatred between Friends characters that were once friends become enemies
High Status Revealed high status of one character is revealed, opening new possibilities
I am what I am character defiant about who they are/their identity
Ill-fated Imprudence recklessness or curiosity that results in loss
Imprisoned a character is confined against their will
Inconstancy one partner is inconstant in their love of the other
Initial situation description of initial situation
Involuntary Crimes of Love character unknowingly commits adultery or incest
Jealousy a character is justifiably jealous of another
Judgement Deferred to Authority conflict between characters to be solved by a decision of someone in authority
Lack Fulfilled an important lack suffered by a character earlier in the story is fulfilled
Lack a member of a group lacks something important in their life
Lesson Learned a character realizes they were wrong because of an event
Loss of Loved Ones a character experiences the loss of a loved one
Love Triangle a third individual is involved in a love relationship
Love shift person that loved one character shifts their love to another character
Madness a madman slays, injures or brings disgrace onto a character
Maturation immature character evolves to maturity, usually as a result of a challenging incident
Metamorphosis character transformed into a beast (literal/metaphorical)

Misfortune character suffers a cruel master or misfortune
Mistaken Jealousy a character is mistakenly jealous of another
Mistaken murder character unwittingly kills a friend
Misunderstanding Arises a misunderstanding arises that constitutes an obstacle to the intentions of some characters
Misunderstanding Cleared a misunderstanding that constituted an obstacle is cleared
Moral Dilemma Failure character faces a moral dilemma and makes the wrong decision
Moral Dilemma Triumph character faces a moral dilemma and makes the right decision
Murderous Adultery a betrayed husband or wife kills one or both adulterers
Obstacles to Love relationship prevented by social norms
One-sided Love character loves another but is not loved back
Parent Convinced parent or guardian that opposed a relationship is now convinced
Persuasion character has to obtain something through eloquence and diplomacy
Poverty character suffers extreme poverty
Provision of a Magical Agent a character acquires the use of a magical agent
Punishment a character is punished
Pursuit a character is pursued
Recognition a character that was in disguise is recognized
Reconciliation two characters that had been at odds get back together
Reconnaissance a character makes an attempt at reconnaissance
Recovery of a Lost One a character recovers a loved one the had lost
Remorse a culprit suffers remorse for a crime or love fault
Repentance Rewarded a character that has repented is rewarded
Repentance a character that has behaved badly repents
Rescue one character rescues another character imprisoned against their will
Rescue from pursuit a character that was being pursued manages to get away
Return a characters returns to to a place where they had been earlier in the story
Revenge character enacts revenge
Revolt character conspires to revolt against a tyrant
Reward character receives a reward
Riches character achieves material fortune
Rivalry a character and an antagonist of balanced power clash
Sacrifice for Family character makes sacrifices for happiness of a relative
Sacrifice for Passion character makes sacrifices for a vice or passion
Sacrifice for an Ideal character sacrifices life, love or well-being to a cause
Sacrifice of Loved Ones character sacrifices a loved one for a necessity or vow
Shame of Loved One a character discovers the shame of a loved one
Solution a task is resolved
Someone Leaves one of the members of a group goes away, the story follows those who remain
Struggle two characters join in direct combat or open confrontation
Succumbing to Temptation a character succumbs to temptation
Summary scene that summarises a particular moment in the story, though nothing much actually happens
Temptation Resisted a character resists temptation
Tested a character is tested, interrogated, attacked, asked to perform a service...
The Enigma a combat of the intelligence to find a person or object
Transfiguration a character is given a new appearance

Transformation because of a crisis, a character's behaviour or sense of self is transformed
Trickery a character attempts to deceive another to get something they want
Underdog a character and an antagonist of balanced power clash, and the character is at disadvantage
Unfounded Claims a character presents unfounded claims
Unrecognized Arrival unrecognized, a character arrives in a new place
Unrelenting Guardian parent or guardian of one partner opposes a desired relationship
Useful Information character receives information relevant to their intentions
Victory focal character achieves victory over another
Villainy a character causes harm or injury to another character
Warning/Forbidding Disregarded a character disregards a warning or a prohibition received earlier in the story
Wedding a character is married

B Appendix B

This appendix presents the way in which the plot elements were grouped into subsets that fell under one of 6 generic labels:

BEGINNINGS – STEPS ON THE JOURNEY – CHANGES Initial situation, Summary, Aspiration, Call to Action, Cross-Dressing, Departure, Deliverance, Disconnected from Reality, Discovery, Disguise, Epiphany, Escape, Guidance, High Status Revealed, Maturation, Metamorphosis, Pursuit, Reconnaissance, Rescue, Return, Someone Leaves, Transfiguration, Transformation, Unrecognized Arrival

BIG DECISIONS – CHOICES Character's Reaction, Decision to Take Action, Deception to Fit In, Erroneous Judgement, Ill-fated Imprudence, Moral Dilemma Triumph, Moral Dilemma Failure, Mistaken Jealousy, Sacrifice for an Ideal, Sacrifice for Family, Sacrifice for Passion, Sacrifice of Loved Ones, Succumbing to Temptation, Temptation Resisted, Warning/Forbidding Disregarded, Character Flaw

TRIALS AND TRIBULATIONS OF LOVE – FRIENDSHIP Boy Meets Girl, Boy Loses Girl, Wedding, Class differences, Forbidden Love, Inconstancy, Involuntary Crimes of Love, Adultery, An Enemy Loved, Crimes of Love, Love shift, Love Triangle, Murderous Adultery, One-sided Love, Obstacles to Love, Parent Convinced, Recovery of a Lost One, Couple Wants to Marry

TESTS, TRIAL AND ORDEALS Abduction, Branding, Deception, Difficult task, Disaster, Shame of Loved One, Exposure, Forbidding/Warning, Mistaken murder, Lack, Loss of Loved Ones, Madness, Misfortune, Persuasion, Poverty, Punishment, Recognition, Remorse, Tested, The Enigma, Defeat, Villainy, Imprisoned, Lesson Learned

CONFLICT Ambition, I am what I am, Complicity, Conflict with a God, Cross-Rank Rivalry, Daring Enterprise, Hatred between Friends, Jealousy, Misunderstanding Arises, Revenge, Revolt, Rivalry, Struggle, Judgement Deferred to Authority, Trickery, Underdog, Unfounded Claims, Unrelenting Guardian

HELP – REWARD – RESOLUTION Assistance, Bond Strengthened, Useful Information, Lack Fulfilled, Aspiration Achieved, Provision of a Magical Agent, Repentance Rewarded, Reward, Riches, Victory, Misunderstanding Cleared, Reconciliation, Repentance, Solution

Dei Genitrix: A Generative Grammar for Traditional Litanies*

Francesco Galofaro[1] and Magdalena Maria Kubas[2]

1 **Politecnico di Milano, CUBE – Bologna, Italy**
`francesco.galofaro@polimi.it`
2 **Faculty of Polish Studies, University of Warsaw, Poland**
`klara.klarnet@gmail.com`

Abstract

The object of the present paper is a study of two traditional litanies, *Litany of the Saints* and *Litany of Loreto*, in their most ancient attested form. We will design a narrative grammar to show how their litanic structures can be generated. We propose the notion of n-selection, a narrative rule which can't be reduced to syntax or semantics: it depends on culture. We will test the grammar on the *Litany to the Divine Mercy*, written by Saint Faustina Kowalska in the XXth century. Our purpose is to identify the rules of the genre which allow its reproduction in more recent versions.

1998 ACM Subject Classification J.5 [Art and Humanities] Literature

Keywords and phrases Litany, generative grammar, computational narratology, digital humanities, semiotics

Digital Object Identifier 10.4230/OASIcs.CMN.2016.12

1 Introduction

Litany is a form of Christian prayer with different functions inside and outside the liturgy, set for public or private use. It exists (or used to) in Western and Eastern rites and its origin is almost as ancient as the Christianity itself. Western litany consists of a variable list of invocations, intoned by an officiant, to whom worshippers answer with a fixed formula: "Saint Ambrose – pray for us"; "Speculum iustitie – ora pro nobis" ... Being the litany a strong vehicle of the *devotio populi*, popular devotion, we register an increasing number of texts over the centuries. Nevertheless, all litanies seem subjected to precise rules: for example, the names of the Saints are listed by their type (martyrs, doctors, virgins ...) and age; the name *Mary* is replaced by rich periphrasis grouped by semantic affinity. Furthermore, all of them can be grouped in two classes: the first addresses a list of Saints, the second only one receiver: the *Holy Name of Jesus*; the *Most Precious Blood*; *Saint Anthony*. The endurance of the two structures let us think that a single narrative grammar can explain how the identity of a text is preserved during times, thus representing diachronic changes. For this reason, we design our grammar only on the most ancient versions of the *Litany of the Saints* (further mentioned as LS) and the *Litany of Loreto* (or LL): these two texts seem prototypical. The first is addressed to many Saints, while the second has a single receiver (the Virgin).

* This article was prepared in conjunction with the research project Litanic Verse in the Culture of European Regions financially supported by the National Science Centre of Poland (DEC-2012/07/E/HS2/00665).

© Francesco Galofaro and Magdalena Maria Kubas;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 12; pp. 12:1–12:8
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1.1 Historical outline

The litany as a genre of discourse has a complex history rooted among the first communities of Eastern Christians, dating from the fourth century in Antioch, and the medieval Byzantine Church. According to Lapidge [12] and Sadowski [17], quite a mature version of LS reached England between the VIIth and the VIIIth century. As far as LL is concerned, its roots can be found in *The Akathist Hymn*, a Byzantine poetic and devotional composition, translated in Latin at the beginning of the IXth century in Venice: it is lexically linked with the *Letanie beatae Mariae virginis* said in the Basilica of Saint Mark in the XIIth century [13] [15] which on its turn manifests a number of concordances with LL. This way, the most ancient Western tradition of litany has been codified through Latin translations from languages of the early Christianity and the Byzantine Greek. LS is probably slightly older than LL, but if we assume that litany is a convergence of previously existing discursive forms (not only Christian) then LL has a number of common traits with LS. Between 1587 and 1608 they were both accepted within the canon of Catholic prayers by a papal bull, which excluded other popular and regional traditions. In the centuries to come, others were written and admitted (or not) to public or private use. For example, in the XXth century Saint Faustyna Kowalska wrote the auteur *Litany to the Divine Mercy* [10]. Sadowski [17] describes three qualities of the litanic genre:

- ektenial;
- polyonymic;
- chairetismic.

The *ektenia* is a supplicatory prayer of the Orthodox rite. In Western tradition only the formula “Kyrie eleison, Christe eleison” was preserved, and it opens each litany. Coming to the second quality, the greek *poly-onoma* means “many names”. This quality develops the trend of laudatory listing of names (i.e. of Saints) or antonomasias (i.e. periphrasis, attributes), that might be even older than Christian culture [18]. Finally, a chairetism is an invocation typical of *The Akathist Hymn* (“Hail! By whom gladness will be enkindled; Hail! By whom the curse will be quenched ... ” – [13]). The litanies of any epoch merge this three components manifesting explicitly some of them, while some others remain implicit.

2 Textual analysis

The most ancient version of LL is published in [13]. The codex dates from the end of the XIIth century. LS is edited in [12], pp. 93–97, and the manuscript dates to the mid-XIth century. The two litanies can be subdivided in three parts. The first is an *introduction* in which we find the traditional Greek invocation Kyrie; then we have the *body* of the litany, which consists of the chairetismic formulas (*Virgo suavis, ora pro nobis; Virgo fidelis, ora pro nobis;*) finally, we have an *envoy* which closes the text, in which we can find the initial Kyrie, and/or the Agnus Dei, and/or an oration.

2.1 Further articulation of the body

As far as the body is concerned, the main difference between LL and LS consists in the fact that the second one can be subdivided in two parts. In the first, worshippers address the Saints. In the second part the receiver of the supplication is God. The object of the two prayers is different too: in the second part, worshippers pray Christ to free them (*libera nos*) and to hear them (*audi nos*), whereas in the first part we find a “second-degree prayer”: worshippers pray a mediator to pray for them (*Sancte Johannes ora pro nobis*). The structure of the first part of LS is similar to LL, since the Virgin is a mediator, too. The list of Saints in LS is hierarchically ordered: Holy Mary, Angels and Archangels, Patriarchs and Prophets, Apostles and Disciples, Martyrs, Confessors, Monks and Hermits, Virgins. During centuries, new categories have been added to this list: Doctors, Priests and Religious, Laity. Some martyr women have been shifted from the category of the Virgins to the Martyrs, after men. Coming to the LL, the antonomasias of the Virgin are subdivided in groups too: *Sancta, Mater, Magistra, Virgo, Vas,* some short lists of various epithets which are semantically related each other, and eventually *Regina*. The symmetry between the two structures is really interesting.

2.2 Why a narrative grammar?

Generative grammars have been used to analyse narrative structures since the seminal work of Lakoff [11]. We can't simply use Chomsky's syntactic grammar because it does not include the notion of *narrative selection* (n-selection), as we are going to explain. First let us consider the two sentences:

1. *Mary wondered the time;
2. *Bill drank the schrimps;

According to generative linguistics, (1) is not correct because it violates a syntactic rule (a *c-selection rule*), whereas (2) is not correct because it violates a semantic rule (a *s-selection rule*) according to which the verb *to drink* implies a liquid object [22]. Now, let us consider these two sentences:

3. *O Lord, pray for us;
4. *Mirror of justice, save us;

Both (3) and (4) are correct from a linguistic point of view, but they are unacceptable for cultural reasons: in fact, *Mirror of justice* is an antonomasia of the Holy Virgin, which can't save worshippers because only God can. On the other hand, who should God pray? Umberto Eco [2] would say that these are not linguistic rules, but *encyclopaedic knowledge*. Nevertheless, a *n-selection* has the strength of a rule: for every version of the text, in every epoch, there are not litanies in which worshippers ask Christ to pray, nor litanies in which the Holy Virgin is asked to save them. As Eco [3] wrote, «every prophet sees only what his culture has taught him to see and allows him to imagine». *N-selections* are the kind of rules which can be caught by a narrative grammar.

2.3 The structure of the invocation

We are going to compare the different invocations of LS and LL. To describe the deep structure which is implied by both models, we'll refer to the notion of actantial structure first formulated by Tesnière [19] and generalized to narrative by Greimas [8]. According to Tesnière the core of the enunciate is the verb, which he calls *stemma* (T). The stemma selects an actantial structure with a maximum of three actants: the *subject*, the *object* (O), the *indirect object* (I). Finally the stemma admits some *circumstants* (C), i.e. the indirect complements. In our case, the subject is further articulated in a *name* (N) and an *antonomasia* (A). Given these categories, let us analyse the structure of the invocations to God (Table 1, Table 2).

As we can see, the main difference between the two kinds of invocation to God concerns the number of enunciates: in the case of the stemma "Deliver us" we deal with a single enunciate with two actants and a circumstant, whereas in the case of the stemma "hear us" we deal with two enunciates; they would become three if we'd analyze also the embedded structure of the first *O* ("that Thou would spare us"). Let us compare now the structure of the invocation to the Holy Virgin and to the Saints (Table 3):

Table 1 Narrative structure of the invocation to God (simple form).

N	A	T	O	I	C
ø	O Lord	deliver	us	ø	from the snares of the devil
ø	O Lord	deliver	us	ø	in the day of judgement

Table 2 Narrative structure of the invocation to God (double form).

N	A	T	I	O	C
We	ø	beseach	Thee	that Thou would spare us	ø
ø	ø	hear	ø	us	ø

Table 3 Narrative structure of the invocation to the Holy Virgin and the Saints.

N	A	T	O	I	C
Saint Mary	ø	pray for	us	ø	ø
ø	Holy Virgin of Virgins	pray for	us	ø	ø
Saint Felicity	ø	pray for	us	ø	ø
All ye holy Martyrs	ø	pray for	us	ø	ø

We deal with a single enunciate with two actants and no circumstances. This structure is similar to the simple kind of invocation to God. The difference between the simple and the double structure reflects their different historical roots: as we said, the Byzantine *ektenia* is the model of the double structure, whereas the model of the simple structure can be found in a Syriac litany of the VII century, probably a translation from a Greek text, discovered by Baumstark and Lapidge [12]. Both the models are present in LS. We find their blend in more recent litanies, such as the *Litany of the Holy Name of Jesus* and the *Litany of the Most Precious Blood of Jesus*, which adopt the scheme of the simple invocation.

3 Dei Genitrix: the grammar

For reasons related to space we will propose a grammar (DG: Dei Genitrix, Table 4) for the simple structure. We will present a similar grammar for the double structure derived from the *ektenia* in a future publication. The grammar consists of two alphabets: Z, non-terminal symbols (upper case letters) and T, terminal symbols (lower case letters), with the initial symbol $I \in Z$. It also contains a set of rules of substitution $\alpha \to \beta$ where α is a non-terminal symbol and β can be a string of both terminal and non-terminal symbols. Starting from the initial symbol S, a derivation applies the rules of substitution till we get a string made of only terminal symbols. On their turn, terminal symbols are variables which can be interpreted by an appropriate member of a subset of litanic elements. Basically, each enunciate belongs to one of the two possible types (W/Z) of stemmas: "to pray for" (p) and "to deliver" (d). Each of them selects a different kind of actantial structure. The W-type is a *mediation structure*: worshippers pray a mediator to pray God. It selects the subject: one of the Saints' names (s) or the name of the Virgin (n) and one of her antonomasias (a). Since one of these two can be omitted, we inserted the empty string ϵ in these two sets. The object (o) of the W-type is always manifested by the pronoun "us". The Z-type is a "petition structure": worshippers address directly God. This structure selects two actants and a circumstant. The subject is represented by the fixed antonomasia (l), "O Lord". The object (Q) is manifested once again by the pronoun "us" (o). The circumstant (C) is a member (c) of the set of

Table 4 Dei Genitrix Grammar.

Symbolic repertoire	Rules
$Z := \{S, W, P, O, Z, D, Q, C\}$	1) $S \to kW$
$I := \{S\}$	2) $W \to sP$
$T := \{k, s, a, n, p, o, d, l, c, e\}$	3) $W \to nA$
	4) $A \to aP$
Introduction k = (Kyrie eleison . . .);	5) $P \to pO$
Subjects s = (Saint Steven; Saint Swithun; Saint Eormenhilda . . .);	6) $O \to oW$
n = (Saint Mary, ϵ); l = (O Lord); a = (ϵ, Mother of God, Gem of mercy . . .)	7) $O \to oZ$
Stemmas p = (pray for); d = (deliver);	8) $O \to oe$
Objects o = (us);	9) $Z \to lD$
Circumstants c = (From the future rage; Through the Cross	10) $D \to dQ$
and Thy passion . . .);	11) $Q \to oC$
Envoy e = (Agnus Dei, Kyrie eleison . . .)	12) $C \to cZ$
	13) $C \to ce$

Table 5 Dei Genitrix derivation of LS.

Derivation	Interpretation of the derived string in LS
S	
kW (rule 1);	(k) kyrie eleison, kriste eleison, kyrie eleison;
ksP (rule 2);	(s) Saint Guthlac
kspO (rule 5);	(p) pray for
kspoW (rule 6);	(o) us
ksposP (rule 2);	(s) Saint Eormenhilda
ksposp O (rule 5);	(p) pray for
kspospoZ (rule 7);	(o) us
ksposp olD (rule 9);	(l) O Lord
ksposp oldQ (rule 10);	(d) deliver
ksposp oldoC (rule 11);	(o) us
ksposp oldoce (rule 13)	(c) from the snares of the devil;
	(e) Kyrie eleison, Christe eleison, Kyrie eleison;

the complements (locative, separation, instrumental. . .). Rule (1) generates the introduction of the litany. Rules (1–8) generate the W-structure of the body, whereas rules (9–13) generate the Z-structures. Rule (2) generates the names in the litany of the Saints, whereas rules (3–4) generate the antonomasias of the Virgin. Rule (6) assures the recursive generation of the number of the needed W-verses. Rule (8) ends the litany with the *envoy* (e) without generating Z-verses, as it happens in LL. On the contrary, rule (7) ends the series of W-verses and starts the recursive generation of Z-verses. Rules (8 and 13) generate the envoy. We derive (Table 5) some litanic lines belonging to LS to exemplify how the grammar works.

In Table 6 we see how the grammar can be used to generate strings as 'knaponapoe', corresponding to LL. We see also how the grammar can generate different litanies. For example, to generate the *Litany to the Divine Mercy* (DML) it is sufficient to substitute the symbolic repertoire with the one represented in Table 7 preserving the same narrative structure.

Further developments of the grammar concern the double-structure which can be found in the most ancient forms of LS, which proves its relation with the Greek *ektenia*.

Table 6 Dei Genitrix derivation of LL.

Derivation	Interpretation in LL	Interpretation in DML
S		
kW (rule 1);	(k) Kyrie eleison . . .	(k) Kyrie eleison . . .
knA (rule 2);	(n) Saint Mary	(n) Divine Mercy
knaP (rule 4);	(a) ϵ	(a) crown of all God's handwork
knapO (rule 5);	(p) pray for	(p) we trust in
knapoW (rule 6);	(o) us	(o) you
knaponA (rule 3);	(n) ϵ	(n) Divine Mercy
knaponaP (rule 4);	(a) Singular vessel of devotion	(a) sweet relief for anguished hearts
knaponapO (rule 5);	(p) pray for	(p) we trust in
knaponapoe (rule 8);	(o) us	(o) you
	(e) Kyrie eleison . . .	(e) Lamb of God, who . . .

Table 7 The Litany to the Divine Mercy – XXth century.

Symbolic repertoire
Introduction k = (Kyrie eleison . . .);
n = (Divine Mercy); a = (greatest attribute of God, incomprehensible mystery, better than the heavens . . .);
Stemmas p = (we trust in);
Objects o = (you);
Envoy e = (Agnus Dei, Kyrie eleison . . .);

Another implementation concerns semantic categories to explain both the thematic groups of Saints and the coherent sets of antonomasias of the Virgin. We consider them as *isotopies* [7]. First we will try to design a regular grammar which generates each group; then we will simply sum the two grammars, since it is a theorem that the sum of two regular grammars is a regular grammar [9].

4 Discussion

Meister [14] defines computational narratology as a methodological instrument which aims to construct models which can be extended to larger bodies of text, providing empirical testing of their predictions in actual corpora, and precise and consistent explication of concepts. The computer-suitable form [21] helps to develop applications aimed to Digital Humanities. In this framework, the design of a generative grammar seems useful to different purposes. First of all, the notion of *n-selection* justifies an inquiry on cultural structures such as the narrative ones. In other terms, narrative grammars are not a simplification of syntax grammars, since they aim to explain forms of textual organisation which are wider and deeper than the enunciates analysed by generative linguistics. A second point concerns the form of the grammar: Dei Genitrix is formulated in Greibach normal form. Each rule has the form $Z \to tZ$, $Z \to t$ or $Z \to tt$, being t a terminal and Z a non-terminal symbol. Thus, it generates a *regular language*, which could be computed by a finite-state automaton [9]. Other forms of narrative structures are self-embedded, thus requiring a Context-Free Grammar with an higher computational power [6]. Though the construction of any grammar allows different solutions and subjective decisions, according to Chomsky's hierarchy regular grammars can not compute CF languages: thus our grammar proves the narrative structure of the litany to be regular beyond every reasonable doubt. This difference in computational power can be interpreted in a morphological framework. The generation of CF structures has a limit related to memory [1]. In fact, CF languages are computed by pushdown automata with an infinite stack of memory. In this perspective, regular structures seem easier to memorize: the origins of litany are related to oral transmission. Notice how Finlayson formalises Propp's model of the morphology of the folktale as a finite-state automaton [4]. Morphology gives us another clue: Propp considers complex forms as the historical developments of simpler ones [16]. The simplicity of litanic structures is an argument on their ancient origins. Thus, computational power is a type of *complexity* which can be considered when dating the different versions of a text. Goethe [20] proposes two notions of forms, considered as structure (*Gestalt*) and development (*Bildung*): our narrative grammar tries to catch this morpho-dynamic dimension. In fact, we saw how our generative grammar can be used to show how posterior litanies share the same structure of LS and LL. Thus, litany is a Gestalt which codifies and transmits sacred knowledge. On the same line, knowledge stored in a litany can be developed (Bildung): during times, the names of the older Saints are continuously forgotten and replaced by newer Saints, such as Saint Maria Goretti or Father Kolbe; the same happens to the Virgin's antonomasias. However, the form is transmitted through the epochs, as it happens with the Byzantine icons. And, to quote Florensky [5], "A window is a window because a region of light opens out beyond it; hence, the window giving this light is not itself *like* the light, nor is it subjectively linked in our imagination with our ideas of light – but the window is that very light in itself, in its ontological self-identity, that very light undivided in itself and thus inseparable from the sun".

References

1. Noam Chomsky. *Syntactic Structures*. Mouton, The Hague, 1957.
2. Umberto Eco. *Semiotics and the Philosophy of Language*. Indiana University Press, Bloomington, 1986.
3. Umberto Eco. Rappresentazioni iconiche del sacro. In Gianfranco Marrone and Nicola Dusi, editors, *Destini del sacro*, pages 111–118, Roma, 2008. Meltemi.
4. Mark A. Finlayson. Deriving narrative morphologies via analogical story merging. In D. Gentner, K. Holyoak, and B. Kokinov, editors, *New Frontiers in Analogy Research: Proceedings of the 2nd international conference on analogy (Analogy'09)*, pages 127–136, Sofia, 2009. New Bulgarian University Press.
5. Pavel Florensky. *Iconostasis*. St. Vladimir's Seminary Press, Crestwood, New York, 1996.
6. Francesco Galofaro. Formalizing narrative structures. glossematics, generativity, and transformational rules. *Signata*, (4):227–246, 2013.
7. Algirdas J. Greimas. *Structural Semantics: An Attempt of Method*. University of Nebraska Press, Lincoln, Nebraska, 1983.
8. Algirdas J. Greimas. *On Meaning*, chapter A Problem of Narrative Semiotics : Objects of Value. University of Minnesota Press, Minneapolis, 1987.
9. John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Pearson Paravia Bruno Mondadori, Milan, 2009.
10. Faustyna Kowalska. *Diary of Saint Maria Faustina Kowalska: Divine Mercy in my soul*. Marian PR, Oak Lawn, Illinois, 2005.
11. George Lakoff. Structural complexity in fairy tales. *The Study of Man*, 1:128–152, 1972.
12. Michael Lapidge. *Anglo-Saxon Litanies of the Saints*. Henry Bradshaw Society, London, 1991.
13. Gilles Gerard Meersseman. *Der Hymnos akathistos im Abendland*. Universitätsverlag, Fribourg, 1958-1960.

14 Jan C. Meister. *Computing Action. A Narratological Approach.* De Gruyter, Berlin, 2009.

15 Alessio Persic. Le litanie mariane 'aquileiesi' secondo le recensioni manoscritte friulane a confronto con la traduzione comune. *Theotokos*, (XII):367–388, 2004.

16 Vladimir Propp. *Readings in Russian poetics: Formalist and structuralist views*, chapter Fairytale transformations. M.I.T. Press, Cambridge, Mass., 1971.

17 Witold Sadowski. *Litania i poezja. Na materiale literatury polskiej od XI do XXI wieku.* WUW, Warsaw, 2011.

18 Witold Sadowski, Magdalena Kowalska, and Magdalena Maria Kubas. *Litanic Verse I: Origines, Iberia, Slavia et Europa Media.* Peter Lang, Warsaw, 2016.

19 Lucien Tesnière. *Éleméntes de syntaxe structurale.* Klincsieck, Paris, 1959.

20 Johann Wolfgang von Goethe. *The Metamorphosis of Plants.* MIT Press, Cambridge, Massachusetts, 2009.

21 Gian Piero Zarri. Representing and managing narratives in a computer-suitable form. In *Computational Models of Narrative: Papers from the AAAI Fall Symposium (FS-10-04)*, pages 73–80. Association for the Advancement of Artificial Intelligence (www.aaai.org), 2010.

22 Niina Ning Zhang. Understanding s-selection. *Studies in Chinese Linguistics*, 2016 (to appear).

What Are Analytic Narratives?*

Philippe Mongin

CNRS and HEC, Paris, France
`mongin@greg-hec.com`

Abstract

The recently born expression "analytic narratives" refers to studies that have appeared at the boundaries of history, political science and economics. These studies purport to explain specific historical events by combining the usual narrative way of historians with the analytic tools that economists and political scientists find in rational choice theory. Game theory is prominent among these tools. The paper explains what analytic narratives are by sampling from the eponymous book *Analytic Narratives* by Bates, Greif, Levi, Rosenthal and Weingast and covering one outside study by Mongin (2008). It first evaluates the explanatory performance of the new genre, using some philosophy of historical explanation, and then checks its discursive consistency, using some narratology. The paper concludes that analytic narratives can usefully complement standard narratives in historical explanation, provided they specialize in the gaps that these narratives reveal, and that they are discursively consistent, despite the tension that combining a formal model with a narration creates. Two expository modes, called *alternation* and *local supplementation*, emerge from the discussion as the most appropriate ones to resolve this tension.

1998 ACM Subject Classification I.2.7 Natural Language Processing, Discourse

Keywords and phrases Analytic narratives, historical explanation, text, form of discourse, narratology

Digital Object Identifier 10.4230/OASIcs.CMN.2016.13

1 Analytic narratives and the problems they raise

Narratives report on their subject matter – stories, or how human actions proceed from and result in events – by obeying definite rules that are both formal and substantial. Competing accounts of these rules exist, and they are all controversial, but they share a common feature, which is to endow narratives with various functions besides the basic one of organizing actions and events in a temporal order (e.g., narratives causally explain the actions and events, make teleological sense of them, turn them into objects of entertainment, and so on). By essence, narratives are synthetic objects, and this feature makes surprising and even paradoxical a recent expression of social sciences that is currently gaining ground, "analytic narratives".

The studies this term refers to emerged at the boundaries of history, political science and economics. They purport to explain specific historical events and claim to do that by combining the usual narrative way of historians with the analytic tools that economists and political scientists find in rational choice theory. The paradox of *a narrative that is also analytic* is compounded by the fact that game theory, a formalized discipline, often provides the analytic tools, whereas narratives are of necessity limited to natural languages. If specialists in narratology had been aware of the new genre (they are not), they would no doubt have expressed worries about the discursive contradiction it may involve. However,

* The author prepared this paper while visiting Wissenschaftskolleg zu Berlin in 2015–16.

© Philippe Mongin;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 13; pp. 13:1–13:13
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

unaware of this objection, the promoters of analytic narratives have strongly argued that, by collaborating on the same explanatory problems, narration and formal analysis deliver better explanations of historical events than each could ever do in isolation. This argument on explanation is at the core of the book, *Analytic Narratives*, by Bates, Greif, Levi, Rosenthal and Weingast [2], which has created the tag and provided the project that this tag expresses with a sophisticated manifesto.[1]

The present paper explains what analytic narratives are by relying on two of the five cases studies collected in the eponymous book, as well as one outside study by this author. At the same time, it explores the two claims just mentioned, i.e., that analytic narratives might involve a discursive contradiction and that they contribute to renew historical explanation. The former claim is more foundational than the latter, but to discuss the explanatory possibilities of analytic narratives turns out to be an easy way of introducing them, so we reverse the theoretical order of examination. This will also permit discussing the thornier issues *in concreto*.

The paper is organized as follows. Section 2 uses the two cases of Bates et al. [2] to evaluate the explanatory record of analytic narratives. Section 3 extends the sample with Mongin's [18] study and carries the preceding discussion to an end. At this stage, we will have selected two schemes of analytical narratives, *alternation* and *local complementation*, as being the most relevant. Section 4 moves from explanation to exposition, raises the discursive problem again, and sketches an argument from narratology to resolve it.

2 Sampling from *Analytic Narratives* and discussing explanation

It would be better to cover *Analytic Narratives* as a whole, but the studies by Greif and Levi are decently representative of the manifesto and its implementation.

Case 1 (Greif). In the Middle-Ages, the city-state of Genoa was first governed by elected consuls (1096-1194, period I), and then by an appointed magistrate, the *podestà*, who was chosen from outside the city (1194-1334, period II). Under consulate, civil peace prevailed from 1096 to 1164 (I.1), and then civil war lasted from 1164 to 1194 (I.2). Under *podesteria*, civil peace prevailed all along. Genoa's main economic activity was long-distance trade in the Mediterranean, and this activity was prosperous concomitantly with civil peace, i.e., for subperiod I.1 and period II, but with a noticeable peak at the end of the former. The main actors of economic and political life were the clans, which appear to have kept their identity and relative influence fixed for much of the period under study. In view of this fact, the time sequence becomes problematic. Why did the clans first cooperate and then fight under the politically unchanged conditions of consulate? Why did they cooperate most efficiently at the end of the first civil peace? How did the institutional move to *podesteria* contribute to reestablish the civil peace that prevailed henceforth and why did it occur when it did? These are Greif's main explanatory questions. He notes that the historians' work fails to answer them satisfactorily and even to raise them in full clarity.

Greif responds by constructing two extensive form games of perfect information involving the clans as strategic players. The first is devised for the consulate regime and is played

[1] The book and complementary article by Bates et al. [3] constitute a manifesto both for a new kind of historical narrative and for a theoretical programme in economics and political science, i.e., the neo-institutionalist programme associated with North [21]. We only consider the former connection.

between two representative clans; it explores their trade-off between maintaining mutual deterrence and collaborating on maritime undertakings. This game explains the changes within period I by using the German Emperor's external threat as a variable parameter. Depending on whether the threat is absent or present, Greif retains a different mutual deterrence equilibrium (MDE) for his game. The presence of the external threat pushes the clans towards *mutually advantageous* MDE by the following mechanism. In general, clans compete to gain control over the consulate, which would guarantee them a higher share of trade benefits, and this competition stabilize peacefully only because they spend on deterrence resources they could more profitably spend on joint trade; this is what MDE formally capture. Now, the controlling clan also incurs the burden of external wars when they happen, so that the external threat changes the clans' *ex ante* costs and benefits of conquering the consulate; this is why MDE with less resources spent on deterrence, and more on joint trade, arise when there is such a threat. The second game, again of the perfect information extensive form, is intended for the next political regime and has thus the *podestà* as a third player. A MDE in this game explains the stabilizing effect of *podesteria*. It relies on the more elementary mechanism that, if one clan challenges the other, the *podestà* can join forces with the challenged clan to defeat the challenging one, and he is motivated to do so by his rewarding scheme and the military means at his disposal. Each time, the subgame perfection concept serves to define what an equilibrium is.[2]

Case 2 (Levi). In the 19th century, several western states changed their regulations of military service, moving from conscription with provisions for buying out the duty to more or less universal conscription. Historians have usually emphasized democratization and military efficiency as being the two likely reasons. However, the latter is not clear technically (a professional army would have dominated all other arrangements), and the former is objectionable in view of the timing of reforms (they often took place either before or later than universal suffrage prevailed). Starting from these objections, Levi compares the changes in France, the USA and Prussia, paying attention not only to the chronological pattern, but also to the spatially and temporally variable pattern of buying out. (There are three distinctive forms, i.e., substitution, replacement, commutation.) She does not mean to displace the previous explanations entirely, but rather to subsume them under her own.

To do so, Levi uses an informal model in the spirit of formal political economy, whereby three main actors contribute to shape national decisions on the conscription regime. They are the army, which only wants military efficiency, the government, which wants it but also pays attention to social and economic considerations, such as employing the population efficiently, and the legislator, who aligns itself on the coalition that emerges among three social groups (traditional elites, middle class, workers). With this informal model at hand, the pattern of reform in each country can be explained by hypothesizing changes in the (informal) parameters. Levi proposes two such changes, i.e., the increased demand for troops from army and government, and the legislator's evolving preferences, both of which push in the direction of universal conscription. To account for the latter, she hypothesizes a change in the politically influential coalition (the pivotal middle class would turn away from the traditional elites and become allied with the workers), and she also mentions increased preference for equality within the social groups themselves. The two main hypotheses from the historical literature appear again, though included within a more systematic explanation. The study includes carefully collected evidence to clarify and support it.

[2] For the game-theoretic concepts of this paper, see, e.g., Osborne and Rubinstein [22].

How analytic narratives explain. There are both significant similarities and dissimilarities between the two cases. They start by pointing out where the historical literature failed in its attempted explanations, and define their own explanatory problems accordingly. The gaps to be filled typically concern fine temporal patterns (e.g., the succession of civil peace and war under apparently unchanged conditions) or fine-grained variations (e.g., between the forms of buying out) that historians did not properly account for. This critical assessment is reflected not only in the explanatory questions, but also in the subsequent modeling choices. These are answered in terms of a *model*, with two basic distinctions to be drawn.

Case 1 includes a formal model and case 2 an informal one, and the model of case 1 draws on game theory alone, whereas that of case 2 exploits more resources. This diversity raises a definitional issue, and we respond to it by taking the following view on analytic narratives: (i) *they require formal models* and (ii) *they can borrow them from any formal branch of rational choice theory*. With restriction (i), we avoid taking the edge off the new methodology. If it just means applying rational choice hypotheses to historical events, it was practiced many times before any talk of analytic narratives, and even by traditional historians themselves.[3] Levi [16], who does not make formalism a condition for analytic narratives, quite consistently claims that they are not a major innovation. The shortcoming of her position is that it deflates the tensions between the modeling and ordinary modes, which make the cases thought-provoking over and beyond their substantial contribution to history. For the sake of the discussion, it seems fruitful to dramatize these tensions, and thus take the *analytic* element also to be a *formal* element. At the same time, (ii) enlarges the toolbox beyond mere game theory, and here we join forces with Levi, whose study draws inspiration from political economy models that do not technically reduce to games. Analytic narratives should be able to accommodate not only *more* complex formal models, as hers would be if it were mathematized, but also *less* complex ones, such as are those of individual decision theory. These are meant to capture the decision makers' uncertainty when they are faced with an imperfectly known phenomenon with which there is no strategic interaction. As a basic example, think of the kind of cost-benefit analysis that might be offered to explain the foundation of Pompeii and Herculaneum, on such prosperous but also dangerous farmland as the slopes of Vesuvius. Mongin [19] cites a more controversial case in Clausewitz, suggesting that a *prima facie* interactive situation can be represented, to a degree of approximation, by a standard expected utility (EU) model of individual decision-making.

From the angle of the philosophy of explanation, the two case studies have much in common: they can be likened to the socalled *hypothetico-deductive* scheme of scientific explanation. The assumptions of the models play the role of theoretical hypotheses, and the conclusions drawn from them play the role of the empirical consequences logically deduced from these hypotheses. According to this familiar scheme, if empirical consequences match the *explanandum*, then theoretical hypotheses can serve as *explanans*, provided however that some other deductions from the same hypotheses match some other facts, so as to discard *ad hoc* explanations. The two cases testify to an effort to avoid adhocness by carrying out loose forms of independent testing. Thus, Greif supports his account of *podesteria* in Genoa by discussing other Italian cities, and Levi actually states her study as a comparison between France and the USA, with Prussia being more like a control case. The effort to secure independent testing would be more successful if the *explanans* contained proper regularities,

[3] The argument should be developed somewhat here. Traditional history massively borrows from commonsense ideas on individual rationality to confer explanatory value on its narratives. How this is done in detail is a matter of philosophical controversy; compare, e.g., Hempel [14] and Dray [7].

in which case not simply the hypothetico-deductive, but also the richer *deductive-nomological* scheme of scientific explanation would apply. The authors clearly believe to have uncovered theoretical patterns that can be transferred elsewhere.[4] However, by distancing themselves explicitly from this scheme, they implicitly recognize that they have not laid their hands on proper regularities.[5] This negative observation is more embarrassing to political scientists and economists than it is to historians. We will return to it after extending our collection of analytic narratives.

3 Sampling from military history and concluding about explanation

This section extends the sample from the history of institutions to that of war events. While revisiting a famous military campaign, Mongin ([18], [20]) offers his study as an analytic narrative and expands on the methodological principles of the new genre.

Case 3 (Mongin). Napoleon's return to power in 1815 tragically ended with a resounding defeat against Wellington and Blücher on the battlefield of Waterloo in Belgium. On June 16th, the campaign began favourably for him, with the French beating the Prussians at Ligny, near Charleroi. On June 17th, Napoleon decided to send a large detachment under Marshal Grouchy against the defeated Prussians, and he took the rest of his army to Waterloo, near Brussels, where the English and Dutch were ready for a defensive battle. On June 18th, the French failed to break through the enemy lines and were eventually crushed when the Prussians came as additional help. Though this battle and the whole campaign have aroused innumerable histories, an explanatory gap remains: why did Napoleon decide to send out Grouchy's detachment? By doing so, he ran the risk of not having it on his side when he would face Wellington, or much worse, Wellington and Blücher together if they managed to join forces. This worst possibility effectively materialized.

To make progress with this explanatory question, Mongin proposes a zero-sum game in normal form with two players, Napoleon and Blücher, allowing for uncertainty in several ways. For one, Napoleon is uncertain about whether Blücher was crushed or only weakened after the Ligny victory, and for another, both Napoleon and Blücher are uncertain of the outcomes of the next battles, with this latter form being itself twofold, i.e., each player is uncertain of both his opponent's strategic choice and the objective circumstances (these are handled by EU calculations). Given suitable parameter restrictions, von Neumann and Morgenstern's minimax solution concept for two-person, zero-sum games delivers a unique equilibrium, which involves pure strategies. This arguably delivers not only an equilibrium, but also rational choice recommendations. From his three pure strategies, keeping the army together, sending out Grouchy for a mere pursuit, and sending him out for preventing Blücher from joining Wellington, Napoleon should choose the last; and from his two strategies, retreating to Germany or joining forces with Wellington, Blücher should choose the latter. That Napoleon effectively chose the interposition strategy can only be conjectured, because his orders were not fully reported, but the game reinforces this hypothesis. The *ex post* failure is not an objection since it could result from the objective circumstances turning unfavorably and Grouchy misapprehending the plan – some historical evidence points in these two directions.

[4] See, e.g., Greif's ([11], ch. 6) "theory of endogenous institutional change" in which he fits Genoa as a particular case.

[5] Bates et al. ([2], p. 11–12) say "covering laws explanation" rather than "deductive-nomological explanation". Hempel ([14], [15]), whose philosophy of explanation is at stake here, gives a more general sense to the former expression, which he means to cover both probabilistic and deductive explanation.

Overall, the study exemplifies how an analytic narrative can be *both formal and hermeneutic*, since assumptions and conclusions are assessed in terms of evidential reports that are always incomplete, equivocal, and given the high stakes, unavoidably biased. The conclusions adjudicate among existing positions, indeed by reinforcing classic pro-Napoleonic arguments against equally classic anti-Napoleonic ones.

More on how analytic narratives explain. Like cases 1 and 2, case 3 starts from the extant historical literature, defines explanatory questions *negatively* from this corpus, reserves the modeling treatment to the major gap found in it. Mongin ([18], [20]) goes so far as to argue that to work on an analytic narrative requires these three conditions; otherwise, there would be reason for breaking up with the traditional narrative mode of historical explanation. This claim presupposes another, i.e., that the traditional narrative mode can be explanatory. Many philosophers of history agree, but some doubt or even disagree; see, e.g., White [24].

Unlike case 1, case 3 represents the game of interest in *normal* form (whereby players move once, simultaneously and independently) instead of the representation in *extensive* form (whereby players make successive moves and special dependencies result from this succession). Bates et al. [2] make a plea for the latter form, as if this were the only one adapted to historical work, but case 3 is sufficient evidence that the former is also applicable. If the extensive form is so acclaimed, this is presumably because its sequential structure seems capable of paralleling concrete sequences of actions, as made available by historical narratives, but we doubt whether this harmonious parallelism can really take place. Greif's games involves abstract moves, such as "to challenge", "to give in" and "to retaliate" in the *podesteria* game, which do not relate in any simple way to the actions that the Genoan clans once made. A similar comment would apply to the other extensive form games in *Analytic Narratives*. If the moves are idealizations, their sequential ordering cannot represent the passing of historical time, and the alleged superiority of extensive forms over normal forms vanishes.[6] Another defense of the latter is more technical. They allow more easily than the former for an elaborate treatment of uncertainty, as the game in case 3 illustrates. Bates and al. [3] honestly recognize that their work remains short of the target from this perspective.[7]

Perhaps even more strikingly, case 3 differs from the first two by the nature of its *explanandum* and proposed *explanans*. It is directly concerned with interactive decisions, and these are made by designated individuals. By contrast, cases 1 and 2 are directly concerned with *events* (states of affairs), like civil war or the prevalence of a conscription type, and only indirectly with interactive decisions, so that an interpretive reduction step is needed; moreover, the hypothetized decision makers, such as clans or the government, are not individuals, let alone designated individuals, but *collectives*. This is not all there is to the comparison. The *explanandum* of case 3 is narrowly circumscribed in time and space, whereas those of cases 1 and 2 extend in time or even time and space. To some extent, we recover here a topos of 20th century historical methodology, i.e., the contrast between traditional narrative history, which centres on the deeds of historical characters, and the more recent trends that deemphasize decisions and designated actors, while also enlarging the spatio-temporal scope of inquiry. Famously, the French *Annales* school has dramatized this contrast.

---

[6] Note that we do not discuss the quality of the idealizations made to define the moves, which is variable across *Analytic Narratives*. We only stress a consequence of the fact that idealizations are (and need to be) made.

[7] Zagare's [26] more sophisticated analysis of the July 1914 crisis makes a step forward by resorting to games of extensive form with incomplete information. The equilibrium concepts for this case are substantially complex.

Work along the older line is more clearly amenable to explanations given in terms of individual actions, hence to analytic narratives, than work according the recent trends. Roughly speaking, historians are more likely to concede case 3 than cases 1 and 2, assuming equal scholarly merit across these studies. By a similar reaction, Mongin [18] recommends exploring more cases in military strategy on the grounds that the assumptions underlying the technical treatment have a better empirical warrant there than elsewhere. At the same time, one must acknowledge that those cases are in some sense too easy, and it is more challenging to apply game- and decision-theoretic tools to cases in which the individual decision structure needs reconstructing and applies to broad time lags. A balance must be struck between theoretical audacity and empirical reasonableness. Case 1 is perhaps the best candidate for a right balance, because the large scope is made acceptable by the stability of the proposed decision structure in terms of clans, and because clans, despite being collectives, can be given the attributes of strategic players with some plausibility.

From the vantage point of the philosophy of explanation, the last case obeys the *hypothetico-deductive* scheme of scientific explanation no less than the previous ones, but leaves even less room for independent testing than these do. This connects with the choice of not only a more specific *explanandum*, but also of a more specific *explanans*. The Waterloo game is usable only on the *explanandum* for which it was devised, whereas the *podesteria* game, having more abstract structure, can in principle be used elsewhere. However, Greif contents himself with informally comparing Genoa with other towns, and the likely reason for this is that he does not have sufficient data outside Genoa. One should expect practical failures of independent testing to be very common with analytic narratives. It follows that the difference in generality between *explanantia* does not entail a strong difference in the explanatory mode.

We can reinforce this point as follows. While it is often impossible test a game as a whole on something else than the *explanandum* it is devised for, this may be possible for some of its individual assumptions. The *podesteria* game shares assumptions with the mutual deterrence games that have been explored fruitfully in international relations,[8] and the Waterloo game shares its two-person, zero-sum assumptions with at least one predecessor.[9] It is conceivable to test these structural assumptions in a variety of contexts, whether historical or not, and even experimentally. That independent testing can proceed in this way erodes the difference between more or less historically specific games.

We can conclude generally that the cases covered here do not exhibit critical variations in their explanatory functioning, despite the preferential disciplinary connections, which may suggest otherwise (with political science and economics for the first two, and history for the third). To use a familiar contrast, *all* analytic narratives are to be located on the "ideographic" rather than the "nomothetic" side of explanation. They are in this sense closer to the hermeneutic way of traditional history than to any construal of scientific explanation, be it the hypothetico-deductive one or any other. However, this first conclusion must be weighed against the opening point of the paper, i.e., that modeling and narration are discrepant expository modes. We now move from the issues of explanation to those of exposition.

[8] See, e.g., Brams ([4], [5]) and Zagare ([25], [26]).

[9] Haywood [13] used this apparatus to revisit the historical account of two campaigns in World War II. Zero-sum games are *prima facie* more relevant to war studies than to international relations studies, but this picture must be refined by taking the war objectives into account, as Clausewitz famously recommends.

4 Analytic narrative as a mode of exposition

Consider how Greif presents case 1. He organizes it chronologically, covering first civil peace under consulate, then civil war under consulate, and finally civil peace under *podesteria*. He subjects each period or sub-period (more particularly the first two) to a ternary exposition, i.e., he first provides a standard narrative that states the facts and introduces the *explananda*, second states a game-theoretic model based on relevant explanatory hypotheses, and third provides *another narrative* that finalizes the explanations. This revised narrative differs from the initial one, both because it borrows theoretical terms from the modeling part (e.g., "mutual deterrence equilibrium") and because it brings out some more historical evidence. This is, however, a narrative all right, so we have an *alternation* scheme.[10]

This exposition does not appear in an earlier, also very substantial piece by Greif [10], which uses a game-theoretic model to explain how 11th century Mediterranean traders entered in binding agreements, despite the unavailability of formal sanction mechanisms. The exposition proceeds through the previous stages one and two, but differs concerning the third, because it ends up, very classically, with a non-narrative discussion of the explanatory achievements of the model. As a secondary difference, the study is limited to a single period.

Consider now how Levi develops case 2. She (i) states the historical problems of conscription, (ii) introduces the theoretical hypotheses of the political economy model, (iii) proceeds to a detailed narrative on how conscription laws changed, and (iv) compares relevant facts from this narrative with the previously stated hypotheses. Ignoring (i), one would have a classic *hypothesis-testing* apparatus, i.e., a statement of hypotheses, followed by a statement of empirical evidence, followed by a conclusive assessment. If hypothesis testing is the aim, the evidence should remain as much as possible independent of the hypotheses, and it becomes consistent to avoid the theoretical language of (ii) when developing (iii). Levi's writing precisely conforms to this requisite. However, her study also clearly obeys an explanatory purpose, for one can fuse (i) and (iii), considering that this states the *explananda*, and that the explanatory hypotheses come in (ii) before being evaluated in (iv). Viewed in this way, case 2 comes close to Greif's [10] more classical piece. Both allow narration to enter only the *explanandum*, not the *explanans*, and they are best described as *analyzed narratives*.

Case 3 exemplifies still another form of exposition. It begins with an ordinary narrative of the Waterloo campaign written like (and to some extent parodying) extant histories; this serves to introduce both the main facts and the (here unique) *explanandum*. Then the study introduces the game-theoretic model that contains the (also unique) explanatory hypothesis, and a reflective stage assesses this hypothesis while bringing out more historical evidence. A distinctive feature of this study, it considers the initial narrative as being essentially satisfactory, except for an explanatory gap it finds at a critical juncture. The model and its assessment are like *a parenthesis in the initial narrative*, which can be resumed once the gap is filled. This scheme differs from that of analyzed narratives, because the narrative plays the role of an *explanans*, not simply of an *explanandum*, and it also differs from alternation, because it is the initial narrative, not a revised one, which has the final explanatory word. This distinctive scheme can be referred to as *local supplementation*. Consistently with a relatively modest purpose, it is more likely to arbitrate between existing sketches of explanations than to provide entirely new explanations.

Conceptually, an analytic narrative is a narrative that is somehow *made analytic*, not an analyzed narrative, and we will thus concentrate on the two expository schemes of alternation

[10] Bates's and Weingast's studies in *Analytic Narratives* [2], which are not reviewed here, also follow an alternation scheme, however with some differences.

and local supplementation. This is not to say, of course, that the analyzed narrative scheme is unimportant for historical explanation; quite the contrary, it is basic to most the work done in economic history and historical political science. We only mean to say that it does not capture what is new and challenging with analytic narratives.

We now have to check whether the two schemes of interest are not subject to discursive inconsistencies. Here we need some basic groundwork from narratology, and to begin with, the distinction between a narrative text and the narrative *form of discourse*.[11] As such, a narrative text can very well include parts that are not narrative, in the form of discourse sense, but rather descriptive, informative, explanatory, argumentative. All these received forms are *prima facie* acceptable within a narrative text, especially when it is factual (nonfictional) as those of history must be.[12] By contrast, it appears that a narrative text cannot accommodate some other received forms, such as the injunctive one.

These observations suggest an argument to conclude that analytic narratives are consistent when they conform to the local complementation scheme. It simply consists in saying that such analytic narratives are narrative *texts*, which include, among their parts in non-narrative *form*, the statements of formal models and their consequences. These parts are clearly not descriptive, but explanatory, and although this may be less obvious, they are also informative and argumentative. To see that, consider how the author of case 3 *informs* the reader of basic mathematical facts about zero-sum, two-person games, and *argues* for the approximate relevance of his model despite the shortcomings he concedes. Here information and argument are substitutes for a *demonstration* that has to be provided somewhere, but certainly not within the same text, which could accommodate the demonstrative form of discourse no more easily than, say, the injunctive form.

However, to distinguish between *prima facie* acceptable and unacceptable discourse forms is only one step in the argument. Narrative texts put implicit constraints of what they can effectively absorb from the former. It is often suggested that quantity matters – non-narrative parts should be brief so that the author would not lose the thread of the narration. But this view is usually held concerning fictional narratives, whereas factual narratives, especially historical ones, appear to be more tolerant. At least for them, what matters is not the absolute length of the non-narrative parts, but rather *how these parts connect with the narrative parts*. An example from military history will bring the point home.

Clausewitz's [6] monographic study of the Waterloo campaign is based on a nearly perfect alternation of narrative and non-narrative chapters. The former state the background facts and the protagonists' strategic moves with extreme temporal precision and refrain from passing judgments; the latter critically evaluate the adequacy of the strategic moves. Despite the high proportion and trenchantly different style of the non-narrative chapters, the book remains a narrative text, because Clausewitz submits his examinations to an implicit rule of temporal consistency. That is, they never extend beyond what has been narrated thus far and they principally relate to what was narrated last. The explanations can thus be *dated* on the same time scale as the events and actions they discuss. In a sense, they are part of the story told by the book, and this is why it remains a narrative text despite widely using the explanatory and argumentative forms of discourse.

[11] Early narratology did not explicitly make this distinction (e.g., Barthes [1]) or expressed it differently (e.g., Genette [8]), but it has become universal henceforth, even though the terminology fluctuates. By contrast, what is not stabilized is the classification of the discourse forms. None has won the day, despite the fact that they are debated since Aristotle's *Poetics*.

[12] Narratology usually draws a significant distinction between factual and fictional narratives, in which case it automatically treats historical narratives as cases of the former; see, e.g., Genette [9]. There are dissenters, however.

Clausewitz sets an illuminating precedent for the local supplementation scheme. The analytic intrusions into the narrative flow could be as lengthy and complex as the topic requires, but they should respect the rule of temporal consistency, and in particular avoid irrelevant anticipations, generalizations and comparisons. Within the constraint imposed by this rule, Clausewitz develops his examinations from the viewpoint of a best informed narrator (who knows more than the protagonists and even knows what their knowledge consists of). This is one way, but there is another, actually more in the spirit of case 3, which consists in idealizing the protagonists' subjective deliberations. Narratology draws a related distinction between the standpoints of the author, the narrator and the characters.[13]

Thus far, we have been discussing how a narrative text can feasibly accommodate non-narrative forms of discourse. This may be sufficient for local complementation, but clearly not for alternation. There, the *revised narrative* aims at combining forms of discourse with each other, and especially, the narrative with the explanatory form. We are led back to an already-mentioned problem, which we cannot pursue in this paper: is the narrative form at all capable of conveying explanations? If it is not, it is impossible to rewrite the initial narrative so as to make it explanatory; but if it is, a new question arises, i.e., whether explanations that originate in theoretical models can properly be expressed by narrative means.

An obvious point remains to be addressed. We have said nothing yet on the complication brought about by the use of formalism. Pre-analytic narratives such as Clausewitz's do not use any, which facilitates the discursive consistency check in their cases. We locate the tension in the *semantics* of the formal language, not in this language by itself. For one thing, the recourse to mathematical symbols can be minimized, and it is effectively modest in the way cases 1 and 3 are written; for another, tables of numerical data, geographical maps or diagrams, which are symbols of a different kind, occur in factual, especially historical, narrative texts, and they are considered acceptable there. But if the mathematical expressions, or the ordinary words used to replace them, have very specialized meanings, a real difficulty arises; think, for instance, of "zero-sum game" or "mutual deterrence equilibrium".[14]

In our view, the solution to this difficulty could be found in the genesis of the formal theories of rational choice. They evolved from informal conceptions of human action that belong to an immemorial source – what cognitive scientists call *folk psychology*. The stock idea is to explain an observable action by what can be hypothesized from the agent's desires, on the one hand, and his or her beliefs, on the other. In individual decision theory, this duality has become utility and probability. Game theory, which historically emerged later, borrows from a more complex background, but its organizing concepts, and even those of equilibrium, are not completely esoteric either. With some effort, those of political economy could also be traced back to refinements of common usage. The fact that analytic narratives do not draw upon formal theories indiscriminately, but on this group specifically, means that there is a common ground between the semantics of their formal concepts and the intuitions that come to the mind of a cultivated, but non-specialized reader (perhaps the typical historian?). This claim could be substantiated by revisiting cases 1 and 3 in semantic detail.

To supersede the semantic objection is important for the local supplementation scheme, but even crucial for the alternation scheme. What primarily separates the revised from the initial narrative in this scheme is the use of terms that come from the formal modeling stage.

[13] Genette [8] is an authoritative reference here.

[14] This difficulty was usefully pointed out in Grenier, Grignon, and Menger [12] and it is further explored in Mongin [18].

If this material brings with it semantic ideas that conflict with those accompanying the (still necessary) use of ordinary words, the revised narrative will not be properly understandable, and it will also be incoherent discursively.

5 Conclusions

We have explained what analytic narratives are by starting from what exists and gradually becoming more normative. We have restricted them in several ways, up to the point of retaining only two operational types. We have gone some way towards giving these types discursive foundations, although more thoroughly for one than the other; the full argument would need a longer paper. A perhaps easier task, we have argued that analytic narratives can improve historical explanation significantly. Meanwhile, we have likened the novel explanations to the traditional ones from the vantage point of the philosophy of scientific explanation.

Another comparison is in view, though again more for a later extension of the present work. How does the analytic narrative project relate to the project of studying narratives from the computational angle? We foreshadow two possible answers. The first one, which is technical, has to do with the computational aspects of the formal models in analytic narratives. The received solution for games in normal form is Nash equilibrium, which has long been studied from this angle.[15] There even exist software packages to compute the Nash equilibria of n-person games in normal form when complete information prevails, one early example, *gambit*, being still in use (see the account by two of the authors in McKelvey and McLennan [17]). When games are in extensive form and perfect information holds, the received concept of solution is subgame perfection; this is a refinement of the Nash equilibrium concept when the games are represented in normal form. This solution is unexceptionally computed by the backward induction algorithm, which has also been studied extensively (see the recent review by Shoham and Leyton-Brown [23], which also covers the computation of Nash equilibrium). To apply backward induction, hand calculations are sufficient on schoolroom examples, but give way to semi-formal heuristics on more realistic examples such as the games of case 1, and in applying these heuristics, there is always the risk of leaving aside relevant solutions. A computer-aided approach may be commendable here.[16]

Second, and much less technically, the transformations that standard narratives incur to become analytic narratives bears some relation to the transformations they incur to become computational narratives. In either case, simplification, abstraction and modeling must take place to prepare for the final stage. There are a number of ways the work in computational narratives fulfils this requisite, e.g., by making a precis of the standard narratives in a fragment of ordinary language, by translating them into a rich logical language with event and action predicates, by converting them to decision-event trees or other diagrammatic objects. From an all too quick review, the list of existing procedures does not seem yet to include the decision- and game-theoretic tools that analytic narratives rely on. Could it not be extended in this new direction? Consider the (limited) standard narrative: "In the morning of June 17, 1815, Napoleon examined the consequences of his victory against Blücher on June 16, the perspectives of his battle against Wellington on June 18, the possible strategies opened to him, and having pondered over his decisions, he called upon Marshall Grouchy

[15] In zero-sum two-person games, Nash equilibrium coincides with von Neumann and Morgenstern's minimax.

[16] Greif precisely uses such heuristics.

to let him know what they were. In the middle of the day, Grouchy left in the north-east direction with an important detachment of the French army". This can be transformed in all the ways mentioned above, but also – and we have argued, fruitfully – in the way explained by this paper. We recommend this as an addition to the list.

Acknowledgements. The author thanks Ben Miller for encouragements, Lorraine Daston, Luca Giuliani, Michael Gordin, and Daniel Schönpflug for useful conversations, and the participants to a workshop ("The Limits and Possibilities of Narrative Explanations", Wissenschaftskolleg zu Berlin, 17–18 March 2016), for helpful comments.

References

1 R. Barthes. Introduction à l'analyse structurale des récits. *Communications*, 8:1–27. An Introduction to the Structural Analysis of Narrative (Engl. version). *New Literary History*, 6:237–272, 1966.
2 R. H. Bates, , A. Greif, M. Levi, J. L. Rosenthal, B. Weingast. *Analytic Narratives*, Princeton, Princeton University Press, 1998.
3 R. H. Bates, A. Greif, M. Levi, J. L. Rosenthal, B. Weingast. The Analytic Narrative Project. *American Political Science Review*, 94:696–702, 2000.
4 S. Brams. *Game Theory and Politics.* New York, Free Press, 1975.
5 S. Brams. *Game Theory and the Humanities: Bridging Two Worlds.* Cambridge, MA, MIT Press, 2011.
6 C. von Clausewitz. *Der Feldzug von 1815 in Frankreich*, in *Hinterlassene Werke*, t.VIII,1835. Or in *Schriften, Aufsätze, Studien, Briefe*, W. Hahlweg (ed.), Göttingen, Vandenhoek and Ruprecht, 1990. French tr. by M. Niessel, *La campagne de 1815 en France.* Paris, Champ Libre, 1973. No Engl. tr. available
7 W. Dray. *Laws and Explanation in History.* Oxford University Press, 1957.
8 G. Genette. Discours du récit. In *Figures III.* Seuil, 1972. Eng. Tr. *Narrative Discourse.* Oxford, Blackwell, 1980.
9 G. Genette. Récit fictionnel, récit factuel. In *Fiction et diction.* Paris, Seuil, 1991.
10 A. Greif. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review*, 83: 525-548, 1993.
11 A. Greif. *Institutions and the Path to Modern Economy.* Cambridge University Press, 2006.
12 J. Y. Grenier, C. Grignon, P. M. Menger (eds.). *Le modèle et le récit.* Paris, Editions de la Maison des sciences de l'homme, 2001.
13 O. G. Haywood Jr. Military Decision and Game Theory. *Journal of the Operations Research Society of America*, 2:365–385, 1954.
14 C. Hempel. The Functions of General Laws in History. *Journal of Philosophy*, 39:35–48, 1942.
15 C. Hempel. *Aspects of Scientific Explanation.* New York, Academic Press, 1965.
16 M. Levi. Modeling Complex Historical Processes with Analytic Narratives. In *Akteure-Mechanismen-Modelle.* R. Mayntz (ed.). Schriften des Max-Planck-Institute für Gesellschatsforschung Köln, volume 42, pages 108–127, 2002.
17 R. McKelvey and A. McLennan. Computational of Equilibrium in Finite Games. In *Handbook of Computational Economics.* H. Amman, D. Kendrick, and J. Rust (eds.), volume 1. Amsterdam, North Holland, 1996.
18 P. Mongin. Retour à Waterloo. Histoire militaire et théorie des jeux. *Annales. Histoire, Sciences Sociales*, 63: 39-69, 2008.
19 P. Mongin. Waterloo et les regards croisés de l'interprétation. In *La pluralité interprétative.* A. Berthoz (ed.). Paris, Odile Jacob, 2009.

20 P. Mongin. A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project. unpublished, 2015.

21 D. C. North. *Institutions, Institutional Change, and Economic Performance*. Cambridge, Cambridge University Press, 1990.

22 M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. Cambridge, Mass., MIT Press, 1994.

23 Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge, Cambridge University Press, 2009.

24 H. White. The Question of the Narrative in Contemporary Historical Theory. *History and Theory*, 23:1–33, 1984.

25 F. C. Zagare. Explaining the 1914 in Europe. An Analytic Narrative. *Journal of Theoretical Politics*, 21:63–95, 2009.

26 F. C. Zagare. *The Games of July: Explaining the Great War*. Ann Arbor, University of Michigan Press, 2011.

Appraisal of Computational Model for Yorùbá Folktale Narrative

O. Deborah Ninan[1], George O. Ajíbádé[2], and Odetunji A. Ọdẹ́jọbí[3]

1 **Department of Computer Science and Engineering, Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀, Nigeria**
jninan@oauife.edu.ng
2 **Department of Linguistics and African Languages, Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀, Nigeria**
solajibade@yahoo.com
3 **Department of Computer Science and Engineering, Ọbáfẹ́mi Awólọ́wọ̀ University, Ilé-Ifẹ̀, Nigeria**
oodejobi@oauife.edu.ng

Abstract

Our effort at developing computational models for African narratives, particularly those of *Yorùbá* folktales, is challenged by the diversity in concepts and methodologies in the discipline. This motivated us to pause and consider the various computational models of narratives in the literature. This is with a view to finding the most appropriate or otherwise adapt a closely related one for the purpose. Thorndyke's story grammar was among the models of narrative in the literature which were appraised, found close in structure and was adapted for the modelling of *Yorùbá* folktales narrative. In conclusion we found that the modified version of Thorndyke's model was appropriate for modelling *Yorùbá* folktales narrative.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Computational, Modelling, Grammar, Folktales, *Yorùbá*

Digital Object Identifier 10.4230/OASIcs.CMN.2016.14

1 Introduction

In indigenous African communities, and amongst the *Yorùbá* people in particular, folktales are an essential tool for educating the youth on the culture and moral values of society. It is believed that the richness of these folktales provides interesting features and characteristics that can be found in most folktales of African origin. *Yorùbá* folktales have been orally and informally recounted and shared from generation to generation. The elders have ensured their safekeeping by handing them down to the younger generations through transmittal, recitation, narration and expression of their functions, customs and continuity of the culture in which they occur. These folktales have been studied and admired for their aesthetics and educational benefits. However, these stories are yet to receive attention from the computing research community. Hence, the following questions: 1. Can these folktales be expressed and formalised using the theories of narrative available in the literature, or 2. In case the answer to question 1 is no, can we adapt these theories for the purpose of computational modelling of these folktales?

This paper is organised as follows. The features of *Yorùbá* folktales and their peculiarities were discussed in Section 2. Section 3 appraises the existing literature in the theory and computational models of narrative. Section 4 discusses the selected model and the necessary

© O. Deborah Ninan, George O. Ajíbádé, and Odetunji A. Ọdẹ́jọbí;
licensed under Creative Commons License CC-BY
7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 14; pp. 14:1–14:11
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Table 1 Structure *Yorùbá* folktales.

Item	Feature	Definition
Opening	*À à lọ́ o.... À à lọ̀*	Announcement
Setting	There was a king in a village	Location and Time
Actors	King, chiefs, elephant, Tortoise, villagers etc.	Human, animals and spirits
Plot	Events, Episode, Songs, Praise singing	Episodes
Theme	Wisdom can excel might	Lesson learned
Proverbs	*Ọgbọ́n ju agbára*	Traditional saying
Closing	*Ìdí Àlọ́ mi rèé gbáńgbáláká*	Validity of the tale

modification for modelling *Yorùbá* folktales narrative. Section 5 summarises and concludes the paper.

2 Structure of *Yorùbá* Folktales

The *Yorùbá* folktales narrative exhibits recurrent, conventionalized, and stylized structural components that are not found in other folktales. The target audience in a standard *Yorùbá* folktale are mostly children aged 3–12. The stories are most often narrated by old women in the community. Elements of standard *Yorùbá* folktales are listed in Table 1. The *Yorùbá* folktale is usually started with an announcement by the narrator, *Ààlọ́ o*, which is greeted with an interactive response from the audience, *Ààlọ̀* which simply means folktale and the audience's reply also means folktale though said in a different intonation to that of the narrator. This usually prepares and put the audience in an expecting state, waiting for the tale that will follow. The narrator usually proceeds to state the title which usually includes the principal actors and the theme before the introduction.

A unique attribute of folktale of *Yorùbá* origin is that they mostly contain **songs** that are sang at particular stages of the narration [10]. 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. Most of the songs are repetitive as the same chorus is used repeatedly. The contents of these songs usually convey requests, reasons, petitions, vital instructions or information [11]. There are terms in the song that do not serve any semantic function but only to serve the purpose of creating a rhythm. An example is:
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ
As seen in the Tortoise and the Elephant tale. See Appendix A.

The songs are the main mechanism for interactivity in the folktale narratives as they invoke the active audience participation and serve as means to sustain attention, reinforce the story and improve audience experience of the narratives. They usually emphasize the actions and expressions of the characters. They make the tale being narrated lively and dramatic. One of the participatory effects is that they are not easily forgotten and help to remember the whole story hence it usually assist in story recall. Included in the songs are words, phrases and some actions which are repeatedly sung by the narrator and responded to by the audience.

This repetition makes the song and consequently the whole story easy to remember. Personally, I know very well that many folktales that I have been told in my early years of learning easily comes back to my memory when I hear the song included in the tale. This has also been confirmed in the literature by [18] citing (Matateyou 1997) and (Ngugi wa Thiong'o 1986) that repetition of the language and rhythm which are two important characteristics of oral storytelling in Africa make it easy to understand and recall the story from memory.

Storytelling has been identified as a means of preserving African traditions as these folktales can be well preserved in the songs included in them. The song sometimes reveals some outstanding event of the folktale and serves the purpose of portraying a good picture of the story [11]. A collection of folksongs quickly brings to remembrance the corresponding *Yorùbá* folktale.

Another component that may be present in *Yorùbá* folktales is **praise singing** This is called *Oríkì*; it is the art of Oratory in *Yorùbá* eulogy or praise singing of human beings, animals, places, things and other objects or subjects of interest. A praise is a vocative statement for a person, a family, a group, or even an entire community. It is an organised poetic words in verses and mostly, nominalized to expound individual/family attributes, qualities, successes, accomplishments combined with commonly shared community attributes.

The **plot** in *Yorùbá* folktales consists of one or a few events arranged in order determined by the storyteller. The events go through the **introduction**, where the setting's characters (animals and/or human), location (on the farm, in the village, market, river etc), and time (usually uncertain, once upon a time, in the olden days etc.) are presented. This will present a scene in equilibrium that was disturbed by the act of one or more members of the community. This proceeds into the **problem** sometimes requiring the need to find cause of the problem or who caused a disruption. An attempt (which may be one or several) is embarked upon until the solution comes in view or is achieved which may lead to the **climax** and eventually the **resolution** of the problem and the **conclusion**. A story's plot consists one or more episodes while an episode is composed of a subgoal, one or more attempts and an outcome. The simplest *Yorùbá* folktales have at least four propositions, representing a setting, beginning, development, and ending, if it is to be considered a story [8].

Themes in the *Yorùbá* folktales, may be conveyed by one or a set of characters. The theme may sometimes be introduced at the beginning with the opening statement so that the audience has a foreknowledge of the purpose of the story, or included in the resolution at the end of the story. Themes in *Yorùbá* folktales usually promote the virtues of compassion, generosity, and humility over the vices of greed, selfishness (as evident in the tortoise story and other similar ones), wickedness and excessive pride.

The tales provide causal explanations or reasons for human experience and natural phenomenon and most often end in **proverbs**. The origin and meaning of most *Yorùbá* proverbs and proverbial sayings are derived from folktale incidents. Each country in Africa has hundreds of proverbs and folktales. Some proverbs are literally easy to understand, while others are more abstract and difficult to unravel. The separation of proverb from folktale in many instances can obscure the nature of African folklore [19].

According to [14] the association of proverbs and moral tales is, in fact very close in the *Yorùbá* tradition, a proverb being normally used to round off and drive home the point made in a tale. Thus proverbs may account for the content and intent of some folktales and *vice versa.*

At the end of the folktale, after eliciting the theme, lesson and the proverb if present, the narrator usually ends with a generic reporting speech announcing that he has delivered the story and the lesson and he had not told a lie. To prove this the story teller declares that his mouth should sound *po, po, po*, otherwise it should not sound. The sound *po po, po*, is an onomatopoeia.

Understanding of some culturally bounded concepts and terminologies are also necessary to comprehend the fundamentals of *Yorùbá* narratives [12]. Ideophones or mimetics, onomatopoeia, metaphor and the informal paraphrases used by the narrator are important components which raise the imagination and interest of the audience in the story. These also enhance the audience understanding and the ability to recall the story.

3 Computational Models of Narrative Reviewed

Narratives have been categorised into formal, somewhat formal, traditional and informal groups [4]. Folktales which are stories orally passed down through generations are included in the examples of traditional and informal narrative. Several works on narrative modelling have been reported. For example, [20], introduced some criteria for representing ambiguous entities in non-fictional and temporal narratives in computer suitable form. [3] developed a computational system capable of automatically creating narrative morphologies from an existing corpus of stories. Analogical Story Merging (ASM) algorithm, based on Bayesian model merging was used. [6] identified the difficulties in intelligently generating narrative texts which will adequately compare with the human generated text. [7] also explained that structural analysis of narratives has influenced the features of how human mind works. These models provided a general view of modelling narratives that does not account for the computational modelling of peculiar features of the *Yorùbá* folktale narratives described in Section 2.

[17] while simplifying the idea of narrative theory suggests that all narratives have five stages. This simplification facilitates the description of the five stage pattern of experience from an initial state of equilibrium through an action disturbing that state to the attempt of resolving the disruption state, the solution state and finally to the terminal state in which equilibrium is re-established. [13] identified 31 functional parts as fundamental parts of objects in his classification, a function being a character action independent of the character, independent too of its manner of fulfillment, but dependent on its consequence. These may not all appear in a single story, (some functions may be omitted) since each tale is a unique selection and combination of functions but nevertheless always appear in the same sequence. [13] concluded that definition of tales was made from his comparison of the themes of the tales and that the result was a morphology. Though more precise these theories do not describe the structure of the narrative well enough to facilitate computational modelling.

A narrative is generated using the mechanism of a language. The power of the grammar of the language will determine the strength of the expression it can generate. Within the Chomsky theory of language, the computational model of a language for generating expressions in a narrative can be formulated using an appropriately defined grammar [1]. Grammars are good for capturing both the local and global coherence of properly structured plots. Analogical similarities can be found across the corpus by looking for rules and compositions of rules that recur across the narrative corpus [2]. Several story grammars ([15], [8], [16]) have been developed, they have also been used and reviewed by several authors and found to be useful tools for language teaching. Context Free Grammar (CFG) has been identified as a popular model for describing languages. This is because it can describe certain features that have a recursive structure and has been used in the study of human languages.

Rumelhart story grammar which was the earliest upon which other story grammars are built, includes some intricate semantic terms and interpretations which makes the rules somehow complex. There is no provision for series of embedded episodes in Rumelhart story rules. Setting, Beginning (a precipitating event), Reaction (the protagonist's Reaction and

Table 2 Thorndyke story grammar.

Rule	Syntax and Semantics
Rule 1:	*Story* → *Setting* + *Theme* + *Plot* + *Resolution*
	Simple stories consist of setting, theme, plot and resolution.
Rule 2:	*Setting* → *Characters* + *Location* + *Time*
	story's setting contains character location and time.
Rule 3:	*Theme* → *Event*(*s*)*optional* + *Goal*
	that the story's theme may have one or more event or none plus a goal.
Rule 4:	*Plot* → *Episode*(*s*)
	A story's plot consists one or more episodes.
Rule 5:	*Episode* → *Subgoal* + *Attempt*(*s*) + *Outcome*
	Episode is composed of a subgoal, one or more attempts and an outcome.
Rule 6:	*Attempt* → *Event*(*s*) | *Episode*
	An attempt is either one or more events or an episode.
Rule 7:	*Outcome* → *Event*(*s*) | *State*
	An outcome of an attempt is either one or more events or a state.
Rule 8:	*Resolution* → *Event* | *State*
	A story's resolution is either an event or a state.
Rule 9:	*Subgoal* | *Goal* → *Desiredstate*
	Subgoal of an episode and the goal of the theme is a desired state.
Rule 10:	*Characters* | *Location* | *Time* → *State*
	Character, location and time are state.

setting a Goal), Attempt (the effort to achieve the Goal), Outcome (the success or failure of the Attempt), and Ending (the long-range consequence of the action sequence or the added emphasis) were described in the [8] as the six major categories of folktales information. This grammar has more rules than all others yet it does not address interactivity. It has been stated that the grammar has difficulty with conversational stories in which one character says something to another, who reacts and says something to the first character, etc. [8]. It also suggests that settings and outcomes will be remembered better than attempts or goals. [16] used a class of simple narrative stories that is described by a generative grammar of plot structures to study the effects of structure and content variables on memory and comprehension of prose passages. Thorndyke story grammar, a context free purely syntactic grammar and independent of story content, identified the underlying structural elements common to a class of narrative discourses and defined plot and theme (See Figure 1). [5] observed that it is only Thorndyke's story grammar that explicitly provides for a complex or embedded plot. The [16] story grammar predicts that, other things equal, the theme-goal statement of a story would be better remembered than any attempt event statement.

Considering the description of the tale according to its component parts and the relationship of these components to each other and to the whole of the computational model, the Thorndyke story grammar is perhaps the most amenable to *Yorùbá* folktales. This motivates its choice in this work for the story representation. It is represented by the set of rules in the Table 2.

In the Thorndyke story grammar, the simple story is composed of setting, theme, plot and resolution which are found in African folktales. African folktales may not strictly follow the order of the elements in Thorndyke grammar. For example the commonest order is *TIME* + *LOCATION* + *CHARACTER* (once upon a time, in a village, there lived ...). The

Table 3 Modified Thorndyke story grammar.

Rule	Syntax and Semantics
Rule 1:	$Story \rightarrow Opening + Setting + Plot + Resolution + Closing$ stories include, opening, setting, plot, resolution and closing statement.
Rule 2:	$Opening \rightarrow Event$ Opening is a call and response statement .
Rule 3:	$Setting \rightarrow Characters + Location + Time$ Story's setting contains characters, location and time.
Rule 4:	$Plot \rightarrow Episode^{+}$ A story's plot consists one or more episodes.
Rule 5:	$Episode \rightarrow Subgoal + Attempt^{+} + Outcome$ Episode is composed of a subgoal, one or more attempts and an outcome.
Rule 6:	$Attempt \rightarrow Event(s) \mid Episode$ An attempt is the effort to achieve the goal,
Rule 7:	$Outcome \rightarrow Event(s) \mid State$ An outcome of an attempt is either one or more events or a state.
Rule 8:	$Resolution \rightarrow Event \mid State$ A story's resolution is either an event or a state. The state may be the desired state or otherwise.
Rule 9:	$Subgoal \mid Goal \rightarrow Desiredstate$ Subgoal of an episode and the goal of the theme is a desired state.
Rule 10:	$Theme \rightarrow Event(s) optional + Goal$ theme may have one or more event or none plus a goal.

order is not strict. The model will not adequately represent *Yorùbá* folktales narrative because most of *Yorùbá* folktales include songs, praise singing and sometimes result or completely describe a proverb.

4 Computational Model of *Yorùbá* Folktale

Thorndyke grammar as stated by [9], was found to be inapplicable for stories without goal structures. In standard *Yorùbá* folktales narrative, the goal structure is embedded in the story theme. Several goals can run concurrently and goals do overlap. A need to modify the Thorndyke story grammar for standard *Yorùbá* suffices. A Thorndyke story grammar representation for the Tortoise and Elephant tale (See Appendix A) is shown in Figure 1.

As shown in Figure 1, the structure cannot represent all the components of *Yorùbá* folktale. In *Yorùbá* folktale, before establishing the setting and introducing the characters, there is a preceding announcement *Ààlọ́ o* and a closing announcement (*Ìdí Ààlọ́ mi rè é gbángbáláká...*). The melodic content and the poetic eulogy present in *Yorùbá* folktales are emphasised and used to drive home important points in the story. It is obvious that these aspects of the *Yorùbá* folktale structure are not represented in any of the story grammar rules developed in the literature. To account for these *Yorùbá* folktale feature, the modified Thorndyke story grammar 2 was described.

The Modified Thorndyke story grammar (see Figure 2) can be represented by the rules in Table 3.

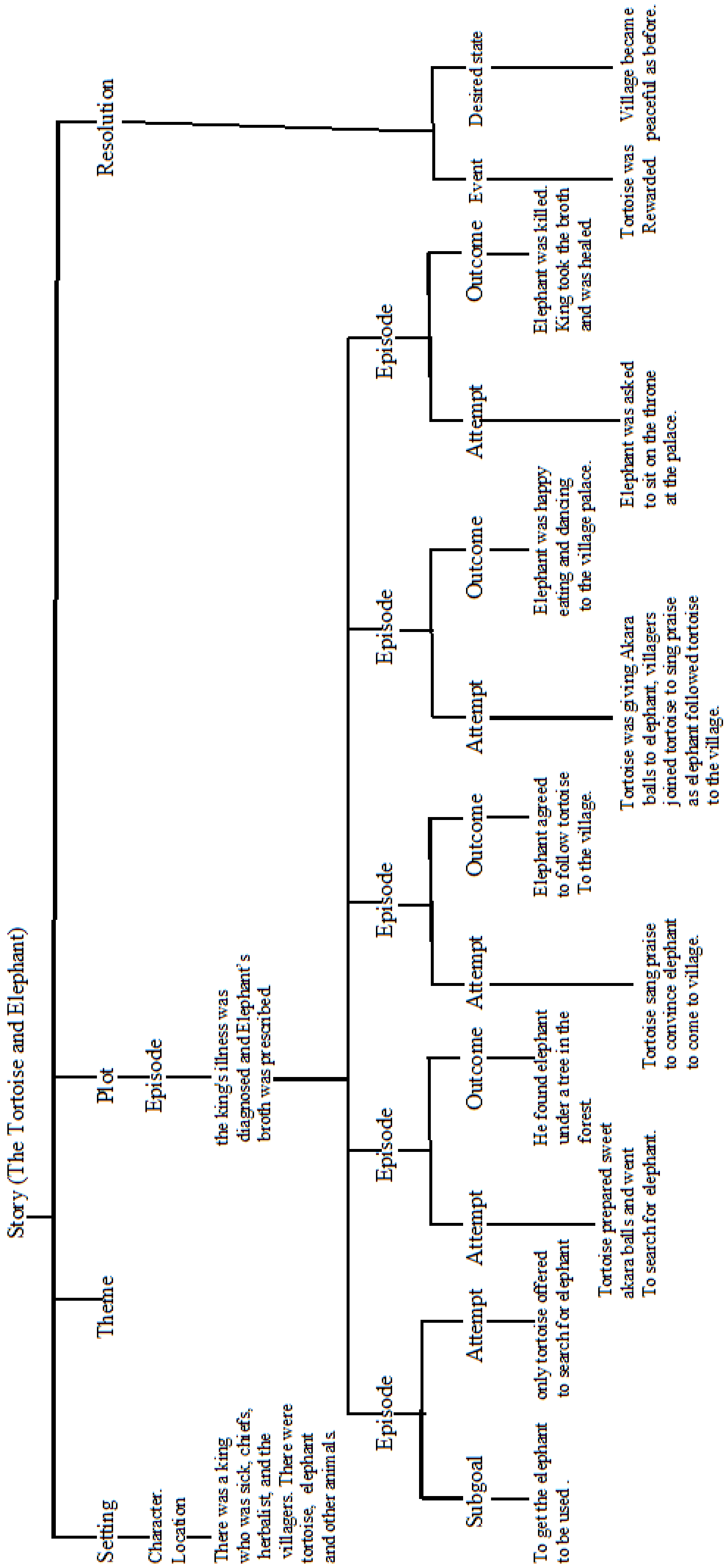


Figure 1 Thorndyke Grammar Representation of Tortoise and Elephant Tale.

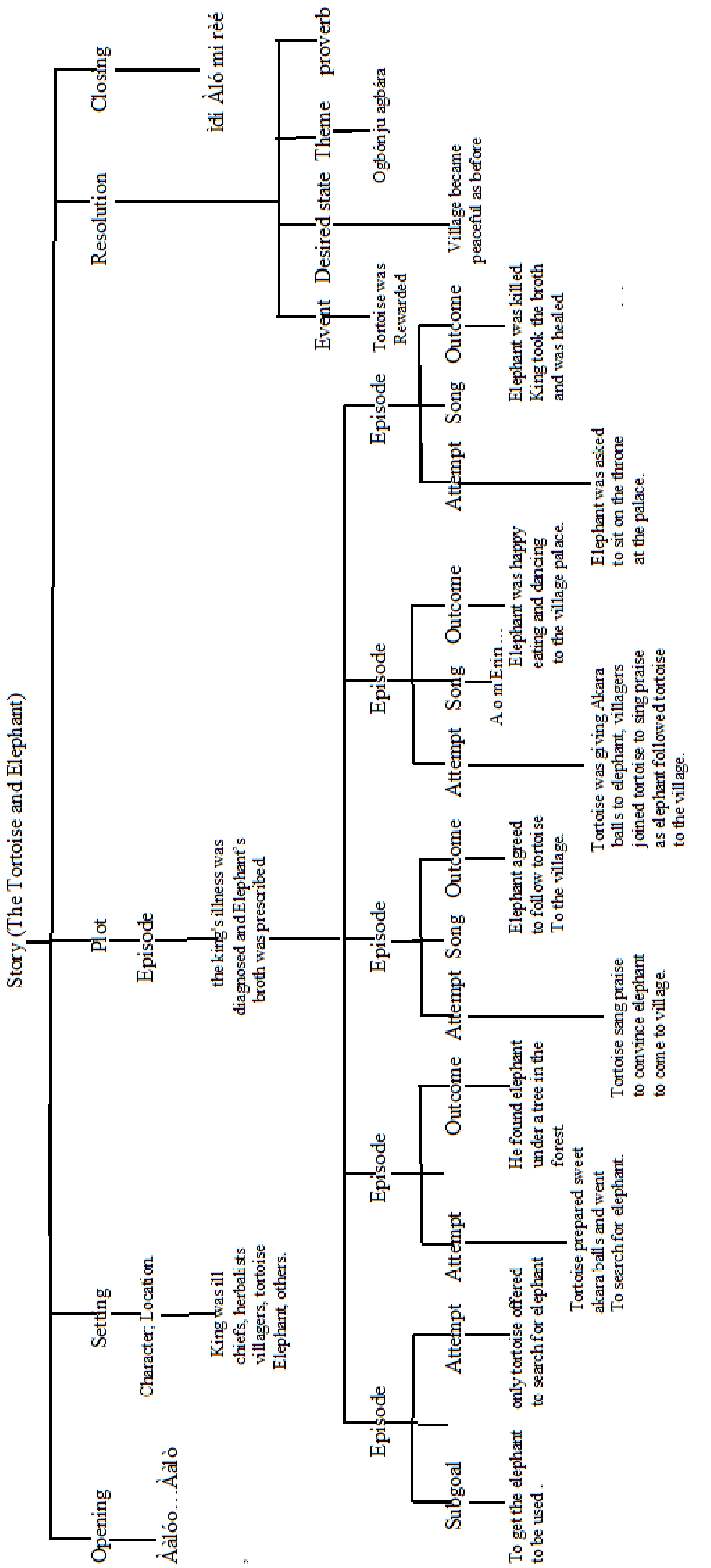


Figure 2 Modified Thorndyke Grammar Representation of Tortoise and Elephant Tale.

5 Conclusion

The story grammars in literature that have been examined were not completely sufficient for the analysis and the computational modelling of *Yorùbá* folktale narrative. Efficient as these story grammars are in analysing and modelling folktales from other sources, the style and the total content structure of the *Yorùbá* folktale narrative are not completely captured by them.

As stated earlier, a simple *Yorùbá* folktale include, opening, setting, plot, resolution and closing statement. Included in these set of rules are attempts in the episodes to achieve some goals which may include events, songs and praise singing. The resolution may include a proverb or an explanation of a norm. The theme of the story is included in the resolution as shown in Figure 2. The semantic explanations in the Rumelhat grammar include several repetitions. Both the [8] and [15] rules could not accommodate interactive opening and closing events and do not also include a resolution.

Among the story grammars appraised, Thorndyke's was found to be closely related in the structure and adaptable for the computational modelling of the *Yorùbá* folktales narrative. A modified Thorndyke story grammar for *Yorùbá* folktale narrative has been presented.

References

1 Noam Chomsky. On certain formal properties of grammars. *Information and control*, 2(2):137–167, 1959.

2 David K. Elson. *Modeling Narrative Discourse*. PhD thesis, Columbia University, 2012.

3 M. A. Finlayson. *Learning narrative structure from annotated folktales*. PhD thesis, Massachusetts Institute of Technology, 2011.

4 M. A. Finlayson, W. Richards, and P. H. Winston. Computational models of narrative: Review of a workshop. *AI Magazine*, 31(2):97, 2010.

5 J. Kwiat. From aristotle to gabriel: A summary of the narratology literature for story technologies. Technical report, Knowledge Media Institute, The Open University, UK, 2008. URL: `http://kmi.open.ac.uk/publications/pdf/kmi-08-01.pdf`.

6 Aznar Carlos León. *A computational model for automated extraction of structural schemas from simple narrative plots*. PhD thesis, Universidad Complutense de Madrid, Servicio de Publicaciones, 2011.

7 Benedikt Löwe et al. Methodological remarks about comparing formal frameworks for narratives. In *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*, pages 10–28, 2011.

8 Jean M. Mandler and Nancy S. Johnson. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9(1):111–151, 1977.

9 Utako K. Matsuyama. Can story grammar speak Japanese? *The Reading Teacher*, 36(7):666–669, 1983. URL: `http://www.jstor.org/stable/20198301`.

10 D. O. Ninan and O. A. Odejobi. Towards a digital resource for african folktales. In Mark A. Finlayson, editor, *CMN'12 Workshop on Computer Models of Narrative*, pages 75–80, 2012. URL: `http://narrative.csail.mit.edu/ws12/proceedings.pdf`.

11 O. D Ninan. *Formal Specification and Computational Modelling of African Folktale Narratives*. Unpublished PhD. Thesis, Obafemi Awolowo University, Nigeria, 2015.

12 Olufemi D. Ninan and Odetunji A. Odejobi. Theoretical issues in the computational modelling of Yorùbá narratives. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models*

of Narrative(CMN'13), pages 153–157. Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany, August 2013. URL: http://www.dagstuhl.de/oasics, doi:10.4230/OASIcs.CMN.2013.i.

13 Vladimir Propp. *Morphology of the Folktale*, volume 9. University of Texas Press, 1971.

14 E. C. Rowlands. The illustration of a Yorùbá proverb. *Journal of the Folklore Institute*, pages 250–264, 1967.

15 David E. Rumelhart. Notes on a schema for stories. *Representation and understanding: Studies in cognitive science*, 211:236, 1975.

16 Perry W. Thorndyke. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive psychology*, 9(1):77–110, 1977.

17 Tzvetan Todorov and Arnold Weinstein. Structural analysis of narrative. In *NOVEL: A forum on fiction*, pages 70–76. JSTOR, 1969.

18 Kudakwashe Tuwe. The African Oral Tradition Paradigm of Storytelling as a Methodological Framework: Employment Experiences for African communities in New Zealand. In *African Studies Association of Australasia and the Pacific (AFSAAP)*, February 2015.

19 John M. Vlach. The fuctions of proverbs in Yorùbá folktales. *Folklore Forum Bibliographic and Special Series*, special series(11):31–41, 1973. URL: http://hdl.handle.net/2022/2576.

20 G. P. Zarri. Representing and managing narratives in a computer-suitable form. In *2010 AAAI Fall Symposium Series*, pages 73–80, 2010. URL: http://www.aaai.org/ocs/index.php/FSS/FSS10/paper/viewPDFInterstitial/2183/2819.

A Tortoise and the Elephant

Tortoise and the Elephant

There was a great King who ruled in a village in *Yorùbá* land a long time ago at a time when animals could talk. The town was not peaceful, there was epidemic and death all over the village. The King also took very ill for a long period of time and was at the point of death. After several attempts by medicine men from within the kingdom to heal the king had failed, the dreaded herbalist who dwelt in the evil forest was consulted. He, after examining the king pronounced that the king would have to take a special brew made of elephant body parts or die within seven days.

The King and his chiefs wondered how they would capture a big and dangerous animal like an elephant. The king after consultation with his chiefs made an announcement throughout the kingdom that anyone who would capture an elephant within seven days would get half of the kingdom and his beautiful daughter as a bride.

The tortoise came forward to accept the challenge. He made a request of the King, that a very deep pit be dug and that the pit should be concealed with raffia and mats and that a throne fit for a king should be set on top of the pit. The tortoise made some *"àkàrà"* balls (bean cakes) and set out into the forest in search of an elephant. He wandered through the forest making inquiries of his fellow animals until the third day when he stumbled on an elephant resting under a tree. Tortoise and the Elephant engaged in a dialog,

Tortoise: Elephant, what are you doing here haven't you heard the news?

Elephant: What news? Do not disturb my siesta tortoise, I do not like gossip.

Tortoise: I can't believe my eyes, a whole King, in the forest under a shade!!.

Elephant: A King, what King?

Tortoise: YOU!! The king is dead and the elders have decided to make you king over the people.

Elephant: (roars with laughter) you must be a joker tortoise, who would want to make an old ugly elephant like me a king?

Tortoise: There is no time for explanations, preparations are already at an advanced stage in the kingdom for your coronation, we must make haste, see, I have proof (he brought out one of the *"àkàrà"* balls and handed one over to the elephant). This *"àkàrà"* is only a small part of the delicacies being prepared for your coronation.

Elephant: (putting the *àkàrà* into his mouth). Tortoise began to sing:

A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
Ní wẹlẹ, ní wẹlẹ Ẹ̀rẹ̀kú – ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
Ní wẹlẹ, ní wẹlẹ Ẹ̀rẹ̀kú – ẹwẹlẹ.
Ní wẹlẹ, ní wẹlẹ Ẹ̀rẹ̀kú – ẹwẹlẹ.
Ní wẹ-ẹ, ní wẹ-ẹ Ẹ̀rẹ̀kú – ẹwẹlẹ.
Ní wẹmu, ní wẹmu Ẹ̀rẹ̀kú – ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.
A ó m' Erin jọba, Ẹ̀rẹ̀kú-ẹwẹlẹ.

And so the tortoise led the elephant all the way to the village handing out the *àkàrà* balls to him at intervals and singing popular coronation songs to him all the way. As the tortoise and the elephant approached the palace, news of the capture of the elephant spread like wild fire, everybody came out of their houses and started following the duo to the palace joyous and joining in tortoise's songs and dancing. This created an atmosphere of festivities reinforcing the belief in the elephant's mind that he was to be made king.

Elephant: Your story must be true the people are really joyous to see me.

Tortoise: You know I wouldn't lie to you, can't you see them singing that your reign shall be long?

As the throne finally came into sight, the elephant lumbered into it majestically amidst dancing and singing. He sat on the throne and instantly the ground gave way beneath him and he fell into the pit.

The king's warriors immediately descended upon him with spears and clubs and butchered him. Once the king had taken a sip of the elephant broth made for him, he became instantly well and fulfilled his promise towards the tortoise.