

Scanning Between the Lines: The Search for the Semantic Story

Lawrence, K. Faith

f.lawrence@ria.ie
Royal Irish Academy

Battino, Paolo

p.battino@dho.ie
Royal Irish Academy

Rissen, Paul

paul.rissen@bbc.co.uk
British Broadcasting Corporation

Jewell, Michael O.

m.jewell@gold.ac.uk
Goldsmiths College, University of London

Lancioni, Tarcisio

lancioni@unisi.it
University of Siena

The panel will present three projects which are exploring the use of metadata to describe the narrative content of media. Computer-assisted textual analysis is now a well known and important facet of scholarly investigation (Potter, 1991; Burrows, 2004; Yang, 2005) however it relies heavily on statistical approaches in which the computer uses character matching to identify reoccurring strings. Although pattern recognition for image and audio search is growing more sophisticated (Downie, 2009), the techniques for annotation of multimedia are subject to the same limitations as those for text in that they cannot go beyond the shape or the waveform into the meaning that those artifacts of expression represent.

This limitation has been addressed in a number of different ways, for example through traditional categorisation and with the use of keywords for theme and motif annotation. New techniques using natural language processing software such as GATE (Auvil, 2007), and IBM's LanguageWare (1641 Depositions Project - <http://www.tcd.ie/history/1641/>) have taken this further allowing a deeper level of meaning to be

inferred from the text through basic entity and relationship recognition. The use of ontologies to support this annotation opens the way for more precise search, retrieval and analysis using the techniques developed in conjunction with semantic and linked web architecture.

The three papers being presented in this panel will address the application of these techniques to both textual and audio-visual media and consider annotation not just of the documents themselves but of the ideas contained within them, how this information might be presented to the user to best effect.

The first paper in this panel by Dr Michael O. Jewell, Goldsmiths College, University of London, focuses on the annotation of narrative in scripts and screenplays. This paper presents the combination of TEI and RDF annotations as a methodology for opening the encoded data up for inference-enhanced exploration and augmentation through linked-data resources.

The second paper from Paul Rissen, BBC, and Dr K Faith Lawrence, Royal Irish Academy, presents work being done at the BBC in the annotation of the narratives within their audio/visual archives. This paper discusses the initiatives within the BBC to make their content more accessible and to allow more personal interaction with the material. Through the use of ontology, the events contained the media object are exposed to exploration, analysis and visualisation.

The final paper by Paolo Battino, Royal Irish Academy, continues the visualisation theme to discuss how narrative annotations might be presented to assist in analysis of texts. Using the example of folktales, this paper considers the graphical representation of plotlines and the possible issues and challenges inherent for visualisation in moving from syntactic to semantic description.

References

Auvil, L., Grois, E., Lloràname, X., Pape, G., Goren, V., Sanders, B., Acs, B., McGrath, R. E. (2007). 'A Flexible System for Text Analysis with Semantic Networks'. *Proceedings of Digital Humanities 2007*.

Burrows, J. (2004). 'Textual Analysis'. *A Companion to Digital Humanities*. Schriebman, S., Siemens, R., Unsworth, J. (eds.). Oxford: Blackwell Publishing Ltd.

Downie, J. S., Byrd, D., Crawford, T. (2009). 'Ten Years Of Ismir: Reflections On Challenges And Opportunities'. *10th International Society for Music Information Retrieval Conference*.

Potter, R. G. (1991). 'Statistical Analysis of Literature: A Retrospective on Computers and the Humanities, 1966–1990'. *Computers and the Humanities*. **25.6**: 401-429.

Yang, H-C., Lee, H-C. (2005). 'Automatic Category Theme Identification and Hierarchy Generation for Chinese Text Categorization'. *Journal of Intelligent Information Systems*. **1**.

Semantic Screenplays: Preparing TEI for Linked Data

Jewell, Michael O.

m.jewell@gold.ac.uk

Goldsmiths College, University of London

Scripts, whether for radio plays, theatre, or film, are a rich source of data. As well as cast information and dialogue, they may include performance directions, locations, camera motions, sound effects, captions, or entrances and exits. The TEI Performance Texts module (<http://www.tei-c.org/P5>) provides a means to encode this information into an existing screenplay, together with more specific textual information such as metrical details.

Meanwhile, Linked Data has become a major component of the Semantic Web. This is a set of best practices for publishing and connecting structured data on the Web, which has led to the creation of a global data space containing billions of assertions, known as the Web of Data (Bizer et al, 2009). Some of the most prominent datasets in this space include DBpedia, with more than 100 million assertions relating to (amongst others) people, places, and films; LMDb (Linked Movie Database), with over three

million filmic assertions; and LinkedGeoData, which has almost two billion geographical assertions.

In this paper, we propose a means to support Linked Data in TEI, thus benefitting from the wealth of information available on top of that which is provided by TEI. We describe the augmentation of TEI documents with RDFa (Resource Description Format in Attributes) to complement the annotated content with URIs and class information, and thence the transformation of this document into triples using our open source `tei2onto` conversion tool. Finally, we provide some case studies that make use of the resultant triples, and show how their compliance with the OntoMedia ontologies (Lawrence et al, 2006) allows for powerful research possibilities.

1. Annoting TEI

1.1. Cast Lists

The first, and simplest, step to adding RDFa attributes to a TEI document begins with the cast list. Listing 1 shows a simple example of a `castItem` element for the role of Jeffrey Beaumont in *Blue Velvet*, portrayed by Kyle MacLachlan. The `about` attribute specifies the object to which the element relates: the actor element refers to the DBpedia entry for Kyle MacLachlan, while the role refers to an object residing within the *Blue Velvet* namespace, created specifically for this screenplay. The **property** attribute defines the predicate that relates the content of the element to the object - in this case, it is the name of the actor or character. When processed by `tei2onto`, actors are specified as **Being** objects, which are subclasses of the Friend of a Friend (FOAF) ontology's Agent class (<http://www.foaf-project.org/>), and roles as Character objects.

The conversion script then analyses `sp` elements for who attributes. These refer to the `xml:id` attributes in the role elements, and thus it is possible to determine the cast present in a scene and the entity speaking a line. The former may be found via the involves predicates of the event, while the latter is represented with the **has-subject-entity** predicate. An OntoMedia Social event is created for each element, with the **precedes** and **follows** attributes describing

the sequence of these events in the screenplay. Listing 2 shows the N3 representation created from a single `sp` element. Listing 2: The N3 extracted from a TEI `sp` element given an annotated `castList` and `valid` who attributes. The `ome` prefix refers to the OntoMedia Expression namespace.

```
<castItem>
  <actor
    about="[dbpedia:Kyle_MacLachlan]"
    property="foaf:name">Kyle MacLachlan</
actor>
  <role
    xml:id="jeffrey_beaumont"
    about="[bv:Jeffrey_Beaumont]"
    property="foaf:name">Jeffrey Beaumont</
role>
</castItem>
```

Listing 1: A TEI `castItem` augmented with RDFa. The `dbpedia`, `foaf` and `bv` prefixes refer to the DBpedia, FOAF, and custom Blue Velvet namespaces.

```
<http://contextus.net/resource/blue_velvet/
event/6>
a ome:Social; ome:follows
<http://contextus.net/resource/blue_velvet/
event/5>;
ome:has-subject-entity bv:Detective_Williams;
ome:involves
bv:Detective_Williams, bv:Jeffrey_Beaumont;
ome:precedes
<http://contextus.net/resource/blue_velvet/
event/7>;
```

Listing 2: The N3 extracted from a TEI `sp` element given an annotated `castList` and `valid` who attributes. The `ome` prefix refers to the OntoMedia Expression namespace.

1.2. Locations

OntoMedia provides an extensive location ontology, and it is thus useful to be able to specify this within a TEI document. `tei2onto` analyses the stage elements for this purpose, specifically when the type attribute is given as `'location'`. The only compulsory attribute is `about` which, as with the `castList` elements, typically refers to an object in the screenplay's namespace. This allows for references to the same location several times throughout a screenplay, or even for other screenplays to reference it (e.g. the Statue of Liberty, or Area 51).

By default, `tei2onto` defines locations as being instances of the `Space` class. This is the highest level `Location` class, and equivalent to the AKT `Location Ontology Abstract-Space` class. To specify a more relevant class, the RDFa

`typeof` attribute may be used. Furthermore, by specifying a `p` sub-element it is possible to use the textual name of the location as the `dc:title` of the object. Listing 3 gives an example of this, with Listing 4 showing the generated location N3 representation for the event from Listing 2 which is set in Room 221.

2. Expression References

Finally, we augment the `rs` element with the RDFa `about` attribute to provide a powerful means of object reference. OntoMedia defines the `refers-to` property as a means to indicate that an Expression object refers to another Expression object. For example, an event may refer to a location, or a character, or even another event. By adding this attribute to a TEI document many interesting queries may be performed. Listing 5 illustrates a (slightly abbreviated) set of examples of this, and Listing 6 contains the full N3 result generated for the two `sp` elements.

```
<stage type="location" about="[bv:Room_221]"
typeof="[loc:Room]">
  <p property="dc:title">INT. ROOM 221 -
POLICE STATION</p>
</stage>
```

Listing 3: A TEI stage direction augmented with RDFa

```
bv:Room_221 a loc:Room;
dc:title "INT. ROOM 221 - POLICE
STATION"^^<rdf:XMLLiteral>.
```

```
<http://contextus.net/resource/blue_velvet/
event/6> a
ome:Social;
loc:is-located-in bv:Room_221.
```

Listing 4: The N3 generated from Listing 3

```
<sp who="#jeffrey_beaumont">
  <speaker>JEFFREY</speaker>
  <l>Is <rs type="person"
about="[bv:Detective_Williams]">Detective
Williams</rs> here?</l>
</sp>
<sp who="#desk_sergeant">
  <speaker>DESK SERGEANT</speaker>
  <l>He's up in <rs type="place"
about="[bv:Room_221]">Room 221</rs>.</l>
</sp>
```

Listing 5: Augmenting the `rs` element to provide location and character references

```
<http://contextus.net/resource/blue_velvet/
event/1> a ome:Social;
```

```

    ome:has-subject-entity bv:Jeffrey_Beaumont;
    ome:involves bv:Desk_Sergeant,
    bv:Jeffrey_Beaumont;
    ome:precedes
<http://contextus.net/resource/blue_velvet/
event/2>;
    ome:refers-to bv:Detective_Williams;
    loc:is-located-in
    bv:Police_Department_Reception.

<http://contextus.net/resource/blue_velvet/
event/2> a ome:Social;
    ome:follows <http://contextus.net/resource/
blue_velvet/event/1>;
    ome:has-subject-entity bv:Desk_Sergeant;
    ome:involves bv:Desk_Sergeant,
    bv:Jeffrey_Beaumont;
    ome:precedes
<http://contextus.net/resource/blue_velvet/
event/3>;
    ome:refers-to bv:Room_221;
    loc:is-located-in
    bv:Police_Department_Reception.

```

Listing 6: The N3 generated from Listing 5

```

PREFIX bv: <http://contextus.net/resource/
blue_velvet/>
PREFIX ome: <http://purl.org/ontomedia/core/
expression#>

```

```

SELECT DISTINCT ?event WHERE
{
  ?event ome:refers-to bv:Detective_Williams.
}

```

Listing 7: SPARQL to retrieve every scene referring to Detective Williams.

As some of the objects in the screenplay refer to external entities, it is also possible to make use of Linked Data. For example, DBpedia has a great deal of information regarding Kyle MacLachlan, who we have specified as the actor playing Jeffrey Beaumont. Listing 8 gives an example of one of these more powerful queries. In this query, every actor starring in Blue Velvet is retrieved and then every other film that they have starred in is retrieved from DBpedia. The directors of these films are then obtained as URIs. Other queries could, for example, find the most common nationality among cast members, or find actors who have been in other films by the same director. Furthermore, the OntoMedia structure could be leveraged to find films in which the same actors have played alongside each other. For series, the character URIs could also be incorporated - for example, to find every episode in which a character has had a scene set in a factory.

3. Conclusion

The **tei2onto** TEI translation tool provides a quick and non-intrusive approach to make use of the Web of Data's Linked Data. Even with the simple addition of about attributes on the cast list, hundreds of assertions about the actors are immediately available. Once location references are provided, it is possible to analyse location usage through TV series, or link them to their real-life counterparts to examine setting information. Finally, the `rs` element allows for interlinking within events: be it to find where characters and locations are introduced, to investigate which characters are the most talked about, or even to find references to characters in an entirely different film.

```

PREFIX bv: <http://contextus.net/resource/
blue_velvet/>
PREFIX dbpedia: <http://dbpedia.org/ontology/>
PREFIX omb: <http://purl.org/ontomedia/ext/
common/being#>

```

```

SELECT DISTINCT ?director WHERE
{
  ?character omb:portrayed-by ?actor.
  ?film dbpedia:starring ?actor;
  dbpedia:director ?director.
}

```

Listing 8: SPARQL to find people who have directed the Blue Velvet actors in other films.

Future revisions of **tei2onto** will include support for TEI person, place, and trait data; an approach to represent references to the original TEI document via XPath; and more specific camera and movement description. Finally, we will be providing annotated TEI and the accompanying N3 and RDF at the Contextus Data Store website (<http://contextus.net/~datastores>), which we hope will provide an entry point into the Semantic Web for narrative researchers.

References

Bizer, C., Heath, T., Berners-Lee, T. (2009). 'Linked Data - The Story So Far'. *International Journal on Semantic Web and Information Systems (IJSWIS)*.

Harris, S., Lamb, N., Shadbolt, S. (2009). '4store: The Design and Implementation of a Clustered RDF Store'. *The 5th*

International Workshop on Scalable Semantic Web Knowledge Base Systems.

Lawrence, K. F., Jewell, M. O., Schraefel, M. C., Prugel-Bennett, A. (2006). 'Annotation of Heterogenous Media Using OntoMedia'. *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*.

Re-imagining the Creative and Cultural Output of the BBC with the Semantic Web

Rissen, Paul

paul.rissen@bbc.co.uk

British Broadcasting Corporation

Lawrence, K. Faith

f.lawrence@dho.ie

Royal Irish Academy

This paper will introduce the work being done at and in conjunction with the BBC into using descriptive metadata to improve production and distribution of content but in such a way that it is positioned within the greater cultural context. With the increased digitisation and release of resources there is also an increased need for associated information that can be computer processed and analysed to adequately index and search the rapidly expanding pool of data. The use of standard metadata formats for description and storage is now part of good practice but is still limited in its scope and application. In this paper we will discuss the ongoing research into semantic description of media content to supplement and expand on current metadata practice to allow more detailed analysis and visualisation of digitised documents and the conceptual links between them.

1. New Media, New Opportunities

For the past twenty years, the World Wide Web has been used by media companies as an enabling technology, allowing them to do the same things as they have always done, but faster and cheaper, whilst making their

content more widely available, and for longer periods of time. In addition, the Web has been used as a promotional and marketing tool, increasing the public awareness of content, and providing direct consumption opportunities. In the UK, this can most obviously be seen through the successes of the BBC's iPlayer (<http://www.bbc.co.uk/iplayer/>), and Channel 4's 4 On Demand (<http://www.channel4.com/-programmes/4od>) services. However, despite these successes, it can be argued that the tendency is for media companies not to have taken full advantage of the creative opportunities offered by web technologies due to concerns regarding control of content, licensing issues, technical limitations and the need to ensure a revenue stream.

There has been much discussion and research within and around the industry on the topic of interactivity analysing of how the web could be used to offer new, more immersive and satisfying experiences to the audiences. The EU's NM2 'New Millennium, New Media' project (<http://www.ist-nm2.org/publications/deliverables/deliverables.html>)

commissioned a number of studies on audience appreciation of media output in addition to paving the way for various experiments into the use of traditional IT-based production tools to enhance existing content offerings (Ursu, 2008). However, we argue that this work has been heavily based in the traditional understanding of media production and as such uses graphical technologies, such as Flash, whereby the audience is still reduced to a primarily passive role in the consumption of media content. This leads to a mis-use of the term 'interactive' to indicate not a more personalised experience but one which is being delivered through a different medium.

2. Interactivity and the Semantic Document

The hype surrounding the term 'Semantic Web' (Berners-Lee et al., 2001) in recent years has led some to doubt its existence and the opportunities it claims to offer. Although commonly acknowledged to be a complex subject, at its heart, the idea of the semantic web, and of the web itself, is simple. Concepts which are of interest, be they people, places, events or cultural movements, are given unique,

permanent identifiers, and links, crucially links with meaning, are drawn between them. In this, its roots can be seen very much in the original design of the Web as put forward by Sir Tim Berners-Lee.

The ideas inherent in the Semantic Web reflect the experience of our own understanding of the world, where, it could be argued, it is the context, i.e. the links we make in our minds, rather than individual objects themselves, which are of the most value. The Semantic Web seeks to replicate this in digital form, and improve upon it, by making these links solid, permanent, and recorded.

At the BBC, three recent projects - BBC/programmes, BBC/music and BBC/wildlifefinder - have sought to apply these principles to parts of our output. These projects sought to increase the value for the content for the user-audience by allowing them to create their own journeys across the resources made available to them on the BBC website and at the same time drawing increased understanding and insight from the knowledge presented on those pages and from selected external sources across the web.

The work to date has concentrated on the structures of production and distribution of BBC content. However, we argue that the real audience value and appreciation can be gained by applying the same approach to the content itself and annotating not only the video/audio files we create, but, more importantly, the narrative structures contained within them. Research into audience and fan studies (for instance Jenkins, 1992; Harris et al., 1998; Baym, 2000; Hills, 2002; Jenkins, 2006) suggests that the audience is creating and expanding a narrative structure within their minds while they are watching media content. While, in reference to this research, this factor was seen as applicable to the content produced by the BBC, its wider relevance should be noted.

Building on this initial work, our research uses a RDF-based triplestore and the Ontomedia ontology to recreate, using semantic web technologies, the users experience of narrative within media. The OntoMedia ontology was designed at the University of Southampton to enable expression of narrative structures within

and between mono- and multimedia documents (Lawrence, 2007, Part IV).

3. The Semantic Viewer

In the context of the BBC's drama output, or by corollary any similar corpus of work, we chose to apply the methodology described above as part of a pilot project. By giving each character, location and significant plot event an identifier, and creating meaningful links between them, in parallel with those found in the media itself, we argue that we can allow audiences to follow their own path through the narrative, to explore stories from different points of view, and to achieve a greater appreciation, and true interaction with, the writers' craft. In the diagram shown below (Fig. 1), the events of the Doctor Who episode Blink (Steven Moffat, 2007) are linked not only the narrative order in which they were experienced within the broadcast but also within the chronological order within the fictional universe containing them and to the orders in which specific characters perceived them. This information, once described with the OntoMedia ontology, is stored within the repository as triples where it can be queried, analysed and visualised.

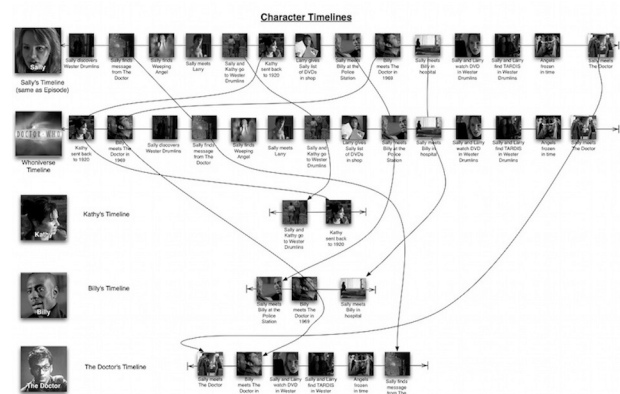


Fig. 1: Timelines For Doctor Who Episode Blink

The potential of this approach can also be seen in greater terms when applied to other areas of a media company such as the BBC's traditional output. Documentaries are often forms of narrative that seek to draw meaningful links between diverse concepts, in order to educate and inform audiences. These same audiences find enjoyment in our coverage of sport precisely because they are able to place what we report in a wider, linked context. Even news coverage, when seen through this lens,

could be transformed, allowing audiences to construct a better understanding of the world we live in, by seeing things from multiple points of view, and constructing their own, informed opinions on events.

4. Conclusion

While the research we discuss in this paper deals with the annotation of narrative within audio/visual media the techniques discussed have a much wider ranging applications. The ontology in use was designed to deal with multiple types of sources and in conjunction with other metadata standards such as CIDOC CRM (<http://cidoc.ics.forth.gr/>), FRBR (<http://www.frbr.org/>) and TEI (<http://www.tei-c.org/>). This allows for the conceptual links between many different types of documents to be made explicit in such a way that a computer could analyse, process and visualise them. While a consumer of BBC content might wish to interrogate the narrative from different perspectives, so might a literary scholar or a historian wish to explore their sources were these techniques applied to heritage materials.

References

- Baym, N.** (2000). *Tune In, Log On: Soaps, Fandom and Online Community*. California: Sage Publications.
- Berners-Lee, T., Hendler, J., Lassila, O.** (2001). 'The Semantic Web'. *Scientific American Magazine*.
- Harris, C., Alexander, A.** (1998). *Theorizing Fandom: Fans, Subculture and Identity*. New Jersey: Hampton Press, Inc.
- Hills, M.** (2002). *Fan Cultures*. London: Routledge.
- Jenkins, H.** (1992). *Textual Poachers: Television Fans and Participatory Culture*. London: Routledge.
- Jenkins, H.** (2006). *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York: New York University Press.
- Lawrence, K. F.** (2007). *The Web of Community Trust - Amateur Fiction Online: A Case Study in Community Focused Design for*

the Semantic Web, Doctoral Thesis, University of Southampton.

Ursu, M. F., Thomas, M., Kegel, I., Williams, D., Lindstedt, I., Wright, T., Leurdijk, A., Zsombori, V., Sussner, J., Myrestam, U., Hall, N. (2008). 'Interactive TV Narratives: Opportunities, Progress, and Challenges'. *ACM Trans. Multimedia Comput. Commun. Appl.*. Tuomola, M. (ed.). **4, 4, Oct.**

Visualization and Narrativity: A Generative Semiotics Approach

Battino, Paolo

p.battino@dho.ie
Royal Irish Academy

Lancioni, Tarcisio

lancioni@unisi.it
University of Siena

Storing text in a digital format opened up a whole range of new possibilities of unconventional visualization techniques, including displaying texts in forms other than textual. Similarly to what happens with pie charts and histograms in representing large amounts of numeric data, word clouds and visual taxonomies allow grasping at a glance some interesting relationships among text's constituents. The power of these graphic representations of e-texts is often based on the possibility of counting occurrences of each word, clustering words, identifying recurring patterns of words.

However, as we move from syntax to semantic, we may run into some serious limitations. Comparatively, there is a limited number of text visualisation tools based on semantics of words. Few tools, if any, exist to account for deeper semantic structures underlying texts. Finding a tool that can rightly account for two sentences with same meaning but different syntax can be hard, not to mention accounting for the narrative structure of plot or the roles of characters.

1. Narrative structures and markup languages

An interesting attempt to account for these structures in e-texts is Proppian Fairy Tale Markup Language (PftML) developed by Scott A. Malec. Based on Vladimir Propp's *Morphology of the Folktale* (1928), *PftML utilizes a Document Type Definition (DTD) to create a formal model of the structure of Russian magic tale narrative and to help standardize the tags throughout the corpus*. According to this approach, a text can be tagged as to encode Propp's "functions," or the 31 fundamental units that the Russian folklorist identified as the recurring building blocks of a Russian magic tale plot. Malec provides an example translated in English, The Swan-Geese tale. The related XML file is shown hereafter (see Listing 1), collapsed to 4 levels of depth (with the exception of the <Preparation> tag, which is fully expanded).

Software could easily parse this XML file, and quickly and reliably help the scholar in verifying some of Propp's theories, for example:

- **Acquisition of Magical Agent** appears three times, none of which directly lead to Victory,
- **Departure** is subsequent to **Villainy**,
- **Pursuit Rescue Of Hero** is composed by three series of **Pursuit Of Hero** + **Rescue Of Hero** (*collapsed under <PursuitRescueOfHero> in the above picture*)

```

<!--
  Propp predicts: gamma-delta-A-B-C-depart-((D-E-neg-F.neg)*3)/(D-E-F)-
-->
<!--
  G-H-I-K-Return-Fr-(DEF)*3-Ra - looks pretty messy, from standpoint of DTD
-->
<!-- -->
<!-- last edited 05.17.2002 -->
-->
<Corpus>
  <Folktale Title="The Swan-Geese" AT="480" NewAfanasievEditionNumber="113"
  ProppConformity="Yes">
    <Move>
      <Preparation>
        <InitialSituation>
          Once upon a time a man and a woman lived with their daughter and small son.
        </InitialSituation>
        <CommandExecution>
          <Command subtype="Interdiction">
            "Dearest daughter," said the mother, "we are going to work. Look after your brother! Don't go
            out of the yard, be a good girl, and we'll buy you a handkerchief."
          </Command>
          <Execution subtype="Violated">
            The father and mother went off to work, and the daughter soon enough forgot what they had told
            her. She put her little brother on the grass under a window and ran into the yard, where she
            played and got completely carried away having fun.
          </Execution>
        </CommandExecution>
      </Preparation>
      <Villainy subtype="Kidnapping"></Villainy>
      <ConsentToCounteraction></ConsentToCounteraction>
      <Departure></Departure>
      <!-- border of segment quite arbitrary here -->
      <!-- trebling of Donors -->
      <DonorFunction subtype="TestOffHero"></DonorFunction>
      <AcquisitionOfMagicalAgent subtype="HelperOffersServices"></AcquisitionOfMagicalAgent>
      <DonorFunction subtype="TestOffHero"></DonorFunction>
      <AcquisitionOfMagicalAgent subtype="HelperOffersServices"></AcquisitionOfMagicalAgent>
      <DonorFunction subtype="TestOffHero"></DonorFunction>
      <AcquisitionOfMagicalAgent subtype="HelperOffersServices"></AcquisitionOfMagicalAgent>
      <Transference subtype="RouteShownToHero"></Transference>
      <StruggleVictory subtype="Competition"></StruggleVictory>
      <LiquidationOfLack subtype="ReleaseFromCaptivity"></LiquidationOfLack>
      <PursuitRescueOfHero></PursuitRescueOfHero>
      <Return></Return>
    </Move>
  </Folktale>
</Corpus>

```

Listing 1: Example of XML Describing The Swan-Geese (Full Version: http://clover.slavic.pitt.edu/~sam/propp/have_a_little_byte/magicgeese.xml)

2. Graphic visualization of narrative structures

Taking Malec's work as a starting point, we produced a graphic representation of the XML file, aimed at highlighting some aspects of the narrative structure of the tale (see Fig. 1)

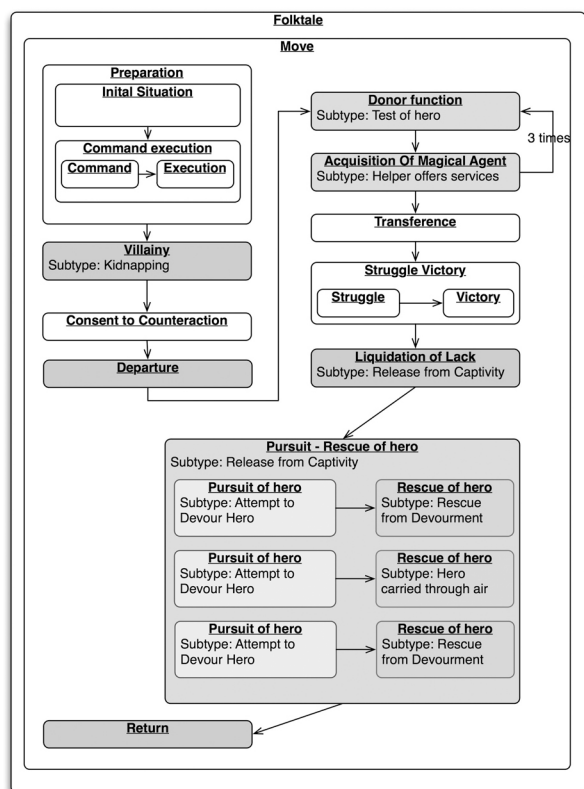


Fig. 1: Diagram Highlighting the Aspects of the Narrative Structure

The diagram above is meant to highlight the following aspects:

- Function nesting (e.g.: **Preparation** includes **Initial Situation** + **Command Execution**)
- Sequencing of “cornerstone” functions (**Villainy** -> **Departure** -> **Liquidation of Lack** -> **Return**).
- Cyclical repetition of same functions (in this case **Donor Function** + **Acquisition of Magical Agent**, and **Pursuit of Hero**).
- In case of repetition, some functions are of same subtypes (in this case the three instances of **Pursuit of Hero** are of same subtype, while **Rescue of Hero** instances are of two sub types, forming a A-B-A sequence).

3. The challenge

When we move from the analysis of words to the analysis of narrative structures we are facing a shift in the unit of analysis: we are no more limited to the elements of expression plane (i.e. words, in case of a text). We are now interested in “functions”, as named by

Propp, or “events”, or “roles”. In other words, we are interested in analysing the meaning of sentences, or even entire paragraphs and chapters. In some cases the actual words used to express the meaning can be almost irrelevant for us, and we would like to “see through” the endless possible variations of expressing the same concept.

In a folktale, if we are looking for that topic event generally called **Villainy**, regardless of whether carried out by villain or villain helper, which words should we look for? And if we are looking for **Liquidation of Lack**, represented in *Sleeping Beauty* by the re-acquisition of consciousness, can we consider these three sentences equivalent?

1. *The Prince kissed the Princess, and the Princess awoke.*
2. *The Princess awoke when kissed by the Prince.*
3. *The kiss given by the Prince awoke the Princess.*

We may say that these three sentences have different *phrase syntax* but same *actantial syntax*, as synthesized by Marsciani & Zinna (1991, pg. 56). That is, they express the same event (the Princess goes from asleep to conscious), with the characters having the same role (the Prince triggers the event), even if the words “Prince”, “Princess”, “kiss”, etc. appear to have different grammatical roles in each sentence. The word *actantial syntax* refers to Tesnière’s theory of *valency grammar*, where the verb is considered central to the sentence, like an atom that attracts a number of “participant roles”, the *actants* (Tesnière, 1959, p. 102). Tesnière explicitly attempts to analyse syntax and semantics separately (Tesnière, 1966 [1959], ch. 97 §3), and his work inspired the *actantial model* subsequently developed by A. J. Greimas (Marsciani & Zinna, 1991, pg. 54-57).

Using the notation proposed by Greimas, we could express the aforementioned

“the Princess goes from asleep to conscious” + “the Prince trigger the event”

by

[S1 -> (S2 ∩ O1)]	where S1 = Subject 1, "the Prince" S2 = Subject 2, "the Princess" O1 = Object of Value 1, "consciousness" ∩ = Union -> = Action
-------------------	--

This is an over-simplification of Greimas' actantial model. However, it is worth noting that Greimas' model is meant to formalise not only the semantics of a sentence, but also the narrative structure of the whole text. On the one hand, Greimas' model is heavily based on Propp theory (Schleifer 1987, pg. 121-126), on the other it goes far beyond the actantial model and seeks to analyse the path of meaning as it goes, in a given text, from deeper structures to surfacing structures, also known as *Generative Trajectory of Discourse* (Greimas & Courtés, 1979, pg. 157).

4. The model: the Generative Trajectory

In an attempt to formalise and graphically represent the narrative structure of a text, the Generative Trajectory proved to be a valuable starting point. This model is well suited for our purpose because:

- It is well-rooted into narratology.
- It seeks for a "*fundamental semantics and grammar*" of narrativity, focusing on the relationships between expression plane and content plane (meaning), as well as on different pertinence levels.
- It already offers some formalism to express and analyse meaning, inspired by structural linguistics.
- It is based on 40 years of semiotic studies and has already proved to be very effective in analysing an impressive variety of texts.

Encoding some elements of this semiotic model into e-texts in the form of tags allowed us to produce a prototype graphic representation of some narrative phenomena.

The elements of Greimas' theory that we took into account are:

- **Multi-level analysis:** signification is articulated in three different pertinence levels, from deep structures to surface structures (Marsciani & Zinna, 1991, pg. 132-133):

- i. **SEMIO-NARRATIVE STRUCTURES:** tags marking *axiologies* and *modalization*, as such describing how deep values are positioned on the *semiotic square*, how these values orientate the *Narrative Programs*, how the *actants* take position within the Narrative Programs.
- ii. **DISCOURSIVE STRUCTURES:** tags marking *thematization* and *figurativisation*, as such describing the actors, places and times that constitute the discourse.
- iii. **TEXTUAL STRUCTURES:** the text itself.

- **Conversion across levels:** Tags at different levels are interrelated, and these relationships constitute the tangible aspect of the "conversion process" across levels, that is the Generative Trajectory going from more abstract (deeper) levels to more concrete (surface) levels, up to the manifest level: the text itself.
- **Narrative Programs Nesting:** Besides the Basic Narrative Program, the one that subsume the whole text, other sub-Programs are taken into account (Array of Narrative Programs, (Hebert, 2006)).
- **Actantial Model.**
- **Semiotic Square:** used to describe articulation and axiology of deep values across levels.

5. Conclusion

In order to give a visual representation to narrative structures of text, we need to formalise the semantic and syntax of these structures. To that end, we relied on the A. J. Greimas' Generative Trajectory theory. Implementing some aspects of this theory in an experimental mark-up language allowed us to generate graphic visualization of the underlying deep semantic structures of texts. Some other aspects of Generative Trajectory could be then implemented, for example the **Canonical Narrative Schema**, especially relevant for the analysis of folktales corpora.

References

Greimas, A. J., Courtés (1979). *Sémiotique, Dictionnaire raisonné de la théorie du langage*. Paris: Hachette.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck, 2nd ed. 1966.

Schleifer, R. (1987). *A. J. Greimas and the Nature of Meaning*. London: Croom Helm.

Propp, V. Y. (1928). *Morfologija skazki*. Leningrad: Academia, English translation: *Morphology of the Folktale*, The Hague: Mouton, 1958; Austin: University of Texas Press, 1968.

Marsciani, F., Zinna, A. (1991). *Elementi di Semiotica Generativa*. Bologna: Esculapio.

Hébert, L. (2006). 'The Actantial Model'. *Signo*. Louis Hébert (dir.) (ed.). <http://www.signosemio.com> (accessed 15 November 2009).