An Augmented Annotation Schema for Fairy Tales Using Proppian Content Descriptors

ECAI 2010 workshop on:
Language Technology for Cultural Heritage, Social Sciences, and Humanities

Thierry Declerck,
Antonia Scheidel

Piroska Lendvai
Motivation

Background
Projects CLARIN, D-SPIN aim to provide an integrated and interoperable research infrastructure of language resources and LT to support eHumanities (among others)

So why start with fairy tales?

• Large, high-quality corpora (Gutenberg project, Afánas'ev collection of Russian folktales, ...)
• Possibilities for comparison of fairy tales across cultures and languages
• Structure has been studied extensively
What makes a Fairy Tale?

1. The Cast: 7 Archetypes

1. The Villain
2. The Princess (and Her Father)
3. The Dispatcher
4. The Hero
5. The Donor
6. The (magical) Helper
7. The False Hero
What makes a Fairy Tale?

1. The Cast: 7 Archetypes

1. The Villain
2. The Princess (and Her Father)
3. The Dispatcher
4. The Hero
5. The Donor
6. The (magical) Helper
7. The False Hero
What makes a Fairy Tale?

2. The Story: 31 Functions

<table>
<thead>
<tr>
<th>Preparation</th>
<th>Complication</th>
<th>Donors</th>
<th>Struggle + Return</th>
<th>Dénouement</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>ε</td>
<td>β</td>
<td>γ</td>
<td>δ</td>
</tr>
<tr>
<td>Initial Situation</td>
<td>Info. sought</td>
<td>Absentation</td>
<td>Interdiction</td>
<td>Interdict. violated</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>ζ</td>
<td>η</td>
<td>Θ</td>
<td>ι</td>
<td>θ</td>
</tr>
<tr>
<td>Info. obtained</td>
<td>Trickery</td>
<td>Fall for Trick</td>
<td>Villainy / Lack</td>
<td>Mediation</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Villainy / Lack</td>
<td>Mediation</td>
<td>Counter-action</td>
<td>Test</td>
<td>Pass Test</td>
<td>Magical Helper</td>
<td>Guidance</td>
<td>Struggle</td>
<td>Victory</td>
<td>Branding</td>
<td>Lack is liquidated</td>
<td>Arrival in Disguise</td>
<td>Difficult Task</td>
<td>Solution</td>
<td>Impostor exposed</td>
<td>Transfiguration</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Example 1: Little Red Riding Hood

Scheme: $\alpha \gamma \delta \ [\varepsilon \zeta]^3 \ [\eta \theta]^3 \ ABC \ IK \ ExU$

The better to eat you with, my dear!
Example 2: The Magic Swan-Geese

\[ \alpha \gamma \beta \delta \ ABC \uparrow [D \neg E \neg F]^3 \ G \ DEF \ HK \downarrow [PrDEF = Rs]^3 \]

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Example 2: The Magic Swan-Geese

Once upon a time a man and a woman lived with their daughter and small son.

"Dearest daughter," said the mother, "we are going to work. Look after your brother! Don't go out into the yard, be a good girl, and we'll buy you a handkerchief."
### Example 2: The Magic Swan-Geese

\[\alpha\gamma\beta\delta\ ABC\uparrow \ [D\neg E\neg F]^3 \ G\ \text{DEF}\ HK\downarrow \ [\text{PrDEF} = \text{Rs}]^3\]

<table>
<thead>
<tr>
<th></th>
<th>(\alpha) Initial Situation</th>
<th>(\varepsilon) Info. sought</th>
<th>(\beta) Absentiation</th>
<th>(\zeta) Info. obtained</th>
<th>(\gamma) Interdiction</th>
<th>(\eta) Trickery</th>
<th>(\delta) Interdict. violated</th>
<th>(\theta) Fall for Trick</th>
<th>(A) Villainy / Lack</th>
<th>(B) Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(\alpha) Initial Situation</td>
<td>0</td>
<td>(\varepsilon) Info. sought</td>
<td>4</td>
<td>(\beta) Absentiation</td>
<td>1</td>
<td>(\zeta) Info. obtained</td>
<td>5</td>
<td>(\gamma) Interdiction</td>
<td>2</td>
</tr>
</tbody>
</table>

Once upon a time a man and a woman lived with their daughter and small son.

"Dearest daughter," said the mother, "we are going to work. Look after your brother! Don't go out into the yard, be a good girl, and we'll buy you a handkerchief."
A Two-Part Problem

Our aim is to annotate fairy tales (semi)automatically.
A Two-Part Problem

Our aim is to annotate fairy tales (semi)automatically.

How?
A Two-Part Problem

Our aim is to annotate fairy tales (semi)automatically.

Using what exactly?

How?
A Two-Part Problem

Our aim is to annotate fairy tales (semi)automatically.

Using what exactly?
Annotation Schema

How?
Strategy

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Annotation Schemes for Fairy Tales

1: PftML (Proppian fairy tale Markup Language)

- Developed by Scott A. Malec
- Faithful to the 31 functions
- Inline XML annotation
  (paragraph / sentence-wise)

**Drawbacks:**
- Not very flexible
- Coarse-grained
Annotation Schemes for Fairy Tales

1: PftML (Proppian fairy tale Markup Language)

- Developed by Scott A. Malec
- Faithful to the 31 functions
- Inline XML annotation
  (paragraph / sentence-wise)

Drawbacks:
- Not very flexible
- Coarse-grained

<Preparation>
  - <InitialSituation>
    Once upon a time a man and a woman lived with their daughter and small son.
  </InitialSituation>
  - <CommandExecution>
    "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
A Closer Look at a Proppian Function

Subfunctions:
β¹: Absentation of Elders
β²: Death of Parents
β³: Absentation of Youth

"Frame":
• Performer of absentation
• Form of absentation
• Motivation

Sources for PftML

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Annotation Schemes for Fairy Tales

2: Our Approach: APftML (Augmented PftML)

- First "Propp complete" annotation scheme
- Will allow semi-automatic annotation of fairy tales

Prototype will be presented at
- CLARIN/DARIAH conference (Oct. 19-20, Vienna)
- and AMICUS workshop (Oct. 21, Vienna)
"Propp complete"?

APfTML

PfTML

Morphology of the Folktale

Proppian "frames"

31 functions

7 characters

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Sources for APftML

APftML

D-SPIN

7 characters

Proppian "frames"

31 functions

Morphology of the Folktale

PftML

annotation standard

pipeline for linguistic annotation

Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales
The parents went off to work, and the daughter soon enough forgot what they had told her.
Annotation of The Magic Swan-Geese

1. Keep Track of Characters

The parents went off to work, and the daughter soon enough forgot what they had told her.
Annotation of The Magic Swan-Geese

2. Keep Track of Functions & "Frames"

Don't go out into the yard

Violation of Interdiction
Interdiction violated
Person performing
Motivation

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.
Annotation of The Magic Swan-Geese

2. Keep Track of Functions & "Frames"

Don't go out into the yard

Violation of Interdiction

- Interdiction violated
- Person performing: she
- Motivation: having fun

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.
Annotation of The Magic Swan-Geese
2. Keep Track of Functions & "Frames"

Don't go out into the yard

Violation of Interdiction
Interdiction violated
Person performing
Motivation

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.
Ongoing Work

- Integration with linguistic and semantic resources (Wiktionary, TEI annotation infrastructure for narratives, WordNet, FrameNet, ProppOnto ontology)
- Implementation of coreference resolution
- Multilingual processing, using multilingual resources
  - Extend ProppOnto with a linguistic model for ontology labels, within project MONNET (Multilingual Ontologies for Networked Knowledge)
...and they lived happily ever after.

Thank you for your attention!

Time for your questions.
Acknowledgements

This work has been partially funded by the projects

**CLARIN & D-SPIN:** Annotation of Fairy Tales,
see [http://www.clarin.eu/external/](http://www.clarin.eu/external/) and
[http://weblicht.sfs.uni-tuebingen.de/](http://weblicht.sfs.uni-tuebingen.de/)

**MONNET:** Multilingual Ontologies, see
Introduction: Vladimir A. Propp: Morphology of the Folktale (1968)

PftML: Scott A. Malec's notes on the development of PftML:

ProppOnto: Federico Peinado, Pablo Gervás, Belén Díaz-Agudo:
A Description Logic Ontology for Fairy Tale Generation (2010)

TEI: The Text Encoding Initiative: http://www.tei-c.org/