

APftML – Augmented Proppian fairy tale Markup Language

Antonia Scheidel
DFKI GmbH, Language Technology Lab
Stuhlsatzenhausweg, 3
66123 Saarbrücken, Germany
Antonia.Scheidel@dfki.de

Thierry Declerck
DFKI GmbH, Language Technology Lab
Stuhlsatzenhausweg, 3
66123 Saarbrücken, Germany
declerck@dfki.de

ABSTRACT

This poster submission presents the actual state of development of a markup scheme that combines narrative and linguistic information for the fine-grained annotation of folktales. The scheme builds on and extends an existing markup language called PftML (Proppian fairy tale Markup Language) and combines this with textual and linguistic annotation standards as proposed by TEI (Text Encoding Initiative) and ISO TC37/SC4 on language resources management. We call our scheme therefore APftML (Augmented Proppian fairy tale Markup Language). While the poster itself will show detailed examples of the application of the annotation scheme to German versions of “Little Red Riding Hood” and “The Magic Swan Geese”, the paper concentrates on describing the resources we have been using, developing and integrating in APftML, which is providing in fact the goal annotation structure of on-going work on the automated semantic annotation of folktales.

1. INTRODUCTION

The work we describe here is part of the projects CLARIN¹ and D-SPIN². While CLARIN is focusing on the establishment of an integrated and interoperable research infrastructure of language resources and technologies that aims at enabling eHumanities research in cooperation with Human Language Technology (HLT), the D-SPIN project, which is the German contribution to CLARIN, is additionally providing for integrated language processing Web services that generate linguistic annotation, which can be concretely used in eHumanities research.

Our particular goal in this context is to integrate linguistic annotation and markup in the fields of folk and fairy tales both in a Markup language and in an automated processing chain. In a first step, which is described in this poster, we propose the combination of standardized linguistic annota-

¹<http://www.clarin.eu/external/>

²<http://weblicht.sfs.uni-tuebingen.de/englisch/index.shtml>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

First International AMICUS Workshop, October 21, 2010, Vienna, Austria.

Copyright 2010 AMICUS project at <http://amicus.uvt.nl>.

tion frameworks with a fine-grained annotation scheme that is implemented in accordance with concepts introduced in [3]³.

As a first example, we chose to annotate German version of “Little Red Riding Hood”. This annotation exercise is planned to be extended to most of the folktales⁴ from the Brothers Grimm’s collection, as they are available within the Gutenberg project⁵. In collaboration with the AMICUS project⁶ we also propose the annotation of a German version of “The Magic Swan Geese” and will extend this exercise to more tales included in [1]⁷, also considering versions of the tales in other languages, like English, Hungarian and Russian.

2. THE RESOURCES

Among the sources for our work, besides Propp’s “Morphology of the Folktale”, we consider Scott Malec’s PftML⁸, the ProppOnto Ontology⁹, FrameNet¹⁰ and the TEI¹¹ and ISO TC37/SC4¹² standards for textual and linguistic annotation. We concentrate here on the resources described by Vladimir Propp and on the PftML scheme, which we extend into APftML (Augmented Proppian fairy tale Markup Language), using TEI and ISO TC37/SC4 annotation standards.

2.1 "Morphology of the Folktale"

In his study of the Russian folktales, Propp aimed at breaking down those tales to smaller and recurrent narrative units, also called narratemes. We summarize here the main outcomes of his studies:

7 Characters. Propp puts forward the notion that the folktale know no more than seven *dramatis personae*: The villain, the donor, the helper, the princess (or “sought-for per-

³See also http://en.wikipedia.org/wiki/Vladimir_Propp.

⁴some of folktales collected by the Grimms deviate too far from the “magic tale” on which Propp based his theory

⁵See <http://www.gutenberg.org/>

⁶<http://amicus.uvt.nl/>

⁷See also http://en.wikipedia.org/wiki/Alexander_Afanasyev

⁸<http://clover.slavic.pitt.edu/sam/propp/theory/propp.html>

⁹<http://www.fdi.ucm.es/profesor/fpeinado/projects/kiids/apps/protopropp/>

¹⁰<http://framenet.icsi.berkeley.edu/>

¹¹<http://www.teic.org/index.xml>

¹²<http://www.tc37sc4.org/>

son”) and her father (sometimes treated as two characters, resulting in a total of 8), the dispatcher, the hero and the false hero.

31 Functions. At the heart of *Morphology of the Folktale* is the introduction and detailed description of 31 “functions”, i.e. (mostly) actions which can be attributed to the dramatis personae of a folktale. According to Propp, every folktale consists of a subset of these 31 functions, arranged in one or more “moves”. The order of the functions is fixed, with a number of scrupulously defined variations. Functions are frequently divided into sub-functions: In the case of function *A: Villainy*, they range from *A(1): The villain abducts a person* to *A(19): The villain declares war*¹³.

A sequence of all the functions from one folktale is called a “scheme” and can be used as a formal representation of the tale (see Fig. 1 for an example).

$$\gamma^1 \beta^1 \delta^1 A^1 C \uparrow \{ [DE^n eg.Fneg.]^3 d^7 E^7 F^9 \} G^4 K^1 \downarrow \\ [Pr^1 D^1 E^1 F^9 = Rs^4]^3$$

Figure 1: Functional scheme for *The Magic Swan-Geese*

150 Elements. In Appendix I of *Morphology of the Folktale*, Propp provides what he calls a “list of all the elements of the fairy tale”. The list contains 150 elements, distributed over six tables:

1. The Initial Situation
2. The Preparatory Section
3. The Complication
4. The Donors
5. From the Entry of the Helper to the End of the First Move
6. Beginning of the Second Move

Some of the 150 elements appear alone, others are grouped under a descriptive heading. If these “element clusters” (as shown in Fig. 2) are counted as one, the appendix contains

¹³It is those subfunctions, which introduce “arguments” to the functions and which contain some linguistic material, that led us to think that a link to the FrameNet resources might be very productive. We also think that the limited linguistic material described by Propp, as well as the linguistic information that can be extracted from our fine-grained annotation of the tales, can “feed” the ProppOnto ontologies with some concrete linguistic information to be associated with their classes. This can facilitate or advance the automated semantic annotation of folktales. We are working here on applying the strategies on combining domain ontologies and complex linguistic information described in the MONNET (Multilingual Ontologies for Networked Knowledge) project to the field of folktales. See also [2] or http://cordis.europa.eu/fp7/ict/language technologies/project-monnet_en.html

56 - as they shall tentatively be called in the following - narratemes¹⁴.

About a third of the narratemes can be mapped directly to functions, such as the aforementioned 30-32. *Violation of an interdiction*. Other narratemes can be combined to form an equivalent to a function (together, narratemes 71-77: *Donors* and 78: *Preparation for the transmission of a magical agent* can presumably be considered as a superset to the information expressed by function *D: First Function of the donor*.

- 30-32. Violation of an interdiction
30. person performing
 31. form of violation
 32. motivation

Figure 2: Example for a narrateme

Another group of narratemes, however, goes beyond the 31 functions: 70. *Journey from home to the donor*, for example, can be seen as filling the gap between the functions \uparrow : *Departure* and *D: First Function of the donor*. The first table (*The Initial Situation*¹⁵) contains a multitude of narratemes dedicated to the circumstances of the hero’s birth and other events/situations which precede the actual adventure.

Furthermore, Table 1 (*The Initial Situation*) includes two “element-clusters”¹⁶ describing the hero and false hero, respectively (see Fig. 3). A closer examination of the appendix reveals such “profiles” for each of the dramatis personae, although sometimes spread over several element clusters.

- 10-15. The future hero
10. nomenclature; sex
 11. rapid growth
 12. connection with hearth, ashes
 13. spiritual qualities
 14. mischievousness
 15. other qualities

Figure 3: Example for an element cluster serving as profile for a character

In *Morphology of the Folktale*, Propp provides an analysis of “The Magic Swan-Geese”, resulting in the scheme shown in Fig. 1 above. It is important to note here that the analysis does not only make use of functions but also of a “list of all the elements of the fairy tale” (given in Appendix I of *Morphology of the Folktale*). For example, Propp annotates the first Donor section from “The Magic Swan-Geese” as shown in the example below:

¹⁴The comment we made in footnote 13 is valid here too

¹⁵Propp makes use of the symbol α : *Initial Situation* to refer to everything that happens before the hero’s parents announce their departure, but it is not a function as such.

¹⁶We suspect that the term “narrateme” may not be applicable to them

She ran and ran until she came upon a a stove.

71, 73

“Stove, stove, tell me: where have the geese flown?”

“If you eat my little rye-cake, I’ll tell.” 76, 78b

“Oh, we don’t even eat cakes made of wheat
in my father’s house.” E¹ neg

where

71 = manner of inclusion into the tale

73 = physical appearance

76 = dialogue with the hero

78b = preparation for the transmission of
a magical agent: request

E¹ neg = the hero does not withstand

a test (insolent answer)

Table 1: Key to Propp’s annotation of *The Magic Swan-Geese*

2.2 PftML

PftML transforms the grammar-like functions, subfunctions and the rules concerning their combination from *Morphology of the Folktale* into a DTD. PftML allows for inline, usually sentence or paragraph-wise XML annotation of fairy tales, as we can see below in the small excerpt of the PftML annotation of *The Magic Swan-Geese* with Proppian functions.

```
<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>
```

The Proppian rules regarding the ways in which functions may be combined are reflected by the DTD design. See, for example, the element `CommandExecution`, which must contain one element of the type `Command` and one `Execution` to make sure that a violation of an interdiction is preceded by the corresponding interdiction. However, this occasionally leads to a lack in flexibility and may bring about unwanted side-effects. Although it is clear from the text that the parents absent themselves from the scene, the tight connection between the interdiction and its violation does not allow in PftML for the function `Absentation subtype="Elders"`, which should have its place between the two, to be marked up.

Also, we have acknowledged before that relying solely on the 31 functions will not allow us to analyze tales to the extent we desire. Seeing that PftML does not go beyond the functions, we will need to find ways to include more

information in PftML - or, as the case may be with APftML, to include PftML in an annotation schema affording more detailed markup on various levels.

3. APFTML

Looking at the annotated excerpt from Propp above, we came to two important findings: Firstly, Propp himself clearly did not limit himself to the 31 functions, but used individual “appendix-elements” as he saw fit. Secondly, although only functions will eventually find their way into a folk-tale’s *scheme*, a deeper analysis of the tale will benefit immensely from the more fine-grained analysis (also at the sub-sentential level) in term of a combination of functions and appendix-elements.

The actual work on APftML¹⁷ is not limited to this extension, but integrates the fairy tale annotation into textual and linguistic annotation standards, like TEI and ISO 37/SC4. For the sake of brevity, we cannot display the full actual annotation here, but give an example of both the TEI and our extension of PftML¹⁸ in the following:

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
xmlns:ht="http://www.w3.org/1999/xhtml">
<teiHeader>
...
<revisionDesc>
<change when="2010-06-16">
Tentative Annotation
</change>
</revisionDesc>
</teiHeader>
<text>
<front>
<docAuthor>
Alexander Afanasiev</docAuthor>
<docTitle>
<titlePart>Die Wilden Schwaene
</titlePart>
</docTitle>
</front>
<body>
<p>
<w xml:id="t1">Es</w>
<w xml:id="t2">war</w>
<w xml:id="t3">einmal</w>
...
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
xmlns:ht="http://www.w3.org/1999/xhtml">
<teiHeader>
<fileDesc>
<titleStmt>
<title>Die Wilden Schwaene</title>
<respStmt>
<resp>collector</resp>
<persName>Alexander Afanasiev</persName>
</respStmt>
</titleStmt>
<publicationStmt>
```

¹⁷The schema and an annotation example (The Magic Swan Geese) are available at <http://www.coli.uni-saarland.de/~ascheidel/APftML.xsd> and <http://www.coli.uni-saarland.de/~ascheidel/APftML.xml>

¹⁸To maintain readability, we include redundant information in our example and show the inline equivalent to the future stand-off annotation

```

    <p>http://www.maerchen-sammlung.de/
    Russische%20M%C3%A4rchen_16/
    Wilde-Schwaene_424.html</p>
  </publicationStmt>
  <sourceDesc/>
</fileDesc>
<revisionDesc>
  <change when="2010-06-16">Tentative Annotation
  </change>
</revisionDesc>
</teiHeader>
<text>
  <front>
    <docAuthor>Alexander Afanasiev</docAuthor>
    <docTitle>
      <titlePart>Die Wilden Schw d'ne</titlePart>
    </docTitle>
  </front>
  <body>
    <p>
      <w xml:id="t1">Es</w>
      <w xml:id="t2">war</w>
      <w xml:id="t3">einmal</w>
      ...
      <w xml:id="t36">Pass</w>
      <w xml:id="t37">gut</w>
      <w xml:id="t38">auf</w>
      <w xml:id="t39">Dein</w>
      <w xml:id="t40">kleines</w>
      <w xml:id="t41">Bruederchen</w>
      <w xml:id="t42">auf</w>
      <w xml:id="t43">und</w>
      <w xml:id="t44">spielt</w>
      <w xml:id="t45">nur</w>
      <w xml:id="t46">auf</w>
      <w xml:id="t47">dem</w>
      <w xml:id="t48">Hof</w>
      <w xml:id="t49">.</w>
      ...
    <Narrateme>
      <Command subtype="interdiction" id="i0">
        Eines Tages sprach die Mutter: Tochterchen,
        wir gehen jetzt auf die Arbeit.
        Pass gut auf Dein kleines Bruederchen auf
        und spielt nur auf dem Hof. Wir bringen Dir
        auch ein schoenes buntes Tuechlein mit.
      </Command>
      <Agent id="p1">die Mutter</Agent>
      <Patient id="p2">Tochterchen</Patient>
      <Content>Pass gut auf Dein kleines Bruederchen
        auf und spielt nur auf dem Hof.
      </Content>
      <Incentive>Wir bringen Dir auch ein schoenes
        buntes Tuechlein mit.</Incentive>
    </Narrateme>
  </Narrateme>
  <Narrateme>
    <Absentation>Als die Eltern gegangen waren
    </Absentation>
    <Agent id="p0, p1">die Eltern</Agent>
  </Narrateme>
  <CommandExecution subtype="violated"
  commandID="i0">
    setzte das Maedchen das kleine Bruederchen
    ins Gras vor dem Haus und lief auf die Strasse,
    um dort mit den anderen Kindern zu spielen.
  </CommandExecution>
  <Agent id="p2">das Maedchen</Agent>
  <Form> setzte das Maedchen das kleine
  Bruederchen ins Gras vor dem Haus und
  lief auf die Strasse</Form>
  <Motivation>um dort mit den anderen
  Kindern zu spielen</Motivation>
</Narrateme> ...

```

We plan also to integrate our work within the FrameNet-like approach to the annotation of semantic roles, since we encountered in the appendix of “Morphology of the Folktale” many descriptions that in fact refer to the semantic roles of lexical units, bearing a distinct resemblance to (FrameNet) frames. The Proppian function/functional narrateme *Interdiction*, for example has its counterpart in FrameNet, Frame “Deny permission”¹⁹ (see Table 2).

	Proppian “Frame”	FrameNet Frame
Name	Interdiction	Deny_permission
Agent role	person performing	Authority
Patient role	receiver of the interdiction (inferred)	Protagonist
Theme role	contents	Action

Table 2: Comparison of a Proppian “element cluster” and FrameNet Frame in regard to the respective definitions of typical semantic roles.

4. CONCLUSIONS

We described ongoing work in extending and partially re-designing an annotation scheme for fairy tales, which integrates both the full “descriptive” power of Vladimir Propp’s work and standards in textual and linguistic annotations, like TEI and ISO TC37/SC4. Examples of this annotation applied to two folk tales will be shown in detail in the poster presentation. As further step in our work, we foresee a multilingual extension, annotating a tale available in different languages (and versions), and an integration of the scheme within more generic semantic resources, like FrameNet and ontologies in the domain of narratives. A test case for the usefulness of our work will lie in the enhanced capability of providing automated comparative studies in the field of folktales.

5. ACKNOWLEDGMENTS

This work has been partially funded by the projects CLARIN & D-SPIN, especially for the linguistic annotation of tales, see <http://www.clarin.eu/external/> and <http://weblicht.sfs.uni-tuebingen.de/> and by the EU FP7 project MONNET – with grant 248458, especially for the topics related to multilingual ontologies, see http://cordis.europa.eu/fp7/ict/language/technologies/project-monnet_en.html

6. REFERENCES

- [1] A. Afanas’ev. *Russian fairy tales*. Pantheon Books, New York, 1945.
- [2] T. Declerck and P. Lendvai. Towards a standardized linguistic annotation of the textual content of labels in knowledge representation systems. In *LREC 2010- The seventh international conference on Language Resources and Evaluation*. ELRA, 2010.
- [3] V. Propp. *Morphology of the folktale*. University of Texas Press, Austin, 1968.

¹⁹http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=118&frame=Deny_permission