A Set of Annotations for supporting a TTS application for Folktales

Thierry Declerck
DFKI GmbH, Saarbrücken, Germany
Stuhlsatzenhausweg, 3
D-66123 Saarbrücken, Germany
E-mail: declerck@dfki.de

Abstract

In this short poster/demo submission we present different layers of annotation for folktales we have been working in the past and which are in the process of being integrated in one set of annotations, which is mediated by a formal representation of the annotation elements in an ontological framework. We list in this short text the various modules of this annotation framework. A main result of this work has been the implementation of a Text-to-Speech (TTS) application for folktales.

1. Introduction

In this short paper submission we present different layers of annotation for folktales we have been working in the past and which are in the process of being integrated in one set of annotations, which is mediated by a formal representation of the annotation elements in an ontological framework. We list in this short text the various modules of this annotation framework. A main result of this work has been the implementation of a Text-to-Speech (TTS) application for folktales.

2. Annotation of Proppian Functions

A first approach to the annotation of folktales was done in the context of cooperation between the European CLARIN project and the Dutch Amicus (Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts) . In this context, we developed an extended annotation scheme for the annotation of folktales with Proppian functions . The scheme includes textual properties, temporal structures, characters, dialog structures, and the Proppian functions (see [1] und [3]). This scheme was later used for supporting a first information extraction system applied to tales (see [2]), and comparing the results of this extraction with manually annotated tales.

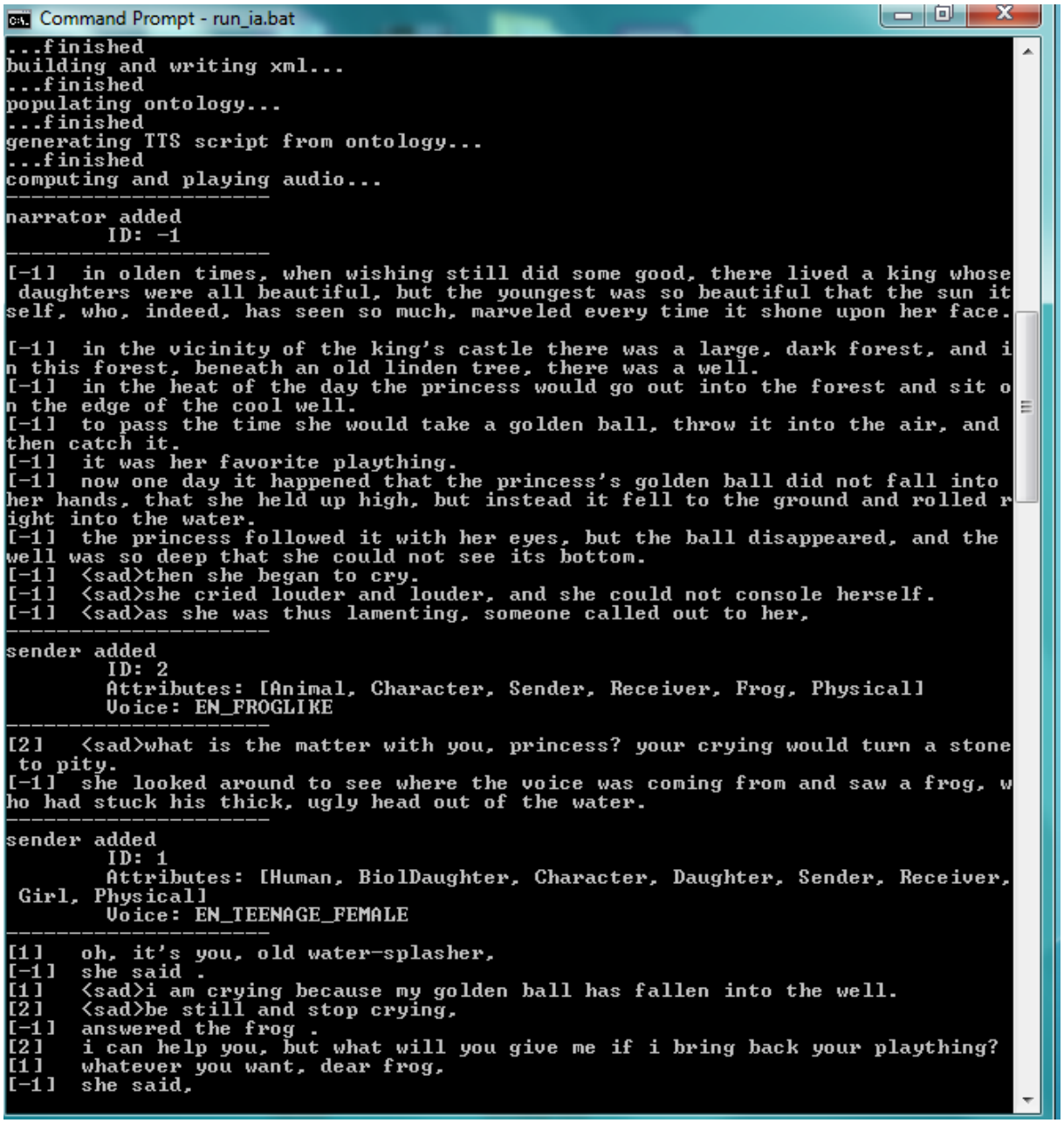
3. Ontology-driven automated Textual Analysis of Tales

Building on the work described in section 2, an automated linguistic analysis of tales was developed. The goal was not only to detect characters of the tales, but also to provide for a co-reference analysis such that the actions in which the characters are involved can be fully specified, and thus helping for an automated detection of Proppian functions, together with the involved personages. Results of the analysis are stored in a database, which has been further developed onto an ontological framework: Adding thus not only an annotation layer but also a formal representation level (s. [5] and [6]). The ontological representation allows also to apply generalisations for the specification of the characters (human vs animals, or supra natural etc.). The system was also able to operate reference resolution of the kind: “daughter” can also be a “sister” etc.

The move to the use of an ontological framework turned out to be very useful, since further work on distinct elements of a tales could be easily integrated. So for example the work described in [7] considered the detection of sentiments expressed by the characters of the tales. Such sentiments (“joy”, “happiness”, “sadness” etc.) could be added in a straightforward manner to instances of characters at the time span in which they occur in the tales.

4. Dialogue Structure Annotation for a TTS Output

In fact, the work in [7] mainly addresses the issue of adding Text to Speech (TTS) functionality to the automatic analysis of the text. The TTS system accesses the instances of the characters in the populated ontology and can retrieve the information on sentiment encoded there and correspondingly model the voice output of the various characters. The output of the ontology –driven automated annotations can be seen in the 2 screen shots displayed just below:


Command Prompt - run_ia.bat
...finished
building and writing xml...
...finished
populating ontology...
...finished
generating TTS script from ontology...
...finished
computing and playing audio...
narrator added
ID: -1
[-1] in olden times, when wishing still did some good, there lived a king whose daughters were all beautiful, but the youngest was so beautiful that the sun itself, who, indeed, has seen so much, marveled every time it shone upon her face.
[-1] in the vicinity of the king's castle there was a large, dark forest, and in this forest, beneath an old linden tree, there was a well.
[-1] in the heat of the day the princess would go out into the forest and sit on the edge of the cool well.
[-1] to pass the time she would take a golden ball, throw it into the air, and then catch it.
[-1] it was her favorite plaything.
[-1] now one day it happened that the princess's golden ball did not fall into her hands, that she held up high, but instead it fell to the ground and rolled right into the water.
[-1] the princess followed it with her eyes, but the ball disappeared, and the well was so deep that she could not see its bottom.
[-1] <sad>then she began to cry.
[-1] <sad>she cried louder and louder, and she could not console herself.
[-1] <sad>as she was thus lamenting, someone called out to her,
sender added
ID: 2
Attributes: [Animal, Character, Sender, Receiver, Frog, Physical]
Voice: EN_FROGLIKE
[2] <sad>what is the matter with you, princess? your crying would turn a stone to pity.
[-1] she looked around to see where the voice was coming from and saw a frog, who had stuck his thick, ugly head out of the water.
sender added
ID: 1
Attributes: [Human, BiolDaughter, Character, Daughter, Sender, Receiver, Girl, Physical]
Voice: EN_TEENAGE_FEMALE
[1] oh, it's you, old water-splasher,
[-1] she said .
[1] <sad>i am crying because my golden ball has fallen into the well.
[2] <sad>be still and stop crying,
[-1] answered the frog .
[2] i can help you, but what will you give me if i bring back your plaything?
[1] whatever you want, dear frog,
[-1] she said,

Very recently, we started also looking at other metadata to be used for annotating folktales, and to see how to integrate those with the Proppian functions. We looked for this at the well-known classification systems of Stith Thompson ([9]), Antti Aarne ([10]) und Hans-Jörg Uther ([11]) and we are starting to integrate those models in our ontology. Additionally we linked the detected characters to WordNet, investigating if this can help for the disambiguation of characters (s [12]). The resulting ontology from the Thompson Motif Index has been presented in ([13]).

As the most recent work we dedicated to the folktale topic was dealing with the implementation of running systems, less attention was given into the extension of the annotation scheme, work which is currently underway and which we would like to present and discuss in the workshop as a poster/demo.

References

[1] Thierry Declerck, Antonia Scheidel, Piroska Lendvai. Proppian Content Descriptors in an Integrated Annotation Schema for Fairy Tales. Language Technology for Cultural Heritage. Selected Papers from the LaTeCH Workshop Series, Theory and Applications of Natural Language Processing, Pages 155-169, Springer, Heidelberg, 2011

[2] Thierry Declerck, Antonia Scheidel. An Information Extraction Approach to the Semantic Annotation of Folktales. In: Sándor Darányi, Piroska Lendvai (eds.): First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts, Vienna, Austria, 10/2010

[3] Antonia Scheidel, Thierry Declerck. APftML - Augmented Proppian fairy tale Markup Language. In: Sándor Darányi, Piroska Lendvai (eds.): First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts, Vienna, Austria, 10/2010

[4] Thierry Declerck, Antonia Scheidel, Piroska Lendvai. Proppian Content Descriptors in an Augmented Annotation Schema for Fairy Tales. In: Caroline Sporleder, Kalliopi Zervanou (eds.): Proceedings of the ECAI 2010 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Lisbon, Portugal, IOS Press, European Coordinating Committee for Artificial Intelligence -- ECCAI, 8/2010

[5] Nikolina Koleva, Thierry Declerck, Hans-Ulrich Krieger. An Ontology-Based Iterative Text Processing Strategy for Detecting and Recognizing Characters in Folktales. In: Jan Christoph Meister (ed.): Digital Humanities 2012 Conference Abstracts, Pages 467-470, Hamburg.

[6] Thierry Declerck, Nikolina Koleva, Hans-Ulrich Krieger. Ontology-Based Incremental Annotation of Characters in Folktales. in: Kalliopi Zervanou, Antal van den Bosch, (eds.): Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2012) , Pages 30-35, Avignon, France, 4/2012

[7] Christian Eisenreich, Jana Ott, Tonio Süßdorf, Christian Willms, Thierry Declerck. From Tale to Speech: Ontology-based Emotion and Dialogue Annotation of Fairy Tales with a TTS Output Proceedings of ISWC 2014, Riva del Garda, Italy, Springer.

[8] Thierry Declerck. Annotationen für die automatisierte Verarbeitung von Märchen. In: Book of Abstracts, DHD 2015, Graz, Austria.

[9] Stith Thompson. 1977. The Folktale. Berkeley: University of California Press

[10] Antti Aarne. 1961. The Types of the Folktale: A Classification and Bibliography. The Finnish Academy of Science and Letters, Helsinki.

[11] Hans-Jörg Uther. 2004. The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson. FF Communications no. 284–286. Helsinki: Suomalainen Tiedeakatemia.

[12] Thierry Declerck, Tyler Klement, Antonia Kostova. Towards a WordNet based Classification of Actors in Folktales. In: Verginica Barbu Mititelu, Corina Forăscu, Christiane Fellbaum, Piek Vossen (eds.): Proceedings of the Eighth Global WordNet Conference, Bucharest, Romania, GWA, Global WordNet Association, 1/2016

[13] Antonia Kostova, Thierry Declerck. Ontologisierung vom Thompson Motif‘s Index. In: Book of Abstracts of DHD2016, Leipzig, Germany