Towards an Integrated Set of Annotations for Folktales

Thierry Declerck
declerck@dfki.de
German Research Center for Artificial Intelligence
University of Vienna, Austria

Introduction

In this poster paper we describe very briefly 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 integrated annotation scheme.

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,* sponsored by a grant from the Netherlands Organization for Scientific Research, NWO Humanities, as part of the Internationalization in the Humanities programme, from 2009-2012). 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 Declerck et al, 2011 and Scheidel and Declerck, 2010). This scheme was later used for supporting a first information extraction system applied to tales (see Declerck and Scheidel, 2010), and comparing the results of this extraction with manually annotated tales.

Building on this work, 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 (see Koleva et al, 2012, and Declerck et al, 2012). 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 Eisenrecih et al (2014) 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. In fact, the work in Eisenrecih et al (2014) 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 (an example using the tale "Frog Prince" can be heard, or preferably downloaded, online).

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 (1977), Antti Aarne (1961) and Hans-Jörg Uther (2004) and we are starting to integrate those models in our ontology. The resulting ontology from the Thompson Motif Index has been presented in Kostova and Declerck (2016). Additionally we linked the detected characters to WordNet, investigating if this can help for the disambiguation of characters (Declerck et al, 2016).

As the most recent work we dedicated to the folktale topic dealt with 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 present as a poster at DH 2017.

Bibliography

Declerck, T., Scheidel, A., Lendvai, P. (2011) 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*

Declerck, T., Scheidel, A. (2010) 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

Scheidel, A., Declerck, T. (2010) 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

Declerck, T., Scheidel, A., Lendvai, P. (2010) 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*

Koleva, N., Declerck, T., Krieger, H-U. (2012). 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.*

Declerck, T., Koleva, N., Krieger, H.-U. (2012). 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

Eisenreich, C., Ott, J., Süßdorf, T., Willms, C., Declerck, T. (2014). 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.

Declerck, T. (2015). **Annotationen für die automatisierte Verarbeitüng von Märchen. In:** *Book of Abstracts, DHD 2015***,** Graz, Austria.

Thompson, S. (1977). *The Folktale.* Berkeley: University of California Press

Aarne, A (1961). *The Types of the Folktale: A Classification and Bibliography.* The Finnish Academy of Science and Letters, Helsinki.

Uther, H-J. (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.

Declerck, T., Klement, T., Kostova, A. (2016). 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

Kostova, A., Declerck, T. (2016). Ontologisierung vom Thompson Motif's Index. In: *Book of Abstracts of DHD2016*, Leipzig, Germany