

# Roles, relations and references: Towards a computation-based distant reading of narrative-semantic roles in large datasets in Dutch

*Puck Wildschut MPhil*

PhD Researcher Radboud University Nijmegen, The Netherlands

Researcher-in-residence National Library of the Netherlands (KB), The Hague, The Netherlands

Corresponding address: p.wildschut@let.ru.nl (primary); Puck.Wildschut@KB.nl (secondary)

Since the rise of Russian Formalism in the early 19<sup>th</sup> century, literary theorists have been interested in finding ways to detect actants (characters) in narratives. The recent rise of computational methods within the humanities offers new ways of tackling this issue. As researcher-in-residence at the National Library of the Netherlands (KB) I am currently involved in a research project<sup>1</sup> that aims to develop a computation-based model for analyzing narrative-semantic roles in large datasets in Dutch. The premise of the project is that actantial roles are not only to be detected on the higher level of motive- and theme-building, but also at the linguistic level of semantic roles. Furthermore, the project aims to not only develop a tool for the detection of actantial roles, but also, and more importantly, for discovering the relationships between those roles as they are encoded in language.

There are three main issues to be dealt with during the tool-development process. First, existing research on semantic role labeling (SRL) almost exclusively focuses on English texts. Since the present projects' focus lies with Dutch, we need to revise existing SRL tools so that they work for Dutch. Some work has been done in this respect already, and features as inspiration for our own tool-development (Stevens, Monachesi, Van den Bosch 2007, Monachesi & Trapman 2006).

Second, the issues of co-reference resolution (Declerck, Koleva, Krieger 2012, Versley, Ponzetto, Poesio, Eidelman, Jern, Smith, Yang, Moschitti 2008, Lee et al. 2013) and animacy detection (Karsdorp, Van der Meulen, Meder, Van den Bosch 2015) needs to be addressed. As natural language processing (NLP) techniques, they are mutually dependent on each other, SRL and syntactic parsing (Idem). We therefore need to develop an integrated approach to actantial role modeling.

Third, and related to this point, the question of parsing comes into view. Automatic parsers are often very inaccurate or even non-existent for many languages (Idem). For Dutch, the automatic parser Alpino is the most advanced option. However, an uninformed approach that does not rely on dependency parsing, is a very real possibility in our tool-development project, following Karsdorp, Van der Meulen, Meder, Van den Bosch 2015. In addition to these main issues, questions of direct and indirect speech (Karsdorp, Van Kranenburg, Meder, Van den Bosch 2012), frame analysis (Reiter et al. 2014) and social interaction analysis (Karsdorp, Kestemont, Schöch, Van den Bosch 2015) play a role.

---

<sup>1</sup> This project is funded by the National Library of the Netherlands as part of their Digital Humanities program. As researcher-in-residence I am assisted by KB researcher Martijn Kleppe and collection-specialist Steven Claeysens, and KB programmers Willem Jan Faber and Juliëtte Lonij. For more information on the Researcher-in-residence project, please visit <http://blog.kbresearch.nl/category/digital-humanities/researcher-in-residence/>

The poster presentation will show the most up-to-date version of the tool we are developing at the KB and the preliminary results of its implementation. Special prominence will be given to how the issues mentioned are integrated in the tool's development. The poster aims to show how literary-linguistic theory and computational practice encourage each other in the development process.

## References

Declerck, Koleva, Krieger (2012) 'Ontology-based incremental annotation of characters in folktales'. *Proceedings of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Avignon, France, 24 April 2012. c 2012 Association for Computational Linguistics, , pp. 30–34.

Karsdorp, Van Kranenburg, Meder, Van den Bosch (2012) 'Casting a spell: Identification and ranking of actors in folktales'. *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, Francesco Mambrini, Marco Passarotti, and Caroline Sporleder (Eds.) Lisbon: Edições Colibri, pp. 39-50.

Karsdorp, Kestemont, Schöch, Van den Bosch (2015) 'The Love Equation: Computational modelling of romantic relationships in French Classical Drama'. *6th Workshop on Computational Models of Narrative (CMN-15) OpenAccess Series in Informatics*, Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard (Eds.) Dagstuhl: Dagstuhl Publishing, pp. 89–107.

Karsdorp, Van der Meulen, Meder, Van den Bosch (2015) 'Animacy detection in stories'. *6th Workshop on Computational Models of Narrative (CMN-15) OpenAccess Series in Informatics*, Mark A. Finlayson, Ben Miller, Antonio Lieto, and Remi Ronfard (Eds.) Dagstuhl: Dagstuhl Publishing, pp. 82–97.

Lee et al. (2013) 'Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules'. *Computational Linguistics*, v. 39, n. 4. Cambridge/London: MIT Press, pp. 885-916.

Monachesi & Trapman (2006) 'Merging FrameNet and PropBank in a corpus of written Dutch'. *Proceedings of the workshop Merging and layering linguistic information. Workshop held in conjunction with LREC 2006, Genoa - Italy, 23 May 2006*, pp. 32-39.

Reiter et al. (2014) 'An NLP-based cross-document approach to narrative structure discovery'. *Literary and Linguistic Computing*, v. 29, n. 4. Oxford: Oxford University Press, pp. 583-605.

Stevens, Monachesi, Van den Bosch (2007) 'A pilot study for semantic role labeling in a Dutch corpus'. *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*, P. Dirix, I. Schuurman, V. Vandeghinste, and F. Van Eynde (Eds.) Leuven, pp. 99-114.

Versley, Ponzetto, Poesio, Eidelman, Jern, Smith, Yang, Moschitti (2008) 'BART: a modular toolkit for coreference resolution'. *Proceedings of the ACL-08: HLT Demo Session (Companion Volume)*, pages, Columbus, June 2008. c 2008 Association for Computational Linguistics, pp. 9–12.