

Behind the Magic: using Word2Vec and Topic Modelling to analyse the Grimm Fairy Tales

In 1928, Vladimir Propp published his influential book *Morphology of the Folk Tale* in which he argues that the morphology of the folktale can be boiled down to 31 plot and 7 character functions. Propp's work has been interpreted as a model for the automated generation of stories, however, this entails implementing Propp's functions instead of testing his hypothesis. The latter is what I want to do. Using word2vec and topic modelling, as implemented in the Gensim library for Python, I want to ascertain the structures of the fairy tales as collected and edited by the Grimm Brothers in the *Kinder-und Hausmärchen* and see whether or not this structure corresponds to what Propp proposed in his book.

Ellen Vanderstraeten recently obtained her Master degree in Digital Text Analysis combined with English Literature from the University of Antwerp where she continues to work as a researcher. Employed at the Centre for Manuscript Genetics, she contributes to the Samuel Beckett Digital Manuscript project which uses XML and TEI encoding to digitize the manuscripts of Samuel Beckett.