

R E -  
T E L L  
- I N G

A Computational-  
Evolutionary  
Perspective

S T O -  
R I E S

Folger Karsdorp

# RETELLING STORIES

## A COMPUTATIONAL- EVOLUTIONARY PERSPECTIVE

FOLGERT KARSDORP

Cover design by Basten Stokhuyzen (<http://www.bstn.nl/>)

ISBN 978-90-9029993-8

© 2016, Folgert Karsdorp

Published under the terms of the Creative Commons Attribution License, CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, provided the original author and source are credited.

# RĚTELLING STORIES

## A COMPUTATIONAL- EVOLUTIONARY PERSPECTIVE

Proefschrift ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op

dinsdag 6 december 2016

om 12:30 uur precies

door

Folgert Bastiaan Karsdorp

geboren op 3 november 1983

te Leidschendam



## PROMOTOREN

Prof. dr. Antal van den Bosch

Prof. dr. Franciska de Jong (Erasmus Universiteit Rotterdam;  
Universiteit Twente; Universiteit Utrecht)

## COPROMOTOREN

Prof. dr. Theo Meder (Meertens Instituut, Amsterdam;  
Rijksuniversiteit Groningen)

Dr. Peter van Kranenburg (Meertens Instituut, Amsterdam)

## MANUSCRIPTCOMMISSIE

Prof. dr. Margot van Mulken

Prof. dr. Karina van Dalen-Oskam (Universiteit van Amsterdam;  
Huygens ING, Amsterdam)

Prof. dr. Tim Tangherlini (University of California,  
Los Angeles, Verenigde Staten)

Prof. dr. Helma van Lierop (Tilburg University)

Dr. Ben Miller (Georgia State University, Atlanta, Verenigde Staten)

*It gets easier. Every day it gets a little easier. But you got to do it every day. That's the hard part. But it does get easier.*

BOJACK HORSEMAN, *Out to Sea*



---

# CONTENTS

Acknowledgments

vii

1. Introduction

1

2. Motif Annotation & Classification

11

3. Character Bias in the Cultural Success of Fairy Tales

49

4. Gradual Accumulation of Modification

89

5. The Structure and Evolution of Story Networks

137

6. General Discussion

167

Bibliography

179

Samenvatting

202

Summary

206



---

## ACKNOWLEDGMENTS

Any story is shaped by the stories that surround it. This dissertation is no different. It could not have been told without the stories of all the wonderful people that have supported me during the past four and a half years.

First and foremost, I would like to thank my supervisor Prof. dr. Antal van den Bosch. During my MA training, I read Antal's book on memory-based learning, which was a paradigm-shifting experience. It sparked my interest in pursuing a Ph.D with a component in language technology. I was very fortunate to be offered a Ph.D position in the Tunes & Tales project, which not only allowed me to expand my knowledge about computational language modeling, but also to learn it from the best. Antal has my deepest gratitude for all his guidance, intellectual encouragement and much-needed advice on presentations, articles, grant proposals, job positions, and different revisions of the manuscript.

I thank my second supervisor Prof. dr. Franciska de Jong for shepherding the manuscript through its final stages. Furthermore, I am grateful to my two co-promoters, Prof. dr. Theo Meder and Dr. Peter van Kranenburg, from whom I benefited immeasurably. Both Theo's profound knowledge of folktales and oral culture, and Peter's invaluable suggestions at numerous occasions have improved the quality of this dissertation considerably.

A number of colleagues at the Meertens Instituut have been generous with their time and expertise: Berit Janssen, Marianne van Zuijlen, Dong Nguyen, Dolf Trieschnigg, and Iwe Muiser. The embedding of the Tunes & Tales project within the eHumanities group of the Royal Dutch Academy of Arts and Sciences has given me the opportunity to engage in interesting discussions regarding the emerging field of Digital Humanities, and I wish to thank Prof. dr. Sally Wyatt, Dr. Andrea Scharnhorst, Jeannette Haagsma, and Anja de Haas-Gruys for all their effort and time. Finally, I thank the LAMA's of the Radboud University

for encouraging me to leave Amsterdam once in a while.

I am enduringly grateful to my friends and family for their emotional support, patience and encouragement. I single out three close friends, whom I owe special thanks. First, I thank Basten Stokhuyzen, who implemented the front-end of MOMFER, designed the wonderful cover of this dissertation, and proved to be a steady friend. I am also greatly indebted to Marten van der Meulen, who has been my research assistant for approximately two years, has co-authored and co-presented several of my articles, and is one of the smartest and funniest persons I know. This dissertation has benefited repeatedly from Marten's wise and thoughtful suggestions, as well as from his enduring patience regarding my endless stream of trial projects. Finally, I thank Lauren Fonteyn, who has been an incisive reader and patient editor of different drafts of this dissertation. Without Lauren's intellectual breadth, critical judgment, and inspiring companionship, I doubt whether this dissertation had been completed at all.

---

## CHAPTER 1

# INTRODUCTION

*“To explain culture, then, is to explain why and how some ideas happen to be contagious.”*

DAN SPERBER, *Explaining Culture*

This study examines the mechanisms underlying the selection and transmission of stories from an evolutionary perspective. Stories, much like human beings, have lives, which start the moment a particular event or experience is formulated in narrative form. Their further ‘survival’ crucially depends on their being re-told: a story that does not acquire new tellers “can have no life beyond the life of the original person who experienced the events and first formulated them as a story”.<sup>1</sup> By being transmitted from one individual to another, stories may propagate, sometimes from generation to generation. Some stories propagate so effectively – or, to use Sperber’s wording, ‘contagiously’ – that they become part of a culture’s central heritage.

When stories are propagated, they tend to get altered and reshaped with each retelling. Experimental studies on story transmission have shown that factors such as memory capacity and other social-cognitive pressures can have severe impact on how accurately stories are transmitted.<sup>2</sup> In other words, stories are commonly subjected to ‘imperfect

1. Charlotte Linde, *Working with the Past: Narrative and Institutional Memory* (New York: Oxford University Press, 2009).

2. Frederic Bartlett, *Remembering. A Study in Experimental and Social Psychology*, Cambridge Books Online (London: Cambridge University Press, 1932); Alex Mesoudi et al., “A bias for social information in human cultural transmission,” *British Journal of Psychology* 97, no. 3 (2010): 405–423; Fritz Breithaupt et al., “Optimal Eventfulness of Narratives,” in *Proceedings of the Computational Models of Narrative Workshop*, vol. 45, OASICS (Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015), 12–22.



replication', and, as such, their transmission leads to their gradual transformation. Interestingly, while such experimental studies merely regard the reshaping of stories as a natural by-product of their being transmitted by multiple agents, some critics have argued that reshaping or transformation is in fact an essential prerequisite for a story's survival; a story *needs* to be recurrently adapted to secure its position in culture.<sup>3</sup> As Geerts and van den Bossche state: "Without all of the transformations, which keep the work available, it would no longer be read and therefore 'die out'".<sup>4</sup> The fact that stories need to be altered to survive even seems to be the case for fairy tales, some of which are among the most 'contagious' stories in present-day Western culture (e.g. "Cinderella" and "Snow White"). Notwithstanding the fact that many fairy tales have been kept in relatively stable form for centuries (at least on an abstract, structural level), the current study wishes to argue that even fairy tales undergo progressive alterations as the result of a process of modification in which changes are gradually accumulated.

Although the complex interplay between adaptation and selection in story transmission has been recognized by various scholars in literary and folklore studies,<sup>5</sup> these claims are, unfortunately, largely programmatic: the current accounts of adaptation and selection in story transmission posit thought-provoking theoretical claims, but they do not systematically nor quantitatively test these claims. Thus, if we wish to arrive at a more concrete and precise understanding of the selection forces that underlie the transmission of stories, we are faced with a challenge: we need to suggest and develop new ways to empirically and

3. John Stephens and Robyn McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature* (New York, London: Garland Publishing Inc., 1998); Fiona Collins and Jeremy Ridgman, eds., *Turning the page, Children's Literature in Performance and the Media* (Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang, 2006).

4. Sylvie Geerts and Sara van den Bossche, "Never-ending Stories. How Canonical Works Live on in Children's Literature," in *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*, ed. Sylvie Geerts and Sara van den Bossche, Ginkgo (Ghent, Belgium: Academia Press, 2014).

5. E.g. Brian Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction* (Cambridge, Massachusetts, London, England: The Belknap Press of Harvard University Press, 2009); Jack Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre* (New York, London: Routledge. Taylor & Francis Group, 2006); Jack Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre* (Princeton, Oxford: Princeton University Press, 2012); Julie Sanders, *Adaptation and Appropriation* (London, New York: Routledge, 2006); Linda Hutcheon, *A Theory of Adaptation* (New York, London: Routledge, 2013).

quantitatively verify such programmatic claims.

With the current study, I aim to tackle this challenge by drawing inspiration from the cultural evolution research program.<sup>6</sup> This program aims to enhance our understanding of cultural selection processes by means of formal models of cultural evolution, which allow us to quantitatively assess changes in cultural variation over time. In recent years, special attention has been devoted to a neutral model of cultural evolution which is comparable to biological drift. In this neutral model, individuals either introduce new variants of a particular cultural trait to the population or copy existing variants proportionally to the frequency of these variants in the population, and more popular variants are more likely to get adopted than less popular ones. This simple model accurately predicts a variety of cultural evolutionary processes, such as the selection of keywords in academic publications,<sup>7</sup> the choice of baby names,<sup>8</sup> and the popularity of particular dog breeds.<sup>9</sup> This random copying model, strikingly, appears to be so accurate that researchers consider it to be the null hypothesis in the description of cultural evolutionary processes. Put differently, it is held that cultural evolution proceeds randomly, unless proven otherwise. Formal models of cultural evolution have at least three major advantages over the programmatic statements that are often employed in literary and folklore studies, which are often based on intuition or research on a small corpus. First, to explain prevalent culture, formal models call for detailed and replicable descriptions and definitions of the mechanisms and processes that underlie cultural change. Second, by adopting a formal approach, researchers can attempt

6. Luigi Luca Cavalli-Sforza and Marcus Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*, Monographs in Population Biology 16 (Princeton, New Jersey: Princeton University Press, 1981); Robert Boyd and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985); Alex Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences* (Chicago, London: University of Chicago Press, 2011).

7. R. Alexander Bentley, "Random Drift versus Selection in Academic Vocabulary: An Evolutionary Analysis of Published Keywords," *PLoS ONE* 3, no. 8 (2008): e3057.

8. Matthew W. Hahn and R. Alexander Bentley, "Drift as a mechanism for cultural change: an example from baby names," *Proceedings of the Royal Society B: Biological Sciences* 270 (2003): S120–S123.

9. Harold A. Herzog et al., "Random drift and large shifts in popularity of dog breeds," *Proceedings of the Royal Society B: Biological Sciences* 271 (2004): S353–S356; Stefano Ghirlanda et al., "Dog Movie Stars and Dog Breed Popularity: A Case Study in Media Influence on Choice," *PLoS ONE* 9, no. 9 (2014): e106565–5.

to isolate and systematically compare forces of selection. Finally, one of the most important characteristics of quantitative, computational models is that they yield specific predictions that can be tested experimentally or against observations in real-world data.<sup>10</sup>

The predictions made by quantitative models, such as, for instance, those built by Cavalli-Sforza and Feldman,<sup>11</sup> have been supported by experimental studies of story transmission. In several of these studies, story transmission is experimentally simulated by means of a ‘transmission chain’ method, which closely resembles the game called ‘Chinese whispers’ or ‘telephone’ in the United States.<sup>12</sup> In these transmission simulations, participants attempt to memorize a story from either reading or listening and pass on their version of the story to the next participant generation. An interesting example of such an experimental study is the one by Eriksson and Coultas who describe an inhibiting effect of multiple ‘parents’ in transmission chains (i.e. chains in which stories are retold on the basis of more than one previous version) on the rate at which stories change.<sup>13</sup> Although experimental studies such as these have yielded a wealth of insights about underlying mechanisms of story transmission, a striking lack in cross-disciplinary communication and exchange of information with literary and folklore research prevents these insights to find their way across fields. Additionally, experimental story transmission studies have focused primarily on finding explanations for the progressive alterations of stories, and not on determinants of binary selection.<sup>14</sup> Put differently, while story transmission studies mainly seek to map out how one and the same story changes through transmission,

10. For an elaborate discussion, see Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

11. Cavalli-Sforza and Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*.

12. Bartlett, *Remembering. A Study in Experimental and Social Psychology*; Alex Mesoudi and Andrew Whiten, “The Multiple Roles of Cultural Transmission experiments in Understanding Human Cultural Evolution,” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, no. 1509 (2008): 3489–3501.

13. Kimmo Eriksson and Julie C. Coultas, “The Advantage of Multiple Cultural Parents in the Cultural Transmission of Stories,” *Evolution and Human Behavior* 33, no. 4 (2012): 251–259.

14. Notable exceptions include, e.g. Chip Heath et al., “Emotional Selection in Memes: The Case of Urban Legends,” *Journal of Personality and Social Psychology* 81, no. 6 (2001): 1028–1041; Kimmo Eriksson and Julie C. Coultas, “Corpses, maggots, poodles and rats: Emotional selection operating in three phases of cultural transmission of urban legends,” *Journal of Cognition and Culture* 14, nos. 1–2 (2014): 1–26.

they do not explain why a story — or in some cases a group or ‘assemblage’ of stories — is recurrently retold and becomes successful, whereas others go extinct altogether.<sup>15</sup> As such, it remains unclear how experimental results should be interpreted in light of real-world story transmission and selection.

One of the main obstacles to study the phenomena observed in real-world story transmission processes is the lack of detailed historical data that specifies the exact paths of transmission between different individuals and generations. The task of identifying such paths of story transmission would be easier to accomplish with diachronic data that describes who learns or copies from whom.<sup>16</sup> Unfortunately, although a number of folktale digitization initiatives which provide large-scale collections of folktales have been undertaken recently,<sup>17</sup> longitudinal data collections of real-world stories are still virtually non-existent. The current study aims to meet this demand by delivering a large diachronic collection of Dutch versions of the fairy tale “Little Red Riding Hood”.

The central question of this study is whether story transmission can be understood and hence should be described as a cultural evolutionary process, and, if so, which evolutionary mechanisms can be identified in story transmission processes. The main focus of the study are the mechanisms that underlie story transmission processes in folktales. Additionally, the goal is to unite the study’s findings, whether they are in line with or contrary to the predictions of evolutionary models, with existing literary and folklore accounts of story transmission and selection. As such, this study contributes to the further synthesis of the disciplines

15. Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*; Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*.

16. Anne Kandler and Adam Powell, “Inferring Learning Strategies from Cultural Frequency Data,” in *Learning Strategies and Cultural Evolution during the Palaeolithic*, ed. Alex Mesoudi and Kenichi Aoki (Tokyo: Springer Japan, 2015), 85–101.

17. James Abello et al., “Computational folkloristics,” *Communications ACM* (New York, NY, USA) 55, no. 7 (2012): 60–70, <http://doi.acm.org/10.1145/2209249.2209267>; Kathryn A. La Barre and Carol L. Tilley, “The elusive tale: leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales,” *Journal of the American Society for Information Science and Technology* 63, no. 4 (2012): 687–701, ISSN: 1532-2890, <http://dx.doi.org/10.1002/asi.21710>; Theo Meder, “From a Dutch Folktale Database towards an International Folktale Database,” *Fabula* 51, nos. 1–2 (2010): 6–22; Theo Meder et al., “Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database,” *Journal of American Folklore* 129, no. 511 (2016): 78–96.

of cultural evolution, literature and folklore. The research's relevance ranges from providing answers to methodological questions regarding how and to what extent evolutionary mechanisms underlying real-world story transmission can be studied from real-world, historical corpus data, to offering more abstract, theoretical reflections on pre-textual relations between stories. In particular, I will address the following four research questions:

*Research question 1:* How should stories be formally represented in order to study real-world story transmission and selection with computational means?

*Research question 2:* Can historical corpus data be employed to reveal content biases in story transmission?

*Research question 3:* Can we describe the diachronic development of stories as a gradual accumulation of modifications and hence as an evolutionary process?

*Research question 4:* Can story transmission and selection be described as a random selection process, and what other mechanisms underlying story transmission and selection can be discerned from historical corpus data?

Research question 1 serves as the methodological base to further investigate the other questions. In the present study, I experiment with a number of different representations of stories, which all have potential advantages and disadvantages for studying story transmission. In particular, I consider the following representations: (i) motif representations (Chapter 2), (ii) semantic representations of the character cast of stories (Chapter 3), (iii) representations based on manual narratological text analysis (Chapter 4), and (iv) 'bag-of-words' representations (Chapter 5).

In folktale research, the relations between stories are often investigated by means of motif representations. Motifs, such as those found in the *Motif-Index* by Stith Thompson,<sup>18</sup> are considered to be the primary building blocks of stories and their constellation defines the tale type of a

18. Stith Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, revised and enlarged (Bloomington: Indiana University Press, 1955–1958).

story as categorized in folktale catalogs such as *The Types of International Folktales*.<sup>19</sup> The current study wishes to contribute to these folklorists' research endeavors by delivering two research instruments which allow researches to analyze large-scale folktale collections and the relations between folktales in terms of their motifs more efficiently.

The first research instrument is aimed at making existing collections of folklore motifs more accessible. To this end, more than fifty years after the first edition of Thompson's seminal *Motif-Index of Folk Literature*, I present an online search engine in Chapter 2, which is tailored to fully disclose the index digitally. This search engine, called MOMFER, greatly enhances the searchability of the *Motif-Index* and provides exciting new ways to explore the collection. This is enabled by the use of modern techniques from the fields of Natural Language Processing and Information Retrieval. The key feature of the search tool is the way in which it allows users to search the *Motif-Index* for semantic concepts, such as 'mythical animals', 'mortality', or 'emotions'. In the first part of Chapter 2, I will explain the motivations for creating the search tool, lay bare its production process, and show in a number of case studies how the search tool can be used to explore the index in innovative ways. The second part of Chapter 2 presents an automated motif identification system. This second research instrument enables scholars to efficiently identify motifs in large-scale folktale collections as well as the relations between stories in terms of their motif representations.

Chapters 3, 4, and 5 more directly address the evolutionary questions of this study. Chapter 3 aims to further our knowledge about how to explain prevalent culture, in which cultural artifacts are differentially preferred and some artifacts are more likely to survive than others. As a case study, I systematically examine the cultural successfulness of fairy tales from the Brothers Grimm collection *Kinder- und Hausmärchen* from a *content*-based perspective (research question 2). I address of which story elements contribute to a story's popularity (or, in other words, which story elements form attractors causing a fairy tale to 'stick' and gain popularity). In particular, I look into the question whether the type of characters in a story correlates with its successfulness, i.e. whether

19. Hans-Jörg Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*, FF Communications, 284–286 (Helsinki: Suomalainen Tiedeakatemia, 2004).

a character type bias is at play in story selection. An essential methodological prerequisite for addressing this question is to find a way to list the character cast of a story. To this end, I first present a linguistically uninformed computational model for animacy classification with which characters can be automatically identified in folktales. I compare the model to a number of linguistically *informed* models that use features such as dependency tags, and show competitive results. The animacy classifier serves as the methodological base to further investigate the question of whether the cultural success of fairy tales can be (partially) ascribed to character type biases. I apply the animacy classifier to a large collection of Dutch folktales, to develop a typology of character types. By mapping the character casts of a selection of successful and unsuccessful fairy tales from *Kinder- und Hausmärchen* to this typology, I provide empirical evidence for several character types that discriminate between successful and unsuccessful tales.

The remaining two chapters (4 and 5) focus on socially informed, *context*-based processes that underlie story transmission and selection. Chapter 4 aims to enhance the understanding of the processes through which stories for children are retold. The study addresses the third research question of this thesis and investigates whether the diachronic development of the world's most famous fairy tale "Red Riding Hood" is to be characterized as a process of gradual accumulation of modification, with new variants of the story most likely deriving from previous versions in temporal proximity. Results suggest that new Dutch versions of "Red Riding Hood" are potentially subjected to age-dependent selection processes (i.e. fashion trends or fads), which means that 'young' story versions are preferred in producing a new version. The analysis presented here is based on a digitized collection of Dutch "Red Riding Hood" retellings.<sup>20</sup> This collection forms the largest longitudinal collection of children's stories today and holds a wealth of information for studies in (children's) literature, history, sociology, folklore, linguistics, and cultural evolution.

Finally, further building upon insights and experimental results from Chapter 4, I pursue a more detailed understanding of the processes through which stories are retold in Chapter 5 (research question 4). A

20. Folger Karsdorp, *Story network data sets*, 2016, doi:10.5281/zenodo.51588, <http://dx.doi.org/10.5281/zenodo.51588>.

collection of story retellings can be considered as a network of stories, in which links between stories represent pre-textual (or ancestral) relationships. In this chapter, I provide a mechanistic understanding of the structure and evolution of such story networks. I construct story networks for two diachronic story collections: (i) the collection of Dutch “Red Riding Hood” retellings from Chapter 4 and (ii) a corpus of paper chain letters. The results confirm and strengthen the analysis presented in Chapter 4 by showing that the formation of these story networks is guided by age-dependent selection processes. Subsequently, I systematically compare these findings with and among predictions of various formal models of network growth to determine more precisely which kinds of attractiveness are also at play or might even be preferred as explicatory models. Carefully studying the structure and evolution of the two story networks, I show that existing stories are differentially preferred to function as a new version’s pre-text given three types of attractiveness: (i) frequency-based and (ii) model-based attractiveness which (iii) decays over time.

Earlier versions of parts of this thesis have appeared as journal articles and papers in conference proceedings. Although much of this material has been substantively extended and revised for the current study, I feel that a study about story transmission should at least make its sources explicit:

Karsdorp, Folgert. “Het is groen en leeft nog lang en gelukkig. Classificatie van volksverhaalgenres op basis van formules.” *Tijdschrift voor Nederlands Taal- en Letterkunde* 129, no. 4 (2013): 274–288.

———. *Story network data sets.*, 2016. doi:10.5281/zenodo.51588. <http://dx.doi.org/10.5281/zenodo.51588>.

Karsdorp, Folgert, and Antal van den Bosch. “The Structure and Evolution of Story Networks.” *Royal Society Open Science* 3 (2016): 160071. <http://dx.doi.org/10.1098/rsos.160071>.

Karsdorp, Folgert, Peter van Kranenburg, Theo Meder, Dolf Trieschnigg, and Antal van den Bosch. “In search of an appropriate abstraction level for motif annotations.” In *Proceedings of the 2012 Computational Models of Narrative Workshop*, 22–26. Turkey, Istanbul, 2012.



- Karsdorp, Folgert, Marten van der Meulen, Theo Meder, and Antal van den Bosch. "Animacy Detection in Stories." In *6th Workshop on Computational Models of Narrative (CMN 2015)*, 45:82–97. OpenAccess Series in Informatics (OASICS). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015. doi:<http://dx.doi.org/10.4230/OASICS.CMN.2015.82>.
- . "MOMFER: A Search Engine of Thompson's Motif-Index of Folk Literature." *Folklore* 126, no. 1 (2015): 37–52. doi:[10.1080/0015587X.2015.1006954](https://doi.org/10.1080/0015587X.2015.1006954).
- Karsdorp, Folgert, and Antal van den Bosch. "Identifying motifs in folktales using topic models." In *Proceedings of BENELEARN 2013*, 41–49. Nijmegen, The Netherlands, 2013.
- Karsdorp, Folgert, Peter Van Kranenburg, Theo Meder, and Antal van den Bosch. "Casting a Spell: Identification and Ranking of Actors in Folktales." In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, 39–50. 2012.

---

## CHAPTER 2

# MOTIF ANNOTATION & CLASSIFICATION

*“A motif is the smallest element in a tale having a power to persist in tradition.”*

STITH THOMPSON, *The Folktale*

### 2.1 INTRODUCTION

In the introduction to his seminal *Motif-Index of Folk Literature* (TMI), Stith Thompson addressed a number of fundamental challenges in folklore research. One crucial problem concerns “the need for a comprehensive classification of the materials in all kinds of traditional narrative”,<sup>1</sup> which became all the more urgent in light of the ever-growing number of tales that were being collected at the time. Furthermore, he felt that the then current type index by Antti Aarne<sup>2</sup> was not sufficient for indexing folktales from outside Europe. Thompson took it upon himself to try to remedy this situation by classifying folktales on the basis of motifs, which resulted in the first edition of the TMI in the 1930s<sup>3</sup> and a revised and enlarged second edition in the 1950s that nearly doubled its scope.<sup>4</sup>

In folktale research, motifs are a key concept in the classification of folktales into tale types. In the authoritative folktale type catalog,

1. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, 8.

2. Antti Aarne, *The Types of the Folktale*, vol. 74, Folklore Fellows Communications (Helsinki: Academia Scientarium Fennica, 1928).

3. Stith Thompson, *Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folk-Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends* (Bloomington: Indiana University Press, 1932–1936).

4. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

*The Types of International Folktales*<sup>5</sup> – a project initiated by Aarne in 1910 in his *Verzeichnis der Märchentypen* (Catalogue of tale types),<sup>6</sup> which was translated and revised by Thompson<sup>7</sup> – motifs form the primary descriptive units, and their configuration defines the folktale type of a story. There is, however, a large body of criticism aimed at the TMI as well as the different editions of the folktale type index, such as, for example, Dundes in relation to Aarne's 1961 edition,<sup>8</sup> and Karsdorp et al. with respect to Uther's most recent edition of the *The Types of International Folktales*.<sup>9</sup> This criticism has focused on the distribution and overlap of tale types and motifs. Furthermore, El-Shamy addresses the lacunae in the collection of sources used.<sup>10</sup> Despite the various problems with respect to the applicability of the Aarne/Thompson folktale index to non-European tales, El-Shamy states that "these can be adequately addressed through its [the folktale index's] relatively open classification system, which allows for adding new items, particularly when compared to other systems".<sup>11</sup> Additionally, Alan Dundes states:

It must be said at the outset that the six-volume *Motif-Index of Folk-Literature* and the Aarne-Thompson tale type index constitute two of the most valuable tools in the professional folklorist's arsenal of aids for analysis. This is so regardless of any legitimate criticisms of these two remarkable indices, the use of which serves to distinguish scholarly studies of folk narrative from those carried out by a host of amateurs and dilettantes. The identification of folk narratives through motif and/or tale type numbers has become an international *sine qua non* among bona fide folklorists.<sup>12</sup>

5. Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*.

6. Antti Aarne, *Verzeichnis Der Märchentypen [Catalogue of folktale types]*, vol. 3, Folklore Fellows Communications (Helsinki: Academia Scientiarum Fennica, 1910).

7. Aarne, *The Types of the Folktale*; Antti Aarne and Stith Thompson, *The Types of the Folktale: A Classification and Bibliography*, vol. 184, FF communications (Helsinki: Academia Scientiarum Fennica, 1961).

8. Alan Dundes, "The Motif-Index and the Tale Type Index: A Critique," *Journal of Folklore Research* 34, no. 6 (1997): 195–202.

9. Folger Karsdorp et al., "In search of an appropriate abstraction level for motif annotations," in *Proceedings of the 2012 Computational Models of Narrative Workshop* (Turkey, Istanbul, 2012), 22–26.

10. Hasan El-Shamy, *Folktales of Egypt* (Chicago: University of Chicago Press, 1980).

11. Hasan El-Shamy, "A Type Index for Tales of the Arab World," *Fabula* 29, nos. 1/2 (1988): 158.

12. Dundes, "The Motif-Index and the Tale Type Index: A Critique," 195.

While many new motif indices<sup>13</sup> as well as many new collections of folktales<sup>14</sup> have since been put forward, relatively few attempts have been made to combine the two.<sup>15</sup> I hypothesize that the lack of combining the two resources is related to a lack of navigability of the TMI. Furthermore, while many of the researchers explicitly integrate their new indices into the TMI,<sup>16</sup> there is no complete and combined edition of all these indices. Because of the problems that printing a book of such magnitude would generate, it is understandable that such an edition is unavailable. These considerations would evaporate for a digital edition. Yet, however valuable a research tool it would be, no digital edition exists at present. Given the TMI's unquestioned status as *the* source for the worldwide coverage of folk narrative, a search engine of the TMI would be of interest to many scholars, ranging from folklorists to narratologists and story generation researchers.

The aims of the present chapter are the following. First, I wish to highlight some of the navigational challenges one has to overcome in order to make use of existing motif indices successfully, taking the TMI as a case study. I agree with Harriet Goldberg that "[t]he utility of an index of folk-motifs clearly lies in its ability to present in an orderly

13. See e.g. Ernest W. Baughman, *Type and Motif-Index of the Folktales of England and North America*, vol. 20, Indiana University Folklore Series (The Hague: Mouton, 1966); Hireko Ikeda, *A Type and Motif Index of Japanese Folk-Literature*, vol. 209, Folklore Fellows Communications (Helsinki: Academia Scientarium Fennica, 1971); Basil F. Kirtley, *A Motif Index of Traditional Polynesian Narratives* (Honolulu: University of Hawai'i Press, 1977); James Wesley Childers, *Tales from Spanish Picaresque Novels: A Motif Index* (New York: State University of New York Press, 1977); Helen Leneva Flowers, *A Classification of the Folk Tale of the West Indies by Types and Motifs* (New York: Arno, 1980); Lena Neuland, *Motif-Index of Latvian Folktales and Legends*, vol. 229, Folklore Fellows Communications (Helsinki: Academia Scientarium Fennica, 1981); Hasan El-Shamy, *Folk Traditions of the Arab World: A Guide to Motif Classification* (Bloomington: Indiana University Press, 1995).

14. El-Shamy, *Folktales of Egypt*; Theo Meder, *De Magische Vlucht. Nederlandse Volksverhalen uit de Collectie van het Meertens Instituut [The magic flight. Dutch folktales from the collection of the Meertens Institute]* (Amsterdam: Uitgeverij Bert Bakker, 2000); Thomas H. Slone, *One Thousand One Papua New Guinean Nights. Folktales from Wantok Newspaper*, vol. 1: Tales from 1972–1985 (Oakland, California: Masalai, 2001); William Crooke and Pandit Pam Gharib Chaube, *Folktales from Northern India* (Santa Barbara, California: ABC-CLIO, 2002).

15. Although see, in recent years Dan Ben-Amos, *Folktales of the Jews. 3 vols.* (Philadelphia: Jewish Publication Society, 2006); Slone, *One Thousand One Papua New Guinean Nights. Folktales from Wantok Newspaper*; For further examples of both motif indices as well as folktale collections, see Hans-Jörg Uther, "Type- and Motif-Indices 1980–1995: An Inventory," *Asian Folklore Studies* 55, no. 2 (1996): 299–317.

16. For a brief and recent example, see Sita Bell, "Anti-Semitic Folklore Motif Index," Master's thesis (2009), <http://%20digitalcommons.usu.edu/etd/299>.

framework those transitory flashes of recognition that we experience upon hearing a familiar story or a familiar narrative component in a new context".<sup>17</sup> To meet this objective, I present *Meertens Online Motif FindER* (MOMFER), which is a free, online search engine designed to increase the accessibility of the TMI.<sup>18</sup> This tool provides various new and unique access points to the index, drawing inspiration from techniques used in the fields of Natural Language Processing and Information Retrieval.<sup>19</sup> MOMFER enables several different search options: besides basic options such as single word, multi-word, and phrase searches, it is also possible to use faceted search, and even semantic search. These options, I suggest, can provide new tools for researchers and can lead to the asking and answering of questions that were hitherto deemed unanswerable. Furthermore, there is great potential for this tool to be used as a framework into which all other existing motif indices will be integrated.

Second, one is confronted with the problem that, although the motif search engine MOMFER provides researchers with a indispensable research tool to more efficiently access and retrieve information from the TMI, it remains rather cumbersome and time-consuming to manually label folktales with their corresponding motifs. This is especially the case in large-scale collections such as the Dutch Folktale Database.<sup>20</sup> In order to reduce the amount of manual labor, then, I will develop a computational system with which motifs can be automatically assigned to folktales. If this automated motif classification system is combined with MOMFER, the analysis of large-scale folktale collections is greatly facilitated, which will aid and enhance future research aimed at mapping out relations between stories and their motifs.

The present chapter is structured as follows: first, I will briefly introduce the TMI to those readers who may be unfamiliar with it, which

17. Harriet Goldberg, *Motif-Index of Medieval Spanish Folk Narratives*, Medieval & Renaissance Texts & Studies (Tempe: AZ, 1998), xiii.

18. Available at <http://www.momfer.ml>

19. For a similar approach that attempts to enrich the motif index with semantic information, cf. Thierry Declerck and Piroska Lendvai, "Linguistic and Semantic Representation of the Thompson's Motif- Index of Folk-Literature," in *Research and Advanced Technology for Digital Libraries*, ed. S. Gradmann et al., vol. 6966, Lecture Notes in Computer Science (Berlin: Springer, 2011), 151–158.

20. Meder, "From a Dutch Folktale Database towards an International Folktale Database"; Meder et al., "Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database."

is followed by a discussion of some important navigational challenges associated with its use (Section 2.2.1). Subsequently, I will describe the creation process and architecture of MOMFER (Section 2.2.2), after which I will proceed to highlight a number of interesting search options by presenting several case studies that use the TMI as a corpus (Section 2.2.3). The second part of this chapter is devoted to the development of the automated system for motif identification (Section 2.3).

## 2.2 MOMFER

### 2.2.1 ACCESSIBILITY OF THE *MOTIF-INDEX*

The TMI<sup>21</sup> contains over forty-five thousand motifs spread out over five volumes. The motifs are hierarchically ordered in a tree structure. There are twenty-three top-level, alphabetically labeled categories, whose contents range from mythological motifs (Category A) to motifs concerning traits of character (Category W) and beyond. Each top-level category is divided into various subcategories: for example, Category D (*Magic*) is divided into, inter alia, *Transformation* (Do–D699) and *Magic Objects* (D800–D1699), which in turn branch out into child motifs, and so on and so forth. At the leaves of the tree, or terminal nodes, we find the most specific instances of a particular motif.<sup>22</sup> Figure 2.1 is an exemplary representation of part of the hierarchical tree.

As Thompson states himself, the motifs in the first five volumes are grouped together as the result “of a gradual evolution, not of any preconceived plan”.<sup>23</sup> This method (or lack thereof) has several potential negative effects. First, as can be seen from Figure 2.2, the distribution of the motifs over the main categories in the index is rather uneven, with categories A – K containing seventy-eight per cent of all motifs in the index. Second, the index lists seventy-two motifs twice under different headwords (i.e. 144 unique motifs). For example, the motif

21. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

22. For an elaboration on the build-up of the index and search strategies, see Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, Volume 1, 19–25; El-Shamy, *Folk Traditions of the Arab World: A Guide to Motif Classification*, 16–17.

23. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, 19.

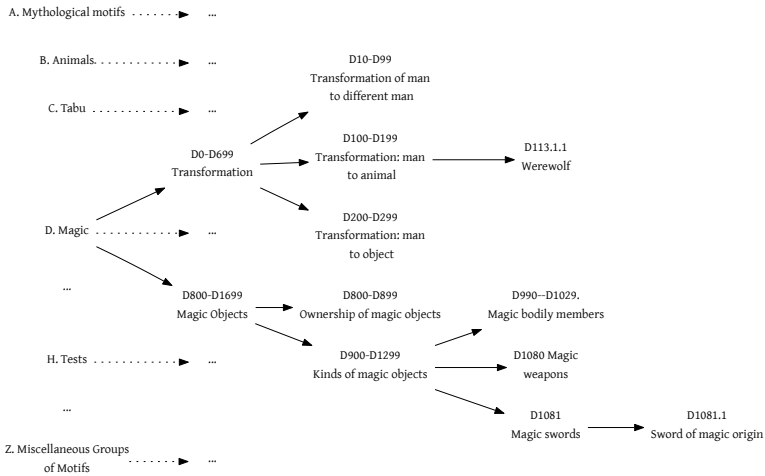


Figure 2.1: Partial view of Thompson's *Motif-Index of Folk Literature*.

*Transformation: utensil to person* has been indexed both as D436.1 and D434.1. Other examples include *One day and one night: object borrowed for a day and a night retained* indexed under K2314.1 and K232.2, and *Well shines at night*, which is categorized at two completely separate branches of the tree: *Magic* (D1645.9) and *Marvels* (F718.5). These cases represent the extreme endpoint on the scale of a more general problem that, as observed by Dundes, many motifs from different branches show a high degree of semantic similarity.<sup>24</sup>

Volume Six of the TMI consists of an alphabetical index to the other five volumes. This index provides a different entrance point to the motif index using an alphabetically sorted list of important terms that appear in the other five volumes. Looking under the term *mice*, for example, we find motifs such as the following (where the headword is left out):

- [mice] army saves kingdom from invasion K632.1;
- [mice] consecrate bishop (lie) X1226.1;
- [mice] and hogs let loose put elephant cavalry to flight K2351.3.

24. Dundes, "The Motif-Index and the Tale Type Index: A Critique."

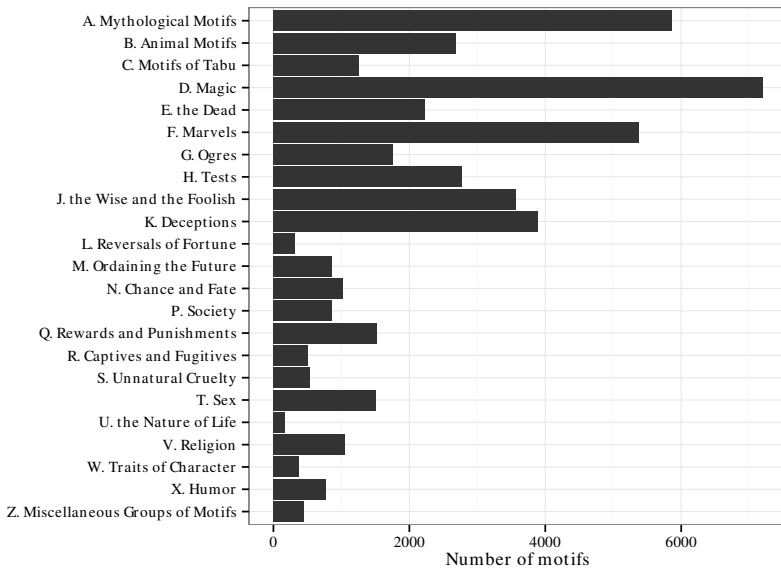


Figure 2.2: Distribution of motifs over main branches of the *Motif-Index of Folk Literature*.

The examples are ordered on different levels. The first level is an alphabetical ordering of the second word of the motif, where the headword is the first word. For instance, as the example above shows, a motif starting with *army* comes before one starting with *consecrate*. However, as can be seen in the same example, this structure is not completely consistent: Thompson ignored function words such as *and*, *with*, and *of*, so that after *consecrate* comes *and hogs*.

After this first set of motifs, a second set of motifs is listed that do not start with the headword but do contain it:

- army of m. B268.6;
- bargain with king of m. M244.1.

While this approach may at first glance seem quite unproblematic, it soon becomes troublesome when entries have a larger number of associated motifs. For example, the word *horse* has over five pages with motifs listed (including *horse's* and *horses*). Here it quickly becomes cumbersome to read through all the motifs in the hopes of finding the appropriate



one. Also, several stratagems have been undertaken by Thompson to save space: these include referring readers to a certain category (as for headword *devil*: “\*G303ff. See, in addition to the following, the extensive list of motifs assembled at G303”) and listing certain sets of motifs together without details pertaining to the content of the motifs (as for headword *cat*: “and witches D1766.4, G211.1.7, G224.11.12, G241.1.4, G225.3, G243.2.2, \*G252, G262.1.1, G262.3.2”). Thompson left room for new motifs – although his reasons for doing so were in some cases, such as the erotic and scatological dimensions, suspect<sup>25</sup> – and this room has been gratefully filled up by subsequent scholars. New motif indices are made compatible with the TMI by integrating newfound motifs into the hierarchical structure of the TMI. For example, Ernest W. Baughman<sup>26</sup> posits the motif *Transformation: man meets and copulates with female snake* as D191.2, as an addition to Thompson’s D191: *Transformation: man to serpent (snake)*. While this shows that a complete motif index is (at least theoretically) possible, the compilation of such an index far exceeds printing possibilities. Even Thompson himself was aware of this, and left out certain parts of folklore because “To have included these would have doubled the size of the index”,<sup>27</sup> and an online version integrating all subsequent motif indices has as yet not been undertaken. Nevertheless, such a complete work is something to be desired, as it would provide a more comprehensive starting point for comparing different narrative traditions, which was, ultimately, the goal of the TMI.<sup>28</sup>

Today, with most of our digital search results only a few clicks away, the time-consuming labor of manually searching through the TMI, hoping to stumble upon the entry we are looking for, appears rather outdated. In response to this, several initiatives have been undertaken to make a version of the TMI available in digital format, both offline and online.<sup>29</sup> The first example of this kind, as far as I am aware, is the CD-ROM edition of the TMI as published by Indiana University Press in 1993.<sup>30</sup> While this version allows users to enjoy some of the benefits

25. Gershon Legman, “Toward a Motif-Index of Erotic Humor,” *The Journal of American Folklore* 75, no. 297 (1962): 227–248.

26. Baughman, *Type and Motif-Index of the Folktales of England and North America*.

27. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, 11.

28. *Ibid.*, 9–10.

29. For an online example, see <http://www.ruthenia.ru/folklore/thompson/>.

30. Stith Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in*

of digital searches (e.g. 'Boolean searches'), it is not readily available, and it is safe to say that it is outdated, since its system requirements are "DOS 2.0 or higher".<sup>31</sup> By presenting the TMI in a digital format, users are able to search through the index using the search facilities provided by their web browser (such as the omnipresent 'Find' function). Although this seems to be quite an improvement over searching through the paper index, in reality the improvements are small: queries remain limited to single expressions and/or phrases. There is one online search engine available at [www.storysearch.symbolicstudies.org](http://www.storysearch.symbolicstudies.org). Unfortunately, this search engine is limited in its applications: searches seem to be confined to one word, and documentation on other options is unavailable. Also, it does not seem to be intended as a search tool, more as a source of inspiration for storytellers. In conclusion, the benefits of the currently available digital and online versions are rather small in comparison with the paper version.

### 2.2.2 MEERTENS ONLINE MOTIF FINDER (MOMFER)

In this section I will introduce MOMFER. First, I will describe the architecture of the search tool, including the sources used and some of the preprocessing steps that were performed. Subsequently, I continue with an in-depth description of the most salient search features implemented in the tool. Figure 2.3 is a representation of the web interface of MOMFER.

#### 2.2.2.1 Preprocessing Steps

As mentioned above, there are a few digitized versions of the TMI in existence. For this study, I constructed a full-text version of the motif index using the edition made available online at <http://www.ruthenia.ru/folklore/thompson/>. This digital edition is based on the revised and

*Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, CD-ROM (Bloomington: Indiana University Press, 1993).

31. Unfortunately, I was unable to acquire a copy of this edition, which is why I have to refrain from a more detailed comparison. However, the fact that no research institution or university library in the Netherlands has a copy available is a case in point. I base myself on a review of this digital edition by Allen Smith, "JAL Guide to Software, Courseware and CD-ROM: Stith Thompson's Motif-Index of Folk Literature," *Journal of Academic Librarianship* 20, no. 4 (1994): 255.

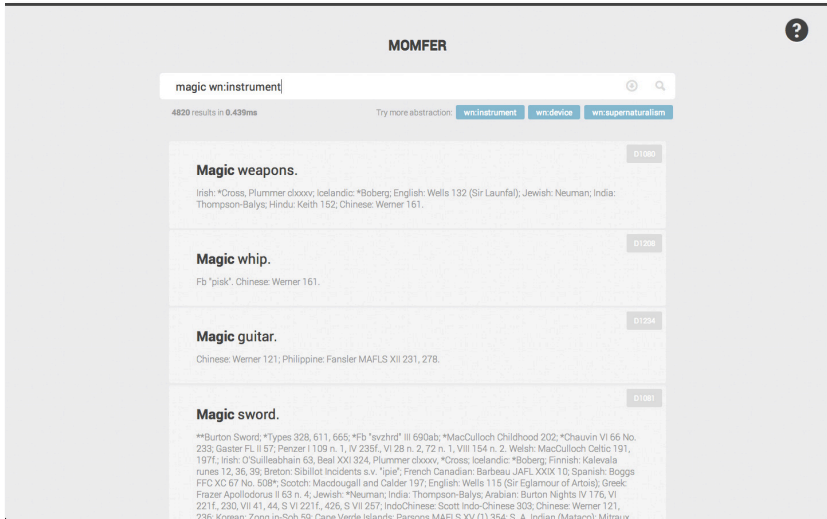


Figure 2.3: Web interface of the Meertens Online Motif FindER.

enlarged edition of the TMI.<sup>32</sup> This edition was subjected to a number of preprocessing steps, which will be described in order. After stripping all of the HTML tags from the digital index, the ‘Stanford CoreNLP toolkit’ (version 3.3.1)<sup>33</sup> was employed to tokenize all motifs into formally bounded sentences. The same toolkit was used to lemmatize and tag all words by part of speech (e.g. nouns, verbs, etc.) and to apply named entity recognition.

Although earlier writers have classified the motifs in the TMI as having a tripartite structure,<sup>34</sup> for the purposes of this study I interpret each motif in the TMI as consisting of four parts: (i) a unique key, denoting the place in the hierarchy (consisting of a upper-level letter and a lower-level number key), (ii) a primary description, (iii) a secondary description (which is optional), and (iv) bibliographical information. After preprocessing a motif string such as “C662. One must eat “death vegetable” whenever one sees it. Otherwise god will be angry. India:

32. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

33. Kristina Toutanova et al., “Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network,” in *Proceedings of HLT-NAACL 2003* (Stroudsburg, PA: Association for Computational Linguistics, 2003), <http://nlp.stanford.edu/kristina/papers/tagging.pdf>.

34. Notably El-Shamy, *Folk Traditions of the Arab World: A Guide to Motif Classification*.

Thompson-Balys.”, it is represented in JSON format as follows:

```
{
  "motif": "C662",
  "description": "One must eat \"death vegetable\" whenever
                 one sees it.",
  "lemmas": [
    "vegetable",
    "death"
  ],
  "locations": [
    "India"
  ],
  "additional_description": "Otherwise god will be angry.",
  "references": "India: Thompson-Balys."
}
```

Here *motif* represents the unique key as given by Thompson; *description* and *additional\_description* provide a verbal characterization of the motif, where the second part is usually an explication of the first part; *references* points to any references and bibliographic information (e.g. collector, collection, and/or location) available for this motif; *lemmas* stores the lemmatized versions of nouns in the description field; *locations* lists all mentions of geographical locations in the reference field. With the help of a number of programming scripts, I extracted these fields for all motifs in the index.<sup>35</sup>

#### 2.2.2.2 *Semantic Expansion*

One of the key features of our search engine is the way in which motifs are semantically expanded to match more generic descriptions. If a user issues a query such as *color animal*, the system should not only return ‘direct hits’ (motifs in which both words occur), but also motifs containing instances of these words with either a higher or a lower specificity, such as B731.2 *Green horse*. In order to do so, all lemmatized nouns and adjectives in the motif descriptions were matched against the semantic

35. The parsed index as well as the source code of MOMFER is available online at <https://github.com/fbkarsdorp/tmi>.

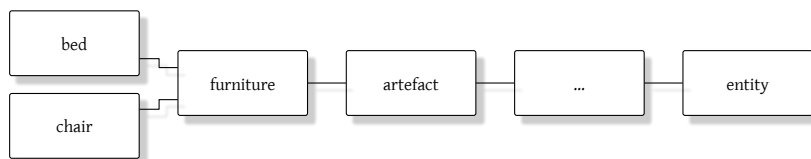


Figure 2.4: Example of hyperonymic and hyponymic relations in WordNet.

lexicon WordNet.<sup>36</sup> WordNet is a large semantic lexicon of English, in which words of various syntactic categories are grouped into sets of related words, or ‘synsets’.<sup>37</sup> The relations between different synsets are encoded by means of super-subordinate relations (or hyperonyms). Thus, a word like *furniture* is linked to both more specific examples (or hyponyms) such as *bed* and *chair* and more general concepts (or hyperonyms) such as *artefact*, and ultimately *entity*, which forms the root node of the network (cf. Figure 2.4). While it is possible to search for such far-removed parents, MOMFER by default returns only the two immediate parental relations, primarily for efficiency reasons and also because a complete expansion would return all motifs in every query, only in different orderings.

By linking motifs with words in other motifs on the basis of hyperonymic relations, the search engine provides a completely new way of accessing the TMI. This enables one to find many related motifs that are otherwise difficult to detect. To give a small example, say we are interested in the various occultists present in the TMI. The umbrella term *occultist* does not occur at all in the index, yet we know the index is packed with motifs about witches, wizards, sorcerers, druids, magicians, enchanters, and so forth. In WordNet all these words are connected through the hyperonym *occultist*. By connecting motifs based on their hyperonymic relations with other motifs, the system allows users to

36. Christiane Fellbaum, “WordNet(s),” in *Encyclopedia of Language and Linguistics*, 2nd ed., ed. Keith Brown (Oxford: Elsevier, 2005), 665–670.

37. While there has been some criticism of the usefulness and limitations of WordNet, our case studies show that, for the purposes of this research tool, synsets and WordNet are in fact valuable methods that improve the retrieval results, cf. Yorick Wilks et al., *Electronic Words: Dictionaries, Computers, and Meanings* (Cambridge, Massachusetts: MIT Press, 1996); Doug Lenat et al., “CYC, WordNet, and EDR: Critiques and Responses,” *Communications of the ACM* 38, no. 11 (1995): 33–38.

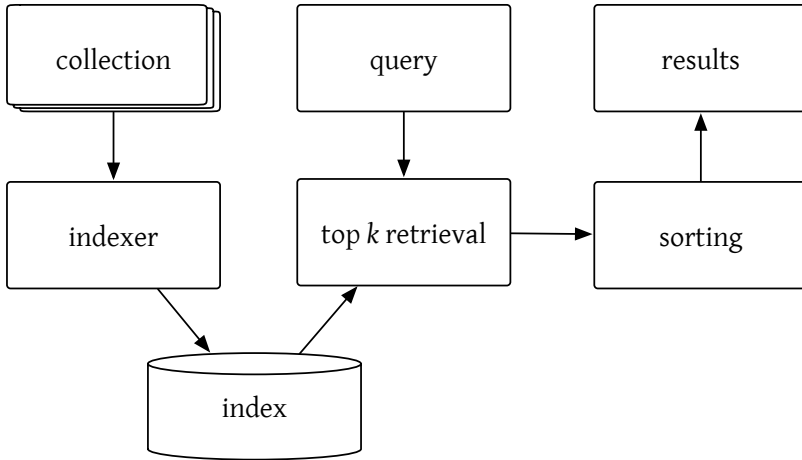


Figure 2.5: MOMFER's Search engine schema.

extract all types of occultists using the single search query `occultist`.<sup>38</sup>

### 2.2.2.3 Indexing Schema

Figure 2.5 provides a detailed view of the indexing schema. All motifs are preprocessed and fed to an indexer. This indexer extracts the various fields described above and stores the result in an index. When a user issues a search query, the index retrieves the top results which, after scoring and sorting, are presented to the user.<sup>39</sup>

The system assumes that not all matches are equally important. If, for example, a user enters the query `book`, motifs that contain the word `book` in the reference field are intuitively less relevant than those in which it is part of the primary description. To reflect this intuition, the system weighs the matches in the different fields in the following descending order: (i) primary description, (ii) additional information, (iii) WordNet expansion, and (iv) bibliographical information and location (references). Thus, a query for the word `book` will first return motifs where the word is

38. When linking the words from the TMI to the entries in WordNet, I did not perform any semantic disambiguation. Each word was linked to the most common sense in WordNet. In a later stage, this heuristic could be further refined.

39. I make use of the programming library *Whoosh*, which is a fast Python-based search engine library. See <http://whoosh.readthedocs.org/>.

found in the primary description and/or additional information, before it turns to motifs matching the WordNet expansion.

#### 2.2.2.4 Query System

MOMFER provides an expressive and rich query system, allowing users to retrieve their results efficiently. In this section I will describe the most important search features implemented in the tool.

Single word search is the simplest form of search, where users search for a single word. The index retrieves all motifs containing the search word and ranks them according to how informative the word is for that particular motif. If, for example, a user searches for the word *devil*, the motif G303 ('Devil') will be ranked highest because it contains no other words than the search term. Subsequently, if a motif mentions a word more than once, it will lead to a high ranking. For example, motif G303.9.4.7.1 (Devil and girl. 'Are you lonely?'; Girl: 'No, devil, with God and angels') has *devil* in both the primary description and in the additional information, leading to a high ranking. As explained above, the system will also look for the two immediate WordNet expansions of the search term and their daughter terms.

Another search feature is multi-word search. By default all queries are represented as Boolean OR statements. This means that when a user searches for two words, for example *cat* and *transformation*, the index will search for all motifs containing either *cat* or *transformation*. Motifs that contain both search words are considered to be better matches and will be ranked higher than those containing only a single word. To force the system to retrieve exclusively motifs containing both search terms, users need to add the Boolean operator AND, as in *cat AND transformation*. Naturally, this can be extended to as many search terms as the user requires.

The Boolean operator AND requires two or more juxtaposed terms to co-occur in the same motif. Phrasal search can be seen as a further restriction of searching with AND where the index attempts to match a list of two or more contiguous words. This functionality is provided by enclosing two terms between double quotation marks, as in "black cat".

Finally, I describe field-specific search. By default, the index searches

for matches of a particular word in all available fields (description, additional description, WordNet expansions, and references) using the field weights as described above. Users can override this default behavior by making explicit in which field they want to search. For example, to search for all motifs that mention a reference to Baughman, one could issue the query `references:Baughman`, where the search term is preceded by the expression `references:.`. Similarly, one could search for all motifs that contain words with the concept ‘color’ as one of its hyperonymic parents, but not the word color itself, using `wn:color`. Users can search the different fields individually using the following keywords, each followed by a colon:

- *motif* (to search for motif IDs, e.g. `motif:A100`);
- *description* (to search for words solely in motif descriptions, e.g. `description:magic`);
- *additional* (to search for words solely in the additional descriptions of motifs, e.g. `additional:dragon`);
- *wn* (to search for concepts available in the WordNet expansions, e.g. `wn:instrument`);
- *references* (to search for words that are part of the references (e.g. sources or countries of origin) in the index, e.g. `references:Thompson`);
- *location* (to search for motifs on the basis of the country of the sources in which a motif appears, e.g. `location:india`).

### 2.2.3 CASE STUDIES

While the goal of this section is expressly to introduce the search engine rather than to present actual research, I nevertheless selected some case studies to highlight the different ways in which MOMFER can be utilized to ask new questions and to enclose different material. All of these case studies are necessarily tentative, as are the hypotheses and conclusions based on them.



### 2.2.3.1 Monster Sighting

Monsters are an important part of folklore, often fulfilling the role of antagonist (cf. Chapter 3). Problematically, however, they are not found in a single category in the TMI, but rather are spread out under different subheadings, such as *Mythological Animals* (Bo–B99), *Marvelous Creatures* (F200–F699), and *Kinds of Ogres* (G10–G399). Furthermore, monsters can reside in less obvious habitats, such as Category C (*Tabus*, e.g. C311.1.4 *Tabu: looking at werewolf*), Category H (*Tests*, e.g. H1174.2 *Task: overcoming dragon*), and Category V (*Religion*, e.g. V236.1 *Fallen angels become fairies (dwarfs, trolls)*). The semantic query expansion implemented in MOMFER provides a convenient solution for such search problems. After a filtering step, the search query `wn:monster` results in 352 motifs containing one or more monsters. Dragons and serpents dominate this result list with 199 and 191 mentions respectively. They are followed by a long tail distribution of monsters such as griffins, werewolves, chimeras, unicorns, and so on and so forth.

Thompson intended his TMI to cover motifs from all over the world. Despite the fact that the references only “give some preliminary guidance in finding examples of the item concerned”,<sup>40</sup> leading to some notable geographical preferences (India) and geographical gaps (e.g. the Arab world),<sup>41</sup> it can be argued that Thompson achieved his goal quite well. The TMI mentions over five hundred different locations ranging from small villages in California to islands around the Coral Sea. This information is unique and can provide us with interesting suggestions about the geographical preferences of certain motifs. Here, I am interested in the geographical distribution of the monsters found in the previous paragraph. Which nations prefer werewolves and where do all those dragons hide out?

For each of the 352 motifs mentioning a monster (according to the classification by WordNet), I extracted all the locations in the references field. The locations were aggregated by country; that is, place names, provinces, and so forth were replaced by the name of their corresponding countries. The geographical distribution is visualized in Figure 2.6.

40. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*, 24.

41. See for commentary and resolution El-Shamy, *Folktales of Egypt*; El-Shamy, *Folk Traditions of the Arab World: A Guide to Motif Classification*.

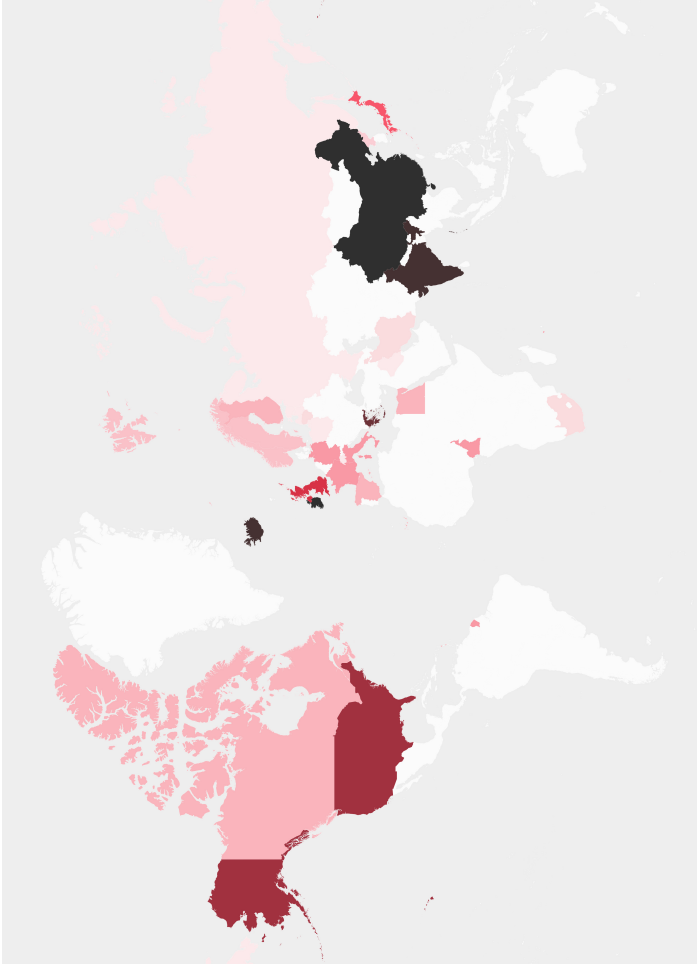


Figure 2.6: Geographical distribution of motifs in Thompson's *Motif-Index of Folk Literature* containing the WordNet concept 'monster'. The color gradient (from black via white to red) represents the frequency of monsters motifs in a country.

The color gradient represents the frequency with which monsters are found in a particular country, ranging from white (zero monsters) to black (most monsters). We can observe a strong preference for sources containing monsters from Ireland, Iceland, India, and, most notably, China. It should be acknowledged that given Thompson's relatively unbalanced data collection, the significance of these findings may be relatively trivial and incomplete, but the results serve as an illustration of the ease with which similar information could be extracted from the TMI and hint at certain trends that could be investigated more thoroughly.

### 2.2.3.2 *Black, White, and Red: Color Term Appearance in the TMI*

Color naming has long been of interest to researchers in such diverse fields as psychology, linguistics, and anthropology. It gained renewed interest with the publication of Brent Berlin and Paul Kay's groundbreaking study *Basic Color Terms: Their Universality and Evolution*.<sup>42</sup> Their main hypothesis was that "color categorization is not random and the foci of basic color terms are similar in all languages".<sup>43</sup> This universalist hypothesis, which states that the addition of colors to a language follows a more or less implicational hierarchy (all languages distinguish the concepts *black* and *white*, languages with three colors add *red* to their arsenal, but no language has only *red* and *white*), has withstood the test of time in key respects, despite the vast amount of critique from more culture-relativistic oriented researchers.

The discoveries of Berlin and Kay also gave impetus to folklore studies. Bolton and Crisp, for example, demonstrated that there is a positive correlation between the relative salience of color categories in folktales and the evolutionary development as proposed by Berlin and Kay.<sup>44</sup> In other words: the higher the color in the hierarchy, the more frequently it occurred in folktales. In a recent study, Hemming proposes an evolutionary source underlying the resonance of the tricolor in symbolic contexts all over the world.<sup>45</sup>

42. Brent Berlin and Paul Kay, *Basic Color Terms: Their Universality and Evolution* (Berkeley: University of California Press, 1969).

43. *Ibid.*, 10.

44. Ralph Bolton and Diane Crisp, "Color Terms in Folk Tales: A Cross-Cultural Study," *Cross-Cultural Research* 14, no. 4 (1979): 231–253.

45. Jessica Hemming, "Red, White, and Black in Symbolic Thought: The Tricolour Folk Motif, Colour Naming, and Trichromatic Vision," *Folklore* 123, no. 3 (2012): 310–329.



Figure 2.7: Visualization of color terms in the *Motif-Index of Folk Literature*. Plates are sized according to the frequency of their corresponding color.

It is to be expected that the importance of color terms will find its reflection in the TMI as well. An example of a motif containing a color name is the motif Z65.1.1 *Red as blood, white as snow, (and black as raven)*, which in Western folklore is most closely associated with the tale of “Snow White” (classified as ATU 709 in Uther 2004). This motif neatly reflects the importance of the tricolor *black*, *white*, and *red*. To (tentatively) investigate whether the TMI as a whole reflects the hierarchy of color terms as proposed by Berlin and Kay, MOMFER was used to search for all motifs that mention a color name using the field-specific search query `wn:color`. This results in 581 motifs containing either chromatic (i.e. blue, green, etc.) or achromatic (i.e. black, white, and gray) colors. Figure 2.7 visualizes the distribution of the color terms in these motifs. The size of the plates represents the frequency with which each color occurs in the TMI. The clear significance of the tricolor black, white, and red is in accordance with findings in previous studies. Again, it must be noted that the current investigation serves only to illustrate how MOMFER can serve as a tool to extract (new) information from TMI. However, it is noteworthy that even in a limited corpus such as the TMI the three colors *black*, *white*, and *red* are indeed the most common.

### 2.2.3.3 *Marvelous Men, Deceptive Women*

Thompson's TMI has been subjected to a number of gender studies. Torborg Lundell in particular points to a range of gender biases present in the index and makes a strong case for some textual revisions to it:<sup>46</sup>

The Motif-Index in general (1) overlooks gender identity in its labelling of motifs, thus lumping male and female actions or characters under the same, male-identified heading or (2) disregards female activity or (3) focuses on male activity at the cost of female.<sup>47</sup>

Leaving any intrinsic commentary aside, I decided to investigate quantitatively how prominent these gender biases are. I extracted all motifs from the TMI mentioning either a female character or a male character. In this example, the advantage of the semantic word expansion of MOMFER becomes apparent. In the previous case study it would be possible manually to extract all motifs containing a color term (although this would be quite a time-consuming exercise), because the list of possible primary colors is finite and quite small. Extracting all mentions of male and female characters, however, is barely possible by hand, because of the endless list of partial synonyms in both categories (e.g. lumberjack, carpenter, etc.). All these words are captured by the semantic expansion mechanism of MOMFER, however, yielding the wanted results.

I issued two queries: `wn: male` and `wn: female`, resulting in 4,178 motifs containing a male character and 3,318 motifs mentioning a female character. These numbers may already indicate some gender biases, since there is no a priori reason to assume that men are more common in stories than women. The gender biases become even more prominent when we look at the distribution of male versus female characters over the main categories in the TMI. Figure 2.8 shows that male characters dominate most motif categories. They are most prominent in *Mythological Motifs* (Category A), *Magic* (Category D), *Marvels* (Category F),

46. Torborg Lundell, "Folktale Heroines and the Type and Motif Indexes," *Folklore* 94, no. 2 (1983): 240–246; Torborg Lundell, "Gender-Related Biases in the Type and Motif Indexes of Aarne and Thompson," in *Fairy Tales and Society: Illusion, Allusion, and Paradigm*, ed. Ruth B. Bottigheimer (Philadelphia: University of Pennsylvania Press, 1989), 149–163; for a different view, see Hasan El-Shamy, "Oral Traditional Tales and the Thousand Nights and a Night: The Demographic Factor," in *The Telling of Stories: Approaches to a Traditional Craft*, ed. Morton Nøjgaard et al. (Odense: Odense University Press, 1990), 83–85.

47. Lundell, "Gender-Related Biases in the Type and Motif Indexes of Aarne and Thompson," 150.

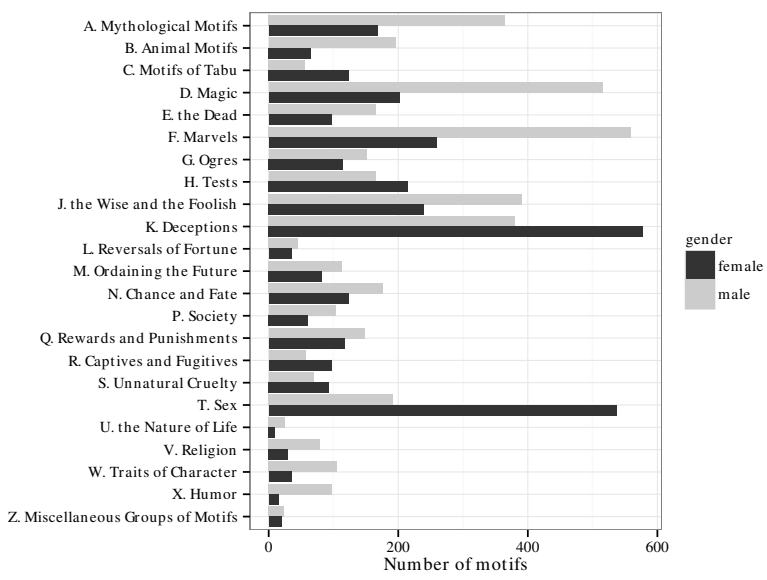


Figure 2.8: Distribution of male versus female entities in *Thompson’s Motif-Index of Folk Literature*.

and the *Wise and the Foolish* (Category J). Women, on the other hand, dominate the categories *Deception* (Category K) and *Sex* (Category T). These results confirm previous investigations and amplify the need for caution for modern scholars when exploring motifs mentioning male and female personas in the index.

2.3 MOTIF CLASSIFICATION

While the motif search engine MOMFER allows researchers to more efficiently access the TMI in order to find the motifs that belong to a particular folktale, it remains somewhat problematic that manually assigning motifs to large-scale collections of folktales is a time-consuming, error-prone, and laborious activity. In recent years, a number of new folktale digitization initiatives have been undertaken on a scale that demands for the development of new ways to enrich these collections at

different facets such as motifs.<sup>48</sup> The remainder of this chapter is devoted to the development of a computational system to automatically identify motifs in folktales, which can be employed to large-scale collections of folktales. Besides reducing manual labor, such a system is relevant for a number of reasons. First, it allows researchers to investigate how folktales have changed through time in terms of their motif material. It is only since the appearance of the Brothers Grimm's version of "Little Red Riding Hood", for example, that the girl and her grandmother are rescued from the wolf's belly (cf. Chapter 4). Extracting motifs from texts also allows researchers to find new relationships between folktales, which could tell us more about their evolutionary lineages and origins.

Only a handful of studies have explored computational systems for motif identification. Voigt et al. for example, have shown that it is possible to automate motif identification in folklore text corpora by automatically grouping texts based on their content similarities.<sup>49</sup> According to Voigt et al., the presence of common motifs can be derived from co-occurrences of keywords in the texts. For folklore researchers, however, such results are not easily interpretable, as these derived motifs are merely represented as principal components to which no recognizable label – such as a motif from the TMI – is assigned. To more directly accommodate folktale researchers' practice of assigning motifs from the TMI to folktales, Antal van den Bosch and I have explored and compared two computational systems on the task of automatically identifying motifs in folktales.<sup>50</sup> We have shown that a relatively simple model that is often used in information retrieval contexts is able to confidently identify motifs in a large number of folktales.

Both computational systems presented by Karsdorp and van den Bosch make the naive assumption that the occurrence of a motif is independent of the occurrence of other motifs. It can be argued, however, that occurrences of motifs in a folktale are – at least partially – mutually

48. Abello et al., "Computational folkloristics"; La Barre and Tilley, "The elusive tale: leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales"; Meder, "From a Dutch Folktale Database towards an International Folktale Database"; Meder et al., "Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database."

49. Vilmos Voigt et al., "Automated Motif Identification in Folklore Text," *Folklore. An Electronic Journal of Folklore* 12 (1999): 126–141.

50. Folger Karsdorp and Antal van den Bosch, "Identifying motifs in folktales using topic models," in *Proceedings of BENELEARN 2013* (Nijmegen, The Netherlands, 2013), 41–49.

predictive. This suggests that we can achieve better performance by taking into account the dependencies between motifs. My aim in this chapter is to further explore and build on these previous models of motif identification while thoroughly investigating and comparing a number of approaches that do incorporate the dependencies between motifs.

In what follows, I will first describe the data resources and data annotations used to construct this system (Section 2.3.1). Subsequently, I will provide a detailed description in Section 2.3.2 of the multi-label classification problem and a number of problem transformation techniques that allow us to more efficiently address this problem. Finally, after having discussed the experimental setup in Section 2.3.3, I will present the empirical results in Section 2.3.4.

### 2.3.1 DATA AND ANNOTATION

Crucially, the construction of an automated classification system requires a large data set of stories that are annotated with motifs. Since no such annotated collection of stories is available, the first step in constructing the system consists in compiling a sufficiently large corpus with motif annotations. In the present study, I have opted to compile this corpus based on a set of folktales from the Dutch Folktale Database,<sup>51</sup> as most of the folktales in the Dutch Folktale Database have been manually linked to the types in Uther's tale type catalog (as far as it was possible to establish *any* such link). These links between particular instances of tales and corresponding tale types provide us with a set of primary descriptive motifs that constitutes a folktale type, and, by definition in its instantiations. Each tale type in Uther's tale type catalog provides a short summary of the most common plot developments along with a sequence of motifs from the TMI. An example of such a tale type description reads as follows:

ATU 327A, **Hansel and Gretel**. A (poor) father (persuaded by the stepmother) abandons his children (a boy and a girl) in the forest [S321, S143]. Twice the children find their way back home, following scattered pebbles [R135]. On the third night, birds eat the scattered peas (bread-crumbs) [R135.1]. The children come upon a gingerbread house which belongs to a witch (ogress) [G401,

51. Meder, "From a Dutch Folktale Database towards an International Folktale Database."



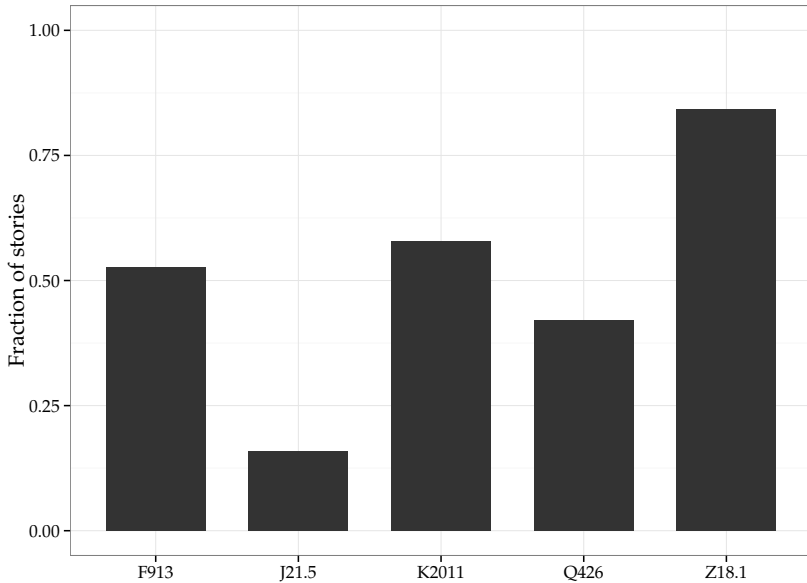


Figure 2.9: Distribution of motifs from tale type ATU 333 “Little Red Riding Hood”. On the basis of 19 instances of tale type ATU 333, the plot displays the fraction of instances in which one of the tale type motifs occurs.

F771.1.10, G412.1]. She takes them into her house. The boy is fattened [G82], while the girl must do house-work. The witch asks the boy to show his finger in order to test how fat he is [G82.1], but he shows her a bone(stick) [G82.1.1]. When the witch wants to cook the boy, the sister deceives her by feigning ignorance and pushes her into the oven [G526, G512.3.2]. [...] The children escape, carrying the witch’s treasure with them. Birds and beasts(angels) help them across water. They return home.

This example shows that the tale type ATU 327A contains twelve motifs, such as S321 *Destitute parents abandon children*, G401 *Children wander into ogre’s house*, G82.1 *Cannibal cuts captive’s finger to test fatness* and G526 *Ogre deceived by feigning ignorance of hero*.

However, particular instances of a tale type do not necessarily contain all motifs listed under their corresponding tale type; they rather comprise a subset of the relevant motifs. Thus, the corpus compilation requires a second step, in which the tale type motifs that were assigned to a

particular instance of the type that did not exhibit them were removed. To give an example of this ‘filtering’ step, in Figure 2.9 I show the distribution of the motifs of tale type ATU 333 “Little Red Riding Hood” for 19 instances in the Dutch Folktale Database. While the tale type ATU 333 enumerates a set of five primary motifs<sup>52</sup>, most instances of the type exhibit only a subset of these motifs. In fact, the distribution shows that none of Red Riding Hood’s tale type motifs mentioned by Uther is present throughout all instances of the tale type.<sup>53</sup>

In total, the complete corpus consists of 1,427 annotated stories. The stories were tokenized using the tokenizer Ucto, which was configured for the Dutch language.<sup>54</sup> The tokenized corpus contains 790,990 word forms (excluding punctuation) and 42,623 word types. On average, each story is about 554 words long ( $\sigma = 742.9$ ). The 1,427 stories exhibit 398 different tale types from Uther’s tale type catalog. The collection contains 2,483 motif annotations, which amounts to about two motifs per story on average ( $\sigma = 1.46$ ). The number of unique motifs in the collection equals 747.

## 2.3.2 MULTI-LABEL RANKING

In this section, I present a detailed formal description of the motif identification task. Let  $d$  denote a story and  $C$  the corpus of annotated stories. Each story  $d_i$  is labeled with a subset of motifs  $M_i \subset M$ , with  $M$  denoting the set of unique motifs in  $C$  (i.e.  $|M| = 747$ ). In the approach presented here, the motif identification task is framed as a ranking problem. This means that I tackle this task by computing a ranking of all unique motifs for each story. In computing this ranking, I aim to allocate the highest positions in these rankings to the most relevant motifs. The reason for approaching the task as a ranking problem rather than a classification problem is the following: because a particular story’s annotated motif subset  $M_i$  can only consist of motifs that are listed under the story’s corresponding tale type, the subset  $M_i$  does not necessarily

52. I.e., J21.5 *Do not leave the highway.*, Z18.1 *What makes your ears so big?*, K2011 *Wolf poses as “grandmother” and kills child.*, F913 *Victims rescued from swallower’s belly.*, and Q426 *Wolf cut open and filled with stones as punishment.*

53. Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson.*

54. Available from <https://language-machines.github.io/ucto/>.

contain all motifs that are potentially relevant for its characterization. It is possible that motifs listed under *different* tale types also prove to be relevant even though they are not part of a story's primary descriptive units. The benefit of a ranking approach is that, instead of yielding a single output (as would be the case with classification), it provides a list of motifs ordered by relevance in which the highest ranks are allocated to the most relevant motifs. Thus, by taking a ranking approach I do not *impose* a classification on a story, but rather, I provide a 'motif recommendation list' from which scholars can select the motifs that apply.

Let us now turn to the formal description of the motif identification task, which is framed as a multi-label ranking problem. Let  $D$  represent the training data consisting of  $N$  annotated stories. Each story  $d_i$  is associated with a feature vector  $\vec{x}_i$  and a motif subset  $M_i \subset M$ , i.e.  $d_i = (\vec{x}_i, M_i)$ . To illustrate this representation, consider the following four versions of ATU 333 "Little Red Riding Hood":

$d_1$	$\vec{x}_1 = (x_1, x_2, \dots, x_V)$	$M_1 = \{F913, Q426\}$
$d_2$	$\vec{x}_2 = (x_1, x_2, \dots, x_V)$	$M_2 = \{F913, Z18.1\}$
$d_3$	$\vec{x}_3 = (x_1, x_2, \dots, x_V)$	$M_3 = \{K2011, Z18.1\}$
$d_4$	$\vec{x}_4 = (x_1, x_2, \dots, x_V)$	$M_4 = \{F913, Q426\}$

Given this representation, the task of the multi-label ranking algorithm is to learn a ranking function  $f$ , which takes as input a story  $d$  for which the motifs are unknown, i.e.  $d = (\vec{x}, ?)$ , and produces a ranking  $R$  which consists of all motifs  $M$  ordered by their relevance to  $d$ , i.e.  $f(d) \rightarrow R$ . In what follows, I will discuss and compare the benefits and downsides of two existing multi-label ranking methods, i.e. 'Binary Relevance Transformation', and 'Label Power Set Transformation'. Subsequently, I will suggest a new multi-label ranking method, which I will call 'Tale Type Transformation'. Each of these ranking methods is a 'problem transformation method'.<sup>55</sup> The goal of these methods is to transform the data in such a way that we can employ single-label learning algorithms,

55. Grigorios Tsoumakas and Ioannis Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining* 3, no. 3 (2007): 1–13; Grigorios Tsoumakas et al., "Mining Multi-label Data," in *Data Mining and Knowledge Discovery Handbook*, ed. Oded Maimon and Lior Rokach (Boston, MA: Springer US, 2010), 667–685.

which are computationally more efficient and easier to apply than multi-label ones.

*PT-1: Binary Relevance Transformation* The ‘Binary Relevance’ method decomposes the multi-label ranking problem into  $|M|$  *binary* ranking problems. The method consists of the following steps: first, for each motif  $m \in M$  a new data set  $D_i$  is constructed in which each document  $d_i$  is associated with a binary label  $y_i \in \{0, 1\}$ .  $y_i$  takes as value 1 if motif  $m_i$  is present in document  $d_i$ , and 0 if otherwise. This transformation process can be illustrated by applying it to the example data presented above, which results in the following four data sets:

$D_1$		$D_2$		$D_3$		$D_4$	
$d_1$	F913	$d_1$	Q426	$d_1$	$\neg$ K2011	$d_1$	$\neg$ Z18.1
$d_2$	$\neg$ F913	$d_2$	$\neg$ Q426	$d_2$	$\neg$ K2011	$d_2$	Z18.1
$d_3$	F913	$d_3$	$\neg$ Q426	$d_3$	K2011	$d_3$	Z18.1
$d_4$	F913	$d_4$	Q426	$d_4$	$\neg$ K2011	$d_4$	$\neg$ Z18.1

The second step consists in training a ranking function  $f$  for each of these transformed data sets and applying these binary ranking functions to a new input story. Each ranking function  $f_{m_i}$  generates a score (for instance, in the case of a probabilistic ranking function, between 0 and 1) which expresses the confidence that motif  $m_i$  is present in one of the stories under consideration. To obtain a final ranking of all motifs, the last step involves aggregating the results of the binary ranking functions into a single list, which is sorted in descending order according to the confidence scores assigned to each motif.

The complexity of the Binary Relevance method scales linearly with the number of unique motifs  $|M|$ . The clearest advantage of the method, therefore, is its low computational complexity. However, this low complexity comes with a cost; the method is blind to any potential relationships between motifs, or, in other words, it treats all motif occurrences as being independent of each other.

*PT-2: Label Power Set Transformation* The second problem transformation method turns the multi-label ranking problem into a multi-class ranking problem by transforming each unique motif subset  $M_i \in M$  into a

distinct class. The following two tables illustrate this transformation process. The table on the left corresponds to the original training data  $D$  and the table on the right represents the transformed data set  $D'$ .

$D$			$D'$	
$d_1$	$M_1 = \{F913, Q426\}$	$\rightarrow$	$d_1$	$A = (F913 \wedge Q426)$
$d_2$	$M_2 = \{F913, Z18.1\}$		$d_2$	$B = (F913 \wedge Z18.1)$
$d_3$	$M_3 = \{K2011, Z18.1\}$		$d_3$	$C = (K2011 \wedge Z18.1)$
$d_4$	$M_4 = \{F913, Q426\}$		$d_4$	$A = (F913 \wedge Q426)$

In this small example set, the Label Power Set method reduces the number of labels for which a prediction has to be made: the original label set consists of four distinct motifs (i.e. F913, Q426, Z18.1, K2011), whereas the transformed label set only consist of three distinct classes (i.e. A, B and C). Because the Label Power Set method considers each unique motif subset to be a distinct class label, it implicitly models dependencies between labels. It should be noted, however, that with large label sets the Label Power Set method often yields an (exponential) increase (rather than a decrease) of the number of labels. Moreover, most labels in the transformed data set are associated with only a few stories. This label sparsity is one of the main drawbacks of the Label Power Set method, because it is difficult to build effective learning algorithms on the basis of a small number of examples.

*PT-3: Tale Type Transformation* In order to counter the drawbacks of the aforesaid transformation methods, I present a new problem transformation method. This method is related to the Label Power Set method, but it does not suffer from the aforementioned label sparsity problem to the same extent. In Section 2.3.1, it was pointed out that each story in the annotated corpus is labeled with a tale type from Uther’s tale type catalog, and that each tale type is associated with a set of distinctive motifs. To tackle the multi-label ranking problem, the problem transformation method I propose aims to exploit the information about tale types by means of a two-tier method. This two-tier method involves the following subsequent steps: first, the original training data  $D$  is transformed to a new data set  $D'$  in which each story  $d_i$  is associated with a single label  $y_i$ , which corresponds to the tale type  $t \in T$  of  $d_i$ , where  $T$  refers to

the complete set of tale types in the training data  $D$ . To illustrate this transformation process, consider the tables below.

$D$			$D'$	
$d_1$	{K335.1.4, K1161, N776}	$\rightarrow$	$d_1$	$y_1 = \text{ATU } 0130$
$d_2$	{L161, F96}		$d_2$	$y_2 = \text{ATU } 0301$
$d_3$	{K1161}		$d_3$	$y_3 = \text{ATU } 0130$
$d_4$	{A2741.1, F1025.1}		$d_4$	$y_4 = \text{ATU } 0295$

The second step consists in training a multi-class ranking function for this transformed data set. When a new, unseen story  $d$  is presented, the algorithm generates a score for each tale type  $t \in T$  which stands for the confidence that the tale type of story  $d$  equals  $t$ . In the case of a probabilistic ranking function, this score represents the probability  $p(t|d)$  that story  $d$  belongs to tale type  $t$ . To arrive at a final ranking of all motifs in  $M$ , we can employ this tale type probability to assign the probability  $p(t|d)$  to each of the motifs associated with  $t$ . However, this would result in a relatively uninformed ranking, as all motifs that belong to a particular tale type would be scored equally. As such, there would be no way to determine which motif is to be preferred over another. To obtain a more fine-grained ranking, I propose to compute the joint probability of a tale type given a story (i.e.  $p(t|d)$ ) and the prior probability  $p(m|t)$  of a motif given a tale type, i.e.  $p(t|d) \times p(m|t)$ . Subsequently, we can rank the motifs according to this product. The prior probability can be computed on the basis of the frequencies of a motif's occurrence with a particular tale type in the training data  $D'$ .

### 2.3.3 EXPERIMENTAL SETUP

*Models* Each of the three problem transformation methods requires a base classifier. In this study, I opted for a Maximum Entropy classifier (also known as a Logistic Regression classifier). I used the probabilistic decision function of this classifier to produce the required rankings, and I employed Scikit-Learn's implementation of this classifier with L2 regularization. In all experiments, the regularization strength parameter

C was set to 500.<sup>56</sup>

*Features* While other, more sophisticated features can be easily adopted, the present approach depends on simple ‘bag-of-words’ features, which have proven to be highly efficient in numerous computational approaches to textual data. Given a corpus C consisting of a vocabulary V (i.e. all unique words in the corpus), bag-of-words models represent documents as histograms over the vocabulary, or, in other words, as vectors consisting of frequencies for each unique word. Given a story  $d$ , the vector representation  $\vec{w}^{(d)}$  is  $(w_1, w_2, \dots, w_{|V|})$ , where  $w_i$  represents the number of occurrences of word  $i$  in story  $d$ . In addition to these bag-of-words features, I make use of so-called ‘character  $n$ -gram features’, which represent counts of substrings of words. By setting  $n$  to four, for example, a word such as *wizard* is represented by the following substrings: wiza, izar, and zard. On the basis of these substrings, a new vector representation is created which represents a histogram over all unique substrings  $s \in S$ , i.e.  $\vec{s}^{(d)} = (s_1, s_2, \dots, s_{|S|})$ , where  $s_i$  represents the number of times the substring  $i$  occurs in story  $d$ .

It is common practice in many computational applications to weigh these frequency vectors for feature importance. As such, I transformed the frequency vectors by means of Parsimonious Language Models as proposed by Hiemstra et al.<sup>57</sup> Parsimonious Language Models represent a probabilistic alternative to traditional tf-idf weighting schemes,<sup>58</sup> and have proven to be highly efficient in Information Retrieval. For each document  $d$ , a parsimonious language model is defined as:

$$P(t_1, \dots, t_n | d) = \prod_{i=1}^n (1 - \lambda) P(t_i | D) + \lambda P(t_i | d) \quad (2.1)$$

where  $D$  represents the complete training collection, and  $t$  represents either a word or a character  $n$ -gram feature. The estimate of  $P(t|d)$  is maximized using an Expectation-Maximization algorithm (cf. Algorithm

56. F. Pedregosa et al., “Scikit-Learn: Machine Learning in Python,” *Journal of Machine Learning Research* 12 (2011): 2825–2830.

57. Djoerd Hiemstra et al., “Parsimonious language models for information retrieval,” in *SIGIR '04 Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (2004), 178–185.

58. Christopher D. Manning et al., *Introduction to Information Retrieval* (New York, USA: Cambridge University Press, 2008).

---

**Algorithm 1: Expectation-Maximization Algorithm**


---

```

for feature  $t_i \in d$  do
  while not converged do
     $e_i = tf(t_i) \cdot \frac{\lambda P(t_i|d)}{(1-\lambda)P(t_i|D) + \lambda P(t_i|d)} ;$ 
     $P(t_i|d) = \frac{e_i}{\sum_i e_i} ;$ 

```

---

1). The hyperparameter  $\lambda$  controls the spread of probability mass to features, i.e. how much weight is assigned to each feature. With  $\lambda < 1$ , the probability mass is moved to fewer features than a maximum likelihood estimate of the documents would, and the probability mass moves to even fewer features as  $\lambda$  approaches 0. This results in a ‘peaky’ distribution of feature weights, with only a few strongly weighted features for each document.

*Evaluation* For the evaluation of the models, I used a  $n$ -fold cross-validation setup. In this setup, all stories are shuffled, and subsequently partitioned into  $n$  subsets of equal size. The model is set to run for  $n$  iterations. At each iteration a different part or ‘fold’ acts as a test set, and the remaining folds are combined to form the training set. In the experiments below,  $n$  is set to 4.

The performance of the models is measured by means of five evaluation metrics, i.e. Average Precision, One Error Score, Is Error score, Margin Score, and Ranking Loss. Each of these metrics addresses a different aspect of the performance quality of the models, as they each ask different questions about the rankings:<sup>59</sup>

*Average Precision:* Are most or all of the target motifs high up in the ranking?

$$AP = \frac{\sum_{k=1}^n (P(k) \times rel(k))}{\text{number of relevant motifs}} \quad (2.2)$$

where  $k$  is the position in the ranked list of  $n$  retrieved motifs.  $P(k)$  represents the precision at  $k$  and  $rel(k) = 1$  if the item at position  $k$  is relevant,  $rel(k) = 0$  otherwise. The Average Precision yields a score between 0 and 1, where 1 indicates a perfect ranking.

59. Tsoumakas et al., “Mining Multi-label Data.”



*One Error*: For what fraction of test documents is the highest-ranked motif *incorrect*?

$$\text{One Error} = \frac{1}{n} \sum_{d=1}^n \text{rel}(1) \quad (2.3)$$

The One Error Score is a value between 0 and 1, where 0 indicates that for all of the test documents the highest rank motif is correct.

*Is Error*: What fraction of rankings is *not* perfect?

$$\text{Is Error} = \frac{1}{n} \sum_{d=1}^n \text{AP}(d) < 1 \quad (2.4)$$

The Is Error Score is a value between 0 and 1, where 0 means that in all generated rankings, all relevant motifs are ranked highest. The Is Error Score is complementary to the Average Precision Score, as it represents the fraction of test documents that has an Average Precision score smaller than 1.

*Margin*: What is the absolute rank difference between the highest ranked irrelevant motif and the lowest ranked relevant motif averaged across test documents?

$$\text{Margin} = \frac{1}{n} \sum_{d=1}^n |k_{\text{highest irrelevant}} - k_{\text{lowest relevant}}| \quad (2.5)$$

The Margin Score is a value from 0 to  $n$ , where  $n$  is the number of possible motifs in the ranking. The closer to 0, the smaller the absolute difference between relevant motifs, and the better the ranking.

*Ranking Loss*: What is the number of times that relevant motifs are ranked lower than irrelevant motifs, first averaged within one document, then across all test documents.

$$L = \frac{1}{n} \sum_{i=1}^n \frac{1}{|M_i| |\bar{M}_i|} |\{(m_a, m_b) : k_a > k_b, (m_a, m_b) \in M_i \times \bar{M}_i\}|, \quad (2.6)$$

where  $\bar{M}_i$  is the complementary set of  $M_i$ . The Ranking Loss score is equivalent to the complement of the area under the ROC curve.<sup>60</sup>

It is a value between 0 and 1, where 0 indicates a perfect ranking.

60. Timothy N. Rubin et al., "Statistical topic models for multi-label document classification," *Machine Learning* 88, no. 1 (2012): 157–208.

	AP	One Error	Is Error	Ranking Loss	Margin
PT-1	.775	.249	.348	.051	23.15
PT-2	.760	.270	.371	.050	22.61
PT-3	<b>.801</b>	<b>.243</b>	<b>.323</b>	<b>.021</b>	<b>8.25</b>

Table 2.1: Motif ranking results. The table provides the Average Precision score, One Error score, Is Error score, Ranking Loss and Margin score for all three problem transformation methods, i.e. PT-1: Binary-Relevance, PT-2: Label Power Set and PT-3: Tale Type Transformation.

#### 2.3.4 RESULTS

Table 2.1 presents the results of the experiments. The table provides the Average Precision score, One Error score, Is Error score, Ranking Loss, and Margin score for the three problem transformation methods set out in Section 2.3.2. Observe that PT-3 outperforms the other two transformations methods across all evaluation metrics. Although all models are able to generate motif rankings with relatively high Average Precision (PT-1: 0.78, PT-2: 0.76, PT-3: 0.8), the difference between the margin score of PT-1 (23.15) and PT-2 (22.61), and that of PT-3 (8.25) is sufficiently large to support the hypothesis that incorporating mutual motif relations via tale type dependencies into the model leads to a performance boost. With respect to the ‘Ranking Loss’ and ‘One Error’ score, PT-3 exhibits a small performance increase compared to the other models: PT-3 generates rankings in which the top ranked motif is correct in 76% of the cases. Finally, the tale type transformation method marginally outperforms the other two transformation methods in producing perfect rankings, with 68% of its rankings having an Average Precision of 1.0.

In what follows, I will provide a more detailed analysis of the experimental results as well as possible explanations for the performance differences between the models. I hypothesize that the Tale Type Transformation method (PT-3) has two main advantages over the other two transformation methods, which leads to the observed difference in performance. First, PT-3 implicitly models dependencies between motifs by incorporating their overarching tale type, whereas in the Binary Relevance model (PT-1) all predictions happen independently of each other.

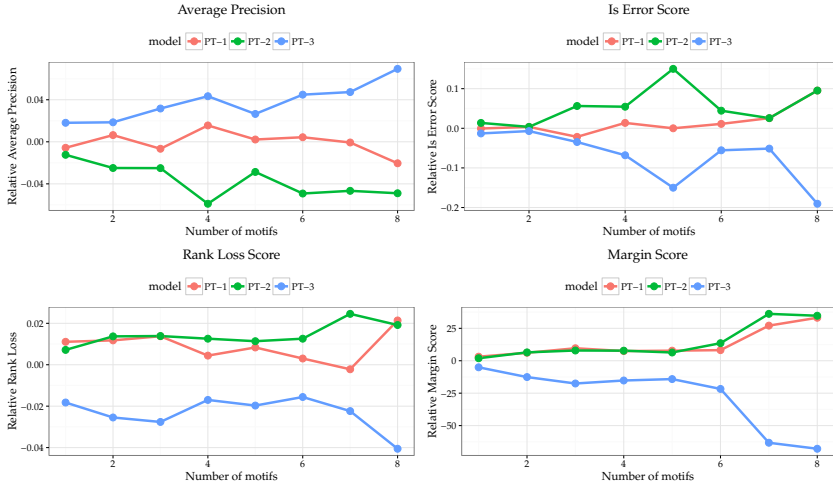


Figure 2.10: Interaction between the number of motifs per document and model performance. In order of appearance, the plots show the ‘Average Precision’, ‘Is Error’ score, ‘Ranking Loss’ score and ‘Margin’ score for subsets of test documents grouped on the basis of how many motifs they contain. The One Error score is not included in the plot, as it highly correlates with the Average Precision.

While the Label Power Set method (PT-2) is also designed to implicitly take some of the dependencies between motifs into account, it suffers from a different problem: PT-2 can lead to severe label sparsity when only a very small number of examples is available for many unique motif subsets. The second advantage of the PT-3 method is that, because of its two-tier architecture, it is less vulnerable to such label sparsity issues.

If these hypotheses hold, we can expect that models in which motif dependencies are taken into account perform even better on stories with an increasing amount of motifs. To empirically ground this hypothesis, I recompute the evaluation metrics for subsets of test documents which are grouped by the number of motifs they contain. Figure 2.10 illustrates the *relative* performance differences between the three models for subsets of stories with an increasing number of motifs. Note that the results for this plot have been centered around zero in order to obtain a clearer picture of the differences. The plot shows that, as the number of motifs

per document increases, the performance difference between PT-3 on the one hand and PT-1 and PT-2 on the other becomes larger. This effect can be observed for all four given evaluation metrics.

Note that, contrary to expectations, the Label Power Set method (PT-2) in Figure 2.10 does not exhibit a performance advantage relative to the more naive Binary Relevance method. In fact, PT-2 consistently performs worse than the other two transformation methods. The main reason why PT-2 lags behind the other methods, I wish to argue, is that it suffers from the aforementioned label sparsity problem: there are not nearly enough examples to build effective classifiers. As the number of unique motifs increases, this sparsity problem becomes more severe because the number of possible motif combinations grows. As such, PT-2 (and also, albeit to a lesser extent, PT-1) exhibits a performance drop on data sets with an increasing number of unique motifs. To this end, it is interesting to compare the performance of the models with respect to the genre distinctions in the tale type catalog by Uther. This catalog lists over two thousand unique tale types, which are divided into the following seven genres:

1. *Anecdotes and Jokes* include classical jokes and riddles. They often end with a punchline and are relatively short;
2. *Animal Tales* are tales in which the protagonist is usually an animal. They are relatively short and contain one or two prominent motifs;
3. *Formula Tales* deal with tales in which “form is all-important”.<sup>61</sup> The tales are often rather short, contain repetition and a small number of motifs;
4. *Tales of Magic* are spatio-temporally unbound stories. The plot often centers around one protagonist who faces a range of challenges in his or her quest towards some desired object. Tales of Magic can be quite detailed and long both in words and in number of motifs;
5. *Realistic Tales* likewise describe adventures and heroes, but in Realistic Tales there is an absence of any form of magic;

61. Stith Thompson, *The Folktale*, 2nd ed. (Berkley, Los Angeles & London: University of California Press, 1977), 229.

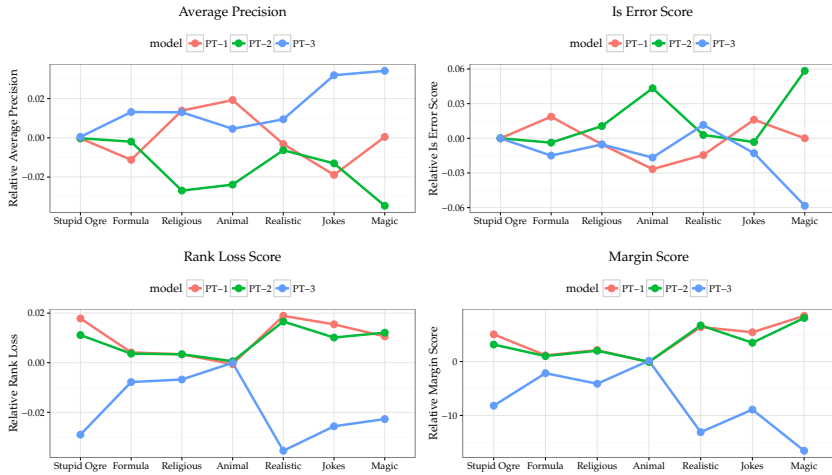


Figure 2.11: Interaction between number of unique motifs per folktale genre and model performance. In order of appearance, the plots show the ‘Average Precision’, ‘Is Error’ score, ‘Ranking Loss’ score and ‘Margin’ score for each folktale genre in Uther’s tale type catalog. The genres are ordered according to the number of unique motifs they contain.

6. *Religious Tales* are comparable to legends. They are spatio-temporally bound stories that occur in the recent past that describe supernatural events and center around religious values;
7. *Tales of the Stupid Ogre* resemble humorous tales about ogres, devils etc. with wondrous aspects.

Figure 2.11 illustrates the *relative* performance between the three models for each of these seven genres. The genres have been ordered according to the number of unique motifs they contain. The plots add visual evidence to the hypothesis that there is an interaction between the performance of the models and the number of unique motifs in a genre. Both the ‘Binary Relevance’ and the ‘Label Power Set’ method undergo a performance drop as the number of unique motifs increases. Such a drop is not observed with the ‘Tale Type Transformation’ method, which can be explained by the fact that the number of possible labels in the first tier of PT-3 is much lower than that of the other two transformation methods. Furthermore, it appears that PT-2 suffers the most if the

number of unique motifs increases, which can again be attributed to its label sparsity problem.

## 2.4 CONCLUSION

In the first part of this chapter, I presented MOMFER, which is a new and fast research tool designed to accompany Thompson's TMI. I set out the production process, documented the query system, and showed how the tool offers new ways to explore the index in three small case studies. A complete collection of all motif indices would be an amazing research tool. MOMFER is a first step towards such an index: with its flat architecture, it enables the inclusion of other motif indices. As Heda Jason notes, proposals for new motifs should be made "only after it has been found that no other motif index has listed the specific content element in question".<sup>62</sup> An integrated tool such as MOMFER could serve as a warrant for such requirements. The tool serves as a starting point for inviting researchers to join in disclosing other indices so that, more than fifty years after the publication of Thompson's great index, we may finally have a complete index of all folklore motifs.

In the second part of this chapter, I developed a computational system for motif identification in folktales. This system, presented in Section 2.3, frames this identification task as a ranking problem. The goal of such a system is to produce a ranking of motifs for a particular story in which the top ranked motifs represent the most relevant motifs for that particular story. An in-depth comparison of three problem transformation methods (i.e. 'Binary Relevance' (PT-1), 'Power Label Set' (PT-2) and 'Tale Type Transformation' (PT-3)) showed that the Tale Type Transformation method (PT-3) can achieve better performance than models that assume that motifs are not mutually predictive (PT-1), because it incorporates dependency relations between motifs through their overarching tale type. Furthermore, it was shown that this model suffers less from the well-known label sparsity problem than existing problem transformation methods such as the Label Power Set method (PT-2). The Tale Type Transformation method is able to quite accurately

62. Heda Jason, *Motif, Type and Genre: A Manual for Compilation of Indices and a Bibliography of Indices and Indexing* (Helsinki: Suomalainen Tiedeakatemia, 2000), 61.

identify motifs in folktales at an Average Precision of 0.8 on unseen documents.

Because the present study adopts a ranking approach rather than a classification approach, it comes with a great benefit for folklore scholars: the approach presented above does not *impose* a particular all-or-nothing classification on a story, but rather, it provides scholars with *recommendations* about which motifs are likely to be applicable to a story. The combination of this recommendation system with the motif search engine MOMFER can greatly facilitate folklore scholars in their future endeavors to analyze large-scale folktale collections and the relations between folktales and their motifs.

---

## CHAPTER 3

# DEVILS, FAIRIES AND DRAGONS: CHARACTER BIAS IN THE CULTURAL SUCCESS OF FAIRY TALES

*"Suppose we get the Professor to tell us a story."*  
*Bruno adopted the idea with enthusiasm. Please do. he cried eagerly.*  
*"Sumfin about tigers—and bumble-bees—and robin-redbreasts, oo knows!"*  
*"Why should you always have live things in stories?" said the Professor.*  
*"Why don't you have events, or circumstances?"*  
*"Oh, please invent a story like that!" cried Bruno.*  
*The Professor began fluently enough. "Once a coincidence was taking*  
*a walk with a little accident, and they met an explanation—a very old*  
*explanation—so old that it was quite doubled up, and looked more like a*  
*conundrum—" he broke off suddenly.*  
*"Please go on!" both children exclaimed.*  
*The Professor made a candid confession. "It's a very difficult sort to invent,*  
*I find. Suppose Bruno tells one first."*  
*Bruno was only too happy to adopt the suggestion.*  
*"Once there were a Pig, and a Accordion, and two jars of Orange-marmalade"*

LEWIS CARROLL, *Sylvie and Bruno Concluded*

### 3.1 INTRODUCTION

With his 2006 monograph *Why fairy tales stick*, Jack Zipes addresses perhaps the most central question in fairy tale transmission research: why does it seem that fairy tales consistently succeed to prevail in culture?<sup>1</sup> Among the most popular stories in present-day Western Culture are fairy tales like "Cinderella", "Snow White" or "Red Riding Hood" – all of

1. Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*.



which are not present-day creations but find their roots in centuries-old tales. In recent research, it has even been suggested that fairy tales might even be much older than their first literary record, evidence suggesting that tales such as “The Smith and the Devil” can be traced back to the Bronze age.<sup>2</sup> Observations such as these have led many to believe that there is something about fairy tales that makes them stick.

But not *all* fairy tales stick. While we have all undoubtedly heard of “Cinderella” or “Red Riding Hood”, it seems that we are much less familiar with the story of “The Poor Boy in the Grave” or “The Jew in the Brambles” – if we are aware of these stories at all. In the 1857 edition of the fairy tale collection *Kinder- und Hausmärchen*, the Brothers Grimm collected and published as much as 210 stories; yet, only a handful of them is included in the present-day canon.<sup>3</sup> The question, then, should not be ‘why fairy tales stick’, but rather ‘*which* fairy tales stick’, and why it is precisely those that prevail.

In the existing literature, some suggestions have been made regarding the appeal of ‘stories that stick’. There are a number of experimental studies of story transmission in a laboratory context using a transmission chain method.<sup>4</sup> In a laboratory setting, stories are passed down either orally or in writing from one participant generation to the next under controlled conditions. Results from such experimental simulation studies suggest that the progressive alterations of stories are primarily controlled by social and cognitive selection pressures.<sup>5</sup> It has been

2. Sara Graça Da Silva and Jamshid J. Tehrani, “Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales,” *Royal Society Open Science* 3, no. 1 (2016), doi:10.1098/rsos.150645.

3. Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*; Vanessa Joosen, “A Translation Far Worse. Canonisation and Adaptation in the Early Dutch and English Translation of the Brothers Grimm’s *Kinder- und Hausmärchen*,” in *Never-ending Stories. Adaptation, Canonisation and Ideology in Children’s Literature*, ed. Sylvie Geerts and Sara van den Bossche, Ginkgo (Ghent, Belgium: Academia Press, 2014).

4. Bartlett, *Remembering. A Study in Experimental and Social Psychology*; Anthony Lyons and Yoshihisa Kashima, “Maintaining stereotypes in communication: Investigating memory biases and coherence-seeking in storytelling,” *Asian Journal Of Social Psychology* 9, no. 1 (2006): 59–71; Mesoudi et al., “A bias for social information in human cultural transmission”; Breithaupt et al., “Optimal Eventfulness of Narratives.”

5. See e.g. Justine Owens and Gordon Bower, “The ‘soap opera’ effect in story recall,” *Memory Cognition* 7, no. 3 (1979): 185–191; Joshua J Diehl et al., “Story Recall and Narrative Coherence of High-Functioning Children with Autism Spectrum Disorders,” *Journal of Abnormal Child Psychology* 34, no. 1 (2006): 83–98; Mesoudi and Whiten, “The Multiple Roles of Cultural Transmission experiments in Understanding Human Cultural Evolution”;

shown, for example, that story elements are preferred to be selected for transmission when they convey minimally counterintuitive concepts, rather than intuitive or counterintuitive concepts.<sup>6</sup> A concept is considered to be minimally counterintuitive when most of its characteristics are in accordance with the ontological assumptions of the concept, yet a few characteristics are not consistent with these assumptions.<sup>7</sup> An example of a minimally counterintuitive concept is that of a ghost; a ghost shares many characteristics with other intentional agents (e.g. consciousness, feelings, beliefs), yet a few of its characteristics, such as its being invisible and ability to move through walls, go against the ontological assumptions about intentional agents. Experimental evidence has shown that such minimally counterintuitive concepts are more memorable and are more accurately transmitted than intuitive or counterintuitive concepts. Furthermore, in a corpus study, Norenzayan et al. have shown that narratives consisting of a few (typically in the range of 2 to 3) minimally counterintuitive concepts have a transmission advantage over other stories.<sup>8</sup> Other examples of selectional preferences include salient story elements, or rather, story elements evoking heavy emotional reactions such as disgust,<sup>9</sup> and a disposition for social information.<sup>10</sup> For instance, stories containing a high degree of ‘gossip-like’ information (such as romantic intrigues between students and professors) are more likely to be accurately transmitted, and therefore have a higher probability of

Eriksson and Coultas, “The Advantage of Multiple Cultural Parents in the Cultural Transmission of Stories.”

6. Justin Barrett and Melanie Nyhof, “Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials,” *Journal of Cognition and Culture* 1, no. 1 (2004): 69–100; Ara Norenzayan et al., “Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives,” *Cognitive Science* 30 (2006): 531–553; M. Afzal Upal, “Memory, Mystery and Coherence: Does the Presence of 2–3 Counterintuitive Concepts Predict Cultural Success of a Narrative?,” *Journal of Cognition and Culture* 11, no. 1 (2011): 23–48.

7. Ara Norenzayan and Scott Atran, “Cognitive and emotional processes in the cultural transmission of natural and nonnatural beliefs,” in *The Psychological Foundations of Culture*, ed. M. Schaller and C. Crandall (Hillsdale, NJ: Lawrence Erlbaum, 2004), 149–169.

8. Norenzayan et al., “Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives.”

9. Eriksson and Coultas, “Corpses, maggots, poodles and rats: Emotional selection operating in three phases of cultural transmission of urban legends.”

10. Matthew B. Reysen et al., “The effects of collaboration on recall of social information,” *British Journal of Psychology* 102, no. 3 (2011): 646–661; Joseph M. Stubbersfield et al., “Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends,” *British Journal of Psychology* 106, no. 2 (2014): 288–307.

persisting in their original form.<sup>11</sup>

However, while such experimental studies in laboratory contexts implicitly affect our understanding of story transmission in the real world, their results cannot be unequivocally taken to apply to real-world story transmission and selection phenomena. Observations regarding real-world data, specifically concerning fairy tale transmission and the canonization of the Brothers Grimm, have been made by Joosen, who suggests that successful fairy tales dominantly revolve around female characters. A study by Koman and Meder remarkably points in a similar direction: of the ten most popular fairy tales in the Dutch language area, seven are centered around female characters and the top four contain a single female protagonist.<sup>12</sup> Yet, female protagonists are not the only recurring character type in the most popular fairy tales; the top three stories all contain a cruel witch or stepmother as an antagonist character and a prince who the protagonists end up marrying. As such, it appears that female protagonists are but one of many possible appealing traits of a story.

Interestingly, neither the experimental research, nor story transmission research in literary and folkloristic studies seems to devote attention to what makes certain stories *unsuccessful*. In both research areas, the main object of interest are the stories that remain after a repeated ‘filter-and-selection’ process in cultural transmission, and the questions asked centered on the traits that distinctively define the persisting stories. However, the distinctive traits of unpopular stories are largely ignored. Of course, by determining popular traits, scholars imply that it is the *lack* of popular characters that makes particular stories less popular and causes them to disappear from cultural memory, but the idea that there might be such a thing as ‘impopularizing’ traits or characters, making a story less interesting or appealing for further transmission, has thus far been consistently overlooked.

It is these questions and remaining hiatuses in the understanding of what determines selection in fairy tale transmission that I will address in the present chapter. I will systematically examine fairy tales from

11. Alex Mesoudi et al., “A bias for social information in human cultural transmission,” *British Journal of Psychology* 97, no. 3 (2006): 405–423.

12. I.e. “Snow White”, “Cinderella”, “Sleeping Beauty”, and “Red Riding Hood”, cf. Ruben A. Koman and Theo Meder, *Resultaten Sprookjesenquête* (Kaatsheuvel, Amsterdam: Efteling; DOC Volksverhaal, 2008).

the Brothers Grimm collection with varying degrees of successfulness. By developing a folktale character typology, and subsequently mapping the characters from the Grimm's fairy tales to this typology, I provide empirical evidence for several character types that discriminate between popular and unpopular tales. The results indicate that popular tales revolve around what can be considered 'minimally counterintuitive characters', such as 'ghosts', 'dwarfs', 'giants' and 'witches'. In addition, we also find that characters in the sphere of close social family relations, such as 'siblings' and 'parents', are commonly found in successful stories. Unsuccessful stories, by contrast, appear to be centered around characters from the religious sphere, comprising 'priests' and 'popes' as well as 'Jesus' and 'God'. Besides biblical and cleric characters, unpopular stories also tend to stage generic or collective characters, which do not represent a single, definite (and, hence, more personal) entity, but a more indeterminate group of a particular kind or type of people, such as 'men', 'thieves' and 'people'. These findings serve to aid our understanding of how the funnel-like selection process observed in the dissemination of the 210 stories in *Kinder- und Hausmärchen* has proceeded, which, in its turn furthers the understanding of the more general question of how to explain prevalent culture, in which some cultural artifacts are more likely to survive than others.

Besides its general theoretical relevance, addressing the question of which characters correlate with popular and unpopular stories in such a large-scale collection as *Kinder- und Hausmärchen* poses an important methodological challenge. The present chapter presents a computational system that automatically extracts the character cast of stories using a mechanism that detects animate behavior of entities in texts. This model, which serves as the methodological base for the further research questions on story transmission addressed here, will be set out and described in detail in the next section (Section 3.2). Section 3.3, then, will return to the central question of this chapter. Here, the animacy detection model will be applied to the Grimm's fairy tale collection, yielding a list of all characters in all stories. Based on these lists, we can determine which character types distinctively occur in popular and unpopular stories. Finally, the interpretation and theoretical implications of these results will be presented in Section 3.4.

### 3.2 ANIMACY DETECTION IN STORIES

For almost all species in the world, the capacity to distinguish animate objects from inanimate objects is essential to their survival. Those objects could be prey, for example, or predators or mates. The fundamental nature that the distinction between animate and inanimate has for humans is reflected in the fact that this division is acquired very early in life: children of less than six months old are well able to distinguish the two categories from one another.<sup>13</sup> Moreover, recent brain research shows that the distinction appears in the organization of the brain.<sup>14</sup> For some researchers, this provides evidence for the idea that the division between animate and inanimate is an innate part of how we see the world.

Although animacy may be a scalar rather than a strictly categorical distinction,<sup>15</sup> the distinction between animate and inanimate is traditionally taken as binary with regard to lexical items: something is either animate (e.g. a human) or not (e.g. a chair). This standpoint has been challenged, however, by researchers from different fields. Firstly, it has long been established in linguistic typology that not all languages award animacy to the same entities in different grammatical categories. As Comrie notes, for example, English, and many other languages, distinguishes between human and not-human in the choice of pronouns, whereas Russian distinguishes between animate (entailing humans and animals) versus non-animate (entailing everything else) in its interrogative pronouns.<sup>16</sup> Another example indicating different subdivisions of animacy in languages is found in Persian. In English, *tree* and *flower* are both inanimate words, but in Persian the word for *tree* is grammatically classified as animate, whereas the word for *flower* is classified as being *inanimate* along with words such as *house* and *chair*.<sup>17</sup> Secondly,

13. John Opfer, "Identifying living and sentient kinds from dynamic information: The case of goal-directed versus aimless autonomous movement in conceptual change," *Cognition* 86, no. 2 (2002): 97–122.

14. Tao Gao et al., "Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus," *The Journal of neuroscience: the official journal of the Society for Neuroscience* 32, no. 41 (2012): 14276–14280.

15. See e.g. the animacy hierarchy in Bernard Comrie, *Language Universals and Linguistic Typology*, 2nd ed. (Chicago: University of Chicago Press, 1989); and research such as Anette Rosenbach, "Animacy and grammatical variation – findings from English genitive variation," *Lingua* 118, no. 2 (2008): 151–171.

16. Comrie, *Language Universals and Linguistic Typology*.

17. Heike Wiese, "Semantics as a gateway to language," in *Mediating between Concepts and*

philosophers such as Daniel Dennett support the view that animacy and aliveness are to be treated as epistemological stances rather than fixed states in the world: not ineffable qualia but behavioral capacity defines our stance towards objects.<sup>18</sup> In other words, depending on whether people *think* that an object is animate, they utilize different cognitive strategies to explain and predict the actions of those objects. Finally, evidence from psycholinguistic research has accumulated to support this view of animacy as a cognitive viewpoint rather than an extra-perceptive absolute. Nieuwland and Van Berkum, for example, show that college student test subjects readily accept animate behavior from inanimate objects within the proper contexts,<sup>19</sup> and Vogels et al. moreover emphasize the relation between animacy and motion, showing that factors such as self-propelment play a crucial role in recognizing and/or awarding animacy to certain objects.<sup>20</sup> This is exemplified in the opening of this well-known story:<sup>21</sup>

A farmer bought a pancake on the market. Once he got home, the farmer was hungry and began to bake the pancake. The farmer tried one of his skillful flipping techniques, but he failed and the pancake fell on the ground. Coincidentally, the door of the kitchen was open and the pancake rolled out to the field, as hard as he could...

Although initially, based on their knowledge of the world, readers will regard the pancake as inanimate, the self-propelled motion verb 'rolled' initiates our shift towards an animate interpretation of the pancake. As readers (or listeners) of a story, we choose to view participating objects at varying levels of abstraction in order to predict their behavior. Dennett defines three levels of abstraction: (i) the 'physical stance', (ii) the 'design stance' and (iii) the 'intentional stance'.<sup>22</sup> The physical stance

*Grammar*, ed. Holden Härtl and Heike Tappe, Trends in Linguistics (Berlin, New York: Mouton de Gruyter, 2003), 211.

18. Daniel Dennett, *The Intentional Stance* (Cambridge, Massachusetts: The MIT Press, 1996).

19. Mante S. Nieuwland and Jos J. A. Van Berkum, "When Peanuts Fall in Love: N400 Evidence for the Power of Discourse," *Journal of Cognitive Neuroscience* 18, no. 7 (2005): 1098–1111.

20. Jorrig Vogels et al., "When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices," *Frontiers in Psychology* 4, no. 154 (2013): 1–15.

21. Taken from: <http://www.verhalenbank.nl/items/show/9636>.

22. Dennett, *The Intentional Stance*.

deals with predictions about objects given their physical properties. The design stance deals with concepts such as purpose, function or design. The intentional stance is concerned with belief, thinking and intentions. These are all cognitive strategies we use to predict and explain the actions of objects in our environment. Interestingly, in the process of reading the opening of the story about the fleeing pancake, readers and listeners experience the transition from one strategy to the next quite clearly. Initially, the pancake is interpreted from a physical stance, or perhaps the more abstract design stance in terms of the purpose (i.e. to stave off hunger). It is only at the last adverbial phrase ‘as hard as he could’ that we start to wonder whether we should adopt to the yet more abstract intentional stance and consider the pancake to be a rational agent.

Given the fundamental nature of the distinction between animate and inanimate, it is perhaps not too surprising that it has proven to be invaluable in a variety of Natural Language Processing tasks dealing with e.g. anaphora resolution and dependency parsing.<sup>23</sup> Existing methods for the automatic labeling of text for animacy are usually rule-based, machine-learning-based, or a hybrid of these methods. Common to most approaches is the fact that they make use of semantic lexicons with information about animacy, as well as syntactic cues in a text. Both feature types are relatively costly to obtain as they require large vocabularies or syntactic parsing systems, which, with the exception of a few languages, are not readily available.

In this chapter, I present a new linguistically uninformed model to automatically label texts for animacy. It is shown that we can do away with features that require syntactic parsing or semantic lexicons while still yielding competitive performance. The focus is on labeling animacy in stories, because stories pose some interesting problems to automatic systems of animacy recognition. As the example of the fleeing pancake already illustrated, in stories any entity may at some point

23. Constantin Orăsan and Richard Evans, “NP animacy identification for anaphora resolution,” *Journal of Artificial Intelligence Research* 29 (2007): 79–103; Heeyoung Lee et al., “Deterministic coreference resolution based on entity-centric, precision-ranked rules,” *Computational Linguistics* 39, no. 4 (2013): 885–916; Lilja Øvrelid and Joakim Nivre, “When word order and part-of-speech tags are not enough – Swedish dependency parsing with rich linguistic features,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)* (2007), 447–451.

exhibit animate behavior, even when they are inanimate in the ‘real’ world. Another example is the *Sorcerer’s Apprentice* sequence in Walt Disney’s famous *Fantasia*, in which brooms display the ability to collect buckets of water. Such examples, in which entities such as pancakes and brooms act as animate beings, make a clear case for developing dynamic, data driven systems that do not rely too much on static and fixed world knowledge, but rather on immediate context.

The remainder of this section is structured as follows. I will start with a short overview of existing techniques for automatically labeling animacy in texts, including the definitions of animacy used in these papers (Section 3.2.1). After a description of the corpus used in this study and how the annotations of the corpus have been established (Section 3.2.2.1), I will give an account of the computational models in Section 3.2.2.5. The empirical results will be reported in Section 3.2.2.6.

### 3.2.1 PREVIOUS WORK

A handful of papers deal with automatic animacy detection. Most approaches make use of rule-based systems or machine learning systems with morphological and syntactic features. Evans and Orăsan present a rule-based system that makes use of the lexical-semantic database WordNet.<sup>24</sup> They label each synset in WordNet for animacy. Using a variety of rules to detect the head of an NP, they use the fraction of synsets in which a particular noun occurs to arrive at a classification for animacy. Orăsan and Evans extend their previous algorithm by first determining the animacy of senses from WordNet on the basis of an annotated corpus.<sup>25</sup> They then apply a *k*-nearest neighbor classifier using a number of lexical and syntactic features and features derived from WordNet to arrive at a final animacy classification.

Øvrelid and colleagues published a variety of studies in which a number of animacy classifiers is presented that make use of syntactic and morphological features.<sup>26</sup> These features include the frequency

24. Richard Evans and Constantin Orăsan, “Improving anaphore resolution by identifying animate entities in texts,” in *Proceedings of the Discourse Anaphora and Reference Resolution Conference* (2000), 154–162.

25. Constantin Orăsan and Richard Evans, “Learning to identify animate references,” in *Proceedings of CoNLL-2001* (Toulouse, France, 2001), 129–136.

26. Lilja Øvrelid, “Animacy classification based on morphosyntactic corpus frequencies:



of analysis of the noun as ‘subject’ or ‘object’, the frequency of the occurrence of a noun in a passive *by*-phrase, and the frequency of the noun as a subject followed by either animate personal pronouns or inanimate personal pronouns. These features are then aggregated for each lemma after which a machine learning system (decision tree or *k*-nearest neighbor classifier) is trained. Bowman and Chopra presents a similar approach by employing a Maximum Entropy classifier which is trained on the basis of three feature types: (1) bag-of-words with and without their corresponding Part-of-Speech tags, (2) internal syntactic features such as the syntactic head and (3) external syntactic features that describe the dependency relation of a noun to a verb (i.e. subject relation, object relation, etc.).<sup>27</sup> This is the only study that makes use of a fully labeled corpus for animacy. Partially related to animacy detection, Karsdorp et al. attempt to extract the cast (i.e. all characters) from a story.<sup>28</sup> Similar to Bowman and Chopra they rely on dependency tags to extract the subjects of direct and indirect speech.

Bloem and Bouma present a model that attempts to generalize the animacy information in a lexical-semantic database of Dutch by augmenting ‘non-ambiguous’ animate entries with contextual information from a large treebank of Dutch.<sup>29</sup> They apply a *k*-nearest neighbor algorithm with distributional lexical features that aim to capture the association using measures like mutual information between a verb or adjective and a particular noun. The idea is that nouns that occur in similar contexts as animate nouns are more likely to be animate than nouns that occur more frequently in contexts similar to inanimate nouns.

Finally, Moore et al. present an approach that combines a number of animacy classifiers in a voting scheme and aims at an interpretable and

Some experiments with Norwegian nouns,” in *Proceedings of the Workshop on Exploring Syntactically Annotated Corpora* (2005), 24–34; Lilja Øvrelid, “Towards robust animacy classification using morphosyntactic distributional features,” in *Proceedings of the EACL 2006 Student Research Workshop* (2006), 47–54; Lilja Øvrelid, “Linguistic features in data-driven dependency parsing,” in *Proceedings of the Conference on Computational Natural Language Learning (CoNLL 2008)* (2008).

27. Samuel Bowman and Harshit Chopra, “Automatic animacy classification,” in *Proceedings of the NAACL - HLT 2012 Student Research Workshop* (2012).

28. Folgert Karsdorp et al., “Casting a Spell: Identification and Ranking of Actors in Folktales,” in *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)* (2012), 39–50.

29. Jelke Bloem and Gosse Bouma, “Automatic animacy classification for Dutch,” *Computational Linguistics in the Netherlands Journal* 3 (2013): 82–102.

correctable model of animacy classification.<sup>30</sup> They combine a variety of classifiers such as the WordNet-based approach of Evans and Orăsan, dictionary sources and systems for named entity recognition.

The approaches mentioned above present us with a number of problems. First, almost all of them rely heavily on costly, linguistically informed features derived from lexical-semantic databases or syntactic parsing. For most languages in the world, however, we cannot rely on these resources, either because they do not exist, or because their performance is insufficient. Second, animacy detection is often seen as a useful feature for a range of Natural Language Processing techniques, such as anaphora resolution and syntactic parsing. The mutual dependence between these techniques and animacy detection, however, is in fact a ‘chicken and egg’ situation.

Another major problem with the approaches above is, as said above, that they are word *type*-based, which means that the models are generally insensitive to different usages of a particular word in particular contexts. In other words, in most of the literature on automatic animacy detection, a static, binary distinction is made between animate and inanimate. Bowman and Chopra for example, define objects as animate if they are alive and have the ability to move under their own will.<sup>31</sup> Orăsan and Evans define animacy in the context of anaphora resolution: something is animate “if its referent can also be referred to using one of the pronouns he, she, him, her, his, hers, himself, herself, or a combination of such pronouns (e.g. his/her)”.<sup>32</sup> However, as was explained above, these definitions are not necessarily in line with current linguistic and neurological research.<sup>33</sup> Similarly, they are not particularly applicable to the rich and wondrous entities that live in the realm of stories. As was shown above, although a pancake is typically not an animate entity, its animacy depends on the story in which it appears, and even within the story the animacy may change. To accommodate this possibility, I therefore choose to define animacy in terms of Dennett’s intentional stance, which is more dynamic, and which ultimately comes down to

30. Joshua Moore et al., “Animacy Detection with Voting Models,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (2013), 55–60.

31. Bowman and Chopra, “Automatic animacy classification.”

32. Orăsan and Evans, “NP animacy identification for anaphora resolution.”

33. Nieuwland and Van Berkum, “When Peanuts Fall in Love: N400 Evidence for the Power of Discourse.”

the question whether “you decide to treat the object whose behavior is to be predicted as a rational agent”.<sup>34</sup> The system for animacy detection I propose aims to be dynamic, data driven, and token based. It may, but cannot rely too heavily on static world knowledge.

### 3.2.2 EXPERIMENT 1

#### 3.2.2.1 Data Collection

To develop the proposed dynamic, data driven system, I use a corpus of Dutch folktales. One of the reasons to use folktales is that, as Vogels et al. note, “[i]n cartoons or fairy tales [...] inanimate entities or animals are often anthropomorphized”,<sup>35</sup> which means that the material could yield interesting cases of unexpected animacy, as is the case with the pancake in *The fleeing pancake* and the broomsticks in *Fantasia*.

The initial corpus consists of 74 Dutch stories from the collection *Volkssprookjes uit Nederland en Vlaanderen*, compiled by Sinninghe.<sup>36</sup> The collection is composed of Dutch retellings of popular and widespread stories, including such tales as *The Bremen Town Musicians* (ATU 130)<sup>37</sup> and *The Table, the Ass, and the Stick* (ATU 563), as well as lesser-known stories such as *The Singing Bone* (ATU 780) and *Cock, Hen, Duck, Pin, and Needle on a Journey* (ATU 210). This last story is again a clear example where otherwise inanimate objects are animated, as it concerns the adventures of several household items, such as a *pin*, a *hackle*, an *egg*, and a *whetstone*. A digital version of the collection is available in the Dutch Folktale Database from the Meertens Institute (corpus SINVSUNV.20e).<sup>38</sup> Using a single collection for our corpus presents us with a helpful homogeneity with regard to the editor, length of the stories and language use. On the other hand, the collection displays diversity as well in the choice of the stories, which contain fairy tales, legends and nursery rhymes.

34. Dennett, *The Intentional Stance*, 17.

35. Vogels et al., “When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices.”

36. Jacques Sinninghe, *Volkssprookjes uit Nederland en Vlaanderen* (Den Haag, The Netherlands: Kruseman, 1978).

37. The ATU numbers refer to the classificatory system for folklore tales, as designed by Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*.

38. See <http://www.verhalenbank.nl>

All together, the corpus consists of 74,504 words, from 5,549 unique words. Using the annotation tool brat (brat rapid annotation tool), an online environment for collaborative editing,<sup>39</sup> two annotators labeled words for animacy, within the context of the story.<sup>40</sup> All unlabeled words were implicitly considered to be inanimate. The following sentence provides an example annotation in which words in gray bold are considered to be animate:

1. **Jij smid**, **jij** bent de sterkste; hou **je** vast aan de bovenste takken, en dan ga **jij kleermaker** aan **zijn** benen hangen en zo gaan **we** maar door...  
*('You, blacksmith, you are the strongest; hold on to the upper branches and then, you, tailor, will grab his legs and so we go on...')*

Because animacy is interpreted within the context of the story, the same lexical item could be labeled differently in different stories. For example, in the above-mentioned example of the pancake, which occurs in story SINVS076, the pancake is tagged consistently as 'animate'. In another story, SINVS042, where at one point a soldier is baking pancakes, the pancakes do not act, and are thus not labeled as 'animate'. The following sentences show how this was employed in practice.

1. Toevallig stond de deur van de keuken open en de **pannekoek** rolde naar buiten, het veld in, zo hard **hij** maar kon.  
*('Coincidentally the door of the kitchen was open and the pancake rolled outside, into the field, as fast as he could.')* (In: SINVS076)
2. Terwijl **hij** de pannekoek bakte, keek **hij** naar het ding, dat uit de schouw gevallen was...  
*('While he was baking the pancake, he looked at the thing, which had fallen from the hearth...')* (In: SINVS042)

This annotation resulted in 11,542 animate tokens of 743 word types, while implicitly yielding 62,926 inanimate tokens from 5,011 unique

39. Available from: <http://brat.nlplab.org>, see Pontus Stenetorp et al., "BRAT: a Web-based Tool for NLP-Assisted Text Annotation," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (2012), 102–107.

40. On the basis of five stories that were annotated by both annotators we computed an inter-annotator agreement score (Cohen's Kappa) of  $K = 0.95$ .

inanimate words. Because of the context-dependent approach, some words, such as *pancake* and *egg*, occurred in both animate types as inanimate types, because they were labeled as both animate and inanimate in some cases in our corpus. It is telling that of the animate tokens 4,627 (40%) were nouns and proper nouns, while only 6,878 of the inanimate tokens (11%) are nouns. This ratio already serves as an important cue for automatic animacy detection systems, because being a noun is strongly associated with being animate. After cleaning and tokenizing all stories by means of the tokenization module of the Python software package Pattern,<sup>41</sup> all stories were fed to the state of the art syntactic parser for Dutch, Alpino.<sup>42</sup> The features for the linguistically informed models were extracted from the resulting syntactic parses (cf. Section 3.2.2.4).

### 3.2.2.2 Problem description

I formulate the problem of animacy detection as a classification problem where the goal is to assign a label at word level, rather than at lemma level. This label indicates whether the word is classified as animate or inanimate.

### 3.2.2.3 Evaluation

Animate words are far outnumbered by inanimate words in the collection (see Section 3.2.2.1). Reporting accuracy scores would therefore provide skewed results, favoring the majority category. The relative rarity of animate words makes evaluation measures such as the well-known *F1*-score more appropriate. For this reason, I report on the Precision, Recall and *F1*-score on both classes for all experiments.<sup>43</sup> The Precision score is defined as the number of ‘true positives’ (i.e. words classified as animate that are truly animate) divided by the sum of the number of ‘true positives’ and the number of ‘false positives’ (i.e. words that are falsely classified as being animate):

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (3.1)$$

41. Tom De Smedt and Walter Daelemans, “Pattern for Python,” *Journal of Machine Learning Research* 13 (2012): 2031–2035.

42. Gosse Bouma et al., “Alpino: Wide-coverage computational analysis of Dutch,” *Language and Computers* 37, no. 1 (2001): 45–59.

43. Cornelis Van Rijsbergen, *Information Retrieval* (London: Butterworths, 1979).

The Recall score is defined as the number of ‘true positives’ divided by the sum of the number of ‘true positives’ and the number of ‘false negatives’ (i.e. words that are false classified as being inanimate):

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (3.2)$$

Finally, the *F1*-score represents the harmonic mean of Precision and Recall, in which Precision and Recall are evenly weighted:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3.3)$$

Furthermore, while in most of the literature on animacy detection results are only presented for the classification of nouns or noun phrases, I will, while reporting on nouns and noun phrases as well, additionally report on the results for all words in a text.

In real-world applications an animacy detection system will most likely be faced with full texts instead of single words. It is therefore important to construct a training and test procedure in such a way that it mimics this situation as closely as possible. If we would, for example, make a random split of 80% of the data for training and 20% for testing on the word level, we run the risk of mixing training data with test data, thereby making it too easy for a system to rely on words it has seen from the same text. Bowman and Chopra fall into this trap by making a random split in their data on the sentence level.<sup>44</sup> In such a setup, it is highly likely that sentences from the same document are present in both the training data and the test data, making their evaluation unrealistic. To circumvent this problem, I split the data at the story level. I make use of 10-fold cross-validation. All stories are shuffled, and, subsequently, partitioned in ten portions of equal size. In ten iterations, each partition acts as a test set, and the other nine partitions are concatenated to form the training set.

#### 3.2.2.4 Features

The model I propose explores a range of different features and feature combinations including lexical features, morphological features, syntactic features and semantic features.

44. Bowman and Chopra, “Automatic animacy classification.”

*Lexical features* The model takes a sliding-window approach where for each focus word (i.e. the word for which we want to predict whether it is animate or not) the model extracts both  $n$  words to the left and  $n$  words to the right, as well as the focus word itself. In all experiments  $n$  is set to  $n = 3$ . In addition to the word forms, each word in a window is accompanied with its lemma as provided by the output of the syntactic parser Alpino.

*Morphological Features* For each word I extract its part-of-speech tag. For reasons of comparability I choose to use the tags as provided by Alpino, instead of a more specialized part-of-speech tagger. Again, a sliding window approach is employed and the part-of-speech tags for three words left and right of the focus word are extracted, as well as the tag of the focus word itself.

*Syntactic Features* For each word and its  $n = 3$  neighbors to the right and to the left, I extract the dependency tag as provided by the syntactic parser Alpino. Animate entities tend to take the position of subject or object in a sentence which is why this feature is expected – and has proven<sup>45</sup> – to perform rather well.

*Semantic Features* The most innovative feature included in the model is concerned with semantic similarity. In his *Philosophische Untersuchungen* Ludwig Wittgenstein already suggests that “Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache”.<sup>46</sup> This is reflected by the well-known insight in computational linguistics that the meaning of words can be approximated by comparing the linguistic contexts in which words appear. In other words: words that often co-appear with the same set of words will have a more similar meaning. Recently, there has been a lot of interest in procedures that can automatically induce so-called ‘word embeddings’ from large, unannotated collections of texts.<sup>47</sup>

45. Karsdorp et al., “Casting a Spell: Identification and Ranking of Actors in Folktales.”

46. ‘The meaning of a word is its use in the language.’ Ludwig Wittgenstein, *Philosophische Untersuchungen*, vol. 203, suhrkamp taschenbuch wissenschaft (Frankfurt am Main: Suhrkamp Taschenbuch Verlag, 1958), PU 43.

47. See for example Tomas Mikolov et al., “Efficient Estimation of Word Representations in Vector Space,” in *Proceedings of Workshop at ICLR* (2013); Jeffrey Pennington et al., “GloVe: Global Vectors for Word Representation,” in *Proceedings of The 2014 Conference on Empirical*

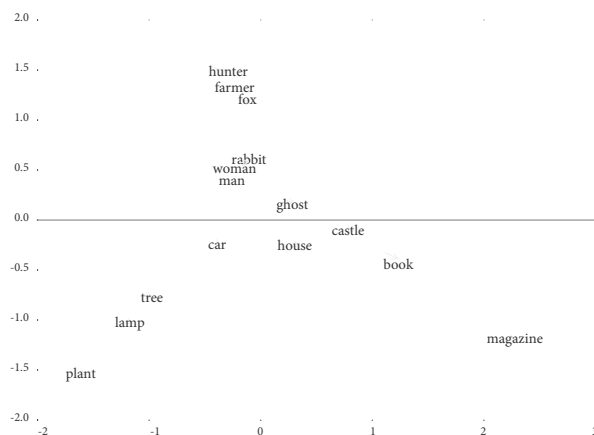


Figure 3.1: Two-dimensional PCA projection of the 300 dimensional word embedding vectors for a number of animate and inanimate words. The horizontal line illustrates the separability between the two classes in the first dimension.

These models typically attempt to learn vector representation with less dimensions than the vocabulary size for each word in the vocabulary, which captures the typical co-occurrence patterns of a word in the corpus. The similarity between words can then be approximated by applying similarity metrics, such as the cosine metric, to these vectors of word embeddings.

I have trained word embeddings with 300 dimensions using the popular skip-gram architecture<sup>48</sup> on the Dutch corpus of COW (COrpora from the Web). COW is a collection of linguistically processed web corpora for English, Dutch, Spanish, French, Swedish and German.<sup>49</sup> The 2014 Dutch corpus contains 6.8 billion word tokens. The idea behind using the word embeddings is that similarities between animate words can be estimated by inspecting the context in which they occur. From this follows, for example, that the word embeddings of an animate word are more similar to those of other animate words, as opposed to the embeddings of inanimate words.

*Methods in Natural Language Processing* (Doha, Qatar, 2014), 1532–1543.

48. Mikolov et al., “Efficient Estimation of Word Representations in Vector Space.”

49. Roland Schäfer and Felix Bildhauer, “Building Large Corpora from the Web Using a New Efficient Tool Chain,” in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)* (Istanbul: ELRA, 2012), 486–493.



To give an illustration of this idea, Figure 3.1 depicts a two-dimensional Principle Component Analysis (PCA) projection of the 300 dimensional word embedding vectors for a number of typically animate and typically inanimate words. The horizontal gray line in the plot illustrates the separability of the animate and inanimate words in the first dimension of the PCA projection. It is interesting to observe that *ghost* is the one closest to all other inanimate entities. Likewise, words such as *castle*, *house* or *car* are often used in figurative language (metonymy), for example to refer to the people owning or living in the castle. Perhaps this ambiguous animacy position is responsible for their position in the first dimension close to real animate entities.

### 3.2.2.5 Models

The model employs a Maximum Entropy classifier with L2 regularization as implemented in the Scikit-Learn Machine Learning Toolkit.<sup>50</sup> In all experiments, the regularization strength parameter  $C$  is set to  $C = 1$ .

I compare nine models which make use of different feature combinations: (1) words, (2) words and Part-of-Speech tags, (3) words, Part-of-Speech tags and lemmata, (4) words, Part-of-Speech tags, lemmata and dependency tags, (5) word embeddings and (6-9) the features in model 1 to 4 with word embeddings.

Although the background corpus is sufficiently large to cover most words in an unseen text, there will always be rare words for which the model does not have learned word embeddings. Therefore, in order to effectively make use of the word embedding vectors, we need a way to deal with out-of-vocabulary items. I adopt a simple strategy where I make use of a primary classifier and a back-off classifier. For models 6 to 9, I augment each word with its corresponding 300 dimension word embeddings vector. In the case of out-of-vocabulary words, the system resorts to a back-off model that contains all features except the word embeddings. For example, a model that makes use of words and word embeddings, will make a prediction on the basis of the word features alone. In case of the model that solely uses the embeddings (model 5), the back-off classifier is a majority vote classifier, which classifies unseen words as inanimate.

50. Pedregosa et al., "Scikit-Learn: Machine Learning in Python."

model	inanimate			animate		
	Precision	Recall	<i>F1</i>	Precision	Recall	<i>F1</i>
1	0.96	0.99	0.98	0.94	0.78	0.85
2	0.97	0.99	0.98	0.94	0.86	0.89
3	0.97	0.99	0.98	0.94	0.86	0.90
4	0.97	0.99	0.98	0.94	0.86	0.90
5	0.98	0.99	0.98	0.93	0.89	0.91
6	0.98	0.99	0.98	0.94	0.90	0.91
7	0.98	0.99	0.99	0.94	0.91	0.93
8	0.98	0.99	0.99	0.94	0.91	0.93
9	0.98	0.99	0.99	0.94	0.92	0.93

Table 3.1: Precision, Recall and *F1*-score for animate and inanimate classes per feature setting for all words. The table presents the results for the following nine models: 1) words; 2) words and Part-of-Speech tags; 3) words, Part-of-Speech tags and lemmata; 4) words, Part-of-Speech tags, lemmata and dependency tags; 5) word embeddings; 6) words and word embeddings; 7) words, Part-of-Speech tags and word embeddings; 8) words, Part-of-Speech tags, lemmata and word embeddings; 9) words, Part-of-Speech tags, lemmata, dependency tags and word embeddings.

### 3.2.2.6 Results

Table 3.1 presents the results for all nine models on the complete data set. For each model I report the Precision, Recall and *F1*-score for animate words and inanimate words.

All models perform well on classifying inanimate words. However, since this is the majority class, it is of course far more interesting to compare the performance of the models on the animate instances. It is interesting to observe that the ‘simple’ *n*-gram word model already performs rather well. Adding more features, such as Part-of-Speech, lemmata, etcetera, only has a positive impact on the recall of the model, while leaving the precision untouched. As can be observed from the table, employing the rather expensive dependency features shows barely any improvement.

The model that only uses word embedding features is one of the best performing models. This is a context-insensitive model that operates on

model	inanimate			animate		
	Precision	Recall	F1	Precision	Recall	F1
1	0.78	0.98	0.87	0.96	0.60	0.74
2	0.86	0.96	0.90	0.93	0.78	0.84
3	0.87	0.96	0.91	0.94	0.80	0.86
4	0.87	0.96	0.91	0.93	0.80	0.86
5	0.90	0.96	0.92	0.93	0.85	0.89
6	0.90	0.97	0.93	0.95	0.85	0.90
7	0.93	0.96	0.95	0.95	0.90	0.92
8	0.93	0.96	0.94	0.95	0.89	0.92
9	0.93	0.96	0.95	0.95	0.90	0.92

Table 3.2: Precision, Recall, and *F1* score for animate and inanimate classes per feature settings for all words tagged as noun. The table presents the results for the following nine models: 1) words; 2) words and Part-of-Speech tags; 3) words, Part-of-Speech tags and lemmata; 4) words, Part-of-Speech tags, lemmata and dependency tags; 5) word embeddings; 6) words and word embeddings; 7) words, Part-of-Speech tags and word embeddings; 8) words, Part-of-Speech tags, lemmata and word embeddings; 9) words, Part-of-Speech tags, lemmata, dependency tags and word embeddings.

the level of the vocabulary, which means that it will predict the same outcome for each token of a particular word type. The high precision and high recall show us that this model has acquired knowledge about which words *typically* group with animate words and which with inanimate words. However, the models that combine the word embeddings with the context sensitive features, such as word *n*-grams or Part-of-Speech tags, outperform the context-insensitive model.

The best performance is achieved by the model that combines the word features, part-of-speech tags and the word embeddings, which has an *F1*-score of 0.93 on animate words and 0.99 on inanimate words. Adding more features does not result in any more performance gain. Table 3.2 presents the results for all nouns and names in the data set. The best performance is again achieved by the model that combines the word features with the part-of-speech tags and word embeddings, resulting in an *F1*-score of 0.92 for animate instances and 0.95 for inanimate instances. Recall from Section 3.2.2.1 that being a noun or proper name already

is an indication for animacy. This could explain the somewhat lower scores for the inanimate class.

### 3.2.3 EXPERIMENT 2

The approach to animacy classification presented in Experiment 1 appears to be successful. In the current section, I employ the classification system to extract all animate entities from unannotated folktales from the Dutch Folktale Database, all of which were not used in the previous experiment.<sup>51</sup> First, it allows us to further the evaluation of the classifier. In a classical evaluation setup – as with the above-described approach – general practice is to train a computational system on some training data. The performance of the system is then evaluated on a held-out test set. Even if we include a separate development set, this setup could run the risk of ‘overfitting’ the data. The annotated corpus contains a reasonably diverse set of stories in terms of genre, yet it is fairly small and rather homogeneous in style. Even though I performed a cross-validation experiment, there is a chance of ‘overfitting’ to the style of the subset of folktales trained on.

#### 3.2.3.1 Data Collection

For the evaluation I make use of a subcollection of folktales from the Dutch Folktale Database. The complete collection consists of about 44,000 folktales,<sup>52</sup> and contains stories from various genres (e.g. fairytales, legends, urban legends, jokes, personal narratives) in standard Dutch and Frisian, as well as a number of dialectal variants. Every entry in the database contains meta-data about the story, including language, collector, place and date of narration, keywords, names, and subgenre. Here, I make use of a subcollection comprising 16,294 stories written in Standard Dutch. The distribution of genres in the subcollection is the following: urban legends ( $n = 2795$ ), legends ( $n = 299$ ), jokes ( $n = 3986$ ), personal narratives ( $n = 693$ ), riddles ( $n = 1626$ ), sagas ( $n = 6045$ ) and

51. <http://www.verhalenbank.nl>

52. Meder, “From a Dutch Folktale Database towards an International Folktale Database”; Meder et al., “Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database.”

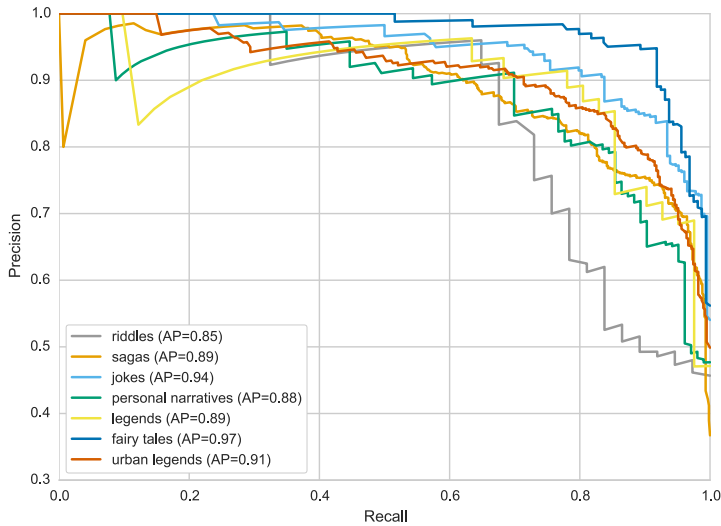


Figure 3.2: Precision-Recall Curves and Average Precision scores per genre.

fairy tales ( $n = 832$ ). I evaluate a random sample of this subcollection ( $n = 212$ ) in which this genre distribution is taken into account.

### 3.2.3.2 Task Description

The proposed definition of animacy allows us to utilize the animacy detection system to extract all characters from a story in a similar vein as Karsdorp et al.<sup>53</sup> The system labels each noun and name in a text for animacy. After removing duplicate words, this produces a set of words that comprises the cast of a story. Without golden standard annotations, however, these character sets can only be evaluated for precision and not for recall. I therefore take an alternative approach, in which for each story I produce a ranking of all its words. The goal here is to allocate the highest positions in these rankings to animate entities.

53. Karsdorp et al., “Casting a Spell: Identification and Ranking of Actors in Folktales.”

### 3.2.3.3 Evaluation

Formulating the task as a ranking problem allows us to evaluate the rankings by means of the *Average Precision* score, which computes the average over precision scores at increasing points of recall:

$$AP = \frac{\sum_{k=1}^n (P(k) \times \text{rel}(k))}{\text{number of relevant items}}, \quad (3.4)$$

where  $\text{rel}(k)$  equals 1 if the item at rank  $k$  is an actual character, zero otherwise, and  $P(k)$  represents the precision at cut-off  $k$ . Naturally, with this evaluation method, the rankings still need to be manually evaluated. By using a rank cutoff and evaluating a sample of all automatically annotated stories, the costly manual labor is reduced to a minimum. All nouns and names in a story are ordered based on the output of the probabilistic decision function of the Maximum Entropy classifier which was used in Experiment 1. After removing duplicate words, this produces a final ranking. The rankings are evaluated with a rank cut-off at 50.

### 3.2.3.4 Results

Figure 3.2 presents the Precision-Recall curve as well as the *Average Precision* score for each genre. The Precision-Recall curve is obtained from computing precision-recall pairs for different probability thresholds. The system performs well, especially on fairy tales ( $AP = 0.97$ ) and jokes ( $AP = 0.94$ ).<sup>54</sup> The lowest performance is measured on riddles ( $AP = 0.85$ ). This lower score is mainly the result of the system's inability to position the word *blondje* ('dumb blond' with a pejorative connotation, which is prevalent in the so-called 'dumb blond jokes') high up the ranking.

## 3.2.4 CONCLUSION

The proposed approach to create a model for animacy classification using linguistically uninformed features proves to be successful. I compared the performance of linguistically informed models (using features

54. An AP of 0.97 means that on average, nearly all actual cast members of a folktale are ranked on top, with the first case of a non-animate entity entering the ranking at about rank 5 or 6 on average.

such as part-of-speech and dependency tags) to models that make use of lower-dimensional, semantic representations of the data. With the exception of the model that solely makes use of these representations, all models benefit from adding these semantic features. The model that requires the least linguistic information (word  $n$ -grams plus word embeddings) outperforms all linguistically informed models (without embeddings). The best results are reported by the model that combines word  $n$ -grams with part-of-speech  $n$ -grams and word embeddings.

### 3.3 CHARACTER TYPE SELECTION BIAS IN FAIRY TALES

In this section I will return to the central questions set out in the introductory section 3.1. My aim is to further explore and build on Joosen's observation that the present-day fairy tale canon is dominated by female protagonists,<sup>55</sup> and more thoroughly investigate the more general question of whether cultural selection and successfulness of fairy tales involves biases towards or *away* from tales that revolve around particular character types. This question will be addressed through a systematic examination of the character cast retrieved from the fairy tales in the 1857 edition of the Grimm's *Kinder- und Hausmärchen*. As will be explained in more detail in Section 3.3.1, the 210 tales in *Kinder- und Hausmärchen* exhibit varying degrees of successfulness. In Section 3.3.2, then, I apply the animacy detection model (set out in Section 3.2) to the Grimm's fairy tales collection, and map out the yielded character lists to a computationally induced folktale character typology. Subsequently, I assign a character type to each of the characters in the Grimm's fairy tale collection, which enables us to employ statistical methods to assess which character types are distinctively associated with either successful or unsuccessful tales (Section 3.3.3).

55. Joosen, "A Translation Far Worse. Canonisation and Adaptation in the Early Dutch and English Translation of the Brothers Grimm's *Kinder- und Hausmärchen*."

## 3.3.1 DATA COLLECTION

The object of study is the Dutch translation of the Brothers Grimm's 1857 edition of *Kinder- und Hausmärchen* by De Vries-Vogel from 1984.<sup>56</sup> The complete collection comprises 210 tales, including nowadays highly popular ones, such as "Cinderella" and "Hansel and Gretel", but also many tales that are relatively unknown, such as "The Prince Who Feared Nothing" and "Maid Maleen". Following Norenzayan et al., each of the collection's tales were considered as popular or unpopular based on their position on a popularity index.<sup>57</sup> In this index, the popularity of fairy tales is measured through web page hits obtained from the search engine Google. Norenzayan et al. propose the following division of successful as opposed to unsuccessful tales:

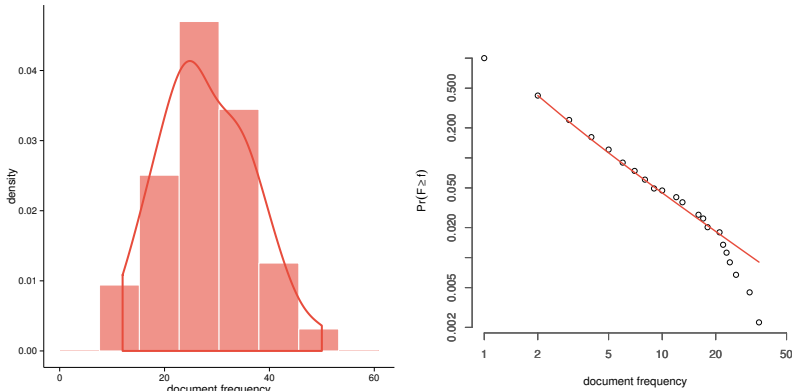
*Culturally successful* ( $N = 21$ ): "The Frog King" ("Iron Henry") (1), "Little Brother and Little Sister" (11), "Rapunzel" (12), "Hansel and Gretel" (15), "The Fisherman and his Wife" (19), "The Brave Little Tailor" (20), "Ashputtelle" ("Cinderella") (21), "Mother Holle" (24), "Little Red Cap" ("Little Red Riding Hood") (26), "The Musicians of Bremen" (27), "The Devil with the Three Golden Hairs" (29), "Little Briar Rose" ("Sleeping Beauty") (50), "King Thrushbeard" (52), "Snow White" (53), "Rumpelstiltskin" (55), "Thousandfurs" (65), "Jorinde and Joringel" (69), "Hans in Luck" (83), "The Singing, Springing Lark" ("Beauty and the Beast") (88), "The Goose Girl" (89), "Snow White and Rose Red" (161).

*Culturally unsuccessful* ( $N = 21$ ): "A Good Stroke of Business" (7), "The Girl Without Hands" (31), "The Magic Table, The Gold Donkey, and the Cudgel in the Sack" (36), "The Knapsack, the Hat, and the Horn" (54), "Frederick and Liza-Kate" (59), "Farmer Little" (61), "Six Who Made Their Way in the World" (71), "The Carnation" (76), "Brother Scamp" (81), "The Golden Children" (85), "The King of the Golden Mountain" (92), "The Spirit in the Bottle" (99), "Bearskin" (101), "Hans My Hedgehog" (108), "The Jew in the Brambles" (110), "The Prince Who Feared Nothing" (121), "The Donkey Lettuce"

56. M. M. De Vries-Vogel, *De sprookjes van Grimm; volledige uitgave* (Weesp, The Netherlands: Unieboek BV – Van Holkema & Warendorf, 1984).

57. Norenzayan et al., "Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives."





(a) Density plot and histogram of the number of characters per story. (b) CCDF of character document frequency.

Figure 3.3: The left subplot (a) displays a density plot and histogram computed over the number of unique characters per story in the  $N = 42$  sub-collection of the Brother's Grimm. The right subplot displays the complementary cumulative frequency distribution of the character story frequency.

(122), “Faithful Ferdinand and Faithless Ferdinand” (126), “Hefty Hans” (166), “The Poor Boy in the Grave” (185), “Maid Maleen” (198).

Although these lists were compiled on the basis of English and German web page hits, the list of successful tales shows a strong correlation with popularity polls in the Dutch language area.<sup>58</sup> For reasons of comparability, then, I will adopt Norenzayan et al.’s classification of successful (i.e. popular) and unsuccessful (i.e. unpopular) tales in this study.

Each of the 42 tales in this collection were submitted to the animacy detection system, which yields a list of characters for each story. After a manual filtering of all classification errors and animate entities of which the syntactic category is not a (proper) noun, we obtain a list of 1,177 characters and 446 unique character types in total. On average, each

58. I.e. the top seven tales in the popularity poll executed by Koman and Meder are all included in the ‘successful group’ (“Snow White” (53), “Cinderella” (21), “Sleeping Beauty” (50), “Little Red Riding Hood” (26), “Hansel and Gretel” (15), “Rumpelstiltskin” (55) and “The Beauty and the Beast” (88)).

tale mentions about 28 characters ( $\sigma = 8.63$ ). Note that this number not necessarily concurs with the actual number of characters in a story, because I did not attempt to resolve any co-references, and multiple words may refer to the same unique character in a story. As can be observed from Figure 3.3(a), the number of characters per story is rather normally distributed. When we investigate the frequency distribution of the characters, we find that a large proportion of characters occurs only in a few stories or in a single story and only a few yet significant number of characters occur in many different stories. In fact, according to the bootstrap procedure proposed by Clauset et al., the distribution exhibits a good fit ( $p = 0.7$ ) to a power-law model with an exponent of  $\alpha = 2.24$  (cf. Figure 3.3(b)).<sup>59</sup>

### 3.3.2 CHARACTER TYPOLOGY

Let us now consider the question whether successful tales revolve around character types that are distinctively different from the character set in unsuccessful tales. In order to address this question, we first need to establish a mapping between each unique retrieved character and a particular character type. A character type is defined as a semantically and/or functionally coherent group of characters. The agglomeration of all character types constitutes a folktale character typology, onto which the unique characters in the Grimm's tales will eventually be mapped. To create this typology, I used the list of folktale characters that have been extracted from the Dutch Folktale Database in Section 3.2.3. The classification errors, duplicates and all entities that occur less than a thousand times in COW were removed, yielding a list of 4,548 unique characters. Rather than manually developing a character typology – which is both labor-intensive as well as prone to subjective assessments – I choose to construct a character typology by means of an automatic procedure, in which character types are formed by means of a cluster algorithm.

The word embeddings that are used as features in the animacy classifier can be employed to describe the similarities and dissimilarity

59. See Aaron Clauset et al., "Power-law distributions in empirical data," *SIAM Review*, no. 4 (2009); and Chapter 5 for a more in depth explanation of power-laws and the bootstrap procedure. doi:10.1137/070710111.

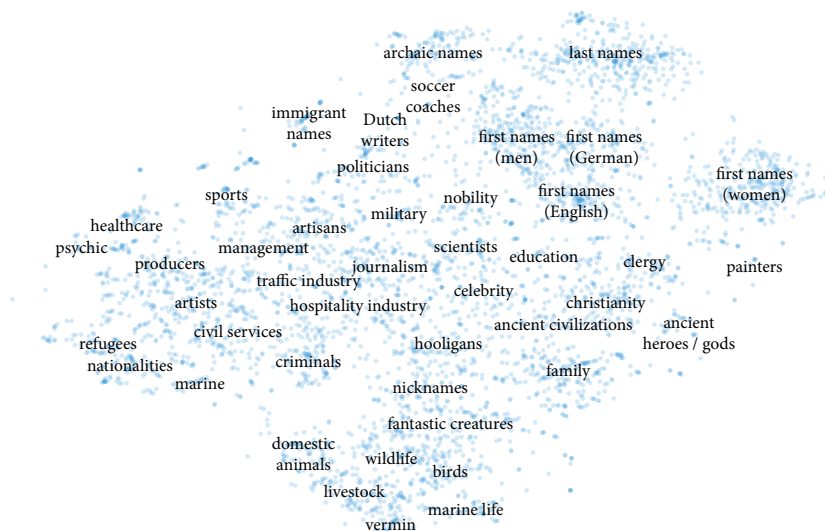


Figure 3.4: Visualization of the characters extracted from Dutch stories of the Dutch Folktale Database based on their word embeddings using t-SNE.

ties between the extracted animate entities. Figure 3.4 presents a two-dimensional semantic map that depicts the (dis)similarities between all extracted animate entities.<sup>60</sup> The dimension reduction was performed using t-Distributed Stochastic Neighbor Embedding (t-SNE).<sup>61</sup> The map discloses a rich diversity of animate entities grouped into highly semantically coherent clusters. The cluster at the bottom of the map, for example, represents a grouping of all kinds of animals. This cluster comprises manifold subtle sub-clusters describing more specific positions in the animal taxonomy, e.g. birds and livestock, marine life, vermin, wildlife, and domestic animals. The groupings to the left are occupied by characters of different professions. There is a large number of characters from the hospitality industry (e.g. waiter and cook), as well as from the transport sector (e.g. chauffeur and train conductor), and from health-care services (e.g. doctors, patients, and surgeons).

Another cluster worth mentioning is the one that is occupied by

60. Readers are invited to view an interactive version of the map at the following web-address: [fbkarsdorp.github.io/animacy-detection](https://fbkarsdorp.github.io/animacy-detection).

61. Laurens Van der Maaten and Geoffrey Hinton, "Visualizing High-Dimensional Data Using t-SNE," *Journal of Machine Learning Research*, 2008, 2579–2605.

characters from Christianity. This cluster is noteworthy because, like the animal cluster, it is clearly structured into various hierarchically ordered clusters with several emerging subgroups. One subgroup entails religious or even Christian professions, such as ‘bishops’ and ‘vicar’. From there, a link via ‘catholics’ and ‘protestants’ leads to the more general ‘believers’ and ‘followers’. This mini-node bifurcates into two different nodes. First, we find a cluster containing words designating followers of different religions, such as ‘Jew’ and ‘Muslim’. This cluster branches off to a ‘religious fringe’ node, containing ‘cult’, ‘satanist’ and ‘Freemasons’. The other cluster connected to the mini-node of ‘believers’ and ‘followers’ is structurally complex, starting with such terms as ‘people’ and ‘believers’ but continuing through ‘Satan’ and ‘Lucifer’ to ‘angels’ and ‘guardian angels’. These words form again a bridge towards more esoteric creatures, such as ‘nature spirits’. Numerous other groupings can be discerned in Figure 3.4, ranging from broad categories, such as ‘fantastic creatures’, ‘criminals’ and ‘hooligans’, to more specialized clusters such as the one containing soccer coaches or people with psychic abilities (e.g. ‘clairvoyants’, ‘mediums’ and ‘fortune-tellers’).

Combined with the strength of t-SNE to position the characters on a two-dimensional map, the word embeddings yield a powerful representation. However, if we want to assign a character type to each character extracted from the fairy tales by the Brothers Grimm, we need to ‘hard-cluster’ the characters into  $k$  types. The above description is only part of the extremely rich network of associations this semantic map displays. Therefore, popular clustering algorithms such as  $k$ -means seem less suitable as they require to fix the number of clusters  $k$  in the data upfront, which is far from straightforward to do. To circumvent the problem of cherry-picking  $k$ , I employ the cluster algorithm DP-means as proposed by Kulis and Jordan, which is a non-parametric Bayesian variant of the  $k$ -means algorithm and based on a Dirichlet Process mixture model.<sup>62</sup> The algorithm behaves similarly to  $k$ -means, yet it does not require that  $k$  is fixed upfront, but it dynamically forms new clusters depending on whether the minimal distance between a data

62. Brian Kulis and Michael I. Jordan, “Revisiting k-means: New Algorithms via Bayesian Nonparametrics,” in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (2012), 513–520.

point and any of the existing centroids exceeds the threshold parameter  $\lambda$ .

Other than the  $k$ -means algorithm, which depends on the initial clustering of the data, the order in which the data is presented to the DP-means algorithm potentially has a severe impact on the resulting clustering. Under the assumption that the quality of word embeddings is higher for words that occur more frequently, I present the data to the cluster algorithm ordered by word frequency in COW in descending order. Note that the cluster penalty parameter  $\lambda$  of DP-means controls for the specificity of the character types with lower values typically producing a larger number and also more detailed character types. In order to keep the typology manageable for interpretation and to reduce the risk of overfitting the data, I choose to set  $\lambda$  at the rather conservative value of 0.7. The algorithm was set to run for 200 iterations. Under these settings, the algorithm produced the 24 clusters depicted in Figure 3.5. The hierarchical tree was computed on the basis of the cosine distances between the centroid representations learned by DP-means.

Similar to the character map in Figure 3.4, the cluster tree discloses multiple semantically coherent clusters of characters. Clusters 11, 7 and 12, for example, deal with various animals, which are grouped under 'wildlife' (e.g. 'lion', 'hippopotamus', 'crocodile'), 'animals' (e.g. 'chicken', 'stork', 'hamster') or 'domestic animals' (e.g. 'poodle', 'guinea pig', 'rabbit'). Another example is the red subtree, which represents an abstract cluster of proper names and entails four sub-clusters with first or last names from different nationalities. The upper subtree branches off into three clusters consisting of 'affective names' (e.g. 'little boy', 'rascal', 'little kid'), 'supernatural entities' (e.g. 'witch', 'wizard', 'elves') and 'family related names' (e.g. 'daughter', 'brother', 'mother'). The above-mentioned grouping of religious characters (predominantly from Christianity) is found in cluster 6, which is connected to cluster 19 consisting of various nobility characters, such as 'duke', 'king', 'queen', and 'prince'.

### 3.3.3 RESULTS

The character typology discussed in the previous section serves as the methodological base for the main research question of this chapter on

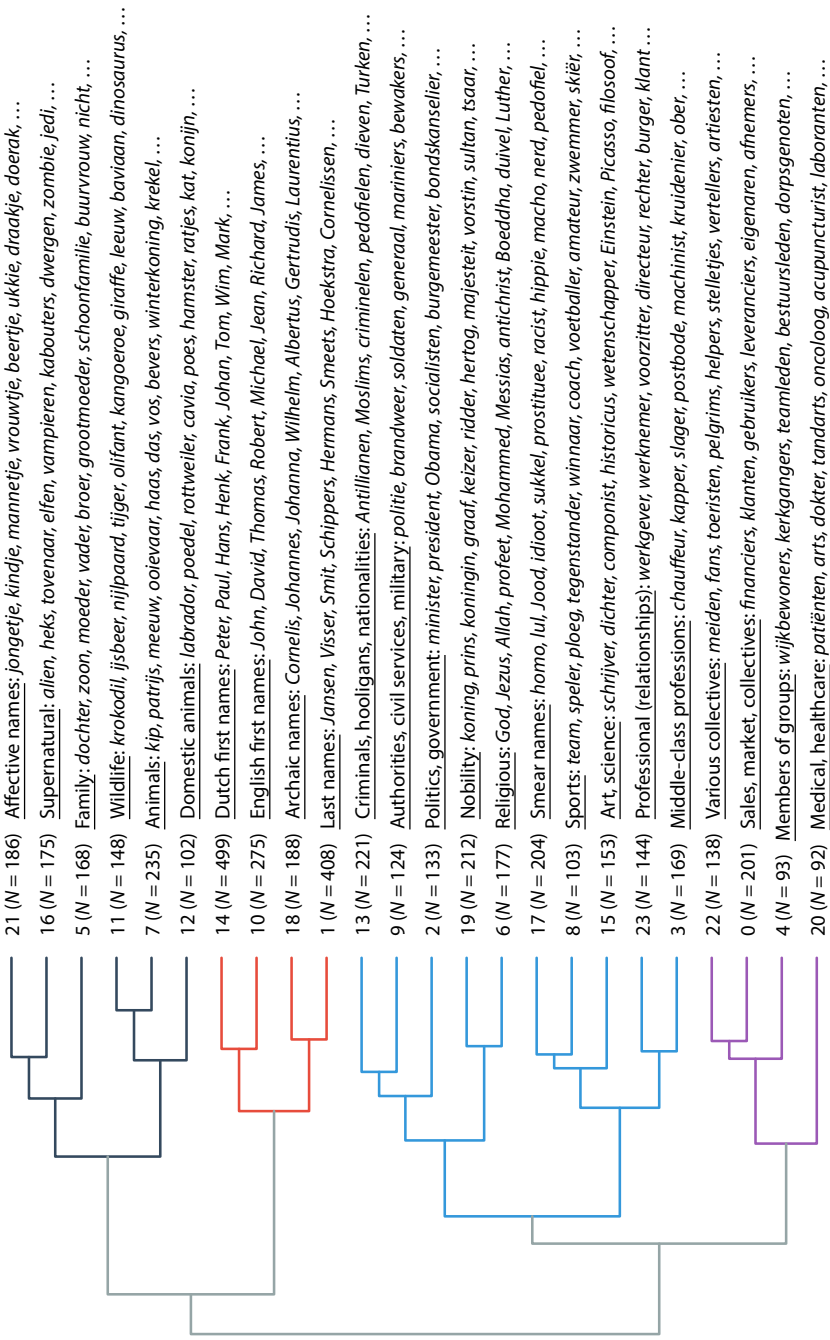


Figure 3.5: Hierarchical cluster tree of the 24 induced character types.

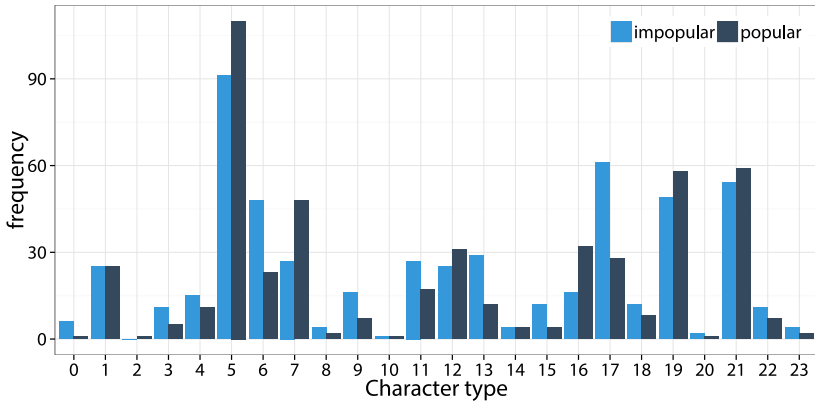


Figure 3.6: Distribution of character types for successful and unsuccessful  $N = 42$  fairy tales from *Kinder- und Hausmärchen* (1857) by the Brothers Grimm.

character type biases in story transmission. All characters extracted from the Brothers Grimm's fairy tale collection were mapped to a type in the character typology. Figure 3.6 displays the distribution of the character types for both successful and unsuccessful tales. It can be observed from the plot that character type 5, which deals with family relationships, dominates both successful and unsuccessful tales, with a preference for popular tales. We find, furthermore, that the religious character type (6) appears to be more frequently used in unsuccessful than in successful tales. Successful tales, by contrast, display a preference for animal characters (type 7). Although frequency counts such as these are insightful and suggestive, to address the question of whether particular characters display *significant* preferences for either successful or unsuccessful fairy tales, we need, however, a more rigorous and statistically sophisticated approach.

The approach I propose aims to quantify whether a particular character type is distinctively associated with either successful or unsuccessful fairy tales by means of Fisher's exact test.<sup>63</sup> Starting from the null hypothesis that the successfulness of tales and a particular character type are independent from each other, we aim to detect deviations from

63. Ronald Fisher, "On the Interpretation of  $\chi^2$  from Contingency Tables, and the Calculation of P," *Journal of the Royal Statistical Society* 85, no. 1 (1922): 87–94.

	character type 5	other types	row totals
successful tales	110	387	497
unsuccessful tales	91	459	550
column totals	201	846	1047

Table 3.3: The distribution of character type 5 in successful and unsuccessful fairy tales from *Kinder- und Hausmärchen* (1857) by the Brothers Grimm.

this null hypothesis, which are indicative for statistically significant distinctive character types. Fisher's exact test has several advantages over other quantities and statistical tests that can be employed to measure association strengths between binomial variables (e.g.  $\chi^2$ -test, pointwise mutual information<sup>64</sup> or Dunning's log-likelihood coefficient<sup>65</sup>). One of the most important characteristics for my purposes, is that being an exact test, Fisher's exact test makes no assumptions about the distribution of the data and provides exact values even for very small sample sizes.

To compute the distinctiveness of a particular character type for successful or unsuccessful tales, the following frequencies need to be collected: (a) the frequency of the character type in successful tales, (b) the frequency of the character type in unsuccessful tales, the summed frequency of all other character types (i.e. the total number of characters minus the frequency of the target character type) in both successful tales (c) and unsuccessful tales (d). These numbers, then, can be entered in a  $2 \times 2$  contingency table and submitted to the Fisher exact test. To illustrate the procedure, in Table 3.3 I provide the required frequencies for character type 5 which deals with family relationships. Character type 5 occurs 110 times in successful tales and 91 times in unsuccessful tales. There are  $497 - 110 = 387$  characters with a different character type in successful tales and  $550 - 91 = 459$  differently classified characters in unsuccessful tales. Submitting these numbers to Fisher's exact test returns a rather small  $p$ -value of 0.013. This number tells us that the probability of observing this or an even more imbalanced ratio by chance is about 1.2%. Since this number is well below the commonly

64. Kenneth Ward Church and Patrick Hanks, "Word Association Norms, Mutual Information, and Lexicography," *Computational Linguistics* 16, no. 1 (1990): 22–29.

65. Ted Dunning, "Accurate methods for the statistics of surprise and coincidence," *Computational Linguistics* 19, no. 1 (1993): 61–74.



used significance level of 5%, we can safely conclude that the observed imbalance is statistically significant; successful tales revolve significantly more often around family characters than unsuccessful tales.

Let us now turn to the main results, which are presented in Table 3.4. The table shows for each character type how often it occurs with successful and unsuccessful tales. The column ‘preference’ displays whether a character type prefers successful or unsuccessful tales, depending on whether it occurs more frequently with successful or unsuccessful tales. The final two columns provide (i) the odds ratio, which is computed as the unconditional Maximum Likelihood Estimate, and (ii) the  $p$ -value corresponding to the preference computed by Fisher’s exact test. The rows colored gray indicate character types that are significantly distinctive for either successful or unsuccessful tales, given a 5% significance level.

Three character types significantly prefer successful tales: (i) character type 5 which deals with family relationships ( $p = 0.013$ ), (ii) the animal character type 7 ( $p = 0.002$ ), and (iii) character type 16 which is occupied by esoteric creatures and characters with supernatural abilities ( $p = 0.005$ ). Note that all three character types also occur with unsuccessful tales, yet they are highly distinctive for successful tales. Unsuccessful tales, by contrast, revolve significantly more often around characters that are grouped with (i) religious characters (type 6,  $p = 0.006$ ), (ii) criminals, hooligans and nationalities (type 13,  $p = 0.012$ ), and (iii) characters that are referred to with smear words (type 17,  $p = 0.001$ ).

The analysis presented here exposed a number of significantly distinctive character types for successful and unsuccessful fairy tales from *Kinder- und Hausmärchen* by the Brothers Grimm. The results are, however, less informative in terms of the predictive power of character types for either successful or unsuccessful tales: To what extent can we predict the successfulness of a tale on the basis of the characters it revolves around? To address this question, I perform a classification experiment where the goal is to assign a label  $y \in \{\text{successful}, \text{unsuccessful}\}$  to each fairy tale. In this experiment, each story  $s$  is represented as a vector  $\vec{a} = (c_1, c_2, \dots, c_C)$ , where  $c_i$  represents the occurrence count of character type  $i$  in story  $s$ , and  $C$  represents the total number of character types in the character typology (i.e. 24). Given that the data collection is rather small, I make use of a ‘leave-one-out’ cross-validation setup in which

type	unsuccessful	successful	preference	odds	<i>p</i>
0	6	1	unsuccessful	5.471	0.080
1	25	25	–	0.899	0.696
2	0	1	successful	0.000	0.475
3	11	5	unsuccessful	2.008	0.145
4	15	11	unsuccessful	1.239	0.370
5	91	110	successful	0.698	0.013
6	48	23	unsuccessful	1.971	0.006
7	27	48	successful	0.483	0.002
8	4	2	unsuccessful	1.813	0.392
9	16	7	unsuccessful	2.097	0.073
10	1	1	–	0.903	0.775
11	27	17	unsuccessful	1.458	0.148
12	25	31	successful	0.716	0.141
13	29	12	unsuccessful	2.250	0.012
14	4	4	–	0.903	0.691
15	12	4	unsuccessful	2.749	0.057
16	16	32	successful	0.435	0.005
17	61	28	unsuccessful	2.089	0.001
18	12	8	unsuccessful	1.363	0.328
19	49	58	successful	0.740	0.085
20	2	1	unsuccessful	1.810	0.538
21	54	59	successful	0.808	0.166
22	11	7	unsuccessful	1.429	0.311
23	4	2	unsuccessful	1.813	0.392

Table 3.4: Character types distinguishing between successful and unsuccessful tales in the Brothers Grimm fairy tales.

	Precision	Recall	F1 score
successful tales	0.76	0.80	0.79
unsuccessful tales	0.81	0.77	0.78

Table 3.5: Classification results for successful versus unsuccessful tales by the Brothers Grimm. The table provides the precision, recall and F1 score for successful and unsuccessful stories.

in each iteration, a single story is removed from the data collection to act as a test item, and the remaining stories are concatenated to form the training data. As a classifier, I employ Scikit-Learn’s implementation of a Maximum Entropy classifier with L2 regularization and the regularization strength parameter  $C$  set to  $C = 1$ .<sup>66</sup>

Table 3.5 presents the classification results. The table reports on the precision, recall and F1 score for the group of successful and unsuccessful tales. Given that a chance-level baseline model would approximately predict a story to be ‘successful’ or ‘unsuccessful’ in 50% of the times, the reported F1 scores of 0.79 and 0.78 for successful and unsuccessful tales respectively, are remarkably high. The classification results support the hypothesis that character types play an important role in the cultural successfulness of fairy tales.

### 3.4 DISCUSSION

Considering the large number of tales in the 1857 edition of *Kinder- und Hausmärchen* that did not become successful parts of present-day culture, it is somewhat inaccurate to ask ‘why fairy tales stick’; the main question addressed in this chapter, then, is that of *which* fairy tales stick and what factors cause their popularity. Taking an evolutionary perspective on story transmission, this question can be translated into which story elements form attractors or biases that accrue a story’s chances of survival into the next generation. In particular, this chapter has further looked into the question of whether there is something like a ‘character bias’ at play in story selection.

66. Pedregosa et al., “Scikit-Learn: Machine Learning in Python.”

This question was addressed in the following manner. First, I developed a computational system with which the character cast of a story can be automatically extracted. This system, which was presented in Section 3.2, aims to extract the cast using a mechanism that detects animate behavior of entities in texts. It was shown that the system produces high-quality results without having to rely on costly linguistic features. Instead, the model employs more parsimonious features, which include a semantic vector representation of words, word  $n$ -grams, and, optionally, part-of-speech tags. After a thorough evaluation of its performance, the model was deployed to extract the character casts from all stories in the Dutch Folktale Database written in Standard Dutch.

As a second step, I developed a folktale character typology on the basis of characters that were extracted from the Dutch Folktale Database (cf. Section 3.3.2). Using the characters' semantic vector representations as a base, I created an interactive folktale character map which describes the similarities and dissimilarities between characters. The map lays bare a rich diversity of semantically homogeneous character groupings and can be used by researchers to explore the Dutch Folktale Database from new perspectives. In order to obtain a hard-clustering of  $k$  character types, I submitted the word embeddings corresponding to the extracted characters to the non-parametric cluster algorithm DP-means. It was shown that this algorithm, combined with the characters' word embeddings, yields a variety of semantically coherent clusters which serve as character types in the character typology.

The animacy detection system and the folktale character typology served as the methodological base to investigate the question of whether cultural selection and successfulness of fairy tales is subject to character type biases. With this methodological base, I was able to turn to the third step, which involved scrutinizing the automatically retrieved character cast of a selection of successful and unsuccessful fairy tales from the Grimm's 1857 edition of *Kinder- und Hausmärchen*. I proposed that, by labeling each character in this collection with a character type from the character typology and subsequently looking at how strongly they are associated with either successful or unsuccessful stories, one can statistically evaluate which character types serve as attractors for either successful or unsuccessful tales.

Building further on this evaluation, I provided empirical evidence

for the existence of a character type bias in the cultural selection of fairy tales. It was shown that (i) characters with names that refer to family relationships, (ii) animal characters, and (iii) characters that exhibit extraordinary or even supernatural powers are significantly distinctive for culturally successful tales. Interestingly, the significant preference for these three character types ties in with findings from experimental studies of story transmission, where it was shown that story selection and story alteration are subject to social and cognitive selection pressures.<sup>67</sup> First, the distinctiveness of characters revolving around family relationships, for example, concurs with experimentally observed transmission biases for social information.<sup>68</sup> Second, there also seems to be a preference for characters that are in line with what in various experimental psychological studies is called ‘minimally counterintuitive concepts’. Characters with extraordinary or supernatural powers, such as witches, are minimally counterintuitive in that most of their characteristics are in accordance with the ontological assumptions we have of humans, but a few of their characteristics are not consistent with these assumptions (e.g. the ability to fly on brooms and brew magic potions). Similarly, animal characters in fairy tales by and large exhibit the appearance and behavior one generally assumes they have, but they are counterintuitive in that they are staged as anthropomorphic beings that have a certain intentionality and the ability to speak.<sup>69</sup> Such concepts have a mnemonic advantage over completely counterintuitive or intuitive concepts, and as such are more likely to be accurately transmitted.<sup>70</sup> Interestingly, these

67. See e.g. Owens and Bower, “The “soap opera” effect in story recall”; Diehl et al., “Story Recall and Narrative Coherence of High-Functioning Children with Autism Spectrum Disorders”; Mesoudi and Whiten, “The Multiple Roles of Cultural Transmission experiments in Understanding Human Cultural Evolution”; Eriksson and Coultas, “The Advantage of Multiple Cultural Parents in the Cultural Transmission of Stories.”

68. See e.g. Mesoudi et al., “A bias for social information in human cultural transmission.”

69. Cf. Justin Barrett, “Coding and Quantifying Counterintuitiveness in Religious Concepts: Theoretical and Methodological Reflections,” *Method & Theory in the Study of Religion* 20, no. 4 (2008): 308–338.

70. See e.g. Barrett and Nyhof, “Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials”; Norenzayan et al., “Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives”; M. Afzal Upal et al., “Contextualizing Counterintuitiveness: How Context Affects Comprehension and Memorability of Counterintuitive Concepts,” *Cognitive Science* 31, no. 3 (2007): 415–439; Harmon-Vukić and D. Jason Slone, “The Effect of Integration on Recall of Counterintuitive Stories,” *Journal of Cognition and Culture* 9, no. 1 (2009): 57–68; Justin Barrett et al., “Counterintuitiveness in Folktales: Finding the Cognitive Optimum,”

findings are in line with suggestions made by Boyer, who has argued that, in the group of counterintuitive concepts, characters (or ‘intentional agents’) enjoy the strongest transmission advantage.<sup>71</sup> The results presented in this chapter make a strong case for the advantageous position of fairy tales that revolve around minimally counterintuitive agents.

By contrast, unsuccessful tales tend to be distinctively characterized by the occurrence of (i) religious characters, (ii) criminals, hooligans and nationalities, and (iii) characters that are referred to with smear words. As impopularizing traits have largely been ignored in story transmission research, it is somewhat more difficult to relate these findings to existing accounts. It is interesting to observe, however, that the most distinctive character types in unsuccessful tales are either part of a religious sphere – a theme that has lost its cultural value in the context of fairy tales over the course of the 20<sup>th</sup> century – or mostly refer to generic groups with a rather negative connotation.

Finally, the findings were further strengthened by the addition of a classification experiment. The experiment classified tales as successful or unsuccessful based on the types of characters that were found in their cast. Results showed that character types serve as relatively strong predictors of a story’s successfulness, which accrues positive evidence that cultural selection of stories involves character bias.

Obviously, character bias alone cannot fully account for the successfulness of particular fairy tales, and, consequently, there is more to investigate concerning story transmission and selection biases. The current study (and other research on the mnemonic advantages of minimally counterintuitive concepts) is primarily concerned with lexical biases, i.e. single words or phrases that impact a story’s successfulness. Taking a slightly different perspective, Loewenstein and Heath investigate the impact of more abstract story content on the propagation of stories.<sup>72</sup> They document evidence that stories with a so-called

*Journal of Cognition and Culture* 9, no. 3 (2009): 271–287; Upal, “Memory, Mystery and Coherence: Does the Presence of 2–3 Counterintuitive Concepts Predict Cultural Success of a Narrative?”

71. Pascal Boyer, *Religion explained. The evolutionary origins of religious thought* (New York: Basic Books, 2001).

72. Jeffrey Loewenstein and Chip Heath, “The Repetition-Break plot structure: A cognitive influence on selection in the marketplace of ideas,” *Cognitive Science* 33 (2009): 1–19; Jeffrey Loewenstein et al., “The Repetition-Break Plot Structure Makes Effective Television Advertisements,” *Journal of Marketing* 75, no. 5 (2011): 105–119.

‘Repetition-Break’ plot structure (i.e. one in which similar plot elements are repeated followed by a contrasting, sudden break) enjoy transmission advantages over stories without such plots. Besides such *content-based* transmission biases, the selection of cultural artifacts such as stories can also be determined by *context-based* biases, in which, for example, social factors impact the success of a story.<sup>73</sup> Publishers, parents, teachers, translators and ‘retellers’ choose to devote special attention to particular tales. The selection of these tales might be motivated by content-based factors, but it might just as well result from the copying of other people’s behaviors. The impact of social, context-based biases on cultural selection and transmission of stories will be addressed further in the next two chapters.

73. For the distinction between context-based and content-based bias, cf. Chapter 5 and Joseph Henrich and Richard McElreath, “The evolution of cultural evolution,” *Evolutionary Anthropology: Issues, News, and Reviews* 12 (2003): 123–135.

---

## CHAPTER 4

# GRADUAL ACCUMULATION OF MODIFICATION IN CHILDREN'S LITERATURE

*"Pas op! Je kent het verhaal toch?"*

CAROLINE ELLERBECK, *Roodkapje*

### 4.1 INTRODUCTION

Many of the stories that occur in children's literature have an interesting property in common with folktales: they are retold and adapted. Clearly, folktales – especially those from oral culture – depend on retellings in order to secure their survival in culture. Each retelling of a story forms the endpoint of a chain of retellings. The continuation of the chain of retellings, or in other words, the lifespan of a story, depends on whether the story will be further retold; if not, its lifespan ends and the story becomes extinct. As Burkert aptly describes: "A tale becomes traditional not by virtue of being created, but by being retold and accepted; transmission means interaction [...]"<sup>1</sup> When it comes to stories in children's literature, which exist mostly in printed or electronic form and, as such, substantially differ from oral folktales, it could be argued that their survival depends first and foremost on the conservation of their printed or electronic records (e.g. in library collections, on a bookshelf or on the Internet). However, scholars have noted that children's stories are retold and adapted as well, and it is stressed that, like oral folktales, children's stories *need* to be recurrently adapted to

1. Walter Burkert, *Structure and History in Greek Mythology and Ritual* (Berkeley, Los Angeles, London: University of California Press, 1982), 2.



secure their position in culture.<sup>2</sup> This is due to the fact that retellings in children's literature have important cultural functions, one of which is to "initiate children into aspects of a social heritage, transmitting many of a culture's central values and assumptions and a body of shared allusions and experiences".<sup>3</sup> Since a culture's central values and ideologies are never fixed but always in transition, adaptations and transpositions are necessary to maintain the relevance of stories for new audiences in changing historical contexts.<sup>4</sup>

The consequence of considering a story as a *re-telling* is that it is positioned in relation to one source or even multiple source texts. In line with Stephens and McCallum, I will refrain from using the term 'source text' and apply the more neutral term 'pre-text'.<sup>5</sup> Oral folktales are generally derived from pre-texts from the immediate previous generation,<sup>6</sup> and each retelling produces a slightly (or very) different story, due to e.g. subconscious factors such as memory capacity<sup>7</sup> or conscious factors such as the creativity of the storyteller.<sup>8</sup> In other words, the oral transmission of folktales is best described as a process of 'gradual accumulation of modifications': oral folktales "are transmitted from individual to individual and generation to generation via social learning mechanisms such as imitation, thus forming lineages of similar traits",<sup>9</sup> with each individual potentially introducing modifications which may be adopted and further transmitted in the next retelling.

When it comes to the transmission of literary texts and children's literature (rather than oral folktales), we may be inclined to believe that transmission is not 'gradually accumulating', but that adaptations in

2. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*; Collins and Ridgman, *Turning the page, Children's Literature in Performance and the Media*; Sylvie Geerts and Sara van den Bossche, eds., *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*, Ginkgo (Ghent, Belgium: Academia Press, 2014).

3. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*, 3.

4. Collins and Ridgman, *Turning the page, Children's Literature in Performance and the Media*.

5. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*.

6. See David C. Rubin, *Memory in Oral Traditions. The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes* (New York, Oxford: Oxford University Press, 1995).

7. Cf. Bartlett, *Remembering. A Study in Experimental and Social Psychology*.

8. Cf. Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction*.

9. Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*, 90.

retellings are derived from some original (printed) text or ‘urtext’ rather than from younger retellings. This would mean that all retellings of, for instance, “Red Riding Hood” are second generation stories that are most likely immediately derived from the version by the Brothers Grimm. However, Stephens and McCallum hypothesize that:

A familiar fairy tale such as *Cinderella* may derive from Perrault’s version, or the Disney film, or British pantomime tradition, but is more apt to borrow freely from amongst these and from versions of them circulating orally. Even where there is a strong pre-text such as Perrault, retellers are most likely to use intermediate versions—to produce a retelling of a retelling.<sup>10</sup>

In other words, these ‘retellings of retellings’ in children’s literature too form a network (rather than a tree) of ‘intermediate’ story versions without a clear ‘root story’ from which all subsequent versions have supposedly sprang. From a theoretical perspective, the effort spent to identify such a ‘first telling’, which has been one of the main occupations of many folklore scholars in the 19<sup>th</sup> century and scholars of the ‘Finnish’ or historical-geographic school, is questionable at best.<sup>11</sup> Retellings of *Cinderella* are influenced by intermediate retellings of the story and the alleged ‘original’ written by Perrault is no different. Perrault’s story also represents some intermediate state, because his version is guided by what Frank, influenced by Bourdieu, calls a *narrative habitus*:

Narrative habitus is a disposition to hear some stories as those that one ought to listen to, ought to repeat on appropriate occasions, and ought to be guided by. [...] [N]arrative habitus involves a *repertoire* of stories that a person at least recognizes and that a group shares. These stories are known against an unseen background of all the stories that person does not know and stories that do not circulate within any particular group.<sup>12</sup>

Perrault’s narrative habitus was formed by – obviously against the background of a much broader literary context – fairy tales written by

10. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Meta-narratives in Children’s Literature*, 4.

11. For a more recent example of such effort, see Jamshid Tehrani, “The Phylogeny of Little Red Riding Hood,” *PLoS ONE* 8, no. 11 (2013).

12. Arthur W. Frank, *Letting Stories Breathe. A Socio-Narratology* (Chicago, London: The University of Chicago Press, 2010), 53.

(mostly female) writers such as Marie-Jeanne L'Héritier, Marie-Catherine d'Aulnoy and Jean de La Fontaine.<sup>13</sup> Perrault had a 'Mother Goose' of his own, so to speak.<sup>14</sup> Stories are dependent on other texts, for which narratologists use the term *intertextuality*. Frank prefers the term *resonance* – a term that emphasizes that intertextuality can be a personal experience – and describes stories as “textures of resonance”.<sup>15</sup> As it turns out “[i]t is quite difficult to invent a tale; even a new creation will inevitably merge with a stream of tales heard before, and thus become a variant of what has already been around”.<sup>16</sup> The quest for a 'first telling' might seem to be a hopeless and theoretically unsound endeavor; determining *the* pre-text(s) of a retelling is not a trivial task either. We can unequivocally select a pre-text or group of pre-texts for merely a handful of retellings; most are indeterminate for which no decisive answer can be provided. Nevertheless, however precisely or imprecisely, retellings are based on pre-texts and some pre-texts are more likely to have formed a source of inspiration for a particular retelling than others.

The aim of the present chapter is to investigate this simple, yet crucial hypothesis of Stephens and McCallum, which holds that retellers of children's literature most likely base their retelling on 'intermediate retellings'. What is meant by such intermediate retellings, however, may cover an indeterminate and considerably large time span. As such, this study aims to come to a more precise understanding of what is meant by 'intermediate'.

In its attempt to determine whether retellings in children's literature are based on intermediate versions or urtexts, and whether the transmission of children's literature can be characterized as a 'gradual accumulation of modifications', the present study can be situated in a large body of research that investigates transmission of cultural traits and cultural evolution. Mesoudi sets out the sufficient conditions to consider cultural change as an evolutionary process, being (i) variation, (ii) competition (differential fitness), and (iii) inheritance.<sup>17</sup> In children's

13. Marina Warner, *From the Beast to the Blonde. On Fairy Tales and their Tellers* (London: Vintage, 1995), 76; Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*, 44.

14. Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*, 44–45.

15. Frank, *Letting Stories Breathe. A Socio-Narratology*, 37.

16. Walter Burkert, *Creation of the Sacred. Tracks of Biology in Early Religions* (Cambridge, Massachusetts, London, England: Harvard University Press, 1996), 70.

17. Alex Mesoudi, "How cultural evolutionary theory can inform social psychology and

literature, variation exists in the sense that multiple retellings of a story exist simultaneously, which may differ on a concrete level (different wordings, character names, etc.) or a more abstract level (story elements, plot structure, cultural framing, etc.). Competition, then, is the result of certain limitations that go hand in hand with culturally acquired information: even if all existing retellings of a story are available to a reteller, this reteller can only produce *one* retelling at a time. This new retelling may inherit traits of multiple existing story versions, but it is impossible to simultaneously produce (or retell) all existing versions at the exact same time. Thus, the theoretical consequence is that each time an author engages in producing a retelling, (all parts of) existing retellings are in competition to be selected as the new version's pre-text. The question is, then, why certain existing retellings are more likely to function as a new version's pre-text than others, or, in other words, what determines a retelling's 'differential fitness'.

In the cultural evolution research program, it has been suggested that such fitness of cultural variants can be generated by an interaction of various psychological, social, and ecological processes.<sup>18</sup> The focus of this study, however, is to examine the extent to which age-dependent selection processes generate differential fitness among existing retellings – or, in other words, whether intermediate retellings have a selection advantage over other retellings. If no such preference for intermediate retellings exists, authors may produce a retelling on the basis of some original text or sample their base material from a uniform distribution over all previous retellings (i.e. a modern retelling is just as likely based on a story from the immediate previous year as on a story from *n* years back in time). In both cases, I am inclined to believe that the process through which stories in children's literature are retold cannot

vice versa." *Psychological Review* 116, no. 4 (2009): 929–952; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

18. See e.g. Joseph Henrich et al., "Five Misunderstandings About Cultural Evolution," *Human Nature* 19, no. 2 (2008): 119–137; Henrich et al.'s account is compatible with ecological perspectives on cultural change, such as Noah Mark, "Birds of a Feather Sing Together," *Social Forces* 77, no. 2 (1998): 453–485; Stanley Lieberman, *A matter of taste: How names, fashions, and culture change* (New Haven, CT: Yale University Press, 2000); Alex Van Venrooij, "A community ecology of genres. Explaining the emergence of new genres in the UK field of electronic/dance music, 1985–1999," *Poetics* 52 (2015): 104–123; For a recent discussion of the notion of 'cultural fitness', see Grant Ramsey and Andreas De Block, "Is Cultural Fitness Hopelessly Confused?," *The British Journal for the Philosophy of Science*, 2015, doi:10.1093/bjps/axv047.

be conceived as gradually accumulating. I thus make the following prediction: if the evolution of stories in children's literature is a gradual accumulation of modifications, retellings are generally derived from versions in temporal proximity.

Because it would be too big a topic to present an in-depth study of evolutionary change in children's literature at large within the scope of this chapter, I restrict this study to a single case, namely, the evolution of the world's biggest iconic fairy tale: "Red Riding Hood". No story has been retold, reinterpreted, recontextualized and reconfigured as often as the story about the little girl in red who meets a wolf in the forest.<sup>19</sup> Currently, the misconception exists that fairy tales for children – especially those written by the Brothers Grimm – have been frozen in time and turned into static cultural artifacts. A single comparison between the version of "Red Riding Hood" by the Brothers Grimm and any of the modern retellings listed in Beckett's *Recycling Red Riding Hood* or in her anthology of international retellings furnishes all the proof that is needed to arrive at quite the opposite conclusion.<sup>20</sup> Beckett speaks of Red Riding Hood as an "inveterate globetrotting" girl.<sup>21</sup> Indeed, the world-wide number of retellings is so vast that a complete account of the evolution of the story is completely beyond comprehension. I therefore further restrict the object of study by concentrating on the evolution of Dutch retellings of "Red Riding Hood" within a timespan of more than two centuries.

Applying cultural evolution theory and methods to historical story transmission is not unprecedented. In particular, the current study is reminiscent of recent research carried out by Tehrani who applies phylogenetic methods to study cross-cultural relationships among retellings of "Red Riding Hood".<sup>22</sup> Crucially, while building upon important methodological insights from Tehrani, the current study differs in terms of data, methods and objectives. First, the large-scale collection of Dutch "Red

19. Jack Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 2nd ed. (New York, London: Routledge, 1993); Sandra L. Beckett, *Recycling Red Riding Hood* (New York, London: Routledge, 2002); Sandra L. Beckett, *Red Riding Hood for all Ages. A Fairy-Tale Icon in Cross-Cultural Contexts* (Detroit, Michigan: Wayne State University Press, 2008).

20. Beckett, *Recycling Red Riding Hood*; Sandra L. Beckett, *Revisioning Riding Hood around the World. An Anthology of International Retellings* (Detroit, Michigan: Wayne State University Press, 2014).

21. Beckett, *Revisioning Riding Hood around the World. An Anthology of International Retellings*.

22. Tehrani, "The Phylogeny of Little Red Riding Hood."

Riding Hood” retellings employed in this study (cf. Section 4.3.1) allows for a more fine-grained tracking of the progression and evolution of the story through time. Second, the approach presented here is more thoroughly embedded in literary theory on retelling stories<sup>23</sup> while at the same time positions itself explicitly and extensively in dialog with the cultural evolution research program.<sup>24</sup> Finally, whereas Tehrani’s goal is phylogenetic reconstruction, the present study is aimed at obtaining a better understanding of specific micro-evolutionary mechanisms underlying story transmission and change.

The remainder of the chapter is structured as follows. I start with a brief history of retellings of “Red Riding Hood” in Section 4.2. I then proceed with a description of the corpus compiled for this study (Section 4.3.1). Subsequently, I provide a detailed account of the data annotations (Section 4.3.2) followed by a section about the computational and statistical methods used (Section 4.4). After presenting the results in Section 4.5, I conclude with a discussion of the main findings of this study against the broader background of cultural evolution.

## 4.2 A BRIEF HISTORY OF RED RIDING HOOD

Girl, forest, wolf. Three words that immediately bring the story of “Red Riding Hood” to mind. The girl with the red cap has been associated with myths about the sunrise and sunset,<sup>25</sup> was the “first love”<sup>26</sup> of Charles Dickens, who wished he could have married her – “I should have known perfect bliss” – and is what Orenstein calls “the age-old star of bedtime drama”.<sup>27</sup> The tale has been told in poems, short stories, picture books, theater, comics and cartoons, video games, film, adult literature, advertisements, and even entire operas are devoted to the

23. E.g. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children’s Literature*; Zipes, *The Trials & Tribulations of Little Red Riding Hood*; Beckett, *Recycling Red Riding Hood*; Sanders, *Adaptation and Appropriation*; Hutcheon, *A Theory of Adaptation*.

24. Cavalli-Sforza and Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*; Boyd and Richerson, *Culture and the Evolutionary Process*; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

25. Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 18.

26. Beckett, *Recycling Red Riding Hood*, xv:xvi.

27. Catherine Orenstein, *Little Red Riding Hood Uncloaked. Sex, Morality and the Evolution of a Fairy Tale* (New York: Basic Books, 2002), 3.

fortunes of the girl in red.<sup>28</sup> As Jack Zipes writes in his seminal book, *The Trials and Tribulations of Little Red Riding Hood*, it is “the most popular and certainly the most provocative fairy tale in the Western world”.<sup>29</sup> Everyone knows “Red Riding Hood”. But do we really?

#### 4.2.1 PERRAULT’S RAPE STORY

The first literary version of the tale, *Le petit Chaperon rouge*, written by Charles Perrault in his 1697 collection *Contes du temps passé avec moralités*, is a retelling based on oral versions.<sup>30</sup> Perrault produced a truncated version and removed many elements dominant in oral tradition. However, some of those elements did survive the popularity of Perrault’s version. In one oral retelling recorded around 1885,<sup>31</sup> a little girl, after meeting with the wolf in the forest, is tricked into eating the flesh and drinking the blood of her grandmother (after which she is scolded by a little cat: “Phooey!... A slut is she who eats the flesh and drinks the blood of her granny.”). After an elaborate undressing scene, the girl manages to escape from the wolf by saying that she has “got to go badly”. The wolf proposes to do it in bed, but the girl refuses. Under the condition that she wears a woolen rope around her foot, the wolf allows her to leave the house. Once outside, she ties the rope to a tree and flees.

Most of these elements are either removed or refined in Perrault’s literary version. In *Le petit Chaperon rouge* Red Riding Hood is devoured by the wolf, and unlike in versions of the story children know today, there is no salvation. Perrault’s collection of tales is dedicated to Elisabeth Charlotte d’Orleans, Louis XIV’s niece and many elements of the story implicitly or explicitly – especially in the rhyming moral at the end – refer to court of the Sun King. As Beckett makes clear: “Perrault turns the oral tale into a parable, particularly adapted for use at the court of Versailles, that warns young ladies to be aware of suave and debonair two-legged wolves who would sweet-talk their way into their beds and

28. Beckett, *Red Riding Hood for all Ages. A Fairy-Tale Icon in Cross-Cultural Contexts*; Zipes, *The Trials & Tribulations of Little Red Riding Hood*.

29. Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 343.

30. *Ibid.*, 20.

31. *Ibid.*, 21-23.



Figure 4.1: *Little Red Riding Hood*, by Gustave Doré (1883).

ruin their reputations”.<sup>32</sup>

Perrault’s version is replete with sexual innuendos, most prominently in the ending moral, but also in the preceding narrative (e.g. “What big arms you have, grandmother!”). Zipes makes a strong case that Perrault transformed a once hopeful oral tale into an aggressive and violent story about a helpless girl who is to be held responsible for her own rape.<sup>33</sup> The etching by Doré from 1867 in Figure 4.1 illustrates Zipes’ thesis. It shows the image of a little girl, seemingly oblivious of the wolf’s intentions, who covers herself with the bed linen and gives the wolf a seductive look. She appears to have no fear, suggesting that “if she were not gullible and disobedient, she could prevent the rapacious wolf from carrying out his designs”.<sup>34</sup>

The oldest Dutch translation in my corpus of “Red Riding Hood” retellings (see section 4.3.1) is entitled *De vertelling van Roodkapje* (“The

32. Beckett, *Red Riding Hood for all Ages. A Fairy-Tale Icon in Cross-Cultural Contexts*, 13.

33. Zipes, *The Trials & Tribulations of Little Red Riding Hood*.

34. *Ibid.*, 39.



## De Vertelling van Roodkapje.



Eene moeder had een bevalig meisje, welke men Roodkapje noemde, omdat zij altijd een rood kapje droeg. De moeder zegt: „kind, gij moet eens naar uwe grootmoeder gaan.”



„Ik zal wat wasfietjes hakken, en die moet gij met dat poeje boter naar uwe grootmoeder brengen, maar vertoef niet op den weg.”



De moeder zende Roodkapje met de wasfietjes en het poeje boter, en zegt: „zie hoe uwe grootmoeder vast!” want zij was ziek.



Roodkapje gaat op weg, en ziet reeds het Dorp in het verlicht, waar hare grootmoeder woonde.



Zij ontmoette op den weg een wolf, die haar vroeg: „waar gaat gij heen?” Zij antwoordde: „naar mijne grootmoeder, om haar deze wasfietjes en het poeje boter te brengen.”



„Woonst zij ver van hier?” vroeg de wolf. „Het is voerbij den molen, dien gij gints ziet,” (zeide Roodkapje,) „het eerste huis in het Dorp.”



Zij hield stieh op den weg bezig om kappelletjes te vangen; de wolf had inustee den korten weg genomen en klopt bij de grootmoeder aan.



De wolf heeft de deur opengedaan, loopt naar het bed en vercheurt de grootmoeder, die ziek lag.



Roodkapje komt en klopt aan. De wolf roept: „wie klopt daar?” Zij zegt: „ik ben het grootmoeder!” „Trek maar aan het riempje van de klink en kom binnen,” zegt de wolf.



De wolf was in het bed gaan liggen; Roodkapje had wantekloppen en kwekshinnen, overofleggen, afdoelft het uwenachtig en boter op de takti neder en kom bij grootmoeder in het bed.”



Roodkapje wist niet dat de wolf in het bed lag, meende dat het hare grootmoeder was. Zij zegt: „grootmoeder! wat hebt gij groote tanden!”



„Dat is, om u des te beter te kunnen opezen,” zeide de wolf; en met eens vercheurde hij Roodkapje.

T. Amsterdam, bij de ERVEN H. RYNDERS, in de Tweede Tien-Duizendst.

Figure 4.2: De vertelling van Roodkapje, Erve H. Rynders, 1831 – 1854.

tale of Little Red Riding Hood') and stems from 1781. This version and most Dutch retellings of "Red Riding Hood" in the 19<sup>th</sup> century adhere to the plot structure of Perrault's story. The suggestive rhyming moral, however, is either set aside or altered by many authors. It is interesting to note that many retellings bear the word *nieuw* ('new') in their title as in the anonymous 1818 version *De Nieuwe Geschiedenis van Roodkapje* ('The new history of Red Riding Hood'). As Beckett remarks, this reminds us "that retellings have always been an attempt to rejuvenate the tale for a contemporary audience".<sup>35</sup> Many of the oldest versions of "Red Riding Hood" in the Netherlands were published as so-called pennycatch prints. Pennycatch prints were the cheapest illustrated printed material – for one or a few pennies – and report on important events or tell popular stories such as "Red Riding Hood" on a single sheet of paper, slightly larger than the current A3-format. Figure 4.2 provides an example one of the pennycatch prints in the collection.

#### 4.2.2 GRIMMS' STORY OF DISCIPLINE

The second classic retelling of "Red Riding Hood" is the one by the Brothers Grimm, *Rothkäpchen*, originally published in 1812 in the collection *Kinder- und Hausmärchen*. The most striking change made by the Brothers Grimm is the rescue scene of Little Red Cap by a hunter. Little Red Cap may not be killed, yet it could be argued that the Brothers Grimm emphasize the image of a naive and helpless girl even more.<sup>36</sup> In contrast with the oral versions, Little Red Cap is unable to save herself and is dependent on the protection of a father figure. The Brothers Grimm have reworked and revised their collection several times in an attempt to adapt the story to fit the emerging Biedermeier image of a child: Little Red Cap must show obedience and good behavior.<sup>37</sup> The emphasis on good behavior is apparent from the added cautionary scene in which Little Red Cap is warned about the dangers of talking to strangers, straying from the path or lingering in the woods. In the 1850 and 1857 edition, the Brothers Grimm stress the importance of good manners when the mother commands the girl: "Und wenn du in ihre

35. Beckett, *Recycling Red Riding Hood*, xvi.

36. Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 32.

37. *Ibid.*, 32–37. For an automatically generated alignment of the seven versions published by the Brothers Grimm, see <http://fbkarsdorp.github.io/grimms>.

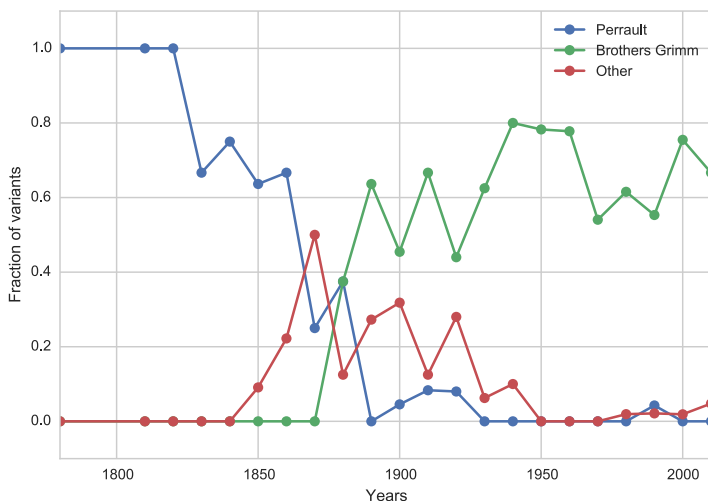


Figure 4.3: Population plot showing the fraction of retellings per decade of “Red Riding Hood” classified as a version of either Perrault, the Brothers Grimm, or ‘other’.

Stube kommst, so vergiß nicht guten Morgen zu sagen und guck nicht erst in alle Ecken herum.”<sup>38</sup> Little Red Cap promises to be obedient, but she is not, and therefore she needs to be punished. At the end of the story, Little Red Cap is aware of her disobedience and concludes that she is to be held responsible for her fate. Some Dutch retellers have emphasized this lesson using a more elaborate moral, as in the version by Simon Jacobus Andriessen (1880): “Die kwaad doet, kwaad ontmoet! en: Wie niet hooren wil, moet voelen!” (‘He that mischief hatches, mischief catches!’ and: ‘He that will not be counseled cannot be helped!’).

Perrault’s tale and the retelling of the Brothers Grimm have been popular in the Netherlands. However, the number of retellings that showed signs of being influenced by the Brothers Grimm directly after *Kinder- und Hausmärchen* was published in 1857 is rather low. Over the

38. ‘And when you enter her parlor, don’t forget to say ‘Good morning’, and don’t peer into all the corners first.’

years the numbers gradually increased before it ‘took off’ in the 1890s and “it virtually dwarfed Perrault’s version”.<sup>39</sup> Figure 4.3 visualizes the competition between Perrault and the Brothers Grimm from the late 18<sup>th</sup> century to the early 21<sup>st</sup> century in the Netherlands. The plot shows for each decade from 1780 to 2010 the fraction of retellings that can be classified as a version derived from either Perrault, the Brothers Grimm or ‘other’. A retelling is classified as a Perrault version if both the grandmother and the girl are eaten by the wolf and they are not rescued. A version counts as a retelling of the Brothers Grimm if, after being swallowed, both the grandmother and the girl are saved from the wolf’s belly. The ‘other’ group shows the fraction of versions in which the grandmother is devoured (without being rescued) while the girl is saved before the wolf is able to lay a hand on her. Without more information about the rest of the narrative, it is hard to say whether these versions adhere to the Perrault or to the Grimm paradigm. They do fit the trend of ‘Victorian censorship’ of the 19<sup>th</sup> century in which scenes considered to be too cruel or too sexual were replaced.

It is interesting to observe that the explosive burst of retellings within the Grimm paradigm is followed by a ‘stabilizing’ period at the beginning of the 20<sup>th</sup> century in which no radical changes to the story are made. As illustrated by Figure 4.4, this is also a period in which a large number of anonymous retellings of “Red Riding Hood” were published, possibly to evoke an image of authenticity as a tale from oral history. Furthermore, quite a number of retellings position themselves explicitly in relation to either Perrault or the Brothers Grimm. An example from my corpus appeared in the collection *Sprookjes Van Moeder De Gans* (‘Fairy tales by Mother Goose’) written by Christine Doorman (1916). Although the title of the collection refers to Perrault, the story of “Red Riding Hood” more closely resembles the plot structure of the Brothers Grimm with its reassuring rescue scene. The beginning of the 20<sup>th</sup> century might be epitomized as a period in which the fairy tale becomes increasingly generic and autonomous, not bound to a particular author but residing in the collective imagination. In this context, Soriano aptly suggests that the fairy tale has become a text “without a text” and a text “without an author”.<sup>40</sup>

39. Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 36.

40. Cited in Beckett, *Recycling Red Riding Hood*, xvii.

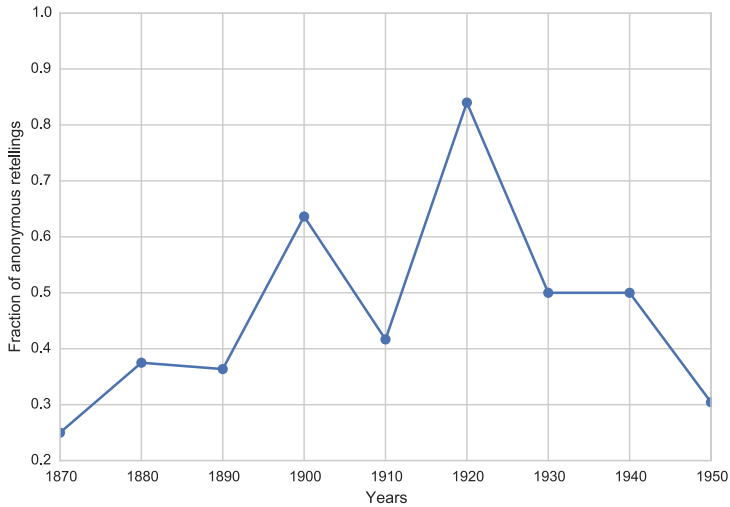


Figure 4.4: Fraction of anonymous (versus authored) retellings of “Red Riding Hood” over the period 1870 – 1950.

#### 4.2.3 TIME FOR CHANGE

We might hypothesize that within the context of a literary environment in which there is a strong tradition of the story, yet it is not linked to a specific author, retellers are able to engage more freely with the material.<sup>41</sup> Starting from the 1940s but especially in recent decades, an increasing number of retellings expose radical innovations, aesthetic experimentation and intertextual references. Because of their limited exposure to cultural heritage, children are generally assumed to be less competent in decoding these intertextualities. Beckett argues that in the case of “Red Riding Hood”, authors *can* use these sophisticated narrative strategies, because they can safely assume children know at least some version of the story, which provides them with the necessary decoding tools.<sup>42</sup>

After World War II, the story of “Red Riding Hood” branched off in too many directions to enumerate in this overview. Zipes distinguishes

41. Beckett, *Recycling Red Riding Hood*.

42. *Ibid.*



Figure 4.5: *Rood Rood Roodkapje*, by Edward van de Vendel; illustration: Isabelle Vandenabeele (2003).

three major branches: (i) retellings in which “Red Riding Hood” becomes increasingly independent, (ii) versions that rehabilitate the wolf and/ or tell his version of the story and (iii) stories that stand out with respect to their aesthetic experimentation.<sup>43</sup> I discuss an example from each of these branches to illustrate some of the radical changes the story has undergone.

*Rood Rood Roodkapje*

The retelling *Rood Rood Roodkapje* (‘Red Red Red Riding Hood’) by the Dutch author Edward van de Vendel with illustrations by the Flemish illustrator Isabelle Vandenabeele, is an exciting example of a retelling in which Red Riding Hood has become self-reliant.<sup>44</sup> Vandenabeele makes use of a drawing style that is reminiscent of the etchings by Gustave Doré, yet much coarser and only in the colors gray, black and red. Red Red Red Riding Hood only wishes for red things, red clothes, red juice, red carpet, red pillows on her bed. She had to choose her own name – the reversal of a dominant motif in “Red Riding Hood”, which is iconic of her more independent status. All her wishes are red, but her days are gray, because she has to walk the gray muddy paths everyday to her gray grandmother. Then one day, she encounters something black. . . The Wolf. By accident she shows the Wolf the way to her grandmother. The Wolf immediately spurts to her grandmother and devours her with a terrifying howling sound. Determined to make up for her mistake, Red Red Red Riding Hood follows the Wolf to her grandmother’s house. No, this time there is no howling

grwarwah  
briaauwah

43. Zipes, *The Trials & Tribulations of Little Red Riding Hood*, 59.

44. Edward Van de Vendel and Isabelle Vandenabeele (illustrations), *Rood Rood Roodkapje* (Wielsbeke: De Eenhoorn, 2003).

# hgroing raaaaaaa

But: chop.

After the restrained slaying scene, we see a calm girl with a somewhat subdued and contemplative look on her face, holding a bloody ax, while the blood flows from her grandmother's doorway (cf. Figure 4.5). Content with the thought that she never has to visit her grandmother again, Red Red Red Riding Hood imagines all the red things she can do in her life. As in the old French oral version of the story in which "Red Riding Hood" outwits the wolf, Van den Vendel and Vandenabeele present a girl who is no longer timid, innocent or powerless, but rather fearless, determined and self-assured.

## *De Wolf die tegen water praatte*

At an abstract level, the story *De Wolf die tegen water praatte* ('The Wolf who talked to water') by the Dutch author Imme Dros, closely resembles traditional retellings of "Red Riding Hood": (i) A girl meets a wolf in the forest, (ii) the wolf swallows both the girl and her grandmother, (iii) they are rescued by a hunter, and (iv) the wolf, filled with stones, drowns in the water.<sup>45</sup> The perspective of the story, the ordering of the different plot elements and especially the meaning attributed to these elements, however, is completely different. The story, reminiscent of Ovid's *Narcissus*, tells us about the wolf Noordenloos ('Northless') whose only friend is his reflection in the water (cf. Figure 4.6). The lonely Noordenloos wishes to return to his former wolf pack in the North together with his imaginary friend Waterwolf. One day, after having a conversation with Waterwolf, Noordenloos meets a girl, who asks for

45. Imme Dros and Henricus Geelen (illustrations), *De Wolf die tegen water praatte* (Bussum: Moon Press, 1991).





Figure 4.6: *De Wolf die tegen water praatte*, by Imme Dros; illustration: Henricus Geelen (1991).

directions to her grandmother's cottage. Following the traditional plot structure, the wolf arrives at grandmother's house before Red Riding Hood does, swallows the grandmother, tricks Red Riding Hood using his famous disguise, and finally, gobbles her up. However, unlike in most traditional retellings, the primary reason for eating Red Riding Hood and her grandmother is not to stave off hunger, but serves to strengthen the wolf for his escape from loneliness to the North. Yet, he only manages to join his family in the North in his own imagination at the moment of his – metaphorically described – death: he was happy. By approaching the story from the wolf's perspective, Dros transforms

the wolf from being primarily ‘Big and Bad’ into a real character with a soul, feelings, thoughts and desires.

*Roodkapje en de zeven geitjes*

The experimental retelling written by Ivo de Wijs with illustrations by Alfons van Heusden (1994) diverges radically from the tradition, both in form and in content. The story is part of a collection entitled *Roodkapje en de zeven geitjes* (‘Red Riding Hood and the Seven Kids’). This rather unusual juxtaposition of characters from two famous fairy tales serves to emphasize the bricolage nature of the stories to come. Each story in the book tells three different versions of a story. On each page, supposedly aligned fragments of the three versions are placed side by side in three columns, each with its own distinctive typography. The reader is supposed to read all three columns before moving on to the next page (see Figure 4.7 for a translation of the first paragraphs of each version).

The first column of *Roodkapje* closely resembles a ‘traditional’ version of “Red Riding Hood”. The second version adheres to the traditional plot, yet it adapts and reworks many well-known motifs. The origin of Little Red Cap’s name, for example, stems from the red roof of her house. Here De Wijs makes playful use of the fact that the noun *kap* in the compound *Roodkapje* can refer to both a cap and a roof in Dutch. Another small change in the opening scene is that Little Red Cap’s father (and not her mother) sends her to her grandmother. Furthermore, the purpose of her trip is to visit her grandmother because of grandmother’s birthday and not because granny has fallen ill. In the forest, Little Red Cap meets the wolf, who is rather fond of the biscuits the girl brought with her. He is not allowed to have any and therefore decides to take a piece of the grandmother. When the wolf arrives at the doorstep of grandmother’s house, grandmother distrusts the visitor because of his low voice. Using a piece of chalk to raise his voice – a motif borrowed from *The Wolf and the Seven Kids* – the wolf finally manages to convince the grandmother to let him in. The traditional formulaic scene has undergone some changes as well. This time Little Red Cap does not marvel about grandmother’s physiognomy, but about the size of her nightcap and her big glasses. The traditional savior of Little Red

Once upon a time there was a little girl who lived with her mother in a small house near the forest. Everyone called the girl Little Red Cap, because she wore a red cap, which her grandmother gave to her. One day her mother said to her: 'Child, your grandmother fell ill. Go pay her a visit. Here is a basket with wine and biscuits. But remember: don't tarry on your way and don't stray from the path in the forest. I hope that grandmother will soon feel better!'

**On the edge of the forest lived a little girl in a small house. Everyone called her Little Red Cap because her house had a red roof. One day her father said to her: 'My dear child, it's your grandmother's birthday. Here is a basket with presents for her: a bottle of wine, a can of biscuits and a bouquet of flowers. Go bring the basket to her, but remember: don't loiter on your way! Say my regards to grandmother and wish her congratulations! And many years to come!'**

Once upon a time there was a little girl called Gretel. Well, her name was Gretel, but she was called Little Red Cap because of her flaming red hair. Little Red Cap lived with her grandmother in an old tower out in the woods. One day grandmother went to the city to buy biscuits and lemonade – and a cough medicine because Little Red Cap wasn't feeling well. 'Come back soon, grandmother,' said Little Red Cap, 'walk briskly!' And grandmother went on her way.

Figure 4.7: First paragraphs of *Roodkapje* by Ivo de Wijs (1994), translation FK.

Cap and her grandmother is replaced by a butcher. After being rescued from the wolf's belly, the butcher, Little Red Cap and the grandmother hide in the clock – again an allusion to *The Wolf and the Seven Kids* – in anticipation of what will happen next. In the climactic scene of the story, the wolf, in search of food, sticks his head into the oven. In this scene, Little Red Cap assumes the role of Gretel from *Hans and Gretel* as she pushes the wolf into the oven.

The third version of the story is truly disorienting. The story is replete with references to other stories, blendings of motifs, reversals of situations, and replacements of characters. In her discussion of the story Beckett described Little Red Cap's identity in this third variant as one "in a constant state of flux".<sup>46</sup> Just as in the second version, yet more overtly, Little Red Cap assumes the roles of various female fairy tale characters. This role-shifting already starts in the introductory scene, in which it explained that because of her red flaming hair, everyone calls the girl Little Red Cap, but her actual name is Gretel. She lives with her grandmother out in the woods in a tower, which immediately brings the story of Rapunzel to mind. The grandmother and Little Red Cap change roles. It is the grandmother who goes into town to buy biscuits, lemonade and a cough medicine for her sick granddaughter, and it is the grandmother who encounters the wolf in the forest. After revealing to the wolf where she is going, the wolf spurts to Little Red Cap's tower – at which moment the reader finds out that the tower is actually the property of Snow White. Unable to find an entrance, the wolf waits for the grandmother to come home to find out how to enter the tower. When the grandmother arrives at the tower, she assumes the role of the witch in *Rapunzel*, claps her hands and commands Little Red Cap to let down her braids. After devouring granny, the wolf climbs up the braids. Little Red Cap is scared by the hairy look of the wolf, but it is not Little Red Cap who is gobbled up, but Rapunzel. The wolf disguises himself as Sinterklaas (wearing a bishop miter) and goes to bed. The savior of the story is a young prince, called Hansel, who, after transforming himself into a raven, flies to the tower. There he assumes his normal form, cuts open the belly of the wolf and saves Cinderella and her grandmother. The dead wolf is thrown out of the window, and by using another magic

46. Beckett, *Recycling Red Riding Hood*, 100.

spell, the prince transforms the grandmother, Sleeping Beauty(!) and himself into ravens, which enables them to escape from the tower. Once they are transformed back into their normal forms, the grandmother and Gretel follow the prince to his homeland. The prince and grandmother get married and they live happily ever after.

The three retellings by De Wijs describe a micro-level development that reflects the development of “Red Riding Hood” on a macro-level. The stories form each other’s pre-text, they echo each other – to use the words of Frank.<sup>47</sup> Reading De Wijs’ retellings can be considered as an individual experience of a process of gradual accumulation of modifications, the same process I hypothesize to be the driving force behind the development of “Red Riding Hood” at the population level. Before I will present the methods to investigate this hypothesis, I will first provide a detailed overview of the data collection used in this study.

## 4.3 DATA COLLECTION AND DATA ANNOTATIONS

### 4.3.1 DATA COLLECTION

The Koninklijke Bibliotheek (National Library) of the Netherlands is in possession of a tremendously rich collection of children’s books. It consists of over 195 thousand books that have been collected over a period of two hundred years. The collection contains about 630 versions of “Red Riding Hood”, of which the oldest version dates from the late 18<sup>th</sup> century and the latest from 2015. Many versions are part of the Special Collections department of the National Library, which contains books and manuscripts that are too old, rare, precious or fragile to be made available through general circulation. For this study, I required a full-text version of the collection, yet only a handful of retellings of “Red Riding Hoods” have been made digitally available. To remedy this problem, I digitized all available versions listed in the catalog of the National Library (with the exception of reprints). The stories were either manually transcribed or by means of OCR followed by manual post-correction. After removing duplicates, the total number of stories

47. Frank, *Letting Stories Breathe. A Socio-Narratology*, 37.

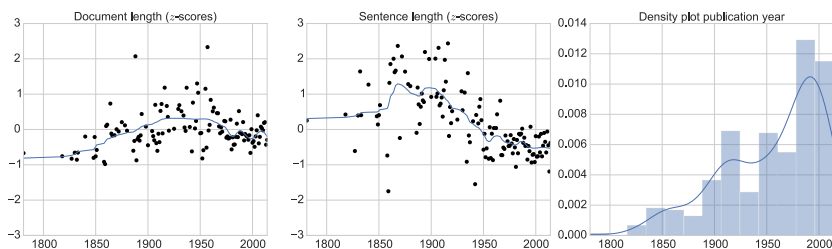


Figure 4.8: Diachronic visualization of some general statistics about the “Red Riding Hood” corpus. The subplots (from left to right) in this figure visualize: the average z-scores of the document length per year (expressed in word forms), the average z-scores of the sentence length per year (expressed in word forms) and a kernel density plot of the year of publication of the stories.

in the digitized version of the collection amounts to 440.<sup>48</sup> The meta data for each story include: the title, author, (optional) collaborator, (optional) illustrator, (estimate of) year of publication, publisher and the dimensions of the book.

I constructed a tokenized version of the “Red Riding Hood” collection using the tokenizer Ucto.<sup>49</sup> The total number of words in the tokenized version amounts to 493,169 (including punctuation). In their lowercased form, 10,683 of these word forms occur uniquely. On average each year in the corpus ranging from 1781 to 2015 is represented by about 3 stories. Stories consist of 1155 word forms on average. Some of the general statistics of the corpus have been diachronically visualized in Figure 4.8. The subplots (in order of appearance) in this figure visualize the average document length per year, the average sentence length per year and the number of retellings of “Red Riding Hood” published in a particular year. For reasons of comparability, I normalized the frequencies in the first two plots to their z-score. As can be observed from the density plot over the publication years of the stories, the corpus is rather skewed towards more recently published versions of “Red Riding Hood”. This bias will be taken into account in the subsequent analyses. Looking at the first subplot, we can observe that the document length of the

48. Karsdorp, *Story network data sets*. See <http://fbkarsdorp.github.io/rrh-browser> for a bibliography of all stories as well as an interactive search engine of the collection.

49. See <https://languagemachines.github.io/ucto/> and corresponding manual [https://github.com/LanguageMachines/ucto/raw/master/docs/ucto\\_manual.pdf](https://github.com/LanguageMachines/ucto/raw/master/docs/ucto_manual.pdf).

stories remains generally stable over time. When we inspect at the average sentence length, however, we note a steady decrease starting at the beginning of the 20<sup>th</sup> century. We might hypothesize that this trend fits a development in which authors more and more adapt their retelling towards their young readers, who generally prefer shorter over longer sentences. However, more materials and research are required to assess whether the observed trend is representative of a development in children's literature at large.

#### 4.3.2 STORY ANNOTATIONS

Variation is one of the three preconditions described by Darwin for biological evolution. Without variation, there is nothing to be selected and hence no change can take place. Because of its central position to his theory, Darwin documents the variation he observed between members of the same species in great detail. An exemplary quote about variation about pigeons reads as follows:

The proportional width of the gape of mouth, the proportional length of the eyelids, of the orifice of the nostrils, of the tongue (not always in strict correlation with the length of beak), the size of the crop and of the upper part of the oesophagus; the development and abortion of the oil-gland; the number of the primary wing and caudal feathers; the relative length of wing and tail to each other and to the body; the relative length of leg and of the feet; the number of scutellae on the toes, the development of skin between the toes, are all points of structure which are variable.<sup>50</sup>

Mesoudi argues that to justify the description of cultural change as a Darwinian evolutionary process, we must show that culture exhibits the same preconditions as biology.<sup>51</sup> An adaptation of Darwin's description of pigeon variation to "Red Riding Hood" could read as follows:

The fabrics of Red Riding Hood's cap (cotton, silk, wool), the contents of her basket (is she carrying wine, bread, waffles, butter, juice, cake, lard), whether or not the girl meets with the wolf in the woods, the surroundings of grandmother's cottage (oaks, a hedge, a mill), whether the wolf eats granny and or Red Riding Hood and in

50. Darwin *The Origin of Species*, cited in Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*, 27.

51. *Ibid.*, 27–34.

what way (swallowing, gobbling, devouring, guzzling), whether the girl and her grandmother are saved and by whom (a woodcutter, a hunter, her father, animals in the forest), whether the wolf is killed and in what way (using an ax, a gunshot, a bat), whether the wolf's belly is filled with stones and who puts the stones in his belly, are all points of the story which are variable.

Like Darwin, I could go on like this for several pages. If we look closely at all the different retellings of the story, the amount of variation is truly overwhelming. To establish a mapping of the variation, I subjected all 440 stories in the corpus to a questionnaire of more than 300 questions. The questionnaire consists of simple "yes-no" questions (e.g. *Does the story explain the origin of Red Riding Hood's name?*), multiple choice questions (e.g. *Where does Red Riding Hood encounter the wolf? (a) on the path, (b) away from the path, (c) somewhere else, (d) they don't meet*) and open questions (e.g. *What is the lemma of the verb used to describe the wolf's eating of Red Riding Hood?*). Most questions in the questionnaire fit concepts from structural text analysis and narratology.<sup>52</sup> The list of potential questions is virtually endless. I have attempted to compile a list of questions of which the answers constitute a detailed and rigorous text analysis. The questions can be classified under the following six categories: (i) genre, (ii) narration, (iii) space, (iv) motif, (v) time and (vi) character.

(i) *Genre* In the category 'genre' we find questions designed to distinguish between kinds of stories in form and content. The collection of "Red Riding Hood" retellings contains the following genres: regular stories, picture books, pop-up books, theater, comics, puzzles, picture story, catchpenny prints and poems. A story with illustrations is not necessarily classified as a picture book. Only stories in which the pictures take central position with an accompanying text are considered to be picture books. As some stories belong to multiple genres, it was allowed to provide multiple answers. As can be observed from Figure 4.9, poems were a popular form for Dutch authors to tell the story in the second half of the 19<sup>th</sup> century, whereas in more recent years, we discover a steady increase in the number of picture books.

52. E.g. Mieke Bal, *Narratology. Introduction to the Theory of Narrative*, 3rd ed. (Toronto, Buffalo, London: University of Toronto Press, 2009); Erica Van Boven and Gillis Dorleijn, *Literair Mechaniek. Inleiding tot de analyse van verhalen en gedichten*, 2nd ed. (Bussum: Uitgeverij Coutinho, 2003).



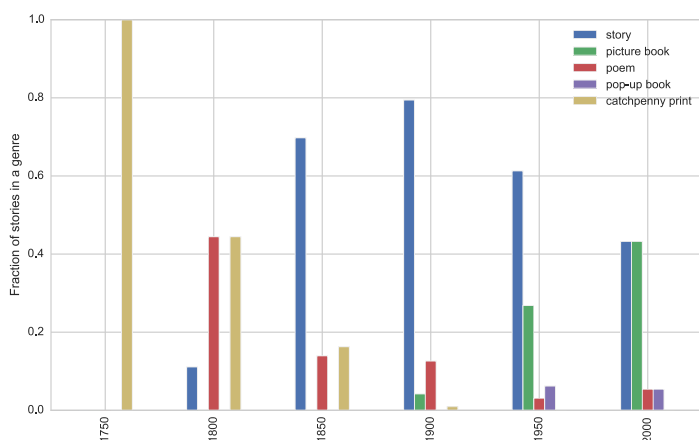


Figure 4.9: Population plot of the five most frequently occurring genres in the “Red Riding Hood” collection. We calculate the fraction of stories published in a particular genre for every 50 years.

*Narration (ii)* The second category of questions deals with the narration of the story. Most importantly, this category addresses the narrator of the story or – in narratological terms – the question of the speech-position from which the narrative contents of a story as a whole originates. Most versions of “Red Riding Hood” are classified as third-person narratives. In recent years retellers of “Red Riding Hood” have experimented with different forms of narration such as the first-person narrative. An interesting example of a first-person retelling is the one by the Dutch author Ivo de Wijs, which is part of the collection *En ze leefden nog... Sprookjes op Rijn*.<sup>53</sup> The potentially hilarious effect of telling “Red Riding Hood” from a first person point of view becomes especially clear after Red Riding Hood is gobbled up by the wolf:

53. Ivo De Wijs and Wouter Tulp (illustrations), *En ze leefden nog... Sprookjes op Rijn*. (Amsterdam: Amsterdam Blue in Green, 2011).

In de maag van het gulzige beest	In the tum of the devouring beast
Zat mijn oma, een tikkie bevreesd	Was my granny, a little afraid
Ze zei: 'Meisje, je komt als geroepen	She said: 'Girl, you're just in time
Want we moeten hier dadelijk uit	Because we have to leave immediately
Stel je voor dat dat monster besluit	What if that the monster descides
Ons gezamenlijk uit te gaan poepen'	to poop us out collectively'

*Space (iii)* The third category of questions deals with space. The questions aim to collect information about e.g. the location or the surroundings of grandmothers house.

*Motif (iv)* The largest group in the questionnaire is category four, which contains various questions about smaller and larger motifs, e.g. *Is Red Riding Hood eaten by the wolf?* or *Is the wolf's belly filled with stones or some other material?*. Category four also queries the main episodes of the story. Examples of questions are: *Does the story contain a cautionary scene in which the mother of Red Riding Hood warns her about the dangers in the forest?* and *Does the text describe a scene in which Red Riding Hood returns back home?* The Grimms' version of "Red Riding Hood" contains a rarely retold and virtually unknown second ending to the story. In this ending – a kind of epilogue or sequel – Red Riding Hood, barely recovered from her previous adventures, visits her grandmother again. In the woods she meets another wolf but this time she shows obedience and proves that she has learned her lesson. Together with her grandmother she manages to kill the wolf by making him fall from the roof into a big stone trough. The Dutch "Red Riding Hood" collection contains thirteen version in which this second episode is retold.

*Time (v)* The category 'time' contains questions that discuss the ordering of events, the passing of time and tension and irony. Examples of questions are: *Does Red Riding Hood encounter the wolf before or after she strays from the path?* and *If the wolf is killed, does that happen before or after Red Riding Hood (and her grandmother) are rescued?*. Questions about irony and tension were included in the category time, because they deal with expectations and anticipations of readers (e.g. 'What will happen next?'). The climactic scene of the wolf in disguise can be read as an example of dramatic irony in which the girl naively believes the wolf to

be her granny. To be able to understand this irony, children must have grown a 'theory of mind', which allows them to reason about others in terms of goals and intentions. Theory of mind is one of the most important concepts in social-psychology and social evolution theory,<sup>54</sup> yet literary and folktale studies have only scantily paid attention to it.<sup>55</sup> In the development of a full-blown theory of mind, children learn about 'beliefs' and 'false beliefs'. In the context of "Red Riding Hood", this knowledge enables children to understand that the girl *believes* that it is her grandmother lying in the bed, yet they know about the actual state of things and hence that this belief is false. Knowledge about (false) beliefs is of crucial importance in our life to make predictions about the behavior and intentions of others. "No wonder," Boyd argues, "that point of view and dramatic irony play such central roles in fiction, or that the gap between appearance and reality is such a wide-spread theme".<sup>56</sup> Retellings of "Red Riding Hood" play with this theme. Sometimes authors explicitly anticipate the potential downer ending at the outset of the story, especially when there is no salvation scene. In other retellings, authors feel the need to make explicit that after being swallowed, the girl and her grandmother are still alive and well in the tummy of the wolf. Yet other versions omit the traditional switch of perspective from Red Riding Hood to the wolf and continue to keep track of the actions of the girl. In such versions, it is not yet clear to the reader who lies in the bed at the moment the girl arrives at her grandmother's cottage. This narrative strategy leads to a reading of the story without false belief and has a stronger effect of surprise.

*Character (vi)* Without the characters there would be no story. The last category contains questions about the participants in the stories. These include questions about the presence or absence of characters (e.g. *Is Red Riding Hood's mother present in the story?*), about their physical properties (e.g. *Does the story describe any physical properties of Red Riding Hood?*), about their clothing (e.g. *Does the text make explicit reference to Red Riding Hood's hood?*), or about their personality traits (e.g. *Is the wolf referred to*

54. See e.g. Michael Tomasello, *The Cultural Origins Of Human Cognition* (Cambridge, Massachusetts: Harvard University Press, 2009).

55. Cf. Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction*.

56. *Ibid.*, 149.

as a “(big) bad wolf”?).

#### 4.3.3 THE RISE OF THE BIG BAD WOLF

In this section, I highlight one group of questions in some detail, viz. those dealing with the way characters are introduced to the story. Writers and storytellers have two basic introduction strategies to their disposal: They can introduce a character by means of (i) *indefinite* reference (using an indefinite article, e.g. Dutch *een*) or (ii) *definite* reference (using a proper name or a definite article, e.g. Dutch *de*).<sup>57</sup> Indefinite reference is generally used to introduce ‘new information’, i.e. information the writer deems inaccessible to the reader. Definite reference, on the other hand, is typically used to refer to ‘given information’ in which case the writer assumes that the reader is able to identify the intended referent. Folktale characters are typically introduced as ‘new’ – hence, their referent is introduced by means of an indefinite article at first mention, as in the following opening sentence from Perrault’s *Le petit Chaperon rouge*: “Once upon a time there was *a* little village girl. . .” Once introduced, the girl becomes given information, i.e. she becomes part of the set of referents that can be assumed to be accessible to the reader. This allows the writer to refer to the girl using definite reference in subsequent parts of the story. In the linguistic literature, the concept ‘discourse reference’ is employed to indicate this transition from indefinite to definite reference, as it depends on the progression of discourse.<sup>58</sup> Referents can also count as given information when they are (assumed to be) part of the shared socio-cultural world knowledge of a writer and a reader. This type of reference is called ‘unique reference’.<sup>59</sup> Such entities can be referred to by means of definite reference even though they have not been introduced in the preceding discourse.

I hypothesize that with the increasing familiarity of the story over

57. See e.g. Franciska M.G. De Jong, “The Compositional Nature of (In)Definiteness,” in *The Representation of (In)definiteness. Current Studies in Linguistics Series*, vol. 14 (Cambridge, Massachusetts: MIT Press, 1987), 270–285; Christopher Lyons, *Definiteness*, Cambridge Textbooks in Linguistics (Cambridge: Cambridge University Press, 1999); Günter Radden and René Dirven, *Cognitive English Grammar*, Cognitive Linguistics in Practice (Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007).

58. Radden and Dirven, *Cognitive English Grammar*, 98.

59. *Ibid.*, 99.

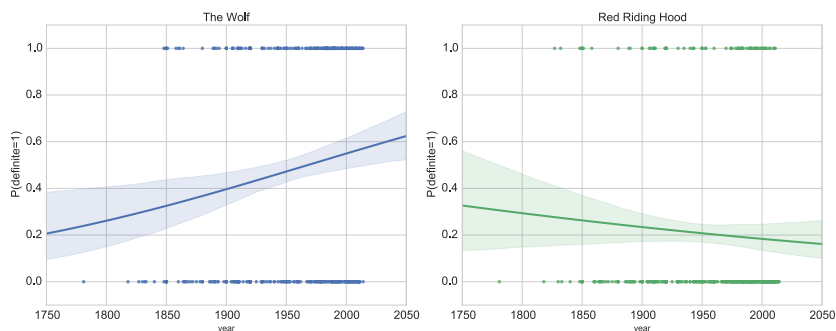


Figure 4.10: Plots showing definite (points with a value of 1.0) and indefinite references (points with a value of 0.0) used to introduce the wolf and Red Riding Hood in the corpus. The line represents the regression model fits to these points, the shade of which shows the confidence intervals.

time and its consequent entrenchment in culture,<sup>60</sup> the characters of “Red Riding Hood” become more and more part of the shared world knowledge of a writer and her or his audience. In other words, they become ‘given information’. Readers have expectations about the course of the story and appearance of certain characters, and writers assume their audience to know at least some version of the story. Indicative of their expectations, readers often act surprised when they discover that there is no rescuer in Perrault’s version of the story. The entrenchment of the characters is exemplified by what Beckett calls fairy tale salads.<sup>61</sup> These are stories in which characters from various popular fairy tales are transferred to new contexts without losing their identity and accompanying literary and cultural expectations.

If my hypothesis holds, we may expect an increase over time in the use of unique reference to introduce the characters of “Red Riding Hood” in favor of discourse reference. More specifically, I predict that the probability of introducing the characters by means of definite reference increases as a function of time. In the following analysis, I limit myself to the introduction of the two main characters: Red Riding Hood and the wolf. The question in the questionnaire corresponding to their introduction reads as follows:

60. Cf. Beckett, *Recycling Red Riding Hood*.

61. Cf. chapter 7 in *ibid*.

*Is Red Riding Hood / the wolf introduced to the story by means of definite or indefinite reference?*

To test for the correlation between time and the increased use of unique reference, I subjected the answers given to these questions to two logistic regression analyses (i.e. one for Red Riding Hood and one for the wolf). The dependent variable in the regression is whether or not definite reference is used, and the predictor variable is the year of publication of a story. Figure 4.10 displays the occurrences of definite and indefinite introductions for the wolf and Red Riding Hood. Points taking the value 1.0 represent introductions by means of definite reference; introductions with indefinite reference take the value 0.0. The curved line represents the fit of the regression model to the data. The shade reflects the confidence intervals of the fit. We can observe that the estimated probability for definite introductions of the wolf increases steadily over time. Analysis of the logistic regression model revealed that the effect of time is significant ( $\beta = 0.006, SE = 0.002, z = 2.887, p < 0.004$ ). For Red Riding Hood, by contrast, the plot shows a slight decrease in the estimated probability of definite introductions over time. This effect, however, is not significant ( $\beta = -0.003, SE = 0.002, z = -1.238, p > 0.2$ ).<sup>62</sup>

How should we interpret the difference between the two model fits? I suggest that the probability of definite introductions of Red Riding Hood is generally low, because her introduction is constrained by the genre-specific opening sentences in which she is introduced (e.g. “Once upon a time. . .” or “In a country far, far away. . .”). Because of their conventional nature, such opening phrases request the use of indefinite reference to introduce characters to a story.<sup>63</sup> Interestingly, in the “Red Riding Hood” collection these formulaic opening phrases have become even more conventional over time<sup>64</sup>, thus making definite introductions

62. The time laps in the collection might introduce a bias into the models. To account for this bias, I performed a logistic regression analysis in which the predictor variable ‘year’ is replaced by a variable that increases by 1 with every subsequent year of publication in the corpus. The results of these analyses were similar to the previous ones, both for the wolf ( $\beta = 0.008, SE = 0.003, z = 2.634, p < 0.008$ ) and for Red Riding Hood ( $\beta = -0.004, SE = 0.004, z = -1.132, p > 0.2$ ).

63. For a more general account of opening formulas in folktales, cf. Folgert Karsdorp, “Het is groen en leeft nog lang en gelukkig. Classificatie van volksverhaalgenres op basis van formules,” *Tijdschrift voor Nederlands Taal- en Letterkunde* 129, no. 4 (2013): 274–288.

64. The analysis of a logistic regression model shows a significant effect of time ( $\beta =$

even less likely. The wolf, on the other hand, is predominantly introduced after the cautionary scene, at which moment the constraints of conventional opening phrases no longer apply. The significant increase over time in the use of unique reference to introduce the wolf reflects the increasing familiarity of the story and the consequent expectations of readers about the appearance of certain characters.

#### 4.3.4 ANNOTATION EVALUATION

All stories in the collection have been subjected to the questionnaire. Using a self-developed annotation application<sup>65</sup>, I annotated the complete collection. To assess the reliability of the annotations, a second annotator filled in the questionnaire for ten percent of the collection, after which discrepancies with my own annotations were discussed. The amount of agreement between the annotators was measured by means of the *F1*-score.<sup>66</sup> The results showed strong agreement between the annotations ( $F1 = 0.95$ ). All categorical variables (resulting from multiple choice and open questions) were converted into binary indicator variables. The resulting  $m \times n$  matrix  $S$  consisted of  $m = 440$  stories, each of which was represented by  $n = 2444$  binary variables.

### 4.4 METHODS

#### 4.4.1 DISTANCE MEASURES

Recall from the introduction that the goal is to identify a set of potential pre-texts for each story in the “Red Riding Hood” collection. In this study, I make the simplifying assumption that stories that are more similar to a particular story in terms of their annotations are more likely to have formed their pre-text than other stories. The distance (or similarity) between two stories can be assessed by means of a distance measure, which summarizes the distance between two stories in terms of a pairwise comparison of the answers given to each of the questions in the questionnaire. Importantly, the choice for a particular distance

0.008,  $SE = 0.002$ ,  $z = 3.733$ ,  $p < 0.000$ ).

65. See <http://github.com/fbkarsdorp/roodkapje>.

66. Cohen’s  $\kappa$ -statistic 1960 is not suitable for this evaluation, because it is impossible to compute the number of true negatives for certain open questions in the questionnaire.

measure can have a severe impact on what is considered to be similar and what is not.<sup>67</sup>

We can distinguish two main groups of binary distance measures: ‘negative match *inclusive*’ and ‘negative match *exclusive*’ measures. Negative match inclusive measures assume that the absence of two features contributes to the similarity between two objects. In the context of story annotations, this implies that the absence of, for example, the rescue scene in two particular stories adds to the similarity between them. Negative match exclusive measures do not include negative matches in their similarity estimation, and thus do not account for the absence of the rescue scene. The potential downside of negative match inclusive measures becomes especially clear in the context of low frequency features. Take the contents of Red Riding Hood’s basket as an example. In only three versions in the collection, Red Riding Hood brings a bottle of milk to her grandmother. Negative match inclusive measures reinforce the similarities between all stories in which the milk is *absent*. Negative match exclusive measures, by contrast, will only reward the similarities between these three versions. More generally, the chance that two stories invoke a negative answer increases as questions become more specific. This does not, however, necessarily imply that the two stories are more similar.

In the subsequent analyses, I will compare three binary distance measures: (i) Manhattan distance, (ii) Jaccard dissimilarity index, and (iii) Cosine distance. All three measures have their own assumptions about what is to be considered similar or distant. In what follows, I provide a detailed description of these measures, with which I hope to clarify some of these assumptions.<sup>68</sup> Let  $a$  represent the number of questions to which two stories  $i$  and  $j$  provide a positive answer. Let  $b$  be the number of questions that are positively answered by story  $i$  and negatively by story  $j$ . Finally, let  $c$  represent the number of questions negatively answered by story  $i$  and positively by story  $j$ . The sum of  $b$  and  $c$  represents the total number of mismatches between two stories and is equivalent to the Manhattan distance (also known as the city

67. Numerous papers have been published in which various binary distance metrics and similarity coefficients have been proposed. For an in-depth comparison, cf. Seung-Seok Choi et al., “A survey of binary similarity and distance measures,” *Journal of Systemics, Cybernetics and Informatics* 8, no. 1 (2010): 43–48.

68. For this description I adopt the terminology used by *ibid*.



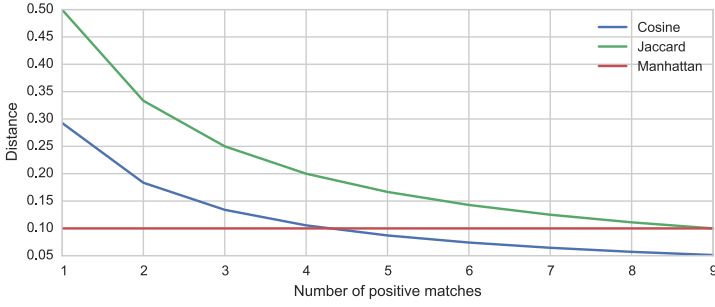


Figure 4.11: Plot showing the differences between the Manhattan distance, Jaccard dissimilarity index and Cosine distance for pairs of vectors with an increasing number of positive matches while keeping the number of mismatches constant at 1. The size of the vectors is 10.

block distance or  $\ell_1$  norm):

$$d_1 = b + c \quad (4.1)$$

The Manhattan distance is an example of a negative match inclusive measure and is based on the assumption that only mismatches contribute to the distance between two stories, and conversely that positive matches and negative matches add to their similarity.

The Jaccard similarity index is an example of a negative match exclusive measure. It is computed as the number of positive matches ( $a$ ) divided by the sum of positive matches and mismatches ( $a + b + c$ ). In set based terms, it computes the fraction of the length of the intersection between two stories over the size of their union. The complement of this fraction represents the dissimilarity between two stories:

$$d_J = 1 - \frac{a}{a + b + c} \quad (4.2)$$

The Cosine distance is an example of a negative match exclusive measure as well. It is computed as follows:

$$d_C = 1 - \frac{a}{\sqrt{(a + b) \times (a + c)}} \quad (4.3)$$

To get an impression of the effect of each of these distance measures,

consider the following four binary vectors, each representing one story:

$$\begin{aligned}\mathbf{a}^{(1)} &= \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \\ \mathbf{a}^{(2)} &= \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \\ \mathbf{a}^{(3)} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \mathbf{a}^{(4)} &= \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}\end{aligned}$$

The Manhattan distance  $d_1(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}) = d_1(\mathbf{a}^{(3)}, \mathbf{a}^{(4)}) = 1$ , because the two pairs contain an equal number of mismatches. The Jaccard dissimilarity index does not consider the negative matches and thus returns a distance of approximately 0.33 between  $\mathbf{a}^{(1)}$  and  $\mathbf{a}^{(2)}$  and 0.16 between  $\mathbf{a}^{(3)}$  and  $\mathbf{a}^{(4)}$ . As expected, the Cosine distance between  $\mathbf{a}^{(1)}$  and  $\mathbf{a}^{(2)}$  (0.18) is also larger than between  $\mathbf{a}^{(3)}$  and  $\mathbf{a}^{(4)}$  (0.08). The magnitude of the difference, however, is larger for the Jaccard dissimilarity index than for the Cosine distance. This effect can also be observed from Figure 4.11, which shows the differences between the three distance measures for pairs of vectors with an increasing number of positive matches, while keeping the number of mismatches constant at 1.

#### 4.4.2 TIME SPAN NORMALIZATION

The goal is to produce a ranking of texts for each story in the collection, in which the top items represent the most likely pre-texts. Using one of the above-mentioned distance measures, we can compute the distance between a story  $i$  published in year  $t_i$  and all potential pre-texts. The set of potential pre-texts for story  $i$  is defined as all stories published in or before year  $t_i$ , i.e.  $\{j | \forall j \in \{1, 2, \dots, m\} \wedge t_j \leq t_i \wedge i \neq j\}$ . To produce a final ranking, the pre-texts are sorted according to their distance to  $i$  in ascending order.

Given a ranking of potential pre-texts, we can evaluate the timespan between a particular retelling and its  $k$  most likely pre-texts. Simply taking the difference between year  $t_i$  and  $t_j$ , however, does not suffice, because older stories have a smaller set of potential pre-texts than stories published in later years, and this potentially introduces a bias in the results towards shorter timespans. Since the oldest story in the collection stems from 1781, the maximum timespan for a story published in, for

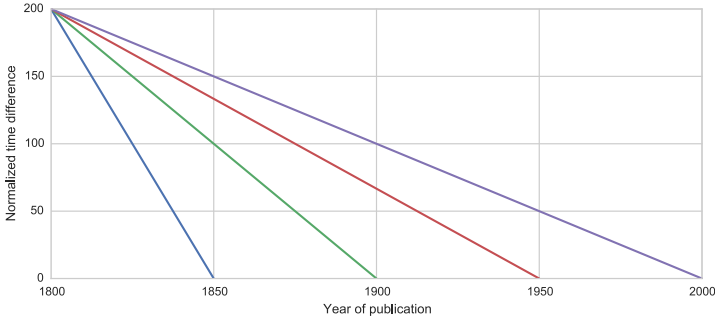


Figure 4.12: Visual illustration of the time difference normalization method, cf. Eq. 4.4.

example, 1850 is only 69 years, whereas for stories published in the 21<sup>st</sup> century, the maximum timespan is more than 200 years. To account for this bias, I propose the following equation, which normalizes the time differences with respect to the total timespan of the collection:

$$\Delta t_{ij} = |t_i - t_j| \frac{t_{\max} - t_{\min}}{t_i - t_{\min}}, \quad (4.4)$$

where  $t_i$  and  $t_j$  refer to the year of publication of story  $i$  and  $j$  respectively.  $t_{\min}$  and  $t_{\max}$  correspond to the publication years of the oldest and youngest story in the collection. Figure 4.12 illustrates the effect of applying the normalization. In this plot, four hypothetical stories published in the years 1850, 1900, 1950 and 2000 each select a story from 1800 as their most likely pre-text. Without the normalization method, the time differences between the stories and this pre-text will be 50, 100, 150 and 200 years respectively. As can be observed from the plot, the normalization method adjusts all time differences to 200 years.

#### 4.4.3 RELIABILITY ASSESSMENT

The performance of the proposed methodology to rank pre-texts depends on a number of factors. We have already seen that the choice for a particular distance measure can have a strong effect on the final rankings. The number of variables and the selection of particular variables can have a similar impact on the results. The dependence on all these factors

brings the reliability of the results into question. In what follows, I propose two ways to assess the reliability of the analyses.

As a first way to assess the consistency of the results, I perform a bootstrap analysis. The general idea behind this analysis is to approach the data from various angles for a large number of trials. In each trial we randomly select (with replacement) 10% of the features from the complete feature space. Using these randomly chosen features, a ranking of  $k$  most likely pre-texts can be established for each story in the collection. Each trial potentially returns different results. The final results are obtained by applying a majority voting mechanism, which selects those results that have reached mode consensus over all trials. In the analyses I will present below, the bootstrap analysis was set to run for 5000 trials.<sup>69</sup>

As a second reliability assessment, I compare the results to those obtained by using a different feature representation. I choose to employ a bag-of-words model, which represent documents as histograms over the vocabulary, i.e. as vectors of occurrence counts of words. For a document  $d$  with a vocabulary size  $V$  the vector representation  $\mathbf{w}^{(d)}$  is given by  $(w_1, w_2, \dots, w_V)$ , where  $w_i$  represents the occurrence count of word  $i$  in  $d$ . These representations can be used to compute the distance between two documents, e.g. by means of the Cosine distance. A clear advantage of bag-of-word models is the obviation of manual annotations for low-level features, such as lexical features. This comes with a price, however, because the many dimensions of word vector spaces may obfuscate the interpretability of the results. Another disadvantage of bag-of-words, at least for the purposes of the present study, is their sensitivity to spelling differences. The “Red Riding Hood” collection contains retellings written according to spelling conventions at various points in history. As a result, the rankings of pre-texts on the basis of bag-of-words models might reflect these spelling reforms, whereas we are more interested in rankings based on the actual contents of the stories. Despite this potential bias, I believe that bag-of-words models function as a useful background check to assess the reliability of the results obtained through ranking pre-texts on the basis of the answers given to the questionnaire (cf. Section 4.3.2).

69. For more information about the bootstrap procedure, see Phillip I. Good, *Resampling Methods. A Practical Guide to Data Analysis*, 3rd ed. (Boston, Basel, Berlin: Birkhäuser, 2006).

I weigh the word frequency vector spaces by means of the well-known ‘term frequency-inverse document frequency’ (tf-idf) weighting scheme.<sup>70</sup> This model weighs the occurrence counts of words in a document (tf) against the background of how many documents in the corpus contain those words (idf). The tf-idf transformed vector spaces put more weight on words with high term frequencies (i.e. words occurring often in a particular document) and relatively low document frequencies. Words with high tf-idf weights can be considered to be topical words, whereas words with low weights occur either too often (e.g. function words) or too infrequently (e.g. spelling mistakes) to be of topical value. In the experiments below, I compute the distance between stories by means of the Cosine distance. Again, I apply a bootstrap procedure in which for 5000 trials, I sample 10% of the vector space to produce a ranking of potential pre-texts for each story in the collection.

## 4.5 RESULTS

Table 4.1 presents the results of the analyses. For both feature representations (i.e. the questionnaire and the bag-of-words vectors) the table lists the mean ( $\mu$ ), median ( $\hat{x}$ ), and standard deviation ( $\sigma$ ) of the timespan between stories and their most likely pre-text. In addition, the table shows the results obtained from applying the bootstrap procedure to each of these two feature representations. The bootstrap procedure produced slightly smaller numbers for both feature representations. Since these bootstrapped numbers represent the mode consensus computed by approaching the data from various angles over a large number of trials, they are to be considered to be the most reliable results. The three distance measures (Manhattan distance ( $d_1$ ), Jaccard dissimilarity index ( $d_j$ ), and Cosine distance ( $d_C$ )) generated nearly equivalent numbers, which is an indication of the robustness of the results. On average, the adjusted timespan between a retelling and its pre-text is about 30 years. However, as will be discussed in more detail below, the timespans distributions are rather skewed. Therefore, rather than using the mean  $\mu$ , it is more appropriate to use the median  $\hat{x}$  as a measure of location, which amounts to approximately 17 years. This number for

70. Manning et al., *Introduction to Information Retrieval*.

model	$d_1$	$d_J$	$d_C$
	$\mu / \hat{x} / \sigma$	$\mu / \hat{x} / \sigma$	$\mu / \hat{x} / \sigma$
questionnaire	35 / 21 / 38	33 / 18 / 38	33 / 18 / 38
questionnaire (bootstr.)	31 / 18 / 33	30 / 17 / 32	30 / 17 / 32
bag-of-words	–	–	21 / 14 / 23
bag-of-words (bootstr.)	–	–	18 / 12 / 20

Table 4.1: Rounded mean ( $\mu$ ), median ( $\hat{x}$ ) and standard deviation ( $\sigma$ ) of timespans adjusted according to Eq. 4.4. Results are given for the questionnaire with the complete feature space, the bootstrap model applied to the questionnaire, the bag-of-words model and the bootstrap model applied to the bag-of-words model. The table shows results obtained by the Manhattan distance ( $d_1$ ), Jaccard dissimilarity index ( $d_J$ ) and the Cosine distance ( $d_C$ ).

the bootstrapped bag-of-words model is about five years less, which indicates that, solely on the basis of their vocabulary, stories seem to select pre-texts in slightly closer temporal proximity. However, without further research, we cannot rule out the possibility that this difference is a reflection of spelling reforms and variation.

It should be noted that the numbers listed in Table 4.1 do not conclusively resolve the question whether retellings are based on intermediate retellings and whether the development of “Red Riding Hood” in the Netherlands can be described as a process of ‘gradual accumulation of modification’ (cf. Section 4.1). There is no ‘ground truth’ indicating which text formed the pre-text of a particular retelling, either because we cannot definitely retrieve it or because it is doubtful whether this concept of a ground truth is applicable to the idea of a pre-text at all. Thus, all we have at our disposal to investigate relations between stories is statistical distributions of timespans. The challenge, then, is to reject alternative hypotheses about the origins of these distributions.

The first hypothesis is that the reported median timespans express artifacts of the data collection or the specific methods employed. Recall from Section 4.3.1 that the distribution of publications is rather skewed towards more recent publication years, and, as such, older retellings have a smaller pool of potential pre-texts to sample from than more recent retellings. More specifically, one might argue that the timespan

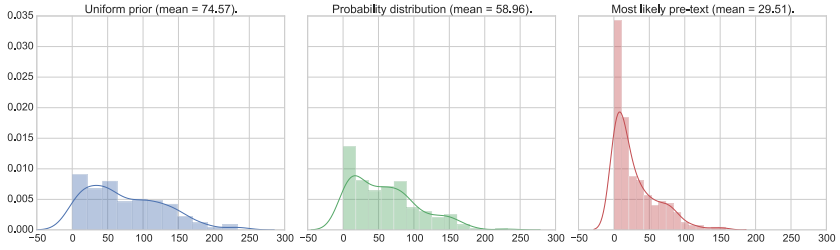


Figure 4.13: Kernel density plots of timespan distributions. The left plot is generated using a uniform prior over previous generations. The plot in the middle represents a non-uniform random sample from the probability distribution over years of publication. The third plot presents the distribution over timespans that was estimated using the bootstrap procedure with  $k = 1$  and the Cosine distance  $d_C$ .

distribution computed on the basis of the similarities between retellings and their pre-texts is not significantly different from a distribution in which pre-texts have been selected without any textual information, but, for example, on the basis of a uniform prior over previous generations. In such a distribution, a pre-text stems just as likely from the immediate previous year as from  $n$  years back in time. To estimate such a distribution, I randomly select a year of publication for each story in the collection from the range  $[t_{\min}, t_i]$ , where  $t_i$  is the year of publication of a story  $i$  and  $t_{\min}$  is the publication year of the first story in the collection (i.e. 1781).

Of course, a uniform prior over previous years of publications might be too naive and too easily rejected. I therefore test a second competing hypothesis, which investigates whether the extracted timespan distribution is significantly different from a distribution in which pre-texts are selected on the basis of a probability distribution over the observed years of publication in the collection. In this distribution, years in which many stories have been published have a higher probability than years in which only a few stories were attested. As such, this hypothesis more directly addresses the potential bias towards shorter timespans between stories and pre-texts resulting from the skewed distribution of publication years in the collection. To estimate this distribution of timespans, I randomly select a year of publication from the range  $[t_{\min}, t_i]$  for each story  $i$  in the collection on the basis of the probabilities associated with

each publication year in the collection.

Figure 4.13 presents the kernel density plots of the three timespan distributions. The right subplot displays the estimated text-based timespan distribution (computed using the bootstrap procedure and the Cosine distance). The left subplot displays the density of a random sample generated from a uniform prior over previous years of publication. A two-sample Kolmogorov-Smirnov test rejects the hypothesis that these two samples are drawn from the same distribution ( $D = 0.404, p < 0.0001$ ). This is confirmed by a Wilcoxon rank-sum test:  $z = 12.902, p < 0.0001$ . The subplot in the middle represents a random sample generated on the basis of the probability distribution over years of publication in the collection. We can also safely reject the hypothesis that this sample is the same as the text-based timespan distribution (Kolmogorov-Smirnov test:  $D = 0.312, p < 0.0001$ ; Wilcoxon rank-sum test:  $z = 9.643, p < 0.0001$ ).

Note that the subplots in Figure 4.13 display an increasingly skewed timespan distribution with pre-texts being published in temporal proximity of retellings. In what follows, I will formally characterize this skewed distribution. The Gamma distribution covers a wide range of skewness, and, as such, it appears to be a good candidate to fit the text-based timespan distribution. The probability density function of the Gamma distribution is given by:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)} \quad (4.5)$$

for  $x \geq 0$ , and  $\alpha, \beta > 0$ . The shape parameter is represented by  $\alpha$ , and  $\beta$  is the scale parameter. I perform a maximum likelihood estimation of the distribution parameters  $\alpha$  and  $\beta$ . The parameters resulting in the smallest sum of square errors between the text-based distribution and the fitted distribution are given by  $\alpha = 0.84$  and  $\beta = 22.7$ . To statistically assess the similarity between the fitted distribution and the text-based timespan distribution, I subject the two distributions to a two-sample Kolmogorov-Smirnov test and a Wilcoxon rank-sum test. Both statistical assessments yield non-significant results at the five percent level ( $D = 0.3, p > 0.1$  and  $z = 1.257, p > 0.2$ ). Therefore, we cannot reject the null hypothesis that the distributions are the same, and hence, the text-based timespan distribution can be considered to be



	Kolmogorov-Smirnov		Wilcoxon rank-sum	
	KS	p	z	p
1850 – 1900	0.55	< 0.004	1.677	> 0.09
1900 – 1950	0.3	> 0.2	0.703	> 0.4
1950 – 2000	0.25	> 0.4	0.839	> 0.4
2000 – 2015	0.4	> 0.05	0.947	> 0.3

Table 4.2: Kolmogorov-Smirnov tests and Wilcoxon rank-sum tests for the null hypothesis that the text-based timespan distributions in four consecutive time periods are gamma-distributed.

Gamma-distributed. Table 4.2 presents the results for the same analysis on four consecutive time periods. With the exception of the Kolmogorov-Smirnov test for the period 1850 – 1900, no test is significant at the five percent level. This suggests that the Gamma distribution is applicable to the complete collection.

## 4.6 DISCUSSION

The results presented in this article strongly suggest that retellings of “Red Riding Hood” are most similar to intermediate retellings. First, I have shown that such ‘intermediate retellings’ can be refined as retellings that are published in close temporal proximity. Acknowledging that there is no definite way to ascertain that a particular author based her or his retelling on one specific pre-text,<sup>71</sup> I suggest that the focus should be shifted to textual resonances,<sup>72</sup> which form a stream of previous retellings. It was shown that this stream gradually changes in a cumulative way, with retellings modifying and adapting other retellings. As such, the transmission process of literary versions of “Red Riding Hood” in the Netherlands can be considered a cultural evolutionary process: in producing a new retelling, authors select from a variety of competing existing retellings and potentially introduce innovations, which can eventually replace existing story elements. If these innovations are, in their

71. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Meta-narratives in Children’s Literature*.

72. Cf. Frank, *Letting Stories Breathe. A Socio-Narratology*.

turn, adopted in further retellings, the cycle of variation-competition-inheritance leads to a gradual accumulation of modifications. Further examining the selection biases, this study indicates that retellings of “Red Riding Hood” are evidently not ‘second generation’ stories that are simply based on one of the ‘classic’ versions written by either Perrault or the Brothers Grimm, and it is also unlikely that retellers sample their base material from a uniform distribution over all previous generations. Instead, it appears that more recent versions, at a median of about 17 years, are more likely to inspire retellings than versions from older dates.

Thus, the present study shows that age-dependent selection criteria can – at least partially – explain the differential fitness among competing retellings of “Red Riding Hood”. Given this insight, we might wonder what this tells us about the development in children’s literature at large and even cultural change in general. A key virtue of the cultural evolutionary approach as advocated in the present study is the use of idealized explanatory models of cultural change.<sup>73</sup> Such models are deliberately kept simple in order to facilitate isolating and manipulating single variables under highly controlled conditions.<sup>74</sup> Yet, despite their simplicity, these models appear to have great informative strength in that they are able to explain unanticipated, complex population-level effects resulting from the actions of individuals.<sup>75</sup> The value of the current study lies in the fact that it considers the evolution of a cultural artifact using a particular methodology that can easily be compared to more parsimonious explanatory models of cultural change, and as such allows us to assess the explanatory strength of a model involving age-dependent selection.

In recent years, several publications within the field of cultural evolution have shown that various population-level effects can in fact often be explained by means of an unbiased process in a neutral model of

73. Cf. e.g. Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*; Peter J. Richerson and Robert Boyd, *Not By Genes Alone: How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2005); Tim Lewens, *Cultural Evolution. Conceptual Challenges* (Oxford: Oxford University Press, 2015).

74. Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

75. Richard McElreath et al., “Applying evolutionary models to the laboratory study of social learning,” *Evolution and Human Behavior* 26 (6 2005): 483–508; Lewens, *Cultural Evolution. Conceptual Challenges*.

selection.<sup>76</sup> This neutral model offers a parsimonious explanation for cultural change as it merely assumes individuals either copy the behavior from a randomly selected individual from the previous generation or they innovate and introduce a new cultural trait into the population. Whether or not individuals introduce an innovation is controlled by an innovation rate parameter  $\mu$ . With smaller values of  $\mu$  the probability of innovation decreases. It has been shown that this simple model provides accurate predictions for a variety of cultural changes, such as the choice of baby names,<sup>77</sup> the selection of keyword in academic publications<sup>78</sup> and the popularity of dog breeds.<sup>79</sup> In fact, this random copying model, equivalent to biological genetic drift, appears to be so powerful that researchers in the cultural evolution research program consider it to be the null-hypothesis in the description of cultural evolutionary processes.

I take it to be a highly intriguing question whether and to what extent the evolutionary process of “Red Riding Hood”, and children’s literature in general, can also be explained under a neutral model of evolution in which retellers base their material on randomly selected retellings from previous generations. Under the neutral model, no single retelling of a story is more valuable than others, and whether or not a particular story element is adopted in a retelling is proportional to its popularity in previous generations. It is clear, however, that a neutral copying model does not suffice for explaining the strong preference for selecting temporally proximate pre-texts.<sup>80</sup> While under the neutral model, any retelling can become the dominant pre-text for further retelling, the present study has clearly shown the need for a mechanism that explains the age-bias for selecting more recently produced versions.

Interestingly, the apparent age-bias (as represented by the discovered gamma-distributed selection of pre-texts with a strong lopsidedness

76. See e.g. R. Alexander Bentley et al., “Random drift and culture change,” *Proceedings of the Royal Society B: Biological Sciences* 271, no. 1547 (2004): 1443–1450; Alex Mesoudi and Stephen J. Lycett, “Random copying, frequency-dependent copying and culture change,” *Evolution and Human Behavior* 30, no. 1 (2009): 41–48.

77. Hahn and Bentley, “Drift as a mechanism for cultural change: an example from baby names.”

78. Bentley, “Random Drift versus Selection in Academic Vocabulary: An Evolutionary Analysis of Published Keywords.”

79. Herzog et al., “Random drift and large shifts in popularity of dog breeds.”

80. Cf. Alberto Acerbi et al., “The Logic of Fashion Cycles,” *PLoS ONE* 7, no. 3 (2012): e32541–9.

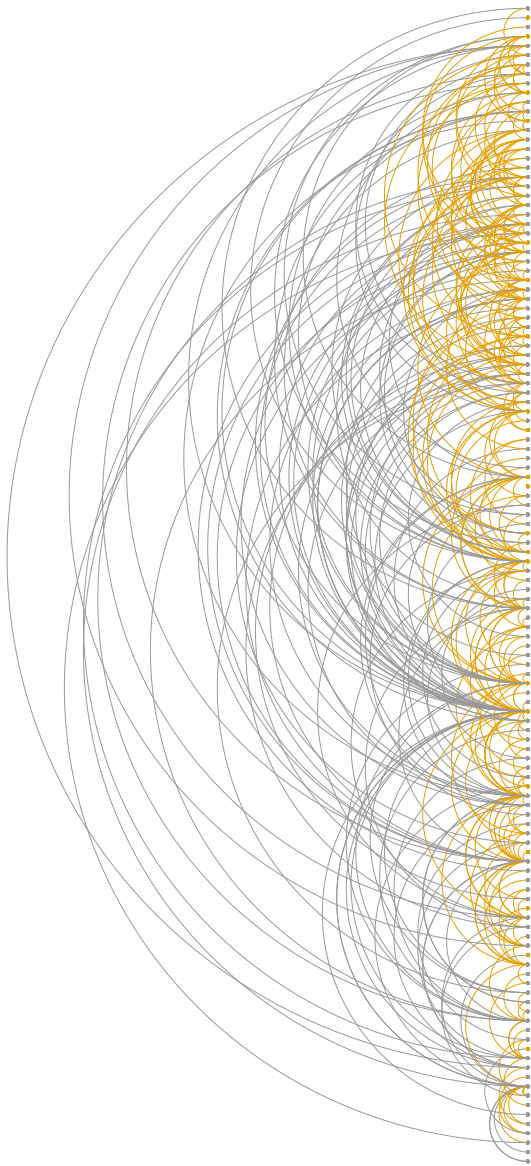


Figure 4.14: Arc diagram of pre-text selection. Nodes represent years of publication observed in the collection. Arcs between two years  $i$  and  $j$  are created if a story from year  $i$  selects a story from year  $j$  as its pre-text. Arcs colored yellow represent time spans shorter than the mean time span; gray arcs represent time spans longer than the mean. A node colored yellow indicates that in its corresponding year at least one retelling selects a pre-text with the same year of publication as the retelling.

towards texts in temporal proximity) could be interpreted as a kind of 'cultural amnesia'. Figure 4.14 presents an alternative visualization of this gamma distribution. In this so-called 'arc diagram', nodes represent years of publication observed in the collection. Arcs between two years  $i$  and  $j$  are created if a story from year  $i$  selects a story from year  $j$  as its pre-text. Arcs colored yellow represent time spans shorter than the mean time span; gray arcs represent time spans longer than the mean. A node colored yellow indicates that in its corresponding year at least one retelling selects a pre-text with the same year of publication as the retelling. The gamma distribution, then, resembles a kind of 'memory parameter'. It is an interesting question how this memory parameter influences the rate of change of a cultural artifact.<sup>81</sup> One might hypothesize that with an increasing number of possible generations to sample from, the rate of change decreases. With more generations, it is more difficult for an innovation to 'catch on' because its likelihood of being selected is reduced. This inhibiting effect of a larger (textual) community on change is addressed by Anderson in the context of oral culture:

"[I]f details become sufficiently well fixed, generations of storytellers and listeners in an oral culture will automatically correct them a good deal of the time. A brilliant parody of the mischievous grandfather attempting to tell wrong versions of a well-known fairytale furnishes all the proof that is needed: the child will not tolerate a tale of 'Little Green Riding Hood' going through the wood and meeting a giraffe, or meeting a wolf that asks 'what's ten times eight?'."<sup>82</sup>

Although it is doubtful whether a child will indeed not tolerate such a story, she will immediately recognize it to be a retelling of "Red Riding Hood", which essentially proves the same point.

Bentley et al. study the effect of such a memory parameter in the context of the neutral model of evolution.<sup>83</sup> They generalize the neutral

81. Cf. Charles Perreault, "The Pace of Cultural Evolution," *PLoS ONE* 7, no. 9 (2012): e45150.

82. Graham Anderson, *Fairytale in the Ancient World* (London, New York: Routledge, 2000), 19.

83. R. Alexander Bentley et al., "Evolving social influence in large populations," *Behavioral Ecology and Sociobiology* 65, no. 3 (2011): 537–546.

model by adding a memory parameter  $m$ . This parameter controls how many generations an individual is allowed to look back in time, ranging from only the immediate previous generation ( $m = 1$ ) to all generations ( $m = \text{all}$ ). In the case of  $m = 1$  the model is equivalent to the ‘traditional’ random copying model. In the special case of  $m = \text{all}$ , cultural traits do not go extinct. Bentley et al. show that increasing  $m$  reduces the effect of the innovation rate parameter, and hence has an inhibiting effect on cultural change. Furthermore, they show that adding more generations to the sample space has a conservative effect on the replacement of words in the vocabulary of languages.<sup>84</sup> In the memory-parameterized neutral model, individuals randomly select an individual to copy from using a uniform distribution over  $m$  previous generations. The memory-parameterized model by Bentley et al. is essentially useful, but, as the present study has shown, a uniform distribution over  $m$  previous generations might be too simplistic to account for all processes of cultural change. Because of its simplicity, however, the memory-parameterized model can easily be modified so that it accounts for more skewed distributions in which age acts as a bias in selection.

Simple, idealized models of cultural change, such as the memory-parameterized neutral model, allow researchers to abstract away from idiosyncratic properties of a particular case under investigation, and apply it to both children’s literature at large as well as cultural change in general. The age-dependent selection in the transmission of “Red Riding Hood” also emerges in examples from biology<sup>85</sup> and ties in with other studies in cultural evolution investigating fashion trends.<sup>86</sup> It would be interesting to further test whether the hypothesis that pre-texts in children’s literature tend to come from the recent past, and subsequently examine whether there is a different age-bias in the transmission of other textual genres or cultural artifacts. One of the most important assets of the cultural evolutionary modeling approach, then, is that it

84. R. Alexander Bentley et al., “Population-level neutral model already explains linguistic patterns,” *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1713 (2011): 1770–1772.

85. E.g. Andrea S. Grunst et al., “Age-dependent relationships between multiple sexual pigments and condition in males and females,” *Behavioral Ecology*, 2014, doi:10.1093/beheco/art124.

86. E.g. Acerbi et al., “The Logic of Fashion Cycles”; Kandler and Powell, “Inferring Learning Strategies from Cultural Frequency Data”; Matthias Mauch et al., “The evolution of popular music: USA 1960–2010,” *Royal Society Open Science* 2, no. 5 (2015): 150081.

enables researchers to explore parsimonious mechanisms and generic explanations that intersect such diverse examples of cultural change.

It should be stressed that age-bias does not fully account for the differential fitness of story variants. The preference for temporally proximate pre-texts alone cannot explain why authors choose for one or another retelling from the same time period. To arrive at such insights, then, we should consider “a longish list of psychological, social, and ecological processes that interact to generate the differential “fitness” of cultural variants”.<sup>87</sup> For instance, the growing disapproval of violence in the poetics of children’s stories may have impacted the choices made by retellers (e.g. within the group of temporally proximate retellings, non-violent versions may thus display a transmission advantage over more violent ones). If such socio-cultural factors co-generate differential fitness with age-dependent selection, then the combination of these factors should yield a model of cultural change that accurately reflects the mechanisms underlying the selection of pre-texts.

In my final words, I would like to draw attention to the collection of Dutch “Red Riding Hood” retellings presented in this study. I believe that this collection, which consists of half a million words of “Red Riding Hood” and spans a time period of more than two hundred years, forms a small, yet unique and highly specialized collection that will hold potential for a variety of research activities to all those engaged in the study of children’s literature, folkloristics, linguistics and cultural evolution. One may assume that the texts in the collection target a homogeneous audience and generally tell the same abstract content over and over. Characteristics as these allow researchers to study aspects of micro-variation, linguistic change, literary developments and cultural evolution that are otherwise deemed less attainable.

87. Henrich et al., “Five Misunderstandings About Cultural Evolution.”

---

## CHAPTER 5

# THE STRUCTURE AND EVOLUTION OF STORY NETWORKS

*“I, too, feel the need to reread the books I have already read,” a third reader says, “but at every rereading I seem to be reading a new book, for the first time. Is it I who keep changing and seeing new things of which I was not previously aware? Or is reading a construction that assumes form, assembling a great number of variables, and therefore something that cannot be repeated twice according to the same pattern?”*

ITALO CALVINO, *If on a Winter's Night a Traveler*

### 5.1 INTRODUCTION

In his thought-provoking study *Fairy Tale in the Ancient World*, Graham Anderson quotes the following passage by the Greek geographer Strabo (first century BC/AD), which tells the story of a girl called Rhodopis:

*“They tell the fabulous story (mytheuouisi) that while she was bathing, an eagle seized one of her shoes from her maid and brought it to Memphis, and while the king was dispensing justice in the open air, the eagle arrived over his head and threw the shoe into his lap. The king was aroused by the *rythmos* of the sandal and the strangeness of the event, and sent all around the country in search of the woman who wore it. When she was found in Naucratis she was brought up country to Memphis and became the king’s wife.”<sup>1</sup>*

Does this sound familiar? The ‘seizure of the girl’s shoe’, the ‘slipper test’ and the ‘marriage to the prince’ are all motifs that resonate one of

1. Anderson, *Fairytale in the Ancient World*.



the best known fairy tales in modern times: *Cinderella*.<sup>2</sup> The ‘Cinderella’ story as we know it today is derived from Charles Perrault’s story *Cendrillon* (from *Contes du temps passé avec moralités*, 1697). Perrault’s retelling adds various elements to the story, of which the following two are mentioned in the folktale catalog by Aarne & Thompson: a persecuted heroine (1) and a meeting with the prince in advance of the slipper test (3).<sup>3</sup> Ever since Perrault published his version of the story, *Cinderella* has been retold to new audiences through a variety of channels: books, picture books, films, advertisements, comics, cartoons, and so forth. Yet, these retellings of *Cinderella* do not necessarily derive from Perrault’s version. In fact, as Stephens & McCallum state, it is more likely that retellers “use intermediate versions – to produce a retelling of a retelling”.<sup>4</sup> These ‘retellings of retellings’, I wish to argue, can be considered as the implicit formation of a network of stories, in which links between stories represent pre-textual relationships.<sup>5</sup> A story network represents a stream of retellings in which retellers modify and adapt retellings in a gradual and accumulative way.

The aim of this chapter is to offer new perspectives on the structure and development of such story networks. More specifically, I am interested in the dynamics and mechanisms that underly retellers’ choices for particular story versions to base their retellings on. Certain retellings seem to be more attractive than others, making them more likely candidates for further retelling. Arguably, attractiveness can be defined in two ways: *content-based* and *context-based* attractiveness.<sup>6</sup> *Content-based attractiveness* concerns inherent aspects of a story which increase or

2. In fact, Rhodopis’ story exhibits three of the five main characteristics attributed to the Cinderella story type (as characterized by Aarne & Thompson Aarne and Thompson, *The Types of the Folktale: A Classification and Bibliography* in the folktale catalog *The Types of the Folktale*): help of an animal (2), proof of identity (4) and marriage with the prince (5).

3. Ibid.

4. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children’s Literature*.

5. In evolutionary terms, such networks can be described as lineages or phylogenetic trees. I prefer to use the term network, however, because it does not presume a clear ‘root node’ from which all subsequent story versions have supposedly sprang. Cf. Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*; Tehrani, “The Phylogeny of Little Red Riding Hood.”

6. These types resemble the concepts of content-based and context-based biases in theoretical models of cultural evolution. Cf. Henrich and McElreath, “The evolution of cultural evolution.”

decrease its likelihood of being retold. For instance, Charles Perrault's retelling of "Little Red Riding Hood" was highly popular until the Brothers Grimm published their version of "Rothkäpchen" in the 19<sup>th</sup> century. Zipes' thesis is that the Brothers Grimm "virtually dwarfed Perrault's version" by the end of the 19<sup>th</sup> century, because their emphasis on obedience and good behavior was a better fit for the emerging Victorian image of the child.<sup>7</sup> With *context-based* attractiveness, on the other hand, dispositions for certain stories are not determined by inherent features, but, for example, by social factors, such as popularity or prestige of a particular author. Both content-based and context-based attractiveness have received a wealth of attention in literary and folkloristic studies of story transmission.<sup>8</sup> However, despite being suggestive and thought-provoking, informal verbal arguments such as Zipes' account of "Red Riding Hood", cannot generate specific predictions which can be quantitatively tested and systematically compared to real-world data.

A more parsimonious explanation for the preference of a reteller for particular story versions is that there are no real 'motivations' or selection criteria underlying their choices, or, in other words, that their choice is completely random. A large number of studies in evolutionary anthropology and cultural evolution has shown that social transmission can often be characterized as an *unbiased* process in a neutral model of selection in which changes are reduced to 'random' frequency effects of competing cultural traits.<sup>9</sup> In the case of story transmission, this would mean that stories with high circulation numbers are more readily available and, in the absence of content- and context-based biases, their attractiveness would be entirely proportional to these numbers. However, if we take enticing accounts of story transmission such as the one by Zipes seriously, it seems unlikely that the selection of a particular story for retelling is entirely frequency-based.

In this study, I wish to depart from the hypothesis that a story's attractiveness for further retelling is merely a 'random' frequency effect – or, in other words, is driven by *frequency-based attractiveness* – by

7. Zipes, *The Trials & Tribulations of Little Red Riding Hood*.

8. Geerts and van den Bossche, *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*.

9. Bentley et al., "Random drift and culture change"; Mesoudi and Lycett, "Random copying, frequency-dependent copying and culture change"; Bentley et al., "Evolving social influence in large populations."

systematically investigating the possible influence of other attractiveness factors. First, besides *frequency-based attractiveness*, stories might be differentially preferred given their *temporal attractiveness*, which is a form of context-based attractiveness. For instance, it has been shown for academic citation networks that relatively young research is preferred over older studies and that the probability of being cited decays with time.<sup>10</sup> Following Stephens & McCallum, I investigate whether this process equally applies to story networks and whether retellers prefer more recent story versions over older ones in producing a retelling.<sup>11</sup> Second, a story might also be more (or less) attractive because, for example, its author enjoys high esteem. This type of context-based attractiveness will be termed *model-based attractiveness*. While each of the three types of attractiveness, i.e. *frequency-based*, *temporal*, and *model-based*, could potentially serve as the sole explicatory factor in story transmission, I wish to suggest that these three kinds of attractiveness interact and collectively impact the choice for particular story versions. Thus, explicatory accounts of story transmission need to account for the interaction between all forces of attractiveness in order to arrive at a more adequate and full explanation of retellers' preferences for particular story versions.

In order to investigate these issues, the current study aims to contribute to the development of methodologies that allow us to induce micro-evolutionary mechanisms underlying macro-evolutionary developments from historical, population-level data.<sup>12</sup> The first challenge is

10. Sergei Dorogovtsev and José Fernando Mendes, "Evolution of reference networks with aging," *Physical Review. E, Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics* 62, no. 2 (2000): 1842–1845; Young-Ho Eom and Santo Fortunato, "Characterizing and Modeling Citation Dynamics," *PLoS ONE* 6, no. 9 (2011): e24926–7; Derek De Solla Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science* 27, no. 5 (1976): 292–306; Matjaž Perc, "The Matthew effect in empirical data," *Journal of The Royal Society Interface* 11, no. 98 (2014), doi:10.1098/rsif.2014.0378.

11. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*.

12. Mesoudi and Lycett, "Random copying, frequency-dependent copying and culture change"; Anne Kandler and Stephen Shennan, "A non-equilibrium neutral model for analysing cultural change," *Journal of theoretical biology* 330 (2013): 18–25; Bret A. Beheim et al., "Strategic social learning and the population dynamics of human behavior: the game of Go," *Evolution and Human Behavior* 35, no. 5 (2014): 351–357; Alberto Acerbi and R. Alexander Bentley, "Biases in cultural transmission shape the turnover of popular traits," *Evolution and Human Behavior* 35, no. 3 (2014): 228–236; Sven Isaksson et al., "A Novel Method to Analyze Social Transmission in Chronologically Sequenced Assemblages,

to develop methods to automatically extract story networks from raw texts that express pre-textual relationships. When such story networks are extracted, we can resort to well-studied concepts and methodologies from network theory to describe their topological and macroscopic properties statistically.<sup>13</sup> In this chapter, specific attention will be devoted to the degree distributions of story networks, because they provide information about the connections between stories and their pre-texts. Some story versions are used only once to produce a retelling, whereas others serve as pre-textual context for many other stories and could be called ‘story hubs’. The central question is, then, how we can characterize the distribution with which stories are selected as pre-text, and how such distributions come into being. Following previous models of network growth,<sup>14</sup> the present study investigates a growing network model which combines the three aforementioned kinds of cultural attractiveness. I analyze these forces of attractiveness in isolation as well as the interplay between them and show how their degree distributions behave in relation to those of two empirical story networks.

The main object of study in this Chapter is the collection of Dutch literary *Little Red Riding Hood* retellings introduced in the previous chapter. In Chapter 4, it has been demonstrated that the development of the story about the little girl in red is evolutionary in nature: retellers produce modifications of existing retellings that, in turn, serve as pre-texts for new retellings of the most popular fairy tale of the Western world. Furthermore, it is shown that retellers of “Red Riding Hood” prefer to base their retellings on story versions that are published in close temporal proximity. Yet, as hypothesized above, temporal attractiveness alone cannot explain retellers’ choices for particular retellings from the same time period. The current chapter seeks to acquire a better understanding of which mechanisms possibly underlie the selection of pre-texts by extracting a story network from the data, and subsequently

Implemented on Cultural Inheritance of the Art of Cooking,” *PLoS ONE* 10, no. 5 (2015): e0122092–13.

13. Mark E. J. Newman, “The structure and function of complex networks,” *SIAM REVIEW* 45, no. 2 (2003): 167–256.

14. Price, “A general theory of bibliometric and other cumulative advantage processes”; Dorogovtsev and Mendes, “Evolution of reference networks with aging”; Albert-Laszlo Barabási and Reka Albert, “Emergence of Scaling in Random Networks,” *Science* 286, no. 5439 (1999): 509–512; Eom and Fortunato, “Characterizing and Modeling Citation Dynamics.”

assessing its structure.

There is, however, a considerable difficulty associated with assessing the validity of the extracted story network and its structural properties, as we lack a ‘ground truth’ of which story served as pre-text for a retelling. For this reason, I made the methodological choice of comparing the structure and development of the story network of “Red Riding Hood” to that of a large collection of paper chain letters. This collection consists of over five hundred letters from the 20<sup>th</sup> century and represents one hundred years of cultural copying. Although chain letters are fundamentally different from fairy tales in many respects, they do make an interesting comparison because of their explicit request to replicate and redistribute the contents of the letter – sometimes to a fixed number of people and often within a particular time window. Crucially, because of this request, we can make at least two predictions about the structure and development of a chain letter network. First, it can be expected that chain letters are connected to pre-texts in close temporal proximity. Second, in a perfect chain (i.e. when all successive recipients of a letter adhere to its request), we can expect a graph structure with a relatively uniform degree distribution, in which all stories exhibit approximately equal degree. These expectations we have of the properties of chain letter networks are confirmed by previous studies that have enhanced the understanding of spreading patterns in Internet chain letter networks.<sup>15</sup> Showing that the extracted chain letter network displays structural and developmental properties that are in accordance with our preconceptions about these properties allows us to partially check for the reliability of the employed methods, and hence serves to strengthen the confidence in the validity of conclusions based on the “Red Riding Hood” network and its extracted properties.

The remainder of this chapter is structured as follows. I begin with a description of the data collections used in this chapter (Section 5.2). After having presented the data collections, I proceed in Section 5.3 with a detailed account of the computational and statistical methods used to construct and analyze story networks. In Section 5.4, then, the structure of the two story networks will be analyzed and compared to those of a model of network growth. The final section offers a discussion about

15. David Liben-Nowell and Jon Kleinberg, “Tracing information flow on a global scale using Internet chain-letter data,” *PNAS* 105, no. 12 (2008): 4633–4638.

the main findings of this chapter.

## 5.2 DATA COLLECTIONS

The analysis in the present chapter is based on the diachronic collection of Dutch “Red Riding Hood” retellings presented in Chapter 4.<sup>16</sup> In the remainder of this section, then, I will focus on the description of the chain letter corpus, which is based on Daniel Vanarsdale’s online *Paper Chain Letter Archive*. Vanarsdale’s archive contains over nine hundred letters, most of which have been transcribed from physical letters.<sup>17</sup> In Vanarsdale’s definition, chain letters are letters that explicitly ask the recipients to copy their contents and redistribute them to a (sometimes explicitly given) number of successive recipients. Some chain letters explicitly ask the recipients to make modifications to the letters, for example by adding their name to the existing list of recipients. Vanarsdale classifies his collection of chain letters into nine categories. In this study, I investigate the development of the largest category, Luck chain letters. The Luck chain letter is generally believed to be derived from the ‘Himmelsbrief’ (Letter from Heaven).<sup>18</sup> The earliest attestations of the Himmelsbrief date from the 17<sup>th</sup> century. The letters are supposedly derived from a mysterious letter written in golden ink by Jesus and was delivered to earth by the archangel Gabriel.<sup>19</sup> The letters generally warn against sin, contain prayers and encourage doing what is right according to Christian beliefs. The most characteristic feature of these letters is their demand to make one or more copies of the letter. The recipient is warned that if (s)he does not believe in the contents of the letter and refuses to follow what it teaches, (s)he “will be punished in eternity, and I [Jesus] shall demand your many sins on Judgment Day, and you will have to answer to me for them”.<sup>20</sup>

The earliest examples of Luck chain letters in Vanarsdale’s collection (‘Ancient Prayer’ letters) adhere to the main characteristics of the Him-

16. Karsdorp, *Story network data sets*.

17. Daniel W. VanArsdale, *The Paper Chain Letter Archive*, <http://www.silcom.com/~barnowl/chain-letter/archive!/content.html>, Accessed: 2015-10-09, 2015.

18. Daniel W. VanArsdale, *Chain Letter Evolution*, <http://www.silcom.com/~barnowl/chain-letter/evolution.html>, Accessed: 2015-10-09, 2015.

19. Bill Ellis, *Lucifer Ascending: The Occult in Folklore and Popular Culture* (Lexington: University Press of Kentucky, 2004).

20. Ibid.

melsbrief tradition. The letters typically start with a prayer of which the origin is explained in the next few lines. The recipients are urged to copy the contents of the letter and distribute it to a fixed number of other persons. Those who follow the instructions of the letter will experience good fortune, whereas those who ignore it are threatened, often with death. An example from 1906 reads as follows:

*I received the other day a chain prayer.*

*Oh, Lord Jesus Christ, we implore Thee, O Eternal God, to have mercy upon mankind. Keep us from all sin and take us to be with Thee eternally. Amen*

*This prayer was sent by Bishop Lawrence, recommending it to be rewritten and sent to nine other persons. He who will not say it will be afflicted with some great misfortune. One person who failed to pay attention to it met with a dreadful accident. He who will rewrite it to nine other persons commencing on the day it is received - and sending only one each day will on or after the ninth day experience great joy.*

*Please do not break the chain.*<sup>21</sup>

Vanarsdale classifies the Luck chain letters of the 20<sup>th</sup> century into 12 distinctive chronological types, which he describes as the “mainline – a century long stream of copying”. Most types display clear influences of prior letter types. The latest letter type ‘Death-Lottery’ which predominantly circulated from 1973 until 2005, for example, is a reversal of the ‘Lottery-Death’ letter type. In these later examples of the Luck chain letter, greater emphasis is put on superstitious beliefs and on possible negative effects of breaking the chain. Reflecting modern times, the number of requested copies increases in these letters, as well as the amounts of money people might receive if they obey the letters’ orders. Another key characteristic of these younger letters is the postscript “It works”. Vanarsdale observes that after the first attestation of this postscript in 1979, in a few years time all succeeding letters include it. The following letter illustrates the youngest form:

21. Taken from [le1906-01-06\\_ap!\\_lawrence\\_q9.htm](#) in VanArsdale, *The Paper Chain Letter Archive*.

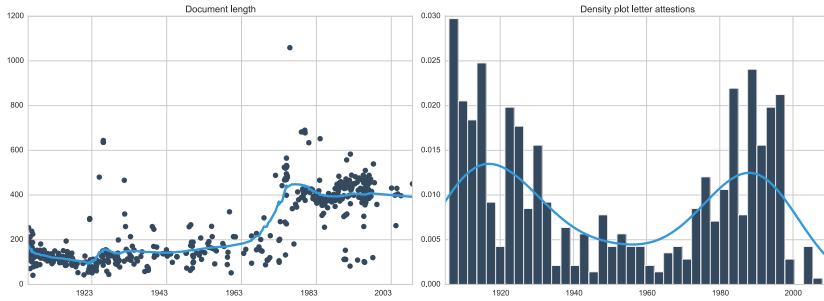


Figure 5.1: Diachronic visualization of some general statistics about the Luck chain letter corpus. The left subplot visualizes the letter length (expressed in word forms), and the right subplot shows a kernel density plot of the number of letters attested per year.

*This letter has been sent to you for good luck. The original is in England. It has been around the world nine times. The luck has now been sent to you. You will receive good luck within four days of receiving this letter, providing you in turn send it on. This is no joke. Send copies to people you think need good luck. Don't send money, as fate has no price. Do not keep this letter. It must leave your hands within 96 hours.*

*An RAF officer received \$70,000. Joe Elliot received \$40,000 and lost it because he broke the chain. While in the Philippines Gene Welch lost his wife six days after receiving this letter and failing to circulate it. However, before her death he received \$7,755,000.*

*The chain comes from Venezuela and was written by Saul Anthony De Coup, a missionary from South Africa. Since the copy must make a tour of the world, you must make 20 copies and send them to your friends and associates. After a few days you will get a surprise. This is true even if you are not superstitious. Constantine Dias received the chain in 1953 and asked his secretary to send out 20 copies. A few days later he won a lottery of two million dollars. Please don't ignore this letter. "It works!"<sup>22</sup>*

For this study, I constructed a full-text version of the Luck chain letter collection.<sup>23</sup> The total number of letters in this version of the collection amounts to 554. The oldest letters stem from 1906 and the youngest from 2008. Each year in the collection is represented by six

22. Taken from `le1985-03_d1_w0_.htm` in *ibid.*

23. Karsdorp, *Story network data sets*.



letters on average. The collection was tokenized using the tokenizer of the *Natural Language Toolkit*.<sup>24</sup> The total number of words amounts to 134,589 (including punctuation). In their lowercased form, 5,256 of these word forms occur uniquely. Letters consist of 243 word forms on average. Figure 5.1 visualizes some general statistics of the corpus. The left subplot visualizes how the letter length changes over time. It can be observed that until the 1980s the letter length has remained generally stable over time. Around 1980 the letters suddenly become significantly longer, which might reflect some severe changes to the tradition that require further investigation. The right subplot shows the density with which letters have been attested each year. Most letters stem from either the beginning or from the last two decades of the 20<sup>th</sup> century.

### 5.3 STORY NETWORK CONSTRUCTION

Story networks consist of stories and links between stories that represent pre-textual relationships. In this study, I make the simplifying assumption that stories that are more similar to each other are more likely to stand in a pre-textual relationship than stories that are more distant. In what follows, I discuss a vocabulary-based representation of stories, the ‘bag-of-words’ representation, that is suitable for computational methods of discovering textual similarity. Subsequently, in Section 5.3.2 I present a clustering algorithm that aims to bootstrap pre-textual relationships between stories from raw collections of texts. Next, I describe how story networks are constructed on the basis of the output of this clustering procedure. Finally, I present the network-theoretic statistics to detect and describe pre-textual relationship characteristics of story networks in Section 5.3.3.

#### 5.3.1 BAG-OF-WORDS MODELS

Bag-of-words models have proven to be invaluable for numerous computational approaches to textual data, such as text classification, textual information retrieval and textual stylometry. Bag-of-words models make the assumption that maintaining word order is unnecessary in determining the relationship between texts. Obviously, this is a crude

24. Steven Bird et al., *Natural Language Processing with Python* (Sebastopol: O’Reilly, 2009).

simplification, yet there is a surprising number of applications in which word order adds barely any additional information. Given a corpus  $C$  consisting of a vocabulary  $V$  (i.e. word types), bag-of-words models represent documents as histograms over the vocabulary, that is, as vectors of occurrence counts of words. For a document  $d$ , the vector representation  $\mathbf{w}^{(d)}$  is given by  $(w_1, w_2, \dots, w_V)$ , where  $w_i$  represents the occurrence counts of word  $i$  in document  $d$ . Consider the following vocabulary:  $\{\text{story, dialogue, network, reader, model}\}$ . Each word in the vocabulary is represented by a unique integer:  $\{1, 2, 3, 4, 5\}$ . A document  $d$  that consists of the words “*model, network, network*” can then be represented by the following vector:  $\mathbf{w}^{(d)} = (0, 0, 2, 0, 1)$ .

It is common practice in many computational applications to weigh the word frequency vectors for term importance. Here, I make use of the well-known ‘term frequency-inverse document frequency’ (tf-idf) weighting scheme, which weighs the frequency of words in a document (tf) against the background of how many documents in the collection contain those words (idf).<sup>25</sup> The transformed vectors put more weight on words that occur relatively often in a particular document (i.e. their term frequency is high) and relatively rarely in the corpus as a whole (i.e. their document frequency is low). Words with high weights can be considered to be topical words that represent the contents of a text. Words with low values occur either too often in the corpus or too rarely in the document to be of topical value to a document.

Using these vector representations, the similarity (or distance) between two stories can be assessed by means of a pairwise comparison of their vector values. I choose to use the Cosine dissimilarity measure to express the distance between two stories in terms of their weighted occurrence count vectors. Given two story vectors  $\mathbf{a}$  and  $\mathbf{b}$ , the Cosine dissimilarity can be computed using the following expression:

$$\delta_C(\mathbf{a}, \mathbf{b}) = 1 - \frac{\sum_{i=1}^V a_i \times b_i}{\sqrt{\sum_{i=1}^V (a_i)^2} \times \sqrt{\sum_{i=1}^V (b_i)^2}} \quad (5.1)$$

where  $a_i$  represents the weighted count of word  $i$  in vector  $\mathbf{a}$  and  $b_i$  the weighted count in vector  $\mathbf{b}$ .

25. Manning et al., *Introduction to Information Retrieval*.

### 5.3.2 BOOTSTRAPPING PRE-TEXTUAL RELATIONSHIPS

Given a story, how can we identify a set of potential pre-texts  $P$  from a collection of stories  $C$ ? Using the story representation discussed in the previous section, we can compute the distance between a story  $i$  attested at time  $t_i$  and all potential pre-texts. Like in Chapter 4, stories are considered to be potential pre-texts if and only if they are attested prior to  $t_i$ , that is  $P = \{j | \forall j \in \{1, 2, \dots, C\} \wedge t_j \leq t_i \wedge i \neq j\}$ , where  $C$  represents the complete set of stories in the collection. This set of potential pre-texts can then be sorted in ascending order to obtain a ranking in which the top items represent the most like pre-texts of story  $i$ .

We can apply a cutoff to these rankings to assign to each story its  $k$  most likely pre-texts. At a cutoff of  $k = 1$ , we only take into consideration the most similar (or least distant) text, whereas at  $k = |P|$  the complete set of potential pre-texts is taken into account. Unfortunately, it is not straightforward to define the value  $k$  in a non-arbitrary way. A more fundamental problem of this approach, however, is the possibility that the set of potential pre-texts does not contain an actual pre-text in the first place. In the most extreme case, the most similar pre-text ( $k = 1$ ) is maximally distant from the story under consideration. A possible way to overcome this problem is to define a threshold value at which stories are considered to be too distant to be related. The exact value of this threshold, however, is sensitive to a multitude of factors, such as the corpus under investigation, its feature representation, and the distance metric used.

The problem we face is essentially an open-set problem: given a set of potential pre-texts, can we decide which of them are similar enough to be considered actual pre-texts of a particular story? This includes the possibility that *none* of the potential pre-texts are similar enough. The problem as formulated here is closely related to the *many candidates* problem in the context of authorship verification.<sup>26</sup> In that problem, the goal is to assess whether the author of some anonymous document is

26. Moshe Koppel and Yaron Winter, "Determining if Two Documents are by the Same Author," *JASIST*, no. 1 (2014): 178–187.

---

**Algorithm 2: Bootstrap Neighbor Clustering**


---

**Data:** story  $i$ , pre-texts  $P = \{j | \forall j \in \{1, 2, \dots, C\} \wedge t_j \leq t_i \wedge i \neq j\}$   
 $S \leftarrow$  an array of size  $|P|$  with assignment counts for each pre-text;  
 $A \leftarrow \emptyset$ ;  
**for** iteration  $t \leftarrow 1$  **to**  $T$  **do**  
     $P^{(t)} \leftarrow$  construct a new representation of  $P$  based on a random  
    sample (e.g. 50%) of the complete feature space;  
     $j \leftarrow \arg \min_{j \in P^{(t)}} d_C(i, j)$ ;  
     $S_j \leftarrow S_j + 1$ ;  
**for** pre-text  $j \leftarrow 1$  **to**  $|P|$  **do**  
    **if**  $\frac{1}{T} S_j > \sigma$  **then**  
         $A \leftarrow A \cup \{j\}$   
**return**  $A$

---

someone among a set of candidate authors, or a yet unknown author.<sup>27</sup> To overcome the sensitive practice of defining ranking cutoff values or distance thresholds, I will describe a procedure that allows us to exclude potential pre-texts from consideration in a more robust and less arbitrary way. The procedure draws inspiration from resampling methods<sup>28</sup> and methods proposed to tackle the author verification problem.<sup>29</sup>

The general idea behind the procedure is to assess the rankings of pre-texts for their robustness by approaching the data from various angles for a large number of trials. If we compute the distances between a story and its potential pre-texts on the basis of a slightly modified feature space (i.e. we leave out a small number of words from the bag-of-words representation), we want the resulting rankings to be consistent with those of the unmodified feature space. We repeat this process and in each trial we compute the distances between each story and its potential pre-texts on the basis of a random sample (e.g. 50%) of the complete feature space. Each trial potentially produces slightly (or very) different rankings. We select from each trial the  $k = 1$  nearest neighbor for each story and compute the fraction of trials  $\mu$  that this potential pre-text was identified as the most similar (or least distant) story. The

<sup>27</sup>. Ibid.

<sup>28</sup>. Good, *Resampling Methods. A Practical Guide to Data Analysis*.

<sup>29</sup>. Koppel and Winter, "Determining if Two Documents are by the Same Author."

result of the procedure is a ranking of potential pre-texts for each story  $i$  in which the positions represent the consistency with which pre-texts have been identified as the nearest neighbor of  $i$  over all trials. Some stories unequivocally select the same nearest neighbor over all trials ( $\mu = 1$ ), whereas for others the selection is more ambiguous, e.g. the most consistent selection does not exceed 20% of the trials ( $\mu = 0.2$ ). Depending on how conservative one wants the selection to be, a fixed cutoff value can be used that expresses the lower bounds of the fraction of trials with which a pre-text has to be selected as the nearest neighbor of a story. At  $\mu = 0$  all potential pre-texts are selected, whereas  $0.5 < \mu \leq 1$  maximally returns one nearest neighbor. In the experiments below,  $\mu$  is set to the rather conservative value of 0.5. Algorithm 2 summarizes the procedure, which I will refer to as *Bootstrap Neighbor Clustering*.

### 5.3.3 STORY NETWORKS AND STATISTICAL METHODS

Given a bootstrapped set of pre-texts  $A$  for story  $i$ , each pre-text selection is represented as a link using the following notation:  $i \rightarrow j$  where  $j$  refers to one of the pre-texts in  $A$ .<sup>30</sup> In network theory, these links are called directed edges between pairs of vertices. Consider the following set of stories, in which each number corresponds to a unique story:

$$V = \{1, 2, 3, 4, 5\}, \quad (5.2)$$

The set of pairs of stories and pre-texts is represented by:

$$E = \{3 \rightarrow 2, 4 \rightarrow 3, 5 \rightarrow 3\}. \quad (5.3)$$

Using  $V$  and  $E$  we create a graph  $G = (V, E)$  that can be visualized as in Figure 5.2. Note that in this hypothetical example, story 1 is a singleton story for which no decisive pre-text could be determined.

The number of incoming edges (i.e. the number of stories that select a particular pre-text) is called *in-degree*. I use the notation  $d_{in}(i)$  to refer to the in-degree of node  $i$ . If we apply this terminology to the network in Figure 5.2, we can observe that node 3 has an in-degree of  $d_{in}(3) = 2$ ,

30. To avoid any misunderstandings, it should be stressed that we do *not* treat stories and pre-texts as two separate classes, as any story can become a pre-text to a later story.

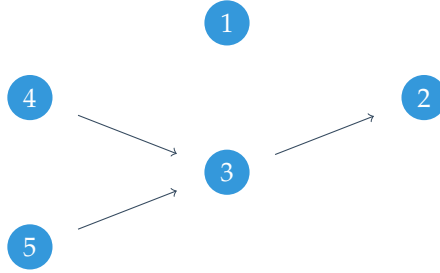


Figure 5.2: Artificial story network.

whereas node 2 has an in-degree of 1.  $Pr(d_{in})$  is used to refer to the in-degree distribution of a network. Node 1, 4 and 5 have an in-degree of 0. Node 2 has one incoming link ( $d_{in}(2) = 1$ ) and node 3 has two incoming edges.  $Pr(d_{in})$  then represents the probability that a randomly chosen node from the network has a particular in-degree (e.g.  $Pr(d_{in} = 1) = (\text{number of nodes with } d_{in} = 1) / (\text{total number of nodes}) = 0.2$ ).

A common approach to characterize degree distributions is to fit the parameters of a probability density function. Many real-world networks such as the Word Wide Web network and (scientific) citation networks display a degree distribution with a so-called ‘heavy tail’ in which a large proportion of the nodes have a small degree and only a few yet significant number of nodes are connected to many nodes.<sup>31</sup> In some networks the degree of the nodes decays exponentially, whereas in others it follows a power-law:

$$Pr(d) \approx d^{-\alpha} \quad (5.4)$$

where values of the parameter  $\alpha$  typically fall in the range  $1.5 \leq \alpha \leq 3$ .<sup>32</sup> Larger values lead to a faster decay in the probability of nodes with a high degree. To illustrate the effect of the exponent, I show in the left subplot of Figure 5.3 the complementary cumulative distribution function (ccdf) of four randomly generated degree distributions using different values of  $\alpha$ . Given that two degree distributions follow a power-law, we can compare the exponents of their distributions. However, as Clauset et al. have shown, many empirical distributions that appear to follow a power-law, are in fact better described using other heavy-tail

31. Newman, “The structure and function of complex networks.”

32. Ibid., table II.

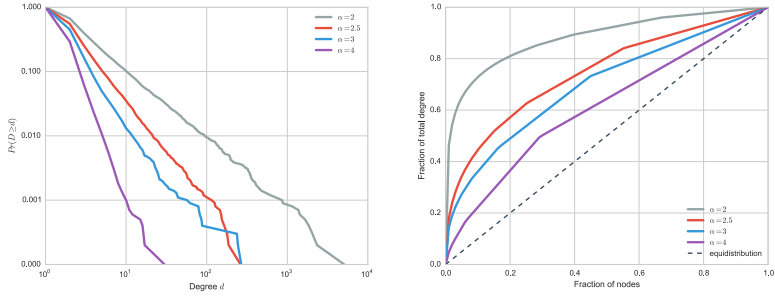


Figure 5.3: Left: Complementary cumulative distribution functions of four power-law degree distributions with exponents of  $\alpha = 2$ ,  $\alpha = 2.5$ ,  $\alpha = 3$  and  $\alpha = 4$ . Right: Lorenz curves corresponding to these power-law distributions. The dashed line represents a degree distribution that exhibits complete equality of the fraction of nodes over the fraction of degree.

distributions such as the log-normal distribution.<sup>33</sup> In this chapter, I apply the techniques developed by Clauset et al. to characterize the degree distributions of story networks.<sup>34</sup>

Heavy tails of degree distributions represent an uneven spread of edges among nodes. This inequality can be represented using a Lorenz curve which was developed by the economist Max Lorenz for representing wealth inequality in a population. The curve displays the fraction of wealth held by the richest fraction of people in a population. Applied to network degree distributions, the curves represent what fraction of edges is held by what fraction of nodes. If all nodes in a network are connected by the same number of edges, the curve forms a straight line, representing total equality. In the situation where a small fraction of nodes holds a large fraction of edges, the curve displays a steep increase, indicating that the edges are spread unevenly among the nodes. To illustrate the Lorenz curve, I show in the right subplot of Figure 5.3 the Lorenz curves of the degree distributions that correspond to the power-law exponents in the left subplot. The dashed line represents a degree distribution that exhibits complete equality of the fraction of nodes held by the fraction of edges. The degree distribution generated using  $\alpha = 2$  displays the strongest form of inequality in which only a

33. Clauset et al., “Power-law distributions in empirical data.”

34. Ibid.

small fraction of the nodes (0.2) holds a large fraction of the edges (0.8).

The degree of inequality observed in Lorenz curves can be conveniently summarized using the Gini coefficient  $G$ . This coefficient can be defined as twice the area between the equidistribution (i.e. the dashed line in Figure 5.3) and an observed Lorenz curve.  $G$  falls in the range  $0 \leq G \leq 1$  where larger values indicate a higher degree of inequality. The Lorenz curve corresponding to the degree distribution generated with  $\alpha = 4$  yields  $G = 0.22$ , whereas the curve corresponding to  $\alpha = 2$  is summarized by  $G = 0.77$ . An advantage of the Gini coefficient is that it allows us to compare networks of different average degree and size.<sup>35</sup>

## 5.4 STORY NETWORK ANALYSIS

To construct story networks for the collection of “Red Riding Hood” retellings and the collection of chain letters, I employ the *Bootstrap Neighbor Clustering* procedure. Figure 5.4 provides a graphical visualization of the two networks. The “Red Riding Hood” network consists of  $n = 427$  nodes and  $m = 439$  edges. The chain letter network consists of  $n = 554$  nodes and  $m = 620$  edges. The color gradient (from black via white to red) in the two networks represents the age of each story. Two important observations can be made from visually inspecting the two networks. First, let us consider the chain letter network. As indicated in the introduction and data section, chain letters are explicitly designed to be replicated and redistributed within a short period of time. To confirm the reliability of the proposed methodology, then, the story network extracted by the *Bootstrap Neighbor Clustering* procedure has to be in accordance with some of our preconceptions about the structure of chain letter story networks. The visualization in Figure 5.4 indicates that this appears to be the case: most stories in the network are connected to other stories with a similar color shade, which means that, in the chain letter network, stories predominantly select stories of approximately the same age as potential pre-texts. Interestingly, the second (“Red Riding Hood”) network exhibits a similar pattern, with each story showing a clear preference to select pre-texts with a similar color shade (i.e. close

35. Jennifer M. Badham, “Commentary: Measuring the shape of degree distributions,” *Network Science* 1, no. 2 (2013): 213–225.



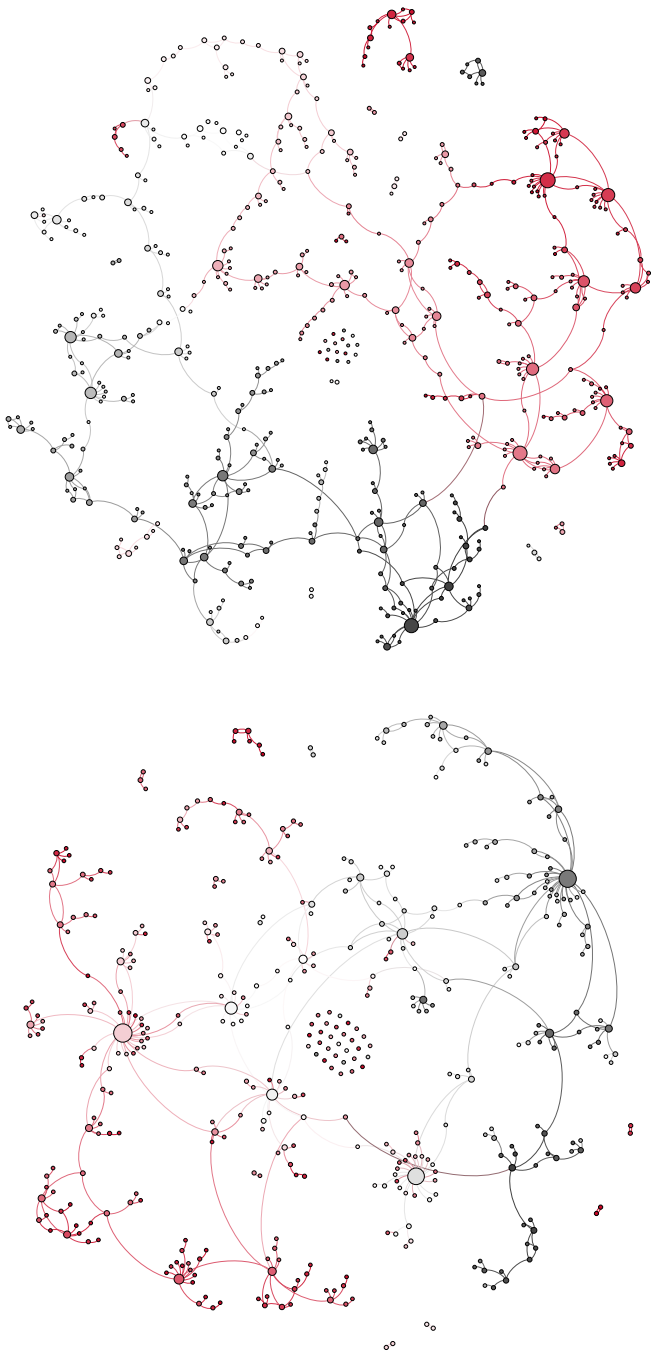


Figure 5.4: Story networks of the collection of chain letters (left network) and “Red Riding Hood” retellings (right network). Nodes represent individual stories and edges represent pre-textual relationships between stories. The color gradient (from black via white to red) represents the age of each story.

temporal proximity). Second, it can be observed that the network derived from “Red Riding Hood” retellings has a structure with a few hubs, i.e. where three or four stories function as pre-text for a large number of stories (which is represented by the size of the nodes in the networks). The chain letter network, on the other hand, displays a more uniform distribution over which stories function as pre-text, which also is in accordance with our expectations about the structure of chain letter networks (cf. Section 5.1).

In what follows, I will statistically characterize these two observations by carefully studying the in-degree distributions of the two story networks. Accurately characterizing the in-degree distribution of the two networks is a fundamental prerequisite to understand which models of network growth potentially underlie the evolution of the two story networks. A power-law characterization of the in-degree distribution, for example, might be accounted for by models of network growth such as the Preferential Attachment model (cf. Section 5.4.2). Subsequently, I will study the development of the two story networks over time. I compare four models of network growth and analyze their in-degree distributions in relation to those of the two story networks.

#### 5.4.1 IN-DEGREE DISTRIBUTION ANALYSIS

Figure 5.5 plots the cumulative complementary distribution function (ccdf) of the in-degree distributions of the story networks on doubly-logarithmic axes. The plots express the probability of stories with at least in-degree  $d_{\text{in}}$ , which is denoted as  $Pr(D \geq d_{\text{in}})$ , where  $D$  represents a random variable drawn from the distribution. The plots clearly show that the vast majority of stories have a small in-degree and that only a few stories are selected as pre-text by a large(r) number of stories.

To obtain a better understanding of the distributions, I perform the rigorous statistical procedure as described by Clauset et al. to detect power-law behavior in distributions.<sup>36</sup> In many real-world datasets, the power-law property only applies to the tail of the distribution, i.e. for values greater than some minimum value  $d_{\text{min}}$ . The method proposed by Clauset et al. estimates a minimum value  $d_{\text{min}}$  by comparing the empirical distribution to the theoretical cumulative distribution function

36. Clauset et al., “Power-law distributions in empirical data.”

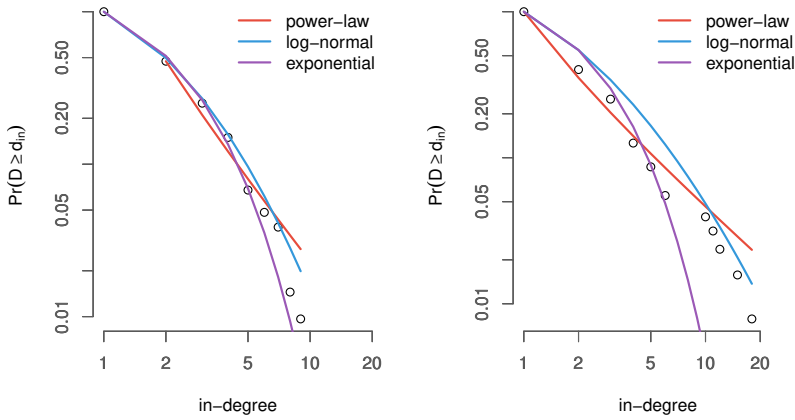


Figure 5.5: Complementary cumulative in-degree distributions of the two story networks (left: chain letters; right: “Red Riding Hood”). The plot provides for both distributions the best fit of a power-law, log-normal and exponential model. The power-law and log-normal model both fit the empirical in-degree distribution of “Red Riding Hood” well. The chain letter in-degree distribution is best described by means of an exponential model.

(cdf). The method aims to optimize the Kolmogorov-Smirnov statistic  $D$  by choosing  $d_i$  as the value for  $d_{\min}$  for which  $D$  is the smallest. Clauset et al. suggest to test whether a dataset actually follows a power law using a goodness-of-fit test and a bootstrapping procedure.<sup>37</sup> The  $p$ -value resulting from this test answers the question whether possible differences between the empirical data and the model are significant or not. If  $p \simeq 0$ , then the model cannot be deemed a plausible fit and other distributions are more appropriate.

At a significance level of 95%, we cannot reject the hypothesis that the in-degree distribution of the “Red Riding Hood” network is generated by a power-law model:  $D = 0.048, p > 0.07$ . By contrast, the power-law model is not a good fit for the in-degree distribution of the chain letter network ( $D = 0.09, p = 0.001$ ). Another way to test for power-law behavior is to directly compare a power-law model to another

37. Clauset et al., “Power-law distributions in empirical data.”

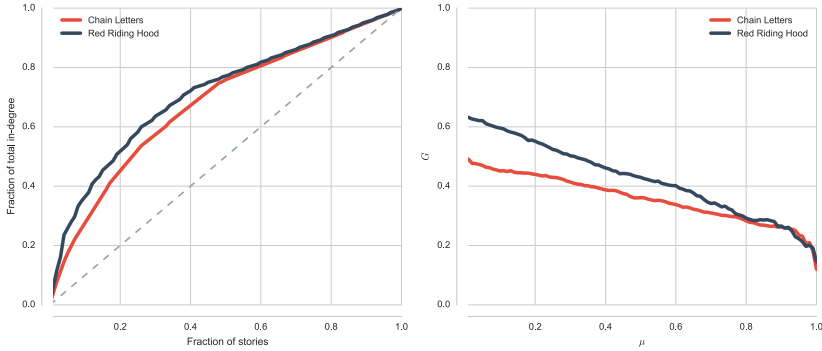


Figure 5.6: The left subplot displays the Lorenz curve of the chain letter and *Red Riding Hood* story network. The gray striped line represents the equidistribution. The right subplot shows the Gini coefficient  $G$  of the in-degree distributions for story networks constructed using  $\mu$  in the range  $0.1 \leq \mu \leq 1$ . It can be observed that the value of  $G$  remains relatively stable as we increase the value of  $\mu$ .

model, such as a log-normal model via a *likelihood ratio test*  $R$ .<sup>38</sup> Using the likelihood ratio test, we can assess whether the in-degree distributions are more appropriately described by means of a log-normal or exponential model. The log-normal model fits the chain letter distribution significantly better than the power-law model for the full range of in-degree values ( $TS = -4.05, p < 0.0001$ ). The next step is to compare the log-normal model to an exponential model. The test results are indecisive as to whether a log-normal model is more appropriate than an exponential model ( $R = 0.079, p > 0.9$ ). Since a log-normal function has more parameters than an exponential function, it seems reasonable – for reasons of parsimony – to characterize the chain-letter in-degree distribution as exponential. Although a log-normal model appears to fit the in-degree distribution of “Red Riding Hood” slightly better than a power-law function, the test results suggest that the two models do not perform *significantly* different ( $R = -1.52, p > 0.1$ ).

In the left subplot of Figure 5.6, I present the Lorenz curves of the two in-degree distributions. The summarizing Gini coefficients of these curves are  $G = 0.36$  for the chain letter network and  $G = 0.43$  for “Red Riding Hood”. Compared to their random counterparts, the story

38. Ibid.

networks exhibit a greater degree of inequality with respect to their in-degree distributions (chain letters:  $G_{\text{random}} = 0.3$ ; “Red Riding Hood”:  $G_{\text{random}} = 0.28$ ).<sup>39</sup> The degree of inequality is especially large in the case of “Red Riding Hood”, suggesting that relatively few stories function as pre-text for many other stories. In the right subplot of Figure 5.6, I visualize the Gini coefficients  $G$  for story networks that were constructed with  $0.1 \leq \mu \leq 1$ . It can be observed that the in-degree distributions display a relatively uneven spread of edges among nodes (i.e. stories) even for high values of  $\mu$ . This indicates that the ‘hubness’ of the networks is a stable characteristic and not too much the result of cherry picking a particular value of  $\mu$ .

#### 5.4.2 STORY NETWORK EVOLUTION

In the previous section, I have shown that the two studied story networks display distinct in-degree distributions. The chain letter network exhibits an in-degree distribution which decays exponentially. The “Red Riding Hood” network, by contrast, exhibits a heavy-tail in-degree distribution that fits a power-law or log-normal model reasonably well. Retellings of “Red Riding Hood” preferentially link to a small number of stories that are pre-texts of many other retellings. In the present section, I turn to the central question of how these distributions come into being. By carefully studying and comparing the structure and evolution of the two story networks to formal models of network growth, I provide empirical evidence for three major conclusions regarding the formation of story networks. First, I provide empirical evidence that stories preferentially select stories in close temporal proximity, which is indicated by a strong lopsidedness towards smaller time-spans in the time-span distributions of stories and their selected pre-texts. Second, I show that the in-degree distribution of the “Red Riding Hood” network is significantly correlated with the age of stories, suggesting that retellings of “Red Riding Hood” are affected by a mechanism of preferential attachment in which slightly older versions are preferred to be selected as pre-text(s) in producing a new version. Finally, I show that stories have individual attractiveness

39. These random graphs are created by randomly rewiring the edges of the story networks while preserving the time constraint, i.e. stories may only link to randomly chosen pre-texts of the same or older age.

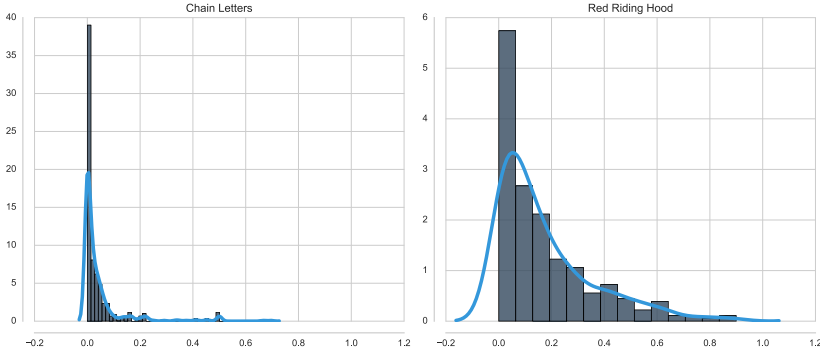


Figure 5.7: Kernel density plots of time-span distributions. The plots display the distributions over time-spans for the chain letter network and the network induced for Red Riding Hood. As can be observed, both time-span distributions display a strong lopsidedness towards smaller time-spans, which indicates a preference for pre-texts in close temporal proximity.

values that lessen over time.

The Preferential Attachment (PA) model has proven to be a reliable model to account for heavy-tail degree distributions observed in real-world networks. The model was invented by Derek de Solla Price in the context of citation networks<sup>40</sup> and simplified and generalized by Albert-László Barabási and Réka Albert to account for undirected networks.<sup>41</sup> The algorithm generates these networks using an attachment mechanism in which new nodes preferentially link to existing nodes with high (in-)degree. A social network is a classic example in which the mechanism of preferential attachment is at play. In a social network, an edge between two persons  $a$  and  $b$  exists if  $a$  knows  $b$  or vice versa. Some persons have many connections and a newcomer to the network is more likely linked to these well-known persons than to persons who are relatively unknown. Similarly, web pages preferentially link to well-known sites, such as Wikipedia, rather than to sites that are less familiar. The mechanism of preferential attachment predicts that the probability of creating a new link between entity  $a$  and  $b$  is proportional to the number of existing connections of  $b$ .

40. Price, "A general theory of bibliometric and other cumulative advantage processes."

41. Barabási and Albert, "Emergence of Scaling in Random Networks."

The consequence of the preferential attachment mechanisms is that nodes that enter the network first will attract more links early on, and will continue to do so. As an explanatory model for story networks, the PA model predicts that stories preferentially select old(er) stories as pre-text rather than new(er) stories. To test this hypothesis, I evaluate for each pair of story and pre-text the time elapsed between their publication dates. Figure 5.7 plots the kernel density plots of the time-span distribution for the chain letter network and the “Red Riding Hood” network. The time-spans have been normalized to  $0 \leq \tau \leq 1$ , where  $\tau$  represents the normalized time-span between a story and a pre-text. Both plots display strong lopsidedness towards smaller time-spans. Chain letters display an even stronger preference for pre-texts in close temporal proximity (median  $\hat{\tau} = 0.01$ ) than the retellings of “Red Riding Hood” (median  $\hat{\tau} = 0.11$ ). Both time-span distributions run counter the prediction of the PA model that stories are predominantly connected to old(er) stories.

To obtain more insight into the relation between the in-degree and time-span distributions of the story networks, I test in Figure 5.8 for the correlation between in-degree and age. As can be observed from the left subplot and confirmed by a Pearson  $\rho$  correlation test, there is no evidence that the in-degree distribution of the chain letter network depends on the age of stories ( $\rho = -0.02, p > 0.6$ ). However, the in-degree distribution of “Red Riding Hood” retellings (right subplot in figure 5.8) is significantly correlated with age ( $\rho = 0.27, p < 0.0001$ ). In other words, retellings of “Red Riding Hood” do display a preference to select older versions as their pre-text(s). Note, however, that although the correlation is significant, its coefficient is rather low and the relatively low slope of the linear fit suggests that age affects in-degree only in a limited, upper-bounded range.

To account for the interaction between in-degree and age, I propose a growing network model, which is based on Price’s model of preferential attachment.<sup>42</sup> In addition to the frequency-based attraction resulting from the preferential attachment mechanism, stories in this new model have individual values of attraction which lessen over time.<sup>43</sup> Similar to

42. Price, “A general theory of bibliometric and other cumulative advantage processes.”

43. Cf. Dorogovtsev and Mendes, “Evolution of reference networks with aging”; Eom and Fortunato, “Characterizing and Modeling Citation Dynamics”; Sophia R. Goldberg et al.,

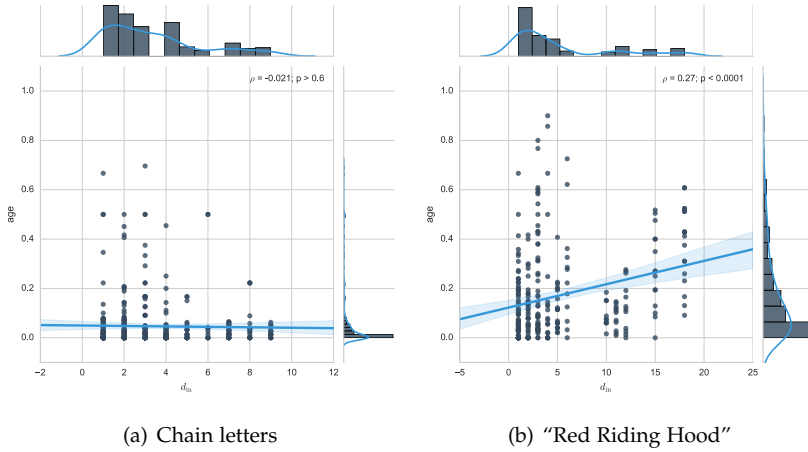


Figure 5.8: Correlation between in-degree and age. The two plots display the correlation between the age of a story and its in-degree for both the chain letter network and the “Red Riding Hood” network. The chain letter in-degree distribution does significantly correlate with the age of the letters ( $\rho = -0.021, p > 0.6$ ). The in-degree distribution of the network of “Red Riding Hood” is significantly correlated with age ( $\rho = 0.27, p < 0.0001$ ).

the preferential attachment algorithm, the model begins with initializing an unconnected network consisting of  $m_0$  nodes. At each succeeding time step a single node is added to the network. With a probability  $(1 - p)$ , a new node connects to  $m \leq m_0$  existing nodes with uniform probability. With a probability  $p$ , the node is connected to  $m \leq m_0$  nodes according to the preferential attachment mechanism. The probability to connect to an existing node is given by:

$$p(i) = \frac{\alpha_i \beta_i}{\sum_{j=1}^n \alpha_j \beta_j}, \quad (5.5)$$

where  $\alpha_i = d_{\text{in}}(i)$  and  $\beta_i$  represents the attractiveness of a node at a particular time step. If we set  $\beta_i$  to represent a constant value, the model reduces to the original preferential attachment algorithm. Following previous studies,<sup>44</sup> I propose to lessen the attractiveness of a node

“Modelling citation networks,” *Scientometrics* 105, no. 3 (2015): 1577–1604.

44. E.g. Eom and Fortunato, “Characterizing and Modeling Citation Dynamics”; Goldberg et al., “Modelling citation networks.”



exponentially over time:<sup>45</sup>

$$\beta_i = \beta_0^i e^{-\tau/\gamma}, \quad (5.6)$$

where  $\beta_0^i$  represents a node's initial attractiveness and  $\tau$  represents the age of a node. The parameter  $\gamma$  acts as an "attention span" parameter which controls the slope of the exponential decay.<sup>46</sup> If  $\alpha_i$  is held constant, the model's growth mechanism is restricted to the temporal attractiveness of nodes. Each node  $i$  is assigned an individual initial attractiveness value  $\beta_0^i$ , which is sampled from a symmetric Dirichlet distribution with hyper-parameter  $\phi$ .<sup>47</sup> Values of  $\phi$  between zero and one generally result in a more 'peaky' attractiveness distribution, in which only a few nodes are highly attractive. Higher values ( $\phi \geq 1$ ) result in a more uniform attractiveness distribution. In all experiments, I fix  $\phi$  to 0.1. By holding the variable  $\alpha_i$  and  $\beta_i$  constant, we can analyze the various types of attractiveness in isolation and their interplay. More specifically, I investigate the following four models of network growth:

1. Preferential Attachment (PA) Model, where  $\beta_i = 1$ ;
2. Temporal Preferential Attachment (T-PA) Model, where  $\beta_0^i = 1$ ;
3. Temporal Attractiveness (TA) Model, where  $\alpha_i = 1$ ;
4. Preferential Attachment Temporal Attractiveness (PA-TA) Model.

Figure 5.9 presents the complementary cumulative in-degree distributions of both story networks as well as those of the four proposed models. I will compare the in-degree distributions on the basis of their corresponding Gini coefficients  $G$ , which are presented in Table 5.1. The four growing network models are probabilistic and therefore results vary from simulation to simulation. The reported Gini coefficients are obtained by averaging 50 simulations.

The PA model generates heavy-tailed in-degree distributions for "Red Riding Hood", of which the summarizing  $G$  values are comparable to

45. Eom & Fortunato add a node's in-degree to its attractiveness at a particular time step. In this study, I choose to weigh a node's in-degree by its attractiveness by *multiplying* the two values. Cf. Eom and Fortunato, "Characterizing and Modeling Citation Dynamics."

46. Goldberg et al., "Modelling citation networks."

47. The choice for sampling from a Dirichlet distribution is largely motivated by the fact that Dirichlet processes are often employed to model data that, like the presented story networks, tend to develop in a 'rich get richer' fashion.

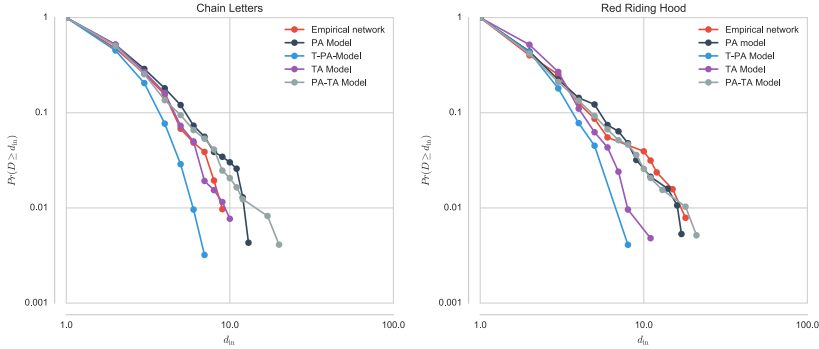


Figure 5.9: Complementary cumulative in-degree distributions of the empirically derived network and simulations of the PA, T-PA, TA, and PA-TA modes. The left subplot shows the in-degree distribution of the chain letter network. The corresponding distributions of the four growing network models were obtained from a single simulation. The right subplot provides the same information for “Red Riding Hood”.

the empirical coefficient. The Gini coefficient of the chain letter network is much smaller than those produced by the PA model. As was expected, the time-span distributions of the PA model are negatively skewed (chain letters: median  $\hat{\tau} = 0.7$ ; “Red Riding Hood”: median  $\hat{\tau} = 0.5$ ) and the oldest nodes (i.e. the nodes that first enter the network) receive most of the incoming links over time. The three growing network models that take the age of nodes into account produce positively skewed time-span distributions in which younger nodes are preferred over older ones (chain letters: T-PA = 0.01, TA = 0.03, PA-TA = 0.04; “Red Riding Hood”: T-PA = 0.06, TA = 0.07, PA-TA = 0.08). The Gini coefficients of the T-PA model are, however, rather small compared to the empirical coefficients. The temporal attractiveness (TA) model generates in-degree distributions that display the most similar degrees of inequality to those of the chain letter network. In the case of “Red Riding Hood”, the best results are obtained by the PA-TA model.

## 5.5 DISCUSSION

In this chapter, I have studied the structure and evolution of story networks. Story networks, defined as non-hierarchical agglomerations

	Gini Coefficient				
	Empirical	PA	T-PA	TA	PA-TA
Chain letters	0.36	0.41	0.29	<b>0.36</b>	0.41
Red Riding Hood	0.43	0.41	0.29	0.37	<b>0.43</b>

Table 5.1: Gini Coefficients of the two story networks (i.e. chain letters and “Red Riding Hood”) and the four growing network models (i.e. PA, T-PA, TA, and PA-TA). The reported Gini coefficients are obtained by averaging 50 simulations for each of the four models of network growth. The empirical chain letter network displays an in-degree distribution which is most similar to those generated by the temporal attractiveness (TA) model. The in-degree distribution of the empirical “Red Riding Hood” network is most similar to the distributions generated by the Preferential Attachment Temporal Attractiveness (PA-TA) model.

of pre-textual relationships, represent streams of retellings in which retellers modify and adapt retellings in a gradual and accumulative way. The first challenge was to develop methods that allow us to automatically extract such story networks from raw text collections. To this end, I have proposed a clustering procedure, termed *Bootstrap Neighbor Clustering*, which approaches the problem of pre-text selection as an open-set problem and attempts to bootstrap pre-textual story networks. I have constructed such story networks for two diachronic collections of retellings: a collection of paper chain letters and a collection of Dutch “Red Riding Hood” retellings.

Using these extracted networks as a base, it was possible to provide a mechanistic understanding of story network growth and, by extension, of retellers’ motivations for choosing particular story versions to base their retelling on. I hypothesized that stories are differentially preferred to function as a retelling’s pre-text given three types of attractiveness: frequency-based, temporal, and model-based attractiveness. To gain more insight into the relations between stories and their pre-texts, I assessed the patterns of connectivity of the two story networks by performing a rigorous statistical analysis of their in-degree distributions. The in-degree distribution of “Red Riding Hood” displays heavy-tail properties that are well characterized by means of a power-law or log-normal model. Such heavy-tailness implies that a large proportion of stories have small in-degree values and only a few, yet significant number

of stories function as pre-textual context for a large number of other stories. The heavy-tail distribution of “Red Riding Hood” contrasts with the relatively uniform in-degree distribution of the chain letter network, which was characterized as being exponential.

In addition, I have demonstrated that retellings of chain letters and “Red Riding Hood” are published in relatively close temporal proximity. The effect of temporal attractiveness is most strongly observed in the chain letter corpus. Its story network displays a chain-like structure that is reminiscent of the letter’s request to be redistributed within a short period of time. It was shown that the time-span distribution of the chain letters does not display a positive correlation with its corresponding in-degree distribution. Retellings of “Red Riding Hood”, on the other hand, do exhibit a significant correlation between in-degree and age. These contrasting results can be linked to an important difference between the chain letter and “Red Riding Hood” retellings: Whereas retellers of “Red Riding Hood” can choose from a vast amount of story versions, chain letter retellers are requested to redistribute and retell one specific version. Moreover, it was shown that, in the retelling of chain letters, the mechanism of preferential attachment has no effect.

To explore which mechanisms potentially underlie the evolution of the two story networks, I have investigated a model of network growth. In addition to a preferential attachment mechanism, this model implements a form of model-based attractiveness which decays exponentially in time. The model that incorporates both preferential attachment and temporal attractiveness (PA-TA) best simulates the in-degree distribution of “Red Riding Hood”. The more parsimonious temporal attractiveness (TA) model sufficed to account for the observed in-degree distribution of the chain letter network. This result concurs with the finding that the in-degree distribution of the chain letter network does not depend on age. Both models of network growth generate positively skewed time-span distributions that are comparable to the observed story network distributions.

This being said, I wish to stress that there is not one true story network. In this chapter, I made a rather conservative and simplifying choice by investigating pre-textual relationships between stories on the basis of their vocabulary. However, the number of dimensions on which stories can be considered (dis)similar is virtually endless. In order

to avoid falling into an 'essentialist' trap, future research should first be directed toward studying different dimensions of story similarity (e.g. topics, motifs, genre) and their effect on the structure of story networks.<sup>48</sup> A second point of future research is to further explore macroscopic properties of story networks. While the present study mainly focused on the in-degree distributions of story networks, one needs to investigate other general principles governing their structure and evolution in order to obtain a more profound understanding of story networks. Many complex real-world networks can be characterized as so-called 'small-world networks', which exhibit two fundamental properties: (i) locally connected groups of nodes and (ii) a short average shortest path length between nodes.<sup>49</sup> I take it to be an interesting question whether story networks display this small-world property. In this scenario, stories would be connected to only a few other stories, while at the same time all stories in the network would be connected to each other through only a few intermediate steps. Another principle governing many real-world networks is a modular structure. Networks with modular structure are hierarchically organized into local groups of densely connected nodes, with a low density of connections *between* groups.<sup>50</sup> A modularity analysis of story networks could reveal that stories with high in-degrees function as bridges between local 'story communities' and integrate them into a single network.<sup>51</sup>

48. Cf. Abello et al., "Computational folkloristics."

49. Duncan J. Watts and Steven H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature* 393, no. 6684 (1998): 440–442; Newman, "The structure and function of complex networks."

50. Newman, "The structure and function of complex networks."

51. Pádraig Mac Carron and Ralph Kenna, "Universal properties of mythological networks," *EPL (Europhysics Letters)* 99, no. 2 (2012): 28002–28007.

---

## CHAPTER 6

# GENERAL DISCUSSION

*"[Darwin] is no longer the authoritative old man with a beard substituting for God."*

GILLIAN BEER, *Darwin's Plots*

In what preceded, I have offered new perspectives on the mechanisms underlying story transmission and selection. In essence, the approach presented here builds on the insights gained from both folkloristic and literary accounts of story transmission.<sup>1</sup> However, while such accounts have undoubtedly yielded a wealth of insightful ideas about the mechanisms at play in story transmission, their arguments and claims often remain programmatic and are based on informal verbal arguments that do not allow for rigorous quantitative evaluations. Recent developments in the cultural evolution research program have shown the benefits of employing *formal* models to further the understanding of cultural selection processes. By positioning itself explicitly and extensively in dialog with the cultural evolution research program,<sup>2</sup> the present study aimed to advance our understanding of story transmission processes by means of such formal models of cultural evolution. The added value of the computational approach to story transmission presented in this study is

1. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*; Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction*; Brian Boyd et al., eds., *Evolution, Literature, and Film. A Reader* (New York: Columbia University Press, 2010); Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*; Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*; Geerts and van den Bossche, *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*.
2. Cavalli-Sforza and Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*; Boyd and Richerson, *Culture and the Evolutionary Process*; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*; Alex Mesoudi, "Cultural Evolution: A Review of Theory, Findings and Controversies," *Evolutionary Biology*, 2015, 1–17, doi:10.1007/s11692-015-9320-0.

that it yields specific, replicable predictions that can be quantitatively and rigorously evaluated against real-world data. At the same time, it also enables us to isolate and systematically compare forces of selection at play in story transmission. In this concluding chapter, I will synthesize the findings presented in Chapters 2 to 5 from (i) a methodological perspective that assesses various ways to formally represent stories in order to study real-world story transmission (Research question 1), and (ii) a theorizing perspective that evaluates what the preceding analyses tell us about the factors that determine story transmission and selection (Research question 2–4).

## 6.1 METHODOLOGICAL CHALLENGES

While the cultural evolution research program has investigated the mechanisms of story transmission in laboratory contexts,<sup>3</sup> few attempts have been made to apply this framework and encompassing research methods to real-world, historical story data.<sup>4</sup> Therefore, the present study needed to resolve a number of methodological challenges prior to addressing its main research questions. Perhaps the most important methodological issue, which is addressed throughout this entire study, is how stories should be represented in order to computationally study real-world story transmission and selection (Research question 1). As there is no simple ‘one solution fixes all’ answer to this question, I explored a number of different data representations, each of which has potential advantages and disadvantages for studying certain aspects of story transmission.

In traditional folktale research, the methods used to describe and study relations between stories revolves primarily around the motifs from Thompson’s *Motif-Index*.<sup>5</sup> In such studies, motifs are used as the primary descriptive units of stories, and their constellation defines their corresponding tale type as described in, for example, Uther’s tale type

3. Mesoudi and Whiten, “The Multiple Roles of Cultural Transmission experiments in Understanding Human Cultural Evolution.”

4. Notable exceptions are: Tehrani, “The Phylogeny of Little Red Riding Hood”; Da Silva and Tehrani, “Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales.”

5. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

catalog.<sup>6</sup> The problem with these traditional approaches is that, in practice, the motifs are not used to identify tale types. Rather, the typical classification strategy in such analyses is an *a priori* assigning of the story under investigation to a single tale type, before any of its motifs are properly identified. As a consequence, the object of investigation is considered solely in light of the motifs that are associated with that particular type. Thus, the classification strategy of predefined tale types and encompassing motifs *encourages* to foreclose particular connections between stories. Neither the concept of a tale type, nor the motifs that supposedly constitute them, is without risk.

Another problem with tale types and predefined motifs is that they inevitably decontextualize stories. As Marina Warner aptly puts it: the folktale type catalog “provides a list of ingredients [i.e. motifs] and recipes [i.e. tale types] with no evocation of their taste or the pleasure of the final dish, nor sense of how or why it was eaten”.<sup>7</sup> In other words, stories are torn from their original contexts and might be generalized to a level too abstract to retain any significance. If we, for instance, would analyze the more than four hundred versions of “Red Riding Hood” from Chapter 4 in terms of the presence or absence of the motifs listed under tale type ATU 333 “Little Red Riding Hood”, most versions would become indistinguishable. As such, the motif classification leaves us without means to explain the observed variation and progression through time.

One of the central claims in this study is that tale types should serve the interpretation of actual stories; not the other way around. Tale types can in fact be dangerous when the classification of stories into tale types becomes a static, authoritative, ‘this or that’ enterprise that mutes the many consonant and dissonant resonances with other stories.<sup>8</sup> The potential danger of foregrounding types instead of stories is amplified by Arthur Frank in the context of stories describing experiences of illness:

“Typologies risk putting stories in boxes, thus allowing and even encouraging the monological stance that the boxes are more real than the stories, and the types are all that need to be known about

6. Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*.

7. Warner, *From the Beast to the Blonde. On Fairy Tales and their Tellers*, XVIII.

8. Cf. Frank, *Letting Stories Breathe. A Socio-Narratology*.



the stories. In a world where simplification is a pretext for knowing, and knowing is a pretext for controlling, typologies are risky.”<sup>9</sup>

In principle, there is nothing wrong with using tale types as analytical tools, as long as we recognize them to be only one out of many possible perspectives. Typologies invite us to make connections among similar stories, which helps researchers to get a grip on stories and enhance their interpretation. However, when the identification of types becomes a goal in itself, we run the risk of remaining blind to the variation exposed within a category, and of foreclosing both existing and potential relations between stories belonging to separate categories. This risk is anything but trivial if we want to explain how stories are created, interpreted, adapted, and retold. The key or Bakhtinian “dialogical trick”, as Frank suggests, is to sustain *openness* and “[t]ypologies should never be considered final”<sup>10</sup> nor should the understanding of which stories fit what tale type.

In contrast with arborescent classification systems which consider tale types to be primary, I advocated an exemplar-based approach of stories. Exemplar-based approaches differ from typology-based approaches in that the object of interest is not the overarching tale type, but lies primarily with the tale ‘token’, i.e. the actual story.<sup>11</sup> Each story is considered to be a unique entity that is related to other story exemplars in various – not necessarily predetermined – ways. In Chapter 5, I conceptualize the relationships between story exemplars as a network that consists of more and less densely connected clusters of stories. As such, tale types can be considered to be emerging implicitly from highly dense clusters of similar stories. What is considered to be a similarity between stories in the present approach is never static, but depends on the perspective one wishes to take or the image one aims to construct. In Chapter 5, for example, similarities between stories were determined on the basis of their vocabulary (using a bag-of-words representation), whereas in the study about “Red Riding Hood” in Chapter 4, more abstract perspectives such as ‘time’, ‘plot’ and ‘irony’ contributed to the conceived similarities. Chapter 3 presented yet another perspective. To study the factors determining the successfulness of fairy tales from

9. Frank, *Letting Stories Breathe. A Socio-Narratology*, 118–119.

10. *Ibid.*, 119, 121.

11. *Ibid.*, 119.

the Brothers Grimm's *Kinder- und Hausmärchen*, I focused on the stories' character cast, which were represented as semantic vectors (i.e. word embeddings). In studying story transmission and selection from an evolutionary perspective, it is of utmost importance to acknowledge that relations between stories are malleable and can only be accounted for if we adopt data-driven representations of stories that do not superimpose predefined categories onto stories.

Each of the data-driven representations explored in the present study have certain advantages and disadvantages. First, the character representations employed in Chapter 3 have the advantage of conveying semantic information. In addition, these representations require minimal manual labor, which greatly facilitates analyses of large-scale data collections. However, these representations only represent stories on a single dimension (i.e. the character cast), and ignore the fact that there are many other dimensions that are (potentially) of equal importance in establishing connections between stories. The bag-of-words representations used in Chapter 5, by contrast, represent stories on *multiple* dimensions. Bag-of-words representations are powerful, widely used representations, which are computationally efficient and – even more so than the character representations in Chapter 3 – require minimal manual labor. Despite the fact that bag-of-words make the rather crude simplification of ignoring many higher-order aspects of texts (e.g. word order, sentences, structure), they provide effective means to expose important relationships between stories.

The bag-of-words representations of Chapter 5 contrast sharply with the detailed and fine-grained representations employed in Chapter 4. The representations in Chapter 4 depend on a rigorous and extensive narratological text analysis consisting of over three hundred questions, which served as the basis for studying the evolution of “Red Riding Hood” in a large corpus of Dutch retellings. This questionnaire consists of various questions concerning high- and low-level aspects of the story, ranging from the way characters are named to aspects of theory of mind. While the bag-of-words representations of Chapter 5 also represent stories in a high-dimensional space, the questionnaire approach has the advantage of enabling researchers to represent stories on multiple dimensions with varying degrees of abstraction and detail. The obvious drawback of these representations is, however, that they require labor-

intensive and subjective manual analysis (both in the choice for particular questions as well as in the answers given to these questions).

Note that many of the questions employed in Chapter 4 resemble motifs from Thompson's *Motif-Index of Folk Literature*,<sup>12</sup> which formed the central representation form of stories in Chapter 2. Questions such as "Is the wolf's belly filled with stones or some other material?" or "Is Red Riding Hood eaten by the wolf?" can be linked to, respectively, motifs Q426 (*Wolf cut open and filled with stones as punishment.*) and K2011 (*Wolf poses as "grandmother" and kills child.*). Yet, the fact that there are some obvious similarities between some of Thompson's motifs listed under tale type ATU 0333 and some parts of the questionnaire does not imply that the list of motifs and the questionnaire should be considered to be on a par. The questionnaire is to be considered as a data-driven superset of tale type ATU 333 and its constituting motifs. That is to say, the questionnaire entails the given motifs and, at the same time, adds numerous other dimensions of variation on which the stories can be compared. Importantly, these dimensions of comparison (i) arise from the collection of stories under investigation rather than being imposed onto the data and (ii) are never finalized but remain open to reconfiguration and addition when new versions of the story are added to the collection.

## 6.2 STORY TRANSMISSION AND THEORY

Using these data-driven representations, I addressed a number of questions concerning the mechanisms underlying story transmission and selection in Chapters 3 to 5. Chapter 3 presented original empirical work that contributes to answering Sperber's fundamental question of how to explain 'contagious' culture,<sup>13</sup> or, more specifically, which stories successfully 'stick' and what content-based factors determine their successfulness. In particular, I tested the hypothesis of whether a 'character bias' (i.e. a disposition for particular character types) is at play in the cultural selection of fairy tales (Research question 2). Taking

12. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

13. Dan Sperber, *Explaining Culture. A Naturalistic Approach* (Oxford: Blackwell Publishing, 1996).

the famous fairy tale collection *Kinder- und Hausmärchen* (1857) by the Brothers Grimm as a case study, I provided empirical evidence for the existence of several character types that affect the successfulness of a story. It was shown that successful fairy tales exhibit a significant preference for (i) characters with names that refer to family relationships, (ii) animal characters and (iii) minimally counterintuitive agents.

It is interesting to regard these findings in light of previous proposals with respect to the role of characters in story transmission. Recall the epigraph of Chapter 2, which recites Thompson stating that a motif is “the smallest element in a tale having a power to persist in tradition”.<sup>14</sup> Thompson’s motifs generally fall into three categories, the first being the tale’s characters. These characters, in order to have the power of persistence Thompson attributes to motifs, must exhibit something “unusual and striking”.<sup>15</sup> Interestingly, these properties typically apply to marvelous creatures such as witches, ogres and fairies, or, in other words, the so-called minimally counterintuitive agents (cf. Chapter 3). The present study provides empirical evidence for Thompson’s hypothesis that such marvelous characters have a power to persist in tradition. At the same time, however, the present study updates Thompson’s proposal by showing that minimally counterintuitive agents are not only persisting motifs, but also serve as popularizing *attractors* in the cultural selection of fairy tales. Moreover, the empirical findings of the current study add further support to hypothesized content-based biases in story transmission (e.g. a disposition for social information<sup>16</sup>). Although these biases have to some extent been investigated through experimental transmission chain studies in the lab,<sup>17</sup> the current study presents a unique account in support of these hypotheses on the basis of historical, real-world data.

Importantly, while current experimental research and research in

14. Thompson, *The Folktale*, 415.

15. Ibid.

16. Reysen et al., “The effects of collaboration on recall of social information”; Stubbersfield et al., “Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends.”

17. See, for instance, Mesoudi et al., “A bias for social information in human cultural transmission”; Barrett and Nyhof, “Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials”; Upal et al., “Contextualizing Counterintuitiveness: How Context Affects Comprehension and Memorability of Counterintuitive Concepts”; Harmon-Vukić and Slone, “The Effect of Integration on Recall of Counterintuitive Stories”; Barrett et al., “Counterintuitiveness in Folktales: Finding the Cognitive Optimum.”

literary and folkloristic studies primarily focuses on detecting positive biases in story transmission (i.e. traits that accrue a story's chances of success), the concept of 'negative biases' has been largely overlooked. To address this hiatus in our understanding of story transmission, the current study homed in on the so-called 'impopularizing' character types of fairy tales. The results suggested that unsuccessful fairy tales from *Kinder- und Hausmärchen* typically revolve around (i) religious characters, (ii) criminals and hooligans and (iii) generic groups with a rather negative connotation. This negative bias away from stories with generic, indefinite groups ties in with the study on the evolution of "Red Riding Hood" (Chapter 4), in which the wolf has become less generic and more individualized over time. Thus, it would be interesting to further pursue the question of whether these findings are part of a more general development in which individualized stories exhibit a transmission advantage in future research.

The progressive individualization of the wolf is only one of the many transformations that "Red Riding Hood" has undergone in the past centuries. While originally intended as a parable intended to warn young ladies of the French court about debonair and sweet-talking rapacious wolves, "Red Riding Hood" has become an increasingly autonomous story and has been subjected to experimentation and reconfiguration. In Chapter 4, I have systematically and quantitatively investigated this development on the basis of a large longitudinal collection of Dutch retellings of "Red Riding Hood". I demonstrated that the development of the story can be characterized as a gradual accumulation of modification: new versions of a story tend to modify and adapt prior retellings, and these prior retellings are published in close temporal proximity to these new versions (Research question 3). The resulting diachronic picture resembles a chain of retellings, in which retellers introduce adaptations and innovations. If these adaptations and innovations are further retold, they may come to gradually replace existing story elements. As such, I have argued, the progressive alteration of "Red Riding Hood" in the Netherlands can be interpreted as a cultural evolutionary process.

Finally, the observed preference of authors to produce new versions of "Red Riding Hood" on the basis of temporally proximate versions can be interpreted as evidence for a second bias in story transmission (i.e. 'age bias'). Yet, although this age-dependent selection mechanism

for “Red Riding Hood” adds important insights to our knowledge about story transmission, it cannot explain why particular story versions of approximately the same age are differentially preferred to function as pre-text for new retellings. To address these issues, Chapter 5 further scrutinized this age bias while testing other biases that might function as explanatory models in story transmission (Research question 4). By extensively and systematically comparing outcomes of computational simulations with real-world observations of story transmission, it was shown that story transmission is affected by a positive frequency bias as well as a model-based bias which reduces in strength over time.

The simulations of network growth employed in Chapter 5 represent simplified mathematical models of story transmission. A major theoretical benefit of employing such simulation models is that they call for detailed and replicable definitions. As such, simulation models force researchers to make their theoretical assumptions explicit. Moreover, simulation models also exhibit an interesting degree of simplicity. In the social sciences, the use of such simplified simulation models has often been fiercely criticized for their inability to capture the complex nature of cultural phenomena.<sup>18</sup> Yet, I wish to underscore that – as already pointed out by Mesoudi – “the fact that simulations are highly simplified is the very reason for their usefulness”.<sup>19</sup> By deliberately keeping the models simple, researchers can easily isolate and manipulate single variables under highly controlled conditions.<sup>20</sup> In recent years, the use of evolutionary models has had revolutionary effects across the social sciences,<sup>21</sup> and I believe that the application of such models could also lead to a wealth of new insights in folkloristic and literary studies.

### 6.3 AN OUTLOOK TO FUTURE RESEARCH

Yet, despite the attested descriptive successes we should refrain from rejoicing too much in technological positivism. Comprehensive, in-depth

18. Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

19. Alex Mesoudi, “The Transmission and Evolution of Human Culture” (2005), 326.

20. Mesoudi, “The Transmission and Evolution of Human Culture”; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*.

21. Mesoudi, “Cultural Evolution: A Review of Theory, Findings and Controversies.”

study of the phenomena observed in real-world story transmission processes still faces a number of challenges. Perhaps the biggest challenge is to develop methods to address the relevant questions considering that there is a lack of data. Longitudinal data collections such as the “Red Riding Hood” corpus presented in Chapter 4 are virtually non-existent. In an ideal-case scenario, we would be in possession of data collections that specify the exact paths of transmission between individuals and generations, i.e. who learns or copies from whom.<sup>22</sup> The rare existing data collections, however, merely represent sparse, incomplete samples of the actual data and none of them provide explicit information about the transmission paths. An additional complication to the matter is that the currently available data only allow us to study the evolution of a small number of individual stories. Making more data available to obtain a more profound understanding of story selection dynamics that transcend those of the individual story would, however, be time-consuming and laborious, and the effort becomes even more impossible to accomplish if we want to broaden the perspective to the comparison of story transmission processes between different cultures.

Given this lack of diachronic data, I consider it to be of crucial importance that we continue to explore different and new methodologies to analyze real-world story transmission. As pointed out by Kandler and Powell, fine-grained individual-level data of cultural change are difficult to obtain, yet many archaeological and anthropological data collections do describe aggregate, population-level outcomes of evolutionary processes.<sup>23</sup> Such outcomes take the form of frequency distributions of variants of a cultural trait at a particular moment in time, or of frequency changes of variants over time. A central question in the cultural evolution program is how such outcomes can be used to explain observed periods of cultural change.<sup>24</sup> For story transmission processes, good examples of population-level outcomes are popularity polls, indices of reprints and bibliographical databases.<sup>25</sup> All of this

22. Kandler and Powell, “Inferring Learning Strategies from Cultural Frequency Data.”

23. Ibid.

24. Mesoudi and Lycett, “Random copying, frequency-dependent copying and culture change”; Kandler and Powell, “Inferring Learning Strategies from Cultural Frequency Data”; Anne Kandler and Stephen Shennan, “A generative inference framework for analysing patterns of cultural change in sparse population data with evidence for fashion trends in LBK culture,” *Journal of The Royal Society Interface* 12, no. 113 (2015): 20150905–12.

25. Koman and Meder, *Resultaten Spreekjesenquête*; Joosen, “A Translation Far Worse.

information is much easier to obtain than large-scale longitudinal full-text corpora, providing us with a useful starting point for addressing the above-mentioned issues. I take it to be an intriguing question how such sparse, population-level data can be used to further identify and map out the underlying acting-forces in story transmission and selection processes.

With the present study, I hope to have revealed and underscored the descriptive and theoretical advantages as well as the further potential of adapting and appropriating computational-evolutionary models to story transmission research. Yet, just as the life span of a story depends on its being picked up and further retold, the continuation of the ideas and models presented above depends on its being adapted and appropriated in new retellings within future research on story transmission, cultural evolution, and folkloristic and literary studies. Hopefully, the present 'story' will contribute to "spark related thoughts, responses, and readings"<sup>26</sup> in retellings to come.





---

## BIBLIOGRAPHY

### PRIMARY SOURCES

- Beckett, Sandra L. *Revisioning Riding Hood around the World. An Anthology of International Retellings*. Detroit, Michigan: Wayne State University Press, 2014.
- Ben-Amos, Dan. *Folktales of the Jews*. 3 vols. Philadelphia: Jewish Publication Society, 2006.
- Crooke, William, and Pandit Pam Gharib Chaube. *Folktales from Northern India*. Santa Barbara, California: ABC-CLIO, 2002.
- De Vries-Vogel, M. M. *De sprookjes van Grimm; volledige uitgave*. Weesp, The Netherlands: Unieboek BV – Van Holkema & Warendorf, 1984.
- De Wijs, Ivo, and Wouter Tulp (illustrations). *En ze leefden nog... Sprookjes op Rijn*. Amsterdam: Amsterdam Blue in Green, 2011.
- De Wijs, Ivo, and Alfons Van Heusden (illustrations). *Roodkapje en de zeven geitjes*. Amsterdam: Leopold, 1994.
- Dros, Imme, and Henricus Geelen (illustrations). *De Wolf die tegen water praatte*. Bussum: Moon Press, 1991.
- Meder, Theo. *De Magische Vlucht. Nederlandse Volksverhalen uit de Collectie van het Meertens Instituut [The magic flight. Dutch folktales from the collection of the Meertens Institute]*. Amsterdam: Uitgeverij Bert Bakker, 2000.
- El-Shamy, Hasan. *Folktales of Egypt*. Chicago: University of Chicago Press, 1980.
- Sinninghe, Jacques. *Volkssprookjes uit Nederland en Vlaanderen*. Den Haag, The Netherlands: Kruseman, 1978.

- Slone, Thomas H. *One Thousand One Papua New Guinean Nights. Folktales from Wantok Newspaper*. Vol. 1: Tales from 1972–1985. Oakland, California: Masalai, 2001.
- Van de Vendel, Edward, and Isabelle Vandenabeele (illustrations). *Rood Rood Roodkapje*. Wielsbeke: De Eenhoorn, 2003.
- VanArsdale, Daniel W. *The Paper Chain Letter Archive*. <http://www.silcom.com/~barnowl/chain-letter/archive/!content.html>. Accessed: 2015-10-09, 2015.

## SECONDARY SOURCES

- Aarne, Antti. *The Types of the Folktale*. Vol. 74. Folklore Fellows Communications. Helsinki: Academia Scientiarum Fennica, 1928.
- . *Verzeichnis Der Märchentypen [Catalogue of folktale types]*. Vol. 3. Folklore Fellows Communications. Helsinki: Academia Scientiarum Fennica, 1910.
- Aarne, Antti, and Stith Thompson. *The Types of the Folktale: A Classification and Bibliography*. Vol. 184. FF communications. Helsinki: Academia Scientiarum Fennica, 1961.
- Abello, James, Peter Broadwell, and Timothy R. Tangherlini. "Computational folkloristics." *Communications ACM* (New York, NY, USA) 55, no. 7 (2012): 60–70. <http://doi.acm.org/10.1145/2209249.2209267>.
- Acerbi, Alberto, and R. Alexander Bentley. "Biases in cultural transmission shape the turnover of popular traits." *Evolution and Human Behavior* 35, no. 3 (2014): 228–236.
- Acerbi, Alberto, Stefano Ghirlanda, and Magnus Enquist. "The Logic of Fashion Cycles." *PLoS ONE* 7, no. 3 (2012): e32541–9.
- Anderson, Graham. *Fairytale in the Ancient World*. London, New York: Routledge, 2000.
- Badham, Jennifer M. "Commentary: Measuring the shape of degree distributions." *Network Science* 1, no. 2 (2013): 213–225.

- Bal, Mieke. *Narratology. Introduction to the Theory of Narrative*. 3rd ed. Toronto, Buffalo, London: University of Toronto Press, 2009.
- Barabási, Albert-Laszlo, and Reka Albert. "Emergence of Scaling in Random Networks." *Science* 286, no. 5439 (1999): 509–512.
- Barrett, Justin. "Coding and Quantifying Counterintuitiveness in Religious Concepts: Theoretical and Methodological Reflections." *Method & Theory in the Study of Religion* 20, no. 4 (2008): 308–338.
- Barrett, Justin, Emily Reed Burdett, and Tenelle Porter. "Counterintuitiveness in Folktales: Finding the Cognitive Optimum." *Journal of Cognition and Culture* 9, no. 3 (2009): 271–287.
- Barrett, Justin, and Melanie Nyhof. "Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials." *Journal of Cognition and Culture* 1, no. 1 (2004): 69–100.
- Bartlett, Frederic. *Remembering. A Study in Experimental and Social Psychology*. Cambridge Books Online. London: Cambridge University Press, 1932.
- Baughman, Ernest W. *Type and Motif-Index of the Folktales of England and North America*. Vol. 20. Indiana University Folklore Series. The Hague: Mouton, 1966.
- Beckett, Sandra L. *Recycling Red Riding Hood*. New York, London: Routledge, 2002.
- . *Red Riding Hood for all Ages. A Fairy-Tale Icon in Cross-Cultural Contexts*. Detroit, Michigan: Wayne State University Press, 2008.
- Beheim, Bret A., Calvin Thigpen, and Richard McElreath. "Strategic social learning and the population dynamics of human behavior: the game of Go." *Evolution and Human Behavior* 35, no. 5 (2014): 351–357.
- Bell, Sita. "Anti-Semitic Folklore Motif Index." Master's thesis, 2009. <http://digitalcommons.usu.edu/etd/299>.

- Bentley, R. Alexander. "Random Drift versus Selection in Academic Vocabulary: An Evolutionary Analysis of Published Keywords." *PLoS ONE* 3, no. 8 (2008): e3057.
- Bentley, R. Alexander, Matthew W. Hahn, and Stephen J. Shennan. "Random drift and culture change." *Proceedings of the Royal Society B: Biological Sciences* 271, no. 1547 (2004): 1443–1450.
- Bentley, R. Alexander, Paul Ormerod, and Michael Batty. "Evolving social influence in large populations." *Behavioral Ecology and Sociobiology* 65, no. 3 (2011): 537–546.
- Bentley, R. Alexander, Paul Ormerod, and Stephen Shennan. "Population-level neutral model already explains linguistic patterns." *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1713 (2011): 1770–1772.
- Berlin, Brent, and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press, 1969.
- Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. Sebastopol: O'Reilly, 2009.
- Bloem, Jelke, and Gosse Bouma. "Automatic animacy classification for Dutch." *Computational Linguistics in the Netherlands Journal* 3 (2013): 82–102.
- Bolton, Ralph, and Diane Crisp. "Color Terms in Folk Tales: A Cross-Cultural Study." *Cross-Cultural Research* 14, no. 4 (1979): 231–253.
- Bouma, Gosse, Gertjan Van Noord, and Robert Malouf. "Alpino: Wide-coverage computational analysis of Dutch." *Language and Computers* 37, no. 1 (2001): 45–59.
- Bowman, Samuel, and Harshit Chopra. "Automatic animacy classification." In *Proceedings of the NAACL - HLT 2012 Student Research Workshop*. 2012.
- Boyd, Brian. *On the Origin of Stories. Evolution, Cognition and Fiction*. Cambridge, Massachusetts, London, England: The Belknap Press of Harvard University Press, 2009.

- Boyd, Brian, Joseph Carroll, and Jonathan Gottschall, eds. *Evolution, Literature, and Film. A Reader*. New York: Columbia University Press, 2010.
- Boyd, Robert, and Peter J. Richerson. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press, 1985.
- Boyer, Pascal. *Religion explained. The evolutionary origins of religious thought*. New York: Basic Books, 2001.
- Breithaupt, Fritz, Eleanor Brower, and Sarah Whaley. "Optimal Eventfulness of Narratives." In *Proceedings of the Computational Models of Narrative Workshop*, 45:12–22. OASICS. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.
- Burkert, Walter. *Creation of the Sacred. Tracks of Biology in Early Religions*. Cambridge, Massachusetts, London, England: Harvard University Press, 1996.
- . *Structure and History in Greek Mythology and Ritual*. Berkeley, Los Angeles, London: University of California Press, 1982.
- Carron, Pádraig Mac, and Ralph Kenna. "Universal properties of mythological networks." *EPL (Europhysics Letters)* 99, no. 2 (2012): 28002–28007.
- Cavalli-Sforza, Luigi Luca, and Marcus Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Monographs in Population Biology 16. Princeton, New Jersey: Princeton University Press, 1981.
- Childers, James Wesley. *Tales from Spanish Picaresque Novels: A Motif Index*. New York: State University of New York Press, 1977.
- Choi, Seung-Seok, Sung-Hyuk Cha, and Charles C. Tappert. "A survey of binary similarity and distance measures." *Journal of Systemics, Cybernetics and Informatics* 8, no. 1 (2010): 43–48.
- Church, Kenneth Ward, and Patrick Hanks. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics* 16, no. 1 (1990): 22–29.

- Clauset, Aaron, Cosma Rohilla Shalizi, and Mark E. J. Newman. "Power-law distributions in empirical data." *SIAM Review*, no. 4 (2009): 661–703. doi:10.1137/070710111.
- Cohen, Jacob. "A Coefficient of Agreement for Nominal Scales." *Educational and psychological measurement* 20, no. 1 (1960): 37–46.
- Collins, Fiona, and Jeremy Ridgman, eds. *Turning the page, Children's Literature in Performance and the Media*. Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang, 2006.
- Comrie, Bernard. *Language Universals and Linguistic Typology*. 2nd ed. Chicago: University of Chicago Press, 1989.
- Da Silva, Sara Graça, and Jamshid J. Tehrani. "Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales." *Royal Society Open Science* 3, no. 1 (2016). doi:10.1098/rsos.150645.
- De Jong, Franciska M.G. "The Compositional Nature of (In)Definiteness." In *The Representation of (In)definiteness. Current Studies in Linguistics Series*, 14:270–285. Cambridge, Massachusetts: MIT Press, 1987.
- De Smedt, Tom, and Walter Daelemans. "Pattern for Python." *Journal of Machine Learning Research* 13 (2012): 2031–2035.
- Declerck, Thierry, and Piroska Lendvai. "Linguistic and Semantic Representation of the Thompson's Motif- Index of Folk-Literature." In *Research and Advanced Technology for Digital Libraries*, edited by S. Gradmann, F. Borri, C. Meghini, and H. Schuldt, 6966:151–158. Lecture Notes in Computer Science. Berlin: Springer, 2011.
- Dennett, Daniel. *The Intentional Stance*. Cambridge, Massachusetts: The MIT Press, 1996.
- Diehl, Joshua J, Loisa Bennetto, and Edna Carter Young. "Story Recall and Narrative Coherence of High-Functioning Children with Autism Spectrum Disorders." *Journal of Abnormal Child Psychology* 34, no. 1 (2006): 83–98.
- Dorogovtsev, Sergei, and José Fernando Mendes. "Evolution of reference networks with aging." *Physical Review. E, Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics* 62, no. 2 (2000): 1842–1845.

- Dundes, Alan. "The Motif-Index and the Tale Type Index: A Critique." *Journal of Folklore Research* 34, no. 6 (1997): 195–202.
- Dunning, Ted. "Accurate methods for the statistics of surprise and coincidence." *Computational Linguistics* 19, no. 1 (1993): 61–74.
- Ellis, Bill. *Lucifer Ascending: The Occult in Folklore and Popular Culture*. Lexington: University Press of Kentucky, 2004.
- Eom, Young-Ho, and Santo Fortunato. "Characterizing and Modeling Citation Dynamics." *PLoS ONE* 6, no. 9 (2011): e24926–7.
- Eriksson, Kimmo, and Julie C. Coultas. "Corpses, maggots, poodles and rats: Emotional selection operating in three phases of cultural transmission of urban legends." *Journal of Cognition and Culture* 14, nos. 1–2 (2014): 1–26.
- . "The Advantage of Multiple Cultural Parents in the Cultural Transmission of Stories." *Evolution and Human Behavior* 33, no. 4 (2012): 251–259.
- Evans, Richard, and Constantin Orăsan. "Improving anaphore resolution by identifying animate entities in texts." In *Proceedings of the Discourse Anaphora and Reference Resolution Conference*, 154–162. 2000.
- Fellbaum, Christiane. "WordNet(s)." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, 665–670. Oxford: Elsevier, 2005.
- Fisher, Ronald. "On the Interpretation of  $\chi^2$  from Contingency Tables, and the Calculation of P." *Journal of the Royal Statistical Society* 85, no. 1 (1922): 87–94.
- Flowers, Helen Leneva. *A Classification of the Folk Tale of the West Indies by Types and Motifs*. New York: Arno, 1980.
- Frank, Arthur W. *Letting Stories Breathe. A Socio-Narratology*. Chicago, London: The University of Chicago Press, 2010.



- Gao, Tao, Brian Scholl, and Gregory McCarthy. "Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus." *The Journal of neuroscience: the official journal of the Society for Neuroscience* 32, no. 41 (2012): 14276–14280.
- Geerts, Sylvie, and Sara van den Bossche, eds. *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*. Ginkgo. Ghent, Belgium: Academia Press, 2014.
- . "Never-ending Stories. How Canonical Works Live on in Children's Literature." In *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*, edited by Sylvie Geerts and Sara van den Bossche. Ginkgo. Ghent, Belgium: Academia Press, 2014.
- Ghirlanda, Stefano, Alberto Acerbi, and Harold Herzog. "Dog Movie Stars and Dog Breed Popularity: A Case Study in Media Influence on Choice." *PLoS ONE* 9, no. 9 (2014): e106565–5.
- Goldberg, Harriet. *Motif-Index of Medieval Spanish Folk Narratives*. Medieval & Renaissance Texts & Studies. Tempe: AZ, 1998.
- Goldberg, Sophia R., Hannah Anthony, and Tim S. Evans. "Modelling citation networks." *Scientometrics* 105, no. 3 (2015): 1577–1604.
- Good, Phillip I. *Resampling Methods. A Practical Guide to Data Analysis*. 3rd ed. Boston, Basel, Berlin: Birkhäuser, 2006.
- Grunst, Andrea S., John T. Rotenberry, and Melissa L. Grunst. "Age-dependent relationships between multiple sexual pigments and condition in males and females." *Behavioral Ecology*, 2014. doi:10.1093/beheco/art124.
- Hahn, Matthew W., and R. Alexander Bentley. "Drift as a mechanism for cultural change: an example from baby names." *Proceedings of the Royal Society B: Biological Sciences* 270 (2003): S120–S123.
- Harmon-Vukić and D. Jason Slone. "The Effect of Integration on Recall of Counterintuitive Stories." *Journal of Cognition and Culture* 9, no. 1 (2009): 57–68.

- Heath, Chip, Chris Bell, and Emily Sternberg. "Emotional Selection in Memes: The Case of Urban Legends." *Journal of Personality and Social Psychology* 81, no. 6 (2001): 1028–1041.
- Hemming, Jessica. "Red, White, and Black in Symbolic Thought: The Tricolour Folk Motif, Colour Naming, and Trichromatic Vision." *Folklore* 123, no. 3 (2012): 310–329.
- Henrich, Joseph, Robert Boyd, and Peter J. Richerson. "Five Misunderstandings About Cultural Evolution." *Human Nature* 19, no. 2 (2008): 119–137.
- Henrich, Joseph, and Richard McElreath. "The evolution of cultural evolution." *Evolutionary Anthropology: Issues, News, and Reviews* 12 (2003): 123–135.
- Herzog, Harold A., R. Alexander Bentley, and Matthew W. Hahn. "Random drift and large shifts in popularity of dog breeds." *Proceedings of the Royal Society B: Biological Sciences* 271 (2004): S353–S356.
- Hiemstra, Djoerd, Stephen Robertson, and Hugo Zaragoza. "Parsimonious language models for information retrieval." In *SIGIR '04 Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 178–185. 2004.
- Hutcheon, Linda. *A Theory of Adaptation*. New York, London: Routledge, 2013.
- Ikeda, Hireko. *A Type and Motif Index of Japanese Folk-Literature*. Vol. 209. Folklore Fellows Communications. Helsinki: Academia Scientarium Fennica, 1971.
- Isaksson, Sven, Alexander Funcke, Ida Envall, Magnus Enquist, and Patrik Lindenfors. "A Novel Method to Analyze Social Transmission in Chronologically Sequenced Assemblages, Implemented on Cultural Inheritance of the Art of Cooking." *PLoS ONE* 10, no. 5 (2015): e0122092–13.
- Jason, Heda. *Motif, Type and Genre: A Manual for Compilation of Indices and a Bibliography of Indices and Indexing*. Helsinki: Suomalainen Tiedeakatemia, 2000.

- Joosen, Vanessa. "A Translation Far Worse. Canonisation and Adaptation in the Early Dutch and English Translation of the Brothers Grimm's *Kinder- und Hausmärchen*." In *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*, edited by Sylvie Geerts and Sara van den Bossche. Ginkgo. Ghent, Belgium: Academia Press, 2014.
- Kandler, Anne, and Adam Powell. "Inferring Learning Strategies from Cultural Frequency Data." In *Learning Strategies and Cultural Evolution during the Palaeolithic*, edited by Alex Mesoudi and Kenichi Aoki, 85–101. Tokyo: Springer Japan, 2015.
- Kandler, Anne, and Stephen Shennan. "A generative inference framework for analysing patterns of cultural change in sparse population data with evidence for fashion trends in LBK culture." *Journal of The Royal Society Interface* 12, no. 113 (2015): 20150905–12.
- . "A non-equilibrium neutral model for analysing cultural change." *Journal of theoretical biology* 330 (2013): 18–25.
- Karsdorp, Folgert. "Het is groen en leeft nog lang en gelukkig. Classificatie van volksverhaalgenres op basis van formules." *Tijdschrift voor Nederlands Taal- en Letterkunde* 129, no. 4 (2013): 274–288.
- . *Story network data sets.*, 2016. doi:10.5281/zenodo.51588. <http://dx.doi.org/10.5281/zenodo.51588>.
- Karsdorp, Folgert, Peter van Kranenburg, Theo Meder, Dolf Trieschnigg, and Antal van den Bosch. "In search of an appropriate abstraction level for motif annotations." In *Proceedings of the 2012 Computational Models of Narrative Workshop*, 22–26. Turkey, Istanbul, 2012.
- Karsdorp, Folgert, and Antal van den Bosch. "Identifying motifs in folktales using topic models." In *Proceedings of BENELEARN 2013*, 41–49. Nijmegen, The Netherlands, 2013.
- Karsdorp, Folgert, Peter Van Kranenburg, Theo Meder, and Antal van den Bosch. "Casting a Spell: Identification and Ranking of Actors in Folktales." In *Proceedings of the Second Workshop on Annotation of Corpora for Research in het Humanities (ACRH-2)*, 39–50. 2012.

- Kirtley, Bacil F. *A Motif Index of Traditional Polynesian Narratives*. Honolulu: University of Hawai'i Press, 1977.
- Koman, Ruben A., and Theo Meder. *Resultaten Sprookjesenquête*. Kaatsheuvel, Amsterdam: Efteling; DOC Volksverhaal, 2008.
- Koppel, Moshe, and Yaron Winter. "Determining if Two Documents are by the Same Author." *JASIST*, no. 1 (2014): 178–187.
- Kulis, Brian, and Michael I. Jordan. "Revisiting k-means: New Algorithms via Bayesian Nonparametrics." In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 513–520. 2012.
- La Barre, Kathryn A., and Carol L. Tilley. "The elusive tale: leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales." *Journal of the American Society for Information Science and Technology* 63, no. 4 (2012): 687–701. ISSN: 1532-2890. <http://dx.doi.org/10.1002/asi.21710>.
- Lee, Heeyoung, Angel Chang, Yves Peirsman, Nathanael Chambers, Mi-hai Surdeanu, and Dan Jurafsky. "Deterministic coreference resolution based on entity-centric, precision-ranked rules." *Computational Linguistics* 39, no. 4 (2013): 885–916.
- Legman, Gershon. "Toward a Motif-Index of Erotic Humor." *The Journal of American Folklore* 75, no. 297 (1962): 227–248.
- Lenat, Doug, George Miller, and Toshio Yokoi. "CYC, WordNet, and EDR: Critiques and Responses." *Communications of the ACM* 38, no. 11 (1995): 33–38.
- Lewens, Tim. *Cultural Evolution. Conceptual Challenges*. Oxford: Oxford University Press, 2015.
- Liben-Nowell, David, and Jon Kleinberg. "Tracing information flow on a global scale using Internet chain-letter data." *PNAS* 105, no. 12 (2008): 4633–4638.
- Lieberson, Stanley. *A matter of taste: How names, fashions, and culture change*. New Haven, CT: Yale University Press, 2000.

- Linde, Charlotte. *Working with the Past: Narrative and Institutional Memory*. New York: Oxford University Press, 2009.
- Loewenstein, Jeffrey, and Chip Heath. "The Repetition-Break plot structure: A cognitive influence on selection in the marketplace of ideas." *Cognitive Science* 33 (2009): 1–19.
- Loewenstein, Jeffrey, Rajagopal Raghunathan, and Chip Heath. "The Repetition-Break Plot Structure Makes Effective Television Advertisements." *Journal of Marketing* 75, no. 5 (2011): 105–119.
- Lundell, Torborg. "Folktale Heroines and the Type and Motif Indexes." *Folklore* 94, no. 2 (1983): 240–246.
- . "Gender-Related Biases in the Type and Motif Indexes of Aarne and Thompson." In *Fairy Tales and Society: Illusion, Allusion, and Paradigm*, edited by Ruth B. Bottigheimer, 149–163. Philadelphia: University of Pennsylvania Press, 1989.
- Lyons, Anthony, and Yoshihisa Kashima. "Maintaining stereotypes in communication: Investigating memory biases and coherence-seeking in storytelling." *Asian Journal Of Social Psychology* 9, no. 1 (2006): 59–71.
- Lyons, Christopher. *Definiteness*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, 1999.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. New York, USA: Cambridge University Press, 2008.
- Mark, Noah. "Birds of a Feather Sing Together." *Social Forces* 77, no. 2 (1998): 453–485.
- Mauch, Matthias, Robert M. MacCallum, Mark Levy, and Armand M. Leroi. "The evolution of popular music: USA 1960–2010." *Royal Society Open Science* 2, no. 5 (2015): 150081.
- McElreath, Richard, Mark Lubell, Peter J. Richerson, Timothy M. Waring, William Baum, Edward Edsten, Charles Efferson, and Brian Paciotti. "Applying evolutionary models to the laboratory study of social learning." *Evolution and Human Behavior* 26 (6 2005): 483–508.

- Meder, Theo. "From a Dutch Folktale Database towards an International Folktale Database." *Fabula* 51, nos. 1–2 (2010): 6–22.
- Meder, Theo, Folgert Karsdorp, Dong Nguyen, Mariët Theune, Dolf Trieschnigg, and Iwe Muiser. "Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database." *Journal of American Folklore* 129, no. 511 (2016): 78–96.
- Mesoudi, Alex. "Cultural Evolution: A Review of Theory, Findings and Controversies." *Evolutionary Biology*, 2015, 1–17. doi:10.1007/s11692-015-9320-0.
- . *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*. Chicago, London: University of Chicago Press, 2011.
- . "How cultural evolutionary theory can inform social psychology and vice versa." *Psychological Review* 116, no. 4 (2009): 929–952.
- . "The Transmission and Evolution of Human Culture," 2005.
- Mesoudi, Alex, and Stephen J. Lycett. "Random copying, frequency-dependent copying and culture change." *Evolution and Human Behavior* 30, no. 1 (2009): 41–48.
- Mesoudi, Alex, and Andrew Whiten. "The Multiple Roles of Cultural Transmission experiments in Understanding Human Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, no. 1509 (2008): 3489–3501.
- Mesoudi, Alex, Andrew Whiten, and Robin Dunbar. "A bias for social information in human cultural transmission." *British Journal of Psychology* 97, no. 3 (2006): 405–423.
- . "A bias for social information in human cultural transmission." *British Journal of Psychology* 97, no. 3 (2010): 405–423.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space." In *Proceedings of Workshop at ICLR*. 2013.

- Moore, Joshua, Christopher Burges, Erin Renshaw, and Wen-tau Yih. "Animacy Detection with Voting Models." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 55–60. 2013.
- Neuland, Lena. *Motif-Index of Latvian Folktales and Legends*. Vol. 229. Folklore Fellows Communications. Helsinki: Academia Scientarium Fennica, 1981.
- Newman, Mark E. J. "The structure and function of complex networks." *SIAM REVIEW* 45, no. 2 (2003): 167–256.
- Nieuwland, Mante S., and Jos J. A. Van Berkum. "When Peanuts Fall in Love: N400 Evidence for the Power of Discourse." *Journal of Cognitive Neuroscience* 18, no. 7 (2005): 1098–1111.
- Norenzayan, Ara, and Scott Atran. "Cognitive and emotional processes in the cultural transmission of natural and nonnatural beliefs." In *The Psychological Foundations of Culture*, edited by M. Schaller and C. Crandall, 149–169. Hillsdale, NJ: Lawrence Erlbaum, 2004.
- Norenzayan, Ara, Scott Atran, Jason Faulkner, and Mark Schaller. "Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives." *Cognitive Science* 30 (2006): 531–553.
- Opfer, John. "Identifying living and sentient kinds from dynamic information: The case of goal-directed versus aimless autonomous movement in conceptual change." *Cognition* 86, no. 2 (2002): 97–122.
- Orăsan, Constantin, and Richard Evans. "Learning to identify animate references." In *Proceedings of CoNLL-2001*, 129–136. Toulouse, France, 2001.
- . "NP animacy identification for anaphora resolution." *Journal of Artificial Intelligence Research* 29 (2007): 79–103.
- Orenstein, Catherine. *Little Red Riding Hood Uncloaked. Sex, Morality and the Evolution of a Fairy Tale*. New York: Basic Books, 2002.

- Øvrelid, Lilja. "Animacy classification based on morphosyntactic corpus frequencies: Some experiments with Norwegian nouns." In *Proceedings of the Workshop on Exploring Syntactically Annotated Corpora*, 24–34. 2005.
- . "Linguistic features in data-driven dependency parsing." In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL 2008)*. 2008.
- . "Towards robust animacy classification using morphosyntactic distributional features." In *Proceedings of the EACL 2006 Student Research Workshop*, 47–54. 2006.
- Øvrelid, Lilja, and Joakim Nivre. "When word order and part-of-speech tags are not enough – Swedish dependency parsing with rich linguistic features." In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, 447–451. 2007.
- Owens, Justine, and Gordon Bower. "The "soap opera" effect in story recall." *Memory Cognition* 7, no. 3 (1979): 185–191.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (2011): 2825–2830.
- Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation." In *Proceedings of The 2014 Conference on Empirical Methods in Natural Language Processing*, 1532–1543. Doha, Qatar, 2014.
- Perc, Matjaž. "The Matthew effect in empirical data." *Journal of The Royal Society Interface* 11, no. 98 (2014). doi:10.1098/rsif.2014.0378.
- Perreault, Charles. "The Pace of Cultural Evolution." *PLoS ONE* 7, no. 9 (2012): e45150.
- Price, Derek De Solla. "A general theory of bibliometric and other cumulative advantage processes." *Journal of the American Society for Information Science* 27, no. 5 (1976): 292–306.



- Radden, Günter, and René Dirven. *Cognitive English Grammar*. Cognitive Linguistics in Practice. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007.
- Ramsey, Grant, and Andreas De Block. "Is Cultural Fitness Hopelessly Confused?" *The British Journal for the Philosophy of Science*, 2015. doi:10.1093/bjps/axv047.
- Reysen, Matthew B., Natalie G. Talbert, Mura Dominko, Amie N. Jones, and Matthew R. Kelley. "The effects of collaboration on recall of social information." *British Journal of Psychology* 102, no. 3 (2011): 646–661.
- Richerson, Peter J., and Robert Boyd. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press, 2005.
- Rosenbach, Anette. "Animacy and grammatical variation – findings from English genitive variation." *Lingua* 118, no. 2 (2008): 151–171.
- Rubin, David C. *Memory in Oral Traditions. The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes*. New York, Oxford: Oxford University Press, 1995.
- Rubin, Timothy N., America Chambers, Padhraic Smyth, and Mark Steyvers. "Statistical topic models for multi-label document classification." *Machine Learning* 88, no. 1 (2012): 157–208.
- Sanders, Julie. *Adaptation and Appropriation*. London, New York: Routledge, 2006.
- Schäfer, Roland, and Felix Bildhauer. "Building Large Corpora from the Web Using a New Efficient Tool Chain." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul: ELRA, 2012.
- El-Shamy, Hasan. "A Type Index for Tales of the Arab World." *Fabula* 29, nos. 1/2 (1988): 150–163.
- . *Folk Traditions of the Arab World: A Guide to Motif Classification*. Bloomington: Indiana University Press, 1995.

- . "Oral Traditional Tales and the Thousand Nights and a Night: The Demographic Factor." In *The Telling of Stories: Approaches to a Traditional Craft*, edited by Morton Nøjgaard, J. de Mylius, I. Piø, and Bengt Holbeck, 63–117. Odense: Odense University Press, 1990.
- Smith, Allen. "JAL Guide to Software, Courseware and CD-ROM: Stith Thompson's Motif-Index of Folk Literature." *Journal of Academic Librarianship* 20, no. 4 (1994): 255.
- Sperber, Dan. *Explaining Culture. A Naturalistic Approach*. Oxford: Blackwell Publishing, 1996.
- Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. "BRAT: a Web-based Tool for NLP-Assisted Text Annotation." In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 102–107. 2012.
- Stephens, John, and Robyn McCallum. *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*. New York, London: Garland Publishing Inc., 1998.
- Stubbersfield, Joseph M., Jamshid J. Tehrani, and Emma G. Flynn. "Serial killers, spiders and cybersex: Social and survival information bias in the transmission of urban legends." *British Journal of Psychology* 106, no. 2 (2014): 288–307.
- Tehrani, Jamshid. "The Phylogeny of Little Red Riding Hood." *PLoS ONE* 8, no. 11 (2013).
- Thompson, Stith. *Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folk-Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends*. Bloomington: Indiana University Press, 1932–1936.
- . *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*. Revised and enlarged. Bloomington: Indiana University Press, 1955–1958.

- Thompson, Stith. *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempli, Fabliaux, Jestbooks, and Local Legends*. CD-ROM. Bloomington: Indiana University Press, 1993.
- . *The Folktale*. 2nd ed. Berkley, Los Angeles & London: University of California Press, 1977.
- Tomasello, Michael. *The Cultural Origins Of Human Cognition*. Cambridge, Massachusetts: Harvard University Press, 2009.
- Toutanova, Kristina, Dan Klein, Christopher Manning, and Yoram Singer. "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network." In *Proceedings of HLT-NAACL 2003*. Stroudsburg, PA: Association for Computational Linguistics, 2003. <http://nlp.stanford.edu/kristina/papers/tagging.pdf>.
- Tsoumakas, Grigorios, and Ioannis Katakis. "Multi-label classification: An overview." *International Journal of Data Warehousing and Mining* 3, no. 3 (2007): 1–13.
- Tsoumakas, Grigorios, Ioannis Katakis, and Ioannis Vlahavas. "Mining Multi-label Data." In *Data Mining and Knowledge Discovery Handbook*, edited by Oded Maimon and Lior Rokach, 667–685. Boston, MA: Springer US, 2010.
- Upal, M. Afzal. "Memory, Mystery and Coherence: Does the Presence of 2–3 Counterintuitive Concepts Predict Cultural Success of a Narrative?" *Journal of Cognition and Culture* 11, no. 1 (2011): 23–48.
- Upal, M. Afzal, Lauren O. Gonce, Ryan D. Tweney, and D. Jason Slone. "Contextualizing Counterintuitiveness: How Context Affects Comprehension and Memorability of Counterintuitive Concepts." *Cognitive Science* 31, no. 3 (2007): 415–439.
- Uther, Hans-Jörg. *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*. FF Communications, 284–286. Helsinki: Suomalainen Tiedekatemia, 2004.

- . "Type- and Motif-Indices 1980–1995: An Inventory." *Asian Folklore Studies* 55, no. 2 (1996): 299–317.
- Van Boven, Erica, and Gillis Dorleijn. *Literair Mechaniek. Inleiding tot de analyse van verhalen en gedichten*. 2nd ed. Bussum: Uitgeverij Coutinho, 2003.
- Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing High-Dimensional Data Using t-SNE." *Journal of Machine Learning Research*, 2008, 2579–2605.
- Van Rijsbergen, Cornelis. *Information Retrieval*. London: Butterworths, 1979.
- Van Venrooij, Alex. "A community ecology of genres. Explaining the emergence of new genres in the UK field of electronic/dance music, 1985–1999." *Poetics* 52 (2015): 104–123.
- VanArsdale, Daniel W. *Chain Letter Evolution*. <http://www.silcom.com/~barnowl/chain-letter/evolution.html>. Accessed: 2015-10-09, 2015.
- Vogels, Jorrig, Emiel Krahmer, and Alfons Maes. "When a stone tries to climb up a slope: the interplay between lexical and perceptual animacy in referential choices." *Frontiers in Psychology* 4, no. 154 (2013): 1–15.
- Voigt, Vilmos, Michael Preminger, László Ládi, and Sándor Darány. "Automated Motif Identification in Folklore Text." *Folklore. An Electronic Journal of Folklore* 12 (1999): 126–141.
- Warner, Marina. *From the Beast to the Blonde. On Fairy Tales and their Tellers*. London: Vintage, 1995.
- Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *Nature* 393, no. 6684 (1998): 440–442.
- Wiese, Heike. "Semantics as a gateway to language." In *Mediating between Concepts and Grammar*, edited by Holden Härtl and Heike Tappe, 197–222. Trends in Linguistics. Berlin, New York: Mouton de Gruyter, 2003.

- Wilks, Yorick, Brian M. Slator, and Louise Guthrie. *Electronic Words: Dictionaries, Computers, and Meanings*. Cambridge, Massachusetts: MIT Press, 1996.
- Wittgenstein, Ludwig. *Philosophische Untersuchungen*. Vol. 203. suhrkamp taschenbuch wissenschaft. Frankfurt am Main: Suhrkamp Taschenbuch Verlag, 1958.
- Zipes, Jack. *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*. Princeton, Oxford: Princeton University Press, 2012.
- . *The Trials & Tribulations of Little Red Riding Hood*. 2nd ed. New York, London: Routledge, 1993.
- . *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*. New York, London: Routledge. Taylor & Francis Group, 2006.

---

## SAMENVATTING

*Retelling Stories* is een onderzoek naar mechanismes die ten grondslag liggen aan de verspreiding en verandering van (volks)verhalen. Een van de centrale vragen in dit onderzoek is of verhaaltransmissie begrepen kan worden als een cultureel evolutionair proces en welke evolutionaire mechanismen daarbij een rol spelen. Om deze vraag te beantwoorden, zoekt deze studie aansluiting bij onderzoek naar computationele modellen van culturele verandering.<sup>1</sup> Door gebruik te maken van deze computationele benadering van culturele verandering, probeert de huidige studie om diachrone verhaalontwikkelingen op een kwantitatieve en formele manier te bestuderen en te karakteriseren. Hoewel een computationeel-evolutionaire benadering van verhaaltransmissie (en de daarbijhorende methodologische vraagstukken) centraal staat, gaat deze studie ook de dialoog aan met *kwalitatieve* benaderingen van verhaaltransmissie<sup>2</sup> met als doel de synthese tussen culturele evolutie, (jeugd)literatuur en verhaalonderzoek te versterken. Na een inleidend hoofdstuk, waarin het methodologische kader en de probleemstelling van de studie worden gepresenteerd, wordt in de daaropvolgende vier hoofdstukken een evenredig aantal onderzoeksvragen behandeld, waarvan de onderzoeksrelevantie zowel methodologisch als theoretisch van aard is.

De eerste onderzoeksvraag vormt het methodologische fundament van de studie en richt zich op de representatie en formalisatie van verhalen om historische verhaaltransmissie en -selectie computationeel

1. Zie bijvoorbeeld: Cavalli-Sforza en Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*; Boyd en Richerson, *Culture and the Evolutionary Process*; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*; Mesoudi, "Cultural Evolution: A Review of Theory, Findings and Controversies".

2. Zie bijvoorbeeld: Stephens en McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*; Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction*; Boyd en andere, *Evolution, Literature, and Film. A Reader*; Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*; Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*; Geerts en van den Bossche, *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*.

te kunnen bestuderen. In hoofdstuk 2 tot en met 5 worden vier verschillende datarepresentaties onderzocht, die elk potentiële voor- en nadelen hebben voor de bestudering van bepaalde aspecten van verhaaltransmissie. Hoofdstuk 2 presenteert een ‘klassieke’ representatie van verhalen die gebaseerd is op de beroemde *Motif-Index* ontwikkeld door Stith Thompson.<sup>3</sup> In deze representatie worden motieven beschouwd als de primaire bouwstenen van verhalen en hun verzameling bepaalt tot welk verhaalttype een verhaal behoort (zoals de verhaaltypes in *The Types of International Folktales*<sup>4</sup>). Deze representatie, waarin verhalen primair worden beschouwd als instantiaties van bepaalde verhaaltypes, staat haaks op de data-gedreven representaties van verhalen in de hoofdstukken 3 tot en met 5. In deze hoofdstukken worden drie representaties voorgesteld waarin niet het verhaalttype het primaire onderwerp van studie is, maar het eigenlijke verhaal of ‘verhaaltoken’. Een belangrijk voordeel van deze token-gebaseerde representaties ten opzichte van de ‘klassieke’, type-gebaseerde representatie, is dat overeenkomsten tussen verhalen niet statisch zijn, maar afhankelijk van het door de onderzoeker gekozen perspectief.

In hoofdstuk 3 tot en met 5 worden de data-gedreven verhaalrepresentaties gebruikt om een aantal onderzoeksvragen te beantwoorden over mechanismes van verhaaltransmissie. Hoofdstuk 3 levert een bijdrage aan het beantwoorden van de fundamentele vraag hoe ‘besmettelijke’ cultuur verklaard kan worden<sup>5</sup> door te onderzoeken welke type(s) verhalen beklijven en welke *verhaalinhoudelijke* factoren hierop van invloed zijn (onderzoeksvraag 2). De resultaten laten zien dat de (im)populariteit van sprookjes in sterke mate beïnvloed wordt door het type personages dat in de verhalen voorkomt. Zo hebben populaire, beklijvende verhalen een significante voorkeur voor personages uit de familiesfeer en personages die ‘minimaal tegenintuïtief’ genoemd kunnen worden.<sup>6</sup> Impopulaire verhalen, daarentegen, kenmerken zich door bijvoorbeeld het opvoeren van generieke (vaak negatief geëvalueerde) personagegroepen en

3. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

4. Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*.

5. Vergelijk: Sperber, *Explaining Culture. A Naturalistic Approach*.

6. Zie bijvoorbeeld Barrett en andere, “Counterintuitiveness in Folktales: Finding the Cognitive Optimum”.

personages uit de religieuze sfeer.

Na dit hoofdstuk over verhaalinhoudelijke factoren van (on)succesvolle verhaaltransmissie volgen twee hoofdstukken over sociaal-geïnformeerde, *context-gebaseerde* processen die ten grondslag liggen aan verhaaltransmissie. Het doel van hoofdstuk 4 is om een beter begrip te ontwikkelen over de processen waarmee kinderverhalen worden herverteld. Het hoofdstuk probeert een antwoord te geven op de derde onderzoeksvraag van deze dissertatie en onderzoekt of de diachrone ontwikkeling van “Roodkapje” in het Nederlandse taalgebied gekarakteriseerd kan worden als een (Darwiniaans) evolutieproces, waarin aanpassingen van het verhaal geleidelijk opeengestapeld worden en nieuwe varianten van het verhaal afgeleid zijn van versies uit het recente verleden. Uit de resultaten blijkt dat nieuwe Nederlandse hervertellingen van “Roodkapje” inderdaad onderworpen zijn aan leeftijdsafhankelijke selectieprocessen (vergelijkbaar met modetrends). Dat betekent ‘jonge’ vertellingen de voorkeur krijgen om een nieuwe vertelling op te baseren.

Hoewel de leeftijdsafhankelijke selectiemechanismen voor “Roodkapje” belangrijke inzichten toevoegen aan onze kennis over verhaaltransmissie, biedt het nog geen verklaring voor de vraag waarom specifieke verhaalversies worden verkozen boven andere versies van ongeveer dezelfde leeftijd om een nieuwe hervertelling op te baseren. Om deze vraag te onderzoeken, neemt hoofdstuk 5 de leeftijdsafhankelijke selectievoorkeur verder onder de loep en onderzoekt daarnaast welke andere selectievoorkeuren een rol spelen in verhaaltransmissie (onderzoeksvraag 4). De resultaten van een systematische vergelijking tussen computationele simulaties met empirische observaties van verhaaltransmissie laten zien dat de differentiële selectie van verhalen sterk beïnvloed wordt door de interactie van een drietal factoren: (i) een frequentievoorkeur (waarbij hervertellers kiezen voor populaire verhalen), (ii) een ‘modelvoorkeur’ (waarbij hervertellers zich laten leiden door bijvoorbeeld de prestige van een auteur of uitgave) en (iii) de eerder besproken leeftijdsvoorkeur.





---

## SUMMARY

*Retelling Stories* is a study about the mechanisms underlying story transmission and selection. A central question of the study is whether story transmission can be understood as – and hence should be described as – a cultural evolutionary process, and which evolutionary mechanisms can be identified in story transmission. To address this question, the approach presented in this study draws inspiration from research on computational models of cultural evolution.<sup>1</sup> By positioning itself explicitly and extensively in dialog with this cultural evolution research program, the current study aims to advance our understanding of story transmission processes by formally and quantitatively characterizing diachronic story developments. While its central focus is to provide a computational-evolutionary approach to story transmission (and encompassing methodological challenges), the study also builds on insights gained from both folkloristic and literary accounts of story transmission.<sup>2</sup> As such, this study contributes to the further synthesis of the disciplines of cultural evolution, literary theory and folklore. After presenting the methodological framework and problem statement in Chapter 1, the subsequent four chapters address a number of research questions regarding both methodological and theoretical issues of story transmission.

The first research question constitutes the methodological base of the study and addresses the question of how stories should be formally represented in order to study real-world story transmission with com-

1. E.g. Cavalli-Sforza and Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*; Boyd and Richerson, *Culture and the Evolutionary Process*; Mesoudi, *Cultural Evolution. How Darwinian Theory can Explain Human Culture & Synthesize the Social Sciences*; Mesoudi, "Cultural Evolution: A Review of Theory, Findings and Controversies."

2. E.g. Stephens and McCallum, *Retelling Stories, Framing Culture. Traditional Story and Metanarratives in Children's Literature*; Boyd, *On the Origin of Stories. Evolution, Cognition and Fiction*; Boyd et al., *Evolution, Literature, and Film. A Reader*; Zipes, *Why Fairy Tales Stick. The Evolution and Relevance of a Genre*; Zipes, *The Irresistible Fairy Tale. The Cultural and Social History of a Genre*; Geerts and van den Bossche, *Never-ending Stories. Adaptation, Canonisation and Ideology in Children's Literature*.

putational means. Chapters 2 to 5 present four different data representations, each of which has potential (dis)advantages for studying particular aspects of story transmission. In Chapter 2, a 'classical' representation is employed, which is based on the seminal *Motif-Index* developed by Stith Thompson.<sup>3</sup> In this representation, motifs are considered to be the basic building blocks of stories, and their constellation defines the story type of a story (e.g. story types from *The Types of International Folktales*<sup>4</sup>). This story representation, in which stories are primarily considered to be mere instances of story types, contrasts sharply with the data-driven story representations proposed in chapters 3 to 5. These chapters present three different story representations, in which the object of interest is not the overarching story type, but lies primarily with the tale 'token', i.e. the actual story. Crucially, the main methodological and theoretical advantage of these data-driven representations is that what is considered to be a similarity between stories is never static, but depends on the perspective one wishes to take or the image one aims to construct.

Using the exemplar-based story representations, chapters 3 to 5 address a number of questions concerning mechanisms underlying story transmission. Chapter 3 contributes to answering the fundamental question of how to explain 'contagious' culture<sup>5</sup> by investigating which stories 'stick' and what *content*-based factors determine their successfulness (Research question 2). Taking the famous fairy tale collection *Kinder- und Hausmärchen* as a case, empirical evidence was given for the existence of a number of character types that affect the popularity of a story. Successful, sticky stories exhibit a significant preference for, for instance, characters referring to family relationships and 'minimally counterintuitive' agents.<sup>6</sup> Unsuccessful stories, by contrast, typically revolve around religious characters and generic groups with rather negative connotations.

Subsequent to this chapter on content-based factors of (un)successful story transmission, the remaining two chapters (4 and 5) focus on socially

3. Thompson, *Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jestbooks, and Local Legends*.

4. Uther, *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*.

5. Cf. Sperber, *Explaining Culture. A Naturalistic Approach*.

6. See e.g. Barrett et al., "Counterintuitiveness in Folktales: Finding the Cognitive Optimum."

informed, *context*-based processes that underlie story transmission. The goal of Chapter 4 is to enhance our understanding of the processes through which children's stories are retold. The chapter aims to answer the third research question of this thesis and investigates whether the diachronic development of "Red Riding Hood" in the Netherlands can be characterized as an (Darwinian) evolutionary process. It is shown that the evolution of the story can be characterized as a 'gradual accumulation of modification' in which new versions of "Red Riding Hood" tend to modify and adapt prior retellings, and these prior retellings are published in close temporal proximity to these new versions (Research question 3).

The observed disposition of authors for temporally proximate stories to produce new retellings can be interpreted as an 'age bias' in story transmission, comparable to fashion trends or fads. Yet, although the presence of age-dependent selection mechanisms adds important knowledge to our understanding of story transmission processes, it cannot explain why certain story versions of approximately the same age are differentially preferred to function as a story's pre-text. To address this issue, Chapter 5 further scrutinizes the age-dependent selection mechanism while testing and evaluating other biases that could serve as explanatory models for story transmission (Research question 4). On the basis of an extensive and systematic comparison between the output of computational simulations and empirical observations of story transmission, it is shown that the differential selection of stories is affected by the interplay of three biases: (i) a positive frequency bias (in which retellers prefer frequently retold stories), (ii) a model-based bias (e.g. author prestige), and (iii) the aforementioned age bias.



---

## CURRICULUM VITAE

Folgert Karsdorp was born in Leidschendam, the Netherlands, on the 3rd of November, 1983. He studied Dutch Language & Culture at Leiden University for which he received his BA in 2007. Subsequently, he studied the research master Linguistics at the same university and obtained his MA in 2009. After working for a year as a research assistant at the Freie Universität Berlin, Karsdorp worked as a (computational) linguist at the Institute for Dutch Lexicology until December 2011. In 2012, he started working as a Ph.D candidate in the Tunes & Tales project at the Meertens Instituut in Amsterdam, which was funded by the Royal Dutch Academy of Arts and Sciences. In April 2016, Karsdorp acquired a research assistant position at the Centre for Language Studies of the Radboud University in Nijmegen. At present, Karsdorp is employed as a postdoctoral researcher at the Meertens Instituut in Amsterdam.