GRIMMS MÄRCHEN

An Intertextual Puzzle of Fairy Tales

Greta Franzini
February 23, 2016
National University of Ireland, Galway
TABLE OF CONTENTS

1. Introduction

2. Research Focus

3. Digital Humanities

4. Historical text reuse

5. Intertextuality

6. Case study

7. TRACER

8. Conclusion
INTRODUCTION
Electronic Text Reuse Acquisition Project (eTRAP)

Early Career Research Group funded by German Ministry of Education & Research (BMBF).

Budget: €1.7M.


Team: 4 core staff; 9 student assistants.

- Interdisciplinary: Classics, Computer Science, German Studies, Mathematics, Philosophy, Software Engineering.
- International: 7 nationalities, 11 languages spoken.
Professional team coaching for **effective group dynamic**:

- Effective communication;
- Making the most of strengths;
- Effective delegation.
RESEARCH FOCUS
Electronic Text Reuse Acquisition Project (eTRAP)

Text reuse = spoken and written repetition of text across time and space.

**Figure 1:** Text reuse types [Author: Marco Büchler].
Current limitations:

- Detecting text reuse across languages;
- Detecting looser forms of text reuse, e.g. allusion;
- Historical texts: language evolution, copy errors, etc.

Specific interests: text reuse detection **at scale** (Big Data) and **historical** text reuse.
DIGITAL HUMANITIES
Ulrike Rieß (*Big Data bestimmt die IT-Welt*):

- **Large amounts** of data that can’t be processed and analysed manually;
- **Less structured** data, e.g. in comparison to databases and data warehouse systems;
- Link[ed data between heterogeneous and distributed resources.**

**Information overload** = large amounts of data (Big Data).

**Information poverty** = noisy, missing, fragmentary, oral data (Humanities Data).

**COMPLEXITY**
Inconsistent survival, documentation, cataloguing.
HISTORICAL TEXT REUSE
Historical text reuse as an **opportunity** to tackle the complexity:

**Humanities**

- Lines of transmissions;
- Transmissions of ideas/thoughts under different circumstances and conditions.

**Computer Science**

- Text decontamination for stylometry and authorship attribution;
- Dating of texts;
- Text Mining, Corpus Linguistics.
"The fundamental methodological fact that historical linguists have to face is that they have no control over their data... The great art of the historical linguist is to make the best of this bad data - 'bad' in the sense that it may be fragmentary, corrupted or many times removed from the actual productions of native speakers.” (Labov, 1972, p. 100)
INTERTEXTUALITY
Mass digitisation is...

- driving the improvement of close reading methods;
- providing new opportunities for distant reading and for text reuse techniques to automatically find data parallels in large textual collections, i.e. intertextuality.

"[...] a text is [...] a multidimensional space in which a variety of writings, none of them original, blend and clash. The text is a tissue of quotations drawn from the innumerable centres of culture... the writer can only imitate a gesture that is always anterior, never original. His only power is to mix writings [...].” (Barthes, 1977, pp. 146-47)

"[...] any text is constructed as a mosaic of quotations [...].” (Kristeva, 1980, p.66)
eTRAP’s Aristotelian approach to intertextuality:

”The whole is greater than the sum of its parts”

Sources and socio-historical context or influencing factors (Gerard Genette’s epitext).
CASE STUDY
Seven editions of *Kinder- und Hausmärchen*: 1812, 1819, 1837, 1840, 1843, 1850, 1857.

Changes in:

- **Size**: from 156 to 211.
- **Content**: gruesome to mild.
- **Style**: Jacob scholarly, Wilhelm figurative.
- **Language**: Variants, diachronic evolution.
Motivation:

- Impact on society
- Global scope
- Big Data
- Interdisciplinary
Two avenues of enquiry:

1. Intertextual relations between Grimm collection and other tales;
2. Intertextual relations between Grimm collection and Grimm network.
Motif Database: to compare *Kinder- und Hausmärchen* to other tales by investigating measurable primitives:

- **Literature**: tracing MOTIFS
- **Cultural Studies**: tracing MEMES
- **Linguistics**: tracing PATTERNS
- **Computer Science**: tracing MINUTIAE
- **Forensics**: tracing FINGERPRINTS

**Motif**: ”[...] minimal thematic unit” (Prince’s *Dictionary of Narratology*)
Why build the database?

- Investigate & record primitives and their changes;
- Nothing like it exists;
- Advance research in folkloristics;
- Improve algorithms to sharpen our understanding of why and how a text is reused.
**Selection:** *Snow White, Puss in Boots, The Fisherman and his Wife.*

Two stages:

1. **Manual**
   - Collection of motifs as text reuse training data;
   - Creation of an ontology for RDF data representation.

2. **(Semi-)Automatic**
   - Detection of motifs in smaller corpora: TRACER;
   - Detection of motifs in larger corpora: Google Search & Books APIs.
Collection of motifs as text reuse training data


<table>
<thead>
<tr>
<th>A. MYTHOLOGICAL MOTIFS</th>
<th>DEU</th>
<th>RUS</th>
<th>ITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATU:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT709. Snow White</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1. MAGIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D006-D099. Ownership of magic objects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D01. Ownership of magic object</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D900-D1299. Kind of magic objects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1163. Magic mirror</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1300-D1599. Function of magic objects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1310. Magic object gives supernatural information</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1311. Magic object used for divination</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1311.2. Mirror answers questions</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1600-D1699. Characteristics of magic objects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1610. Magic speaking objects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D1610.2. Talking mirror</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>E. THE DEAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E0-E196. Resuscitation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E10. Resuscitation by rough treatment</td>
<td>x</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>E21. Resuscitation by withdrawal of wounding instrument</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

**Figure 2:** Snow White motif collection file.
Creation of an ontology for RDF (Resource Description Framework) data representation

RDF Graph Database = VIRTUOSO
Query Language = SPARQL (RDF query language)
**Output:** a searchable database crossing the language barrier.

**Figure 3:** Mockup of user interface.
• Detection of motifs in **smaller** corpora: TRACER;
• Detection of motifs in **larger** corpora: Google Search & Books APIs.
TRACER
TRACER: suite of **700 algorithms**; developed by Marco Büchler.

**Figure 4:** TRACER steps. More than 1M permutations of implementations of different levels are possible.
Figure 5: TRAViz display of the 7 Grimm variants of an *Aschenputtel* (= Cinderella) extract.

Text normalisation.

Two avenues of enquiry:

1. Intertextual relations between Grimm collection and other tales;
2. Intertextual relations between Grimm collection and Grimm network.

Figure 6: SPin’s Grimm Cluster project: http://ernie.spinnet.eu/viewer
INTERTEXTUALITY: CONTEXT

Figure 7: Grimm’s book collection.

Figure 8: 36,000 Grimm letters.
CONCLUSION
CONCLUSION

- Contribution
  - Ontology of motifs
  - Motifs to Aarne-Thompson Motif-Index

- Next steps: short-term
  - Finish collecting motifs
  - Build the ontology

- Feedback
  - DH 2016 Kraków
  - You!
Author
Greta Franzini
✉️ gfranzini@gcdh.de

Grimm Team (in alphabetical order)
Marco Büchler, Emily Franzini, Greta Franzini, Franziska Pannach, Gabriela Rotari, Christian Würker

Visit us
🌐 http://etrap.gcdh.de

"Copying from one is plagiarism, copying from many is research"
-Wilson Mitzner
Thank you!


APPENDICES
Greta Franzini
Executive Board Member & Journal Associate Editor of *Digital Medievalist*.

- Submit to or review for the [Journal](http://www.digitalmedievalist.org/journal/)
- Follow the [website](https://digitalmedievalist.wordpress.com/)
- Join the [Facebook group](https://goo.gl/Jr7EIG)

Questions & comments to: [gfranzini@gcdh.de](mailto:gfranzini@gcdh.de)
All images used in this presentation are in the public domain.
The LaTeX theme this presentation is based on is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Changes to the theme are the work of eTRAP.