ENTERING THE THICK FOREST OF INTERCULTURAL TRANSMISSION OF MOTIFS

Emily Franzini & Marco Büchler
University of Geneva, January 2016
Kinder- und Hausmärchen
1812, 1819, 1837, 1840, 1843, 1850, 1857

Deutsches Wörterbuch
1854

37,000 letters
Marburg Staatsarchiv & Humboldt-Universität zu Berlin
Why the fairy tales of the Brothers Grimm?

- Impact on society
- Big Data
- Global scope needs an international team (8 nationalities, 13 languages spoken)
- Interdisciplinary
What are we working on?

Investigating measurable primitives

- Cultural Studies: tracing **MEMES**
- Literature: tracing **MOTIFS**
- Linguistics: tracing **PATTERNS**
- Computer Science: tracing **MINUTIAE**
Our Projects

DATABASE OF TALE MOTIFS

TEXT REUSE ACROSS GRIMM TALES EDITIONS
Objective = build a database + interface of the motifs of tales, crossing the language barrier

Why?

- Investigate & record primitives and their changes
- Nothing like it exists
- Humanities: research in folklore
- Computer Science: algorithmic improvements
  - sharpening the understanding of why and how a text is reused

Motif = “...minimal thematic unit[s]” (Prince’s Dictionary of Narratology)
Database of Tale Motifs

Motif = “...minimal thematic unit[s]”
(Prince’s Dictionary of Narratology)

Snow White
**Which tales?**

- Snow White
- Puss in Boots
- The Fisherman and his Wife

Languages:
- Italian
- German
- Russian
- French
- Iranian
- English
- Rumanian
- Spanish
Motif = “…minimal thematic unit[s]” (Prince's Dictionary of Narratology)

Aarne-Thompson (AT) Motif-Index

- A. Mythological Motifs
- B. Animal Motifs
- C. Motifs of Tabu
- D. Magic
- E. the Dead
- F. Marvels
- G. Ogres
- H. Tests
- J. the Wise and the Foolish
- K. Deceptions
- L. Reversals of Fortune
- M. Ordaining the Future
- N. Chance and Fate
- P. Society
- Q. Rewards and Punishments
- R. Captives and Fugitives
- S. Unnatural Cruelty
- T. Sex
- U. the Nature of Life
- V. Religion
- W. Traits of Character
- X. Humor
- Z. Miscellaneous Groups of Motifs
**Motif** = “…minimal thematic unit[s]”  
(Prince’s Dictionary of Narratology)
### S. UNNATURAL CRUELTY

#### S0-S99. Cruel relatives

<table>
<thead>
<tr>
<th>S10. Cruel parents</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>null</th>
<th>null</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>S12. Cruel mother</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S12.2. Cruel mother kills child</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S12.2.1. Cruel mother plans death of daughter</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S30. Cruel step- and foster relatives</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S31. Cruel stepmother</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S31.1. Cruel stepmother kills child</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S31.1.1. Cruel stepmother plans death of stepdaughter</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
</tbody>
</table>

#### S100-S199. Revolting murders or mutilations

<table>
<thead>
<tr>
<th>S111. Murder by poisoning</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>null</th>
<th>null</th>
<th>null</th>
</tr>
</thead>
<tbody>
<tr>
<td>S111.3. Murder with poisoned comb</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>S111.4. Murder with poisoned apple</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>S111.4.1. The poisoned apple was rosy-cheeked</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>S111.4.2. The poisoned apple was thrown to the princess</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>S111.4.3. The princess eats a piece of the poisoned apple</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S115. Murder by stabbing</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>x</td>
</tr>
<tr>
<td>S115.2. Murder by sticking needle through head</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
</tbody>
</table>

#### S199. Murder by poisoning

<table>
<thead>
<tr>
<th>S119. Murder by poisoning</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>null</th>
<th>null</th>
<th>null</th>
</tr>
</thead>
<tbody>
<tr>
<td>S119.1. Murder by poisoning</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
<tr>
<td>S300-S399. Abnormal suicides</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
<td>null</td>
</tr>
</tbody>
</table>

#### S322. Children

<table>
<thead>
<tr>
<th>S322.2. Jinni</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>null</th>
</tr>
</thead>
<tbody>
<tr>
<td>S322.4. The anime character uses magic to animate</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>S322.4.2. Evil stepmother orders stepdaughter to be killed</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>S322.4.4. The cruel stepmother kills her stepdaughter</td>
<td>null</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

---

**Aarne-Thompson: 709**

- D1311. Magic object used for divination
  - x | x | x | x | x | x | x | null | null |

- D1311.2. Mirror answers questions
  - x | x | x | x | x | x | x | x | null | null |

---

**Snow White**

- Sum = 0
GRAPH DATABASE (RDF) = VIRTUOSO
QUERY LANGUAGE = SPARQL
Results: 7 Results

Tale: Bella Venezia
Author: Calvino, Italo
Collection: Fiabe Italiane
Date: 1956
VIAF: 181208131
Motif: **Stepmother** (P282)
Url: www.text......

Tale: Schneewittchen
Author: Gebrüder Grimm
Collection: Kinder- und Hausmärchen
Date: 1819
VIAF: 187449723
Motif: **Stepmother** (P282)
Url: www.text......

Tale: Schneewittchen
Author: Gebrüder Grimm
Collection: Kinder- und Hausmärchen
Date: 1837
VIAF: 187449723
Motif: **Stepmother** (P282)
Url: www.text......
What are we working on?

Investigating measurable primitives

- Cultural Studies: tracing MEMES
- Literature: tracing MOTIFS
- Linguistics: tracing PATTERNS
- Computer Science: tracing MINUTIAE
Typical computer scientists’ expectation: plagiarism
Humanists' expectation: **oversimplification**
**ACID** for the Digital Humanities:

- Acceptance
- Complexity
- Interoperability
- Diversity
ACID for the Digital Humanities - Diversity (Reuse Types)

Stability

Purpose

Size of text reuse

Classification

Degree of distribution

Written and oral transmission
ACID for the Digital Humanities - Diversity (Reuse Types)

- Parallel texts
  - Syntactic Text Re-use
    - Idiomatic Text Re-use
      - Idiom
      - Winged Word
      - Verbatim
      - Near Verbatim
    - Quotation
    - Edition
    - Allusion
  - Semantical Text Re-use
    - Paraphrasing
      - Paraphrase
      - Analogy
      - Translation
    - Ghostwriting
    - Summarizing
TRACER

6 steps

Pre-Processing
- Lemma
- Case
- Diacritics
- Synonyms

Training
- String-Features
- Syntactic Features
- Semantic Features

Selection
- 0 mod p
- Winnowing
- Re-Use-Fingerprint

Linking
- Dice
- Containment

Scoring

Post-Processing
- Intention Detection

DB

TEXT REUSE ACROSS GRIMM TALES EDITIONS
ACID for the Digital Humanities - Complexity

Text Re-use Across Grimm Tales Editions
### ACID for the Digital Humanities - Interoperability

<table>
<thead>
<tr>
<th>cit-quote-bibl</th>
<th>blockquote</th>
<th>bibl without quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;cit&gt;</code></td>
<td><code>&lt;quote rend=&quot;blockquote&quot;&gt;</code></td>
<td><code>&lt;p&gt;</code></td>
</tr>
<tr>
<td><code>&lt;quote&gt;</code></td>
<td><code>&lt;line&gt;</code></td>
<td>`[...]a)nti\tou= proe/pinon. kuri/ws ga/r e)sti tou=to propi/nein, to\ e(te/rw) pro\ e(autou= dou=nai piei=n. kai ( *)odusseu's de\ para\ tw=</td>
</tr>
<tr>
<td>du/o ku/nes a)rgoi\ ei/ponto</td>
<td><code>&lt;bibl n=&quot;Hom. Il. 4.92&quot;&gt;Il. 4.92&lt;/bibl&gt;</code></td>
<td><code>&lt;bibl n=&quot;Hom. Od. 13.57&quot;&gt;Od. 13.57&lt;/bibl&gt;</code></td>
</tr>
<tr>
<td><code>&lt;bibl n=&quot;Hom. Od. 2.11&quot;&gt;Od. 2.11</code></td>
<td><code>&lt;line&gt;</code></td>
<td><code>[...]</code></td>
</tr>
<tr>
<td><code>&lt;/bibl&gt;</code></td>
<td><code>&lt;line&gt;</code></td>
<td><code>&lt;/p&gt;</code></td>
</tr>
<tr>
<td><code>&lt;/cit&gt;</code></td>
<td><code>&lt;quote&gt;</code></td>
<td></td>
</tr>
</tbody>
</table>
ACID for the Digital Humanities - Acceptance I
ACID for the Digital Humanities - **Acceptance II**

Text Mining

How to be accepted by humanists if text mining is a **black box**, that can't be looked into?
ACID for the Digital Humanities - **Acceptance III**

**Transparency:** How to provide user-friendly insights into complex mining techniques and machine learning?
Step 0: Searching

In the beginning God created the heavens and the earth.
And the earth was waste and void; and darkness was upon the face of the deep: and the Spirit of God moved upon the face of the waters.
And God said, Let there be light: and there was light.
And God saw the light, that it was good: and God divided the light from the darkness.
And God called the light Day, and the darkness he called Night. And there was evening and there was morning, one day.
And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters.
And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so.
And God called the firmament Heaven. And there was evening and there was morning, a second day.
And God said, Let the waters under the heavens be gathered together unto one place, and let the dry land appear: and it was so.
And God called the dry land Earth: and the gathering together of the waters called he Seas: and God saw that it was good.
And God said, Let the earth put forth grass, herbs yielding seed, and fruit-trees bearing fruit after their kind, wherein is the seed thereof, upon the earth: and it was so.
And the earth brought forth grass, herbs yielding seed after their kind, and trees bearing fruit, wherein is the seed thereof, after their kind: and God saw that it was good.
And God said, Let there be lights in the firmament of heaven to divide the day from the night; and let them be for signs, and for seasons, and for days and years:
And God made the two great lights; the greater light to rule the day, and the lesser light to rule the night: he made the stars also.
And God set them in the firmament of heaven to give light upon the earth,
and to rule over the day and over the night, and to divide the light from the darkness: and God saw that it was good.
And God said, Let the waters swarm with swarms of living creatures, and let birds fly above the earth in the open firmament of heaven.
And God created the great sea-monsters, and every living creature that moveth, wherewith the waters swarmed, after their kind, and every winged bird after its kind: and God saw that it was good.
And God blessed them, saying, Be fruitful, and multiply, and fill the waters in the seas, and let birds multiply on the earth.
And God said, Let the earth bring forth living creatures after their kind, cattle, and creeping things, and beasts of the earth after their kind: and it was so.
Step 0: Searching

Step 1: Preprocessing

Please select a preprocessing strategy:

Unprocessed Sentence: In the beginning God created the heavens and the earth.
Preprocessed Sentence: in the begin god create the heaven and the earth.

Your correction for the processed sentence:

Your comment:

Other users preference
No users have suggested a change in the preprocessing level

next Level
ACID for the Digital Humanities - **Acceptance VI**

### Step 0: Searching

### Step 1: Preprocessing

### Step 2: Featuring

Preprocessed sentence: in the begin god create the heaven and the earth.

<table>
<thead>
<tr>
<th>Position</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>in the</td>
</tr>
<tr>
<td>1</td>
<td>the begin</td>
</tr>
<tr>
<td>2</td>
<td>begin god</td>
</tr>
<tr>
<td>3</td>
<td>god create</td>
</tr>
<tr>
<td>4</td>
<td>create the</td>
</tr>
<tr>
<td>5</td>
<td>the heaven</td>
</tr>
<tr>
<td>6</td>
<td>heaven and</td>
</tr>
<tr>
<td>7</td>
<td>and the</td>
</tr>
<tr>
<td>8</td>
<td>the earth</td>
</tr>
<tr>
<td>9</td>
<td>earth</td>
</tr>
</tbody>
</table>
ACID for the Digital Humanities - **Acceptance VII**

### Step 3: Selecting

Please select a seeding strategy: Local Max Feature Frequency Selector: FreqSel=0.8

**Agenda**
- word: This word belongs to the fingerprint
- word: This word originally doesn't belong to the fingerprint but was selected by the user to belong to the fingerprint
- word: This word originally did not belong to the fingerprint

**Initial configuration:** In the beg the begin god heard create the the heaven and the the earth earth

**Current configuration:** In the the begin god heard create the the heaven and the the earth earth

### Selected features

- **begin**
- **god**
- **create**
- **heaven**
- **earth**

### Not selected features

- **in**
- **the**
- **and**

### Other users preference

<table>
<thead>
<tr>
<th>Feature</th>
<th>Users selected</th>
<th>Users not selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>in the</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>the begin</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>god create</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>create the</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>the heaven</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>heaven and</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>and the</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>the earth</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>earth</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Statistics

<table>
<thead>
<tr>
<th>Feature</th>
<th>Selected Features</th>
<th>Total number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the</td>
<td>27114</td>
<td>32227</td>
</tr>
<tr>
<td>the begin</td>
<td>470</td>
<td>400</td>
</tr>
<tr>
<td>god create</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>create the</td>
<td>27</td>
<td>45</td>
</tr>
<tr>
<td>the heaven</td>
<td>1624</td>
<td>1695</td>
</tr>
<tr>
<td>heaven and</td>
<td>389</td>
<td>380</td>
</tr>
<tr>
<td>and the</td>
<td>31800</td>
<td>40850</td>
</tr>
<tr>
<td>the earth</td>
<td>4775</td>
<td>5222</td>
</tr>
<tr>
<td>earth</td>
<td>1930</td>
<td>1040</td>
</tr>
</tbody>
</table>

next level  

submit changes
Further topics

- Evaluation of text reuse: manual or automatic
- Research on the measurable primitives including eye-tracking and electroencephalogram
- Scaling Text Reuse at any size
http://etrap.gcdh.de/

Copying from one is plagiarism, copying from many…. is research.

-Wilson Mitzner-