Folktale similarity based on ontological abstraction

Marijn Schraagen

Digital Humanities Lab
Utrecht University, The Netherlands

Global WordNet Conference
January 30, 2016
Research task

- Compute pair-wise similarity of folktale texts using WordNet
Research task

- Compute pair-wise similarity of folktale texts using WordNet
- Capture common elements in actors and events at an abstract level
Research task

- Compute pair-wise similarity of folktale texts using WordNet
- Capture common elements in actors and events at an abstract level
- Complement existing folktale classification standards
Data

- Data taken from Dutch Folktale Database
  - http://www.verhalenbank.nl/, in Dutch

"Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, h` e? Ze heette ook niet echt Roodkapje."

Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.
Data

- Data taken from Dutch Folktale Database
  - [http://www.verhalenbank.nl/](http://www.verhalenbank.nl/), in Dutch
- Subcorpus for proof of concept
Data taken from Dutch Folktale Database
- http://www.verhalenbank.nl/, in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words
Data

- Data taken from Dutch Folktale Database
  - [http://www.verhalenbank.nl/](http://www.verhalenbank.nl/), in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words
- Grammatically correct, modern Dutch
Data

- Data taken from Dutch Folktale Database
  - [http://www.verhalenbank.nl/](http://www.verhalenbank.nl/), in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words
- Grammatically correct, modern Dutch
- “Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje.”
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.*
Preprocessing

- Preprocessing using Frog

Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn't it? She was not actually called Little Red Riding Hood.
Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging

Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.
Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs

Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn't it? She was not actually called Little Red Riding Hood.
Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs
- “Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje.”
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.*
Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs
- “Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje.”
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.*
- klein meisje Roodkapje heten. gek naam hè. heten echt Roodkapje
- *little girl Little_Red call. strange name eh. call really Little_Red.*

Universiteit Utrecht
Similarity computation

- Count number of matching terms
Similarity computation

- Count number of matching terms
- Sentence level comparison
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets
- Directed similarity score for each sentence
Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets
- Directed similarity score for each sentence
- First synset used
Similarity computation example

Good day madam, said the princess, what are you doing? Good afternoon basket maker, said the gnome.

source  Good day  madam  said  the  princess  what  does  you  there
Good day madam, said the princess, what are you doing? 
Good afternoon basket maker, said the gnome.

source

Good day  madam  said  the  princess  what  does  you  there

lemma
day  madam  speak  princess  do
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

<table>
<thead>
<tr>
<th>source</th>
<th>Good day</th>
<th>madam</th>
<th>said</th>
<th>the</th>
<th>princess</th>
<th>what</th>
<th>does</th>
<th>you</th>
<th>there</th>
</tr>
</thead>
<tbody>
<tr>
<td>lemma</td>
<td>day</td>
<td>madam</td>
<td>speak</td>
<td>princess</td>
<td>do</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>synset</td>
<td>day</td>
<td>lady</td>
<td>speak</td>
<td>royal daughter</td>
<td>do</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do

Good afternoon basket maker said the gnome
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do
good afternoon basket maker speak gnome
time unit
Similarity computation example

- Good day madam, said the princess, what are you doing?
- Good afternoon basket maker, said the gnome.

Words: day, lady, speak, royal daughter, do, good afternoon, basket maker, speak, gnome, time unit, unit.
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day  lady  speak  royal daughter  do
time unit
unit
something
good afternoon  basket maker  speak  gnome
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
(\frac{1}{4})
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

((1/4)

day lady speak royal daughter do
good afternoon basket maker speak gnome figure
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

((1/4)

day lady speak royal daughter do
figure

good afternoon basket maker speak gnome
maker
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\begin{array}{cccc}
\text{day} & \text{lady} & \text{speak} & \text{royal daughter} & \text{do} \\
\text{figure} & \\
\hline
\text{good afternoon} & \text{basket maker} & \text{speak} & \text{gnome} & \\
& \text{maker} & \\
& \text{figure} & \\
\end{array}
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$\left(\frac{1}{4}\right)$
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left( \frac{1}{4} + \frac{1}{2} \right)
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[ \left( \frac{1}{4} + \frac{1}{2} \right) \]
Similarity computation example

Good day madam, said the princess, what are you doing? 
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\frac{1}{4} + \frac{1}{2} + 1
\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
((\frac{1}{4} + \frac{1}{2} + 1)
\]

- day
- lady
- speak
- royal daughter
- do

- good afternoon
- basket maker
- speak
- gnome

- daughter
- child
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left( \frac{1}{4} + \frac{1}{2} + 1 \right)
\]

day lady speak royal daughter do
good afternoon basket maker speak gnome
daughter child relative
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\begin{align*}
\frac{1}{4} + \frac{1}{2} + 1
\end{align*}
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{4} + \frac{1}{2} + 1 \right)
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[ \left( \frac{1}{4} + \frac{1}{2} + 1 \right] 

day lady speak royal daughter do
daughter child relative member figure
good afternoon basket maker speak gnome maker
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$\left(\frac{1}{4} + \frac{1}{2} + 1\right)$$
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[ \frac{1}{4} + \frac{1}{2} + 1 \]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6})\]

day lady speak royal daughter do
good afternoon basket maker speak gnome
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right)
\]

day lady speak royal daughter do

good afternoon basket maker speak gnome

act
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\begin{aligned}
&\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right) \\
&\text{day} \quad \text{lady} \quad \text{speak} \quad \text{royal daughter} \quad \text{do} \\
&\text{act} \\
&\text{good afternoon} \quad \text{basket maker} \quad \text{speak} \quad \text{gnome} \\
&\text{notify}
\end{aligned}
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right)$

day lady speak royal daughter do

act

good afternoon basket maker speak gnome
notify inform
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right)
\]

day lady speak royal daughter do

good afternoon basket maker speak gnome

act

notify

inform
do
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right)\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}\right)$
Similarity computation example

<table>
<thead>
<tr>
<th>day</th>
<th>lady</th>
<th>speak</th>
<th>royal daughter</th>
<th>do</th>
</tr>
</thead>
<tbody>
<tr>
<td>good afternoon</td>
<td>basket maker</td>
<td>speak</td>
<td>gnome</td>
<td></td>
</tr>
</tbody>
</table>

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left( \frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2} \right)
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2})
\]

**day** lady speak royal daughter do

good afternoon basket maker speak gnome

*no match*
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + 0\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0)\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0)
\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.
\[
\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0
\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0)
\]

Universiteit Utrecht
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0 + \frac{1}{3})$$

day lady speak royal daughter do
good afternoon basket maker speak gnome
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3})\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3}) + 1
\]

\begin{tabular}{llll}
  day & lady & speak & royal daughter & do \\
  good afternoon & basket maker & speak & gnome \\
\end{tabular}
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{2} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0 + \frac{1}{3} + 1)
\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0 + \frac{1}{3} + 1)$$

day lady speak royal daughter do
good afternoon basket maker speak gnome
daughter creature
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)
\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)\)
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0 + \frac{1}{3} + 1)$
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)
\]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[ \left( \frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2} \right) + (0 + \frac{1}{3} + 1) \]
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[(\frac{1}{2} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1)\]
Good day madam, said the princess, what are you doing?

Good afternoon basket maker, said the gnome.

\[ \left( \frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2} \right) + \left( 0 + \frac{1}{3} + 1 \right) \]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + (0 + \frac{1}{3} + 1)
\]
Good day madam, said the princess, what are you doing?  
Good afternoon basket maker, said the gnome.  
\[(\frac{1}{4} + \frac{1}{2} + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1 + \frac{1}{2})\]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\frac{\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + \left(0 + \frac{1}{3} + 1 + \frac{1}{2}\right)}{5}
\]
Good day madam, said the princess, what are you doing? 
Good afternoon basket maker, said the gnome. 
\[ \frac{\left( \frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2} \right) + \left( 0 + \frac{1}{3} + 1 + \frac{1}{2} \right)}{5 + 4} \]
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

\[
\frac{\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}\right) + \left(0 + \frac{1}{3} + 1 + \frac{1}{2}\right)}{5 + 4} = 0.47
\]
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
- Ranking-based method
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair \((A,B)\) the (relative) amount of sentences from \(A\) for which the most similar sentence was found in \(B\)
- Ranking-based method
- Non-symmetrical
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
- Ranking-based method
- Non-symmetrical
- Clusters based on similarity thresholds
Central nodes and clusters visible

Royal protagonists, moral values vs. civilian protagonists, dangerous circumstances

Universiteit Utrecht
Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora
Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora
- Folktales: approaches for story variants
  - Evaluation using variant-tagged folktale corpora
Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora
- Folktales: approaches for story variants
  - Evaluation using variant-tagged folktale corpora
- Current approach: pair-wise document similarity
  - Evaluation less straightforward
Differences with other approaches

- WordNet graph measures
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge
- Current measure: length from current node to first shared node
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge

- Current measure: length from current node to first shared node

- Intended as measure of actor/event relatedness at some level of abstraction, instead of similarity
Evaluation

- Same method, no WordNet, lemma’s only

- Clustering and central nodes less apparent
## Evaluation

- **Similarity measure vs. human ratings**

<table>
<thead>
<tr>
<th>scored term</th>
<th>McNo</th>
<th>McRel</th>
<th>McSim</th>
<th>RgNo</th>
<th>RgRel</th>
<th>RgSim</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>0.64</td>
<td>0.60</td>
<td>0.64</td>
<td>0.54</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>target</td>
<td>0.44</td>
<td>0.39</td>
<td>0.49</td>
<td>0.53</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>lowest</td>
<td>0.59</td>
<td>0.54</td>
<td>0.63</td>
<td>0.53</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>average</td>
<td>0.62</td>
<td>0.56</td>
<td>0.65</td>
<td>0.58</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>highest</td>
<td>0.58</td>
<td>0.53</td>
<td>0.61</td>
<td>0.58</td>
<td>0.54</td>
<td>0.59</td>
</tr>
</tbody>
</table>

- Miller & Charles, Rubenstein & Goodenough word pairs
- No instruction, report similarity, report relatedness
Evaluation

- Similarity measure vs. human ratings

<table>
<thead>
<tr>
<th>scored term</th>
<th>McNo</th>
<th>McRel</th>
<th>McSim</th>
<th>RgNo</th>
<th>RgRel</th>
<th>RgSim</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>0.64</td>
<td>0.60</td>
<td>0.64</td>
<td>0.54</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>target</td>
<td>0.44</td>
<td>0.39</td>
<td>0.49</td>
<td>0.53</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>lowest</td>
<td>0.59</td>
<td>0.54</td>
<td>0.63</td>
<td>0.53</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>average</td>
<td>0.62</td>
<td>0.56</td>
<td>0.65</td>
<td>0.58</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>highest</td>
<td>0.58</td>
<td>0.53</td>
<td>0.61</td>
<td>0.58</td>
<td>0.54</td>
<td>0.59</td>
</tr>
</tbody>
</table>

- Miller & Charles, Rubenstein & Goodenough word pairs
- No instruction, report similarity, report relatedness
- Correlations lower than Postma & Vossen (2014), around 0.8
Evaluation

Similarity measure vs. human ratings

<table>
<thead>
<tr>
<th>scored term</th>
<th>McNo</th>
<th>McRel</th>
<th>McSim</th>
<th>RgNo</th>
<th>RgRel</th>
<th>RgSim</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>0.64</td>
<td>0.60</td>
<td>0.64</td>
<td>0.54</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>target</td>
<td>0.44</td>
<td>0.39</td>
<td>0.49</td>
<td>0.53</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>lowest</td>
<td>0.59</td>
<td>0.54</td>
<td>0.63</td>
<td>0.53</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>average</td>
<td>0.62</td>
<td>0.56</td>
<td>0.65</td>
<td>0.58</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>highest</td>
<td>0.58</td>
<td>0.53</td>
<td>0.61</td>
<td>0.58</td>
<td>0.54</td>
<td>0.59</td>
</tr>
</tbody>
</table>

- Miller & Charles, Rubenstei & Goodenough word pairs
- No instruction, report similarity, report relatedness
- Correlations lower than Postma & Vossen (2014), around 0.8
- Different type of similarity measure
Comparison with Thompson Motif Index

Limited number of (semi-)abstract story elements for many (but not all) folktales

<table>
<thead>
<tr>
<th>ATU</th>
<th>Title</th>
<th>Motif description</th>
<th>Motif code</th>
<th>match level</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>The Wolf &amp; the Seven Kids</td>
<td>Disguise by changing voice</td>
<td>K1832</td>
<td></td>
</tr>
<tr>
<td>333</td>
<td>Little Red Riding Hood</td>
<td>Wolf puts flour on his paw to disguise himself</td>
<td>K1839.1</td>
<td>4</td>
</tr>
<tr>
<td>533</td>
<td>The Speaking Horsehead</td>
<td>Disguise as goose-girl (turkey-girl)</td>
<td>K1816.5</td>
<td>3</td>
</tr>
<tr>
<td>533</td>
<td>The Speaking Horsehead</td>
<td>Imposter forces oath of secrecy</td>
<td>K1933</td>
<td>2</td>
</tr>
<tr>
<td>709</td>
<td>Snow White</td>
<td>Compassionate executioner: substituted heart</td>
<td>K0512.2</td>
<td>1</td>
</tr>
</tbody>
</table>
Evaluation

- Directed motif overlap, 2 most similar documents per node
- Asterisk (*) indicates relation also found by WordNet method
Evaluation

- Directed motif overlap, 2 most similar documents per node
- Asterisk (*) indicates relation also found by WordNet method
- TMI contains different type of relation
Discussion

- Sensible clusters and central nodes found
Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise
Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise
- Use larger and/or more heterogeneous corpus
Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise
- Use larger and/or more heterogeneous corpus
- Address computational efficiency and scalability
Discussion

Questions?