To Exhibit is not to Loiter: A Multilingual, Sense-Disambiguated Wiktionary for Measuring Verb Similarity

Christian M. Meyer

Christian M. Meyer and Iryna Gurevych
24th International Conference on Computational Linguistics
Motivation

“COLING stresses the importance of multilinguality”
(COLING 2012 Submission FAQ)

- According to Internet World Stats, the Web is no longer dominated by the English language\(^1\)
- The digital turn and the advancing globalization raise a strong demand for multilingual applications (MT, CLIR,…)
- These applications require multilingual knowledge
- Where to find such knowledge?

Finding Multilingual Lexical Knowledge

(Machine-readable) Dictionaries?
- Only weakly structured
- Often not freely available

Multilingual Wordnets?
- EuroWordNet, MultiWordNet,… have small coverage
- English-German ILI contains only 16,347 synsets

Wikipedia?
- Information mostly for nouns
- No linguistic knowledge, such as word usage notes

Wiktionary
- Freely available, high coverage, all parts of speech
Contributions

Contribution 1:
Method for disambiguating semantic relations and translations in Wiktionary

Contribution 2:
Multilingual Wiktionary-based resource with disambiguated and newly inferred relations

Contribution 3:
Employing the new resource for monolingual and cross-lingual verb similarity
Disambiguation of Relations

**hang**

*Verb*

*hang* (third-person singular simple present *hangs*, present participle *hanging*, simple past and past participle *hung* or *hanged* (mostly archaic and legal – see *usage*))

7. *(intransitive, informal)* To loiter, hang around, to spend time idly.
   
   Are you busy, or can you hang with me?
   
   I didn’t see anything, officer. I was just hanging.

8. *(transitive)* To exhibit [en object].

9. *(transitive)* To apply [wallpaper or drywall to a wall].
   
   Let’s hang this cute animal design in the nursery.

**Synonyms**

- *(be or remain suspended): be suspended, dangle*
- *(loiter): hang about, hang around, loiter*
- *(exhibit): exhibit, show*
- *(apply [wallpaper to a wall]): put up*

**Translations**

- *(be or remain suspended): be suspended, dangle*
- *(informal: loiter)*
- *(transitive: exhibit [en object])*
Disambiguation of Relations

Automatic Disambiguation

8. (transitive) To exhibit (an object).

exhibit
Verb
(third-person singular simple present exhibits, present participle exhibiting, simple past and past participle exhibited)

1. (transitive) To display or show (something) for others to see, especially at an exhibition or contest.
   He wanted to exhibit his baseball cards.

2. (transitive) To demonstrate.
   The players exhibited great skill.

ausstellen
Verb, transitiv, regelmäßig
Worttrennung:

Bedeutungen:
[1] zur Schau stellen ✓
[2] Bescheinigung etc. ausfertigen ✗
[3] abschalten ✗

Herkunft:
abtrennbares Präfix aus- und Verb stellen

Synonyme:
[1] zeigen
Why Disambiguation Matters

(to) exhibit  (to) hang  (to) loiter

pretty similar?

shortest path contains only two edges
Why Disambiguation Matters

(to) exhibit₁ → (to) hang₇ → (to) loiter₁
(to) hang₈ → (to) hang₇

not similar!

http://www.flickr.com/photos/ragnarokr/4606143238/sizes/o/in/photostream/(CC BY 2.0 by Flickr user Gabi Apg)
http://www.flickr.com/photos/infrogmation/3019561993/sizes/l/in/photostream/(CC BY 2.0 by Flickr user Infrogmation)
Monolingual Features

4. (transitive) To cause (something) to be suspended, as from a hook, hanger or the like.
   Hang those lights from the ceiling.

5. (transitive, law) To execute (someone) by suspension from the neck.
   The culprits were hanged from the nearest tree.

6. (intransitive, law) To hang (someone) by one’s neck from a gallows, a tree, or other raised bar, attached by a rope tied into a noose.
   You will hang for this, my friend.

7. (intransitive, informal) To loiter, hang around, to spend time idly.
   Are you busy, or can you hang with me?
   I didn't see anything, I just hung around.

8. (transitive) To exhibit (an object).

9. (transitive) To apply (wallpaper or drywall to a wall).
   Let's hang this cute animal design in the bathroom.

**Synonyms**
- (display): display, exhibit, show
- (indicate a fact to be true): demonstrate, prove
- (arrive, show up): demonstrate, prove

**Linguistic labels**
- Definition
- Lemma of the relation source
- Inverse relation
- Shared relations
Cross-lingual Features

4. (transitive) To cause (something) to be suspended, as from a hook, hanger or the like.

Hand those lids from the ceiling

(Etwas) ausgesetzt werden, ab dem ein Haken, Aufhänger oder ähnliches verursachen

5. (transitive) To be suspended, as from a hook, hanger or the like.

The lamp was hung from the overhead bracket.

Der Lampe wurde an den Deckenhalter befestigt.

6. (intransitive, law) To be hanged, as from a gallows, a tree, or other raised bar, attached by the neck or other part of the body.

You are to be hanged, or the like.

Du sollst hingerichtet werden, oder so ein wie.

7. (intransitive, informal) To dangle, swing, or hang around, especially idly.

Are you busy, or can you hang with me?

Ich bin nicht beschäftigt, oder kannst du mit mir hängen?

I didn't see anything, officer. I was just hanging.

Ich habe nichts gesehen, Offizier. Ich war nur hängen.
Experimental Setup

Disambiguation Method

- Composition of features using a backoff strategy
- Machine learning classifier (Naïve Bayes, SVM, J48)

Evaluation

1. Comparison with previous method
   Existing dataset of 250 German semantic relations
   Text-Similarity-based system by Meyer&Gurevych (2010)

2. Comparison with gold standard
   Four newly created monolingual and cross-lingual datasets
   Comparison with MFS baseline system
Comparison with Previous Method

![Bar chart showing observed and kappa agreement with raters for MFS, MG10, WKTWSD, and human.]
Gold Standard Datasets

Four newly created monolingual and cross-lingual datasets

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Relations</td>
<td>394</td>
<td>459</td>
<td>204</td>
<td>204</td>
</tr>
<tr>
<td>Annotations</td>
<td>1,117</td>
<td>1,119</td>
<td>614</td>
<td>656</td>
</tr>
<tr>
<td>Agreement</td>
<td>.91</td>
<td>.92</td>
<td>.89</td>
<td>.90</td>
</tr>
<tr>
<td>Kappa</td>
<td>.82</td>
<td>.85</td>
<td>.73</td>
<td>.75</td>
</tr>
</tbody>
</table>

**Annotation example:**

\[ s_i = \text{phenomenal} \]

(colloquial) Very remarkable; highly extraordinary; amazing.

\[ D \]

0

Causing awe or terror; inspiring wonder or excitement.

\[ t_j = \text{awesome} \]

(colloquial) Very remarkable; highly extraordinary; amazing.

1

(informal) Excellent, exciting, remarkable.
Comparison with Gold Standard

F1-score

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFS</td>
<td>0.8</td>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>WKTWSD</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Human</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Inference of Semantic Relations

Haustier₁ \text{translation} pet₁

Katze₁ \text{translation} cat₁

\text{German Relations}:
- has hypernym: 290,019
- Relations: 300,724

\text{English Relations}:
- has hypernym: 26,965
- Relations: 215,353
## Final Resource

<table>
<thead>
<tr>
<th></th>
<th>Our Resource</th>
<th>Multilingual Wordnets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>German</td>
</tr>
<tr>
<td>Lexical entries</td>
<td>379,694</td>
<td>85,574</td>
</tr>
<tr>
<td>Word senses</td>
<td>474,128</td>
<td>73,500</td>
</tr>
<tr>
<td>Semantic relations</td>
<td>215,353</td>
<td>300,724</td>
</tr>
<tr>
<td>...Synonyms</td>
<td>70,199</td>
<td>78,133</td>
</tr>
<tr>
<td>...Antonyms</td>
<td>35,291</td>
<td>33,391</td>
</tr>
<tr>
<td>...Hypo-/Hypernyms</td>
<td>54,494</td>
<td>87,246</td>
</tr>
<tr>
<td>...Other types</td>
<td>55,269</td>
<td>101,954</td>
</tr>
<tr>
<td>Translations EN/DE</td>
<td>79,382</td>
<td>16,347</td>
</tr>
</tbody>
</table>

### Our resource also contains: sense definitions, linguistic labels, example sentences, translations to other languages,…
Verb Similarity


- 130 English verb pairs (EN–EN); six human raters annotating

  0: not at all related
  1: vaguely related
  2: indirectly related
  3: strongly related
  4: inseparably related

- Example: (approve, support) = 3 “strongly related”

Goal: High correlation of computational method with human raters
Cross-lingual Verb Similarity

Three new verb similarity datasets

EN–EN (approve, support) = 3

1. Manually translate verb pairs

DE–DE (annehmen, unterstützen) = 3

2. Mix EN–EN and DE–DE

EN–DE (approve, unterstützen) = 3

DE–EN (annehmen, support) = 3
Evaluation Results

Spearman’s rank correlation

EN-EN
DE-DE
EN-DE
DE-EN
Take-Home Message

Multilingual, Sense-disambiguated Resource based on Wiktionary data

- Automatic disambiguation of semantic relations and translations
- Inference of new semantic relations
- Resource and evaluation datasets publicly available

Case study: Computing Verb Similarity

- Compete with expert-built wordnets in the monolingual setting
- Outperform Wikipedia and expert-built wordnets in the cross-lingual setting

Future work

- Integrate with other resources such as UBY, BabelNet,…
- Explore other disambiguation methods such as CQC
Supplementary Online Material:
http://www.ukp.tu-darmstadt.de/data/lexical-resources/wiktionary/

Kontakt / Contact

Christian M. Meyer
Technische Universität Darmstadt
Ubiquitous Knowledge Processing Lab

Ｈochschulstr. 10, 64289 Darmstadt, Germany
☎ +49 (0)6151 16–7477
✉ +49 (0)6151 16–5455
✉ meyer (at) ukp.informatik.tu-darmstadt.de

Rechtliche Hinweise

Legal Issues
The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.