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

Christian M. Meyer

TECHNISCHE UNIVERSITÄT DARMSTADT

Christian M. Meyer and Iryna Gurevych

24th International Conference on Computational Linguistics
Mumbai, India, December 8–15, 2012.

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

"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?**

[1] See http://www.internetworldstats.com/stats7.htm (2012-11-28)

Finding Multilingual Lexical Knowledge

TECHNISCHE UNIVERSITÄT DARMSTADT

(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

EuroWordNet

Wikipedia?

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

WIKIPEDIA
The Free Encyclopedia

Wiktionary

- Freely available, high coverage, all parts of speech

Wiktionary
[ˈwɪkʃənrɪ] *n.*,
a wiki-based Open Content dictionary

http://www.flickr.com/photos/jwyg/3746351826/ (CC BY-SA 2.0 by Flickr user jwyg)

Contributions

TECHNISCHE UNIVERSITÄT DARMSTADT

Contribution 1:

Method for disambiguating semantic relations and translations in Wiktionary

Publicly available!

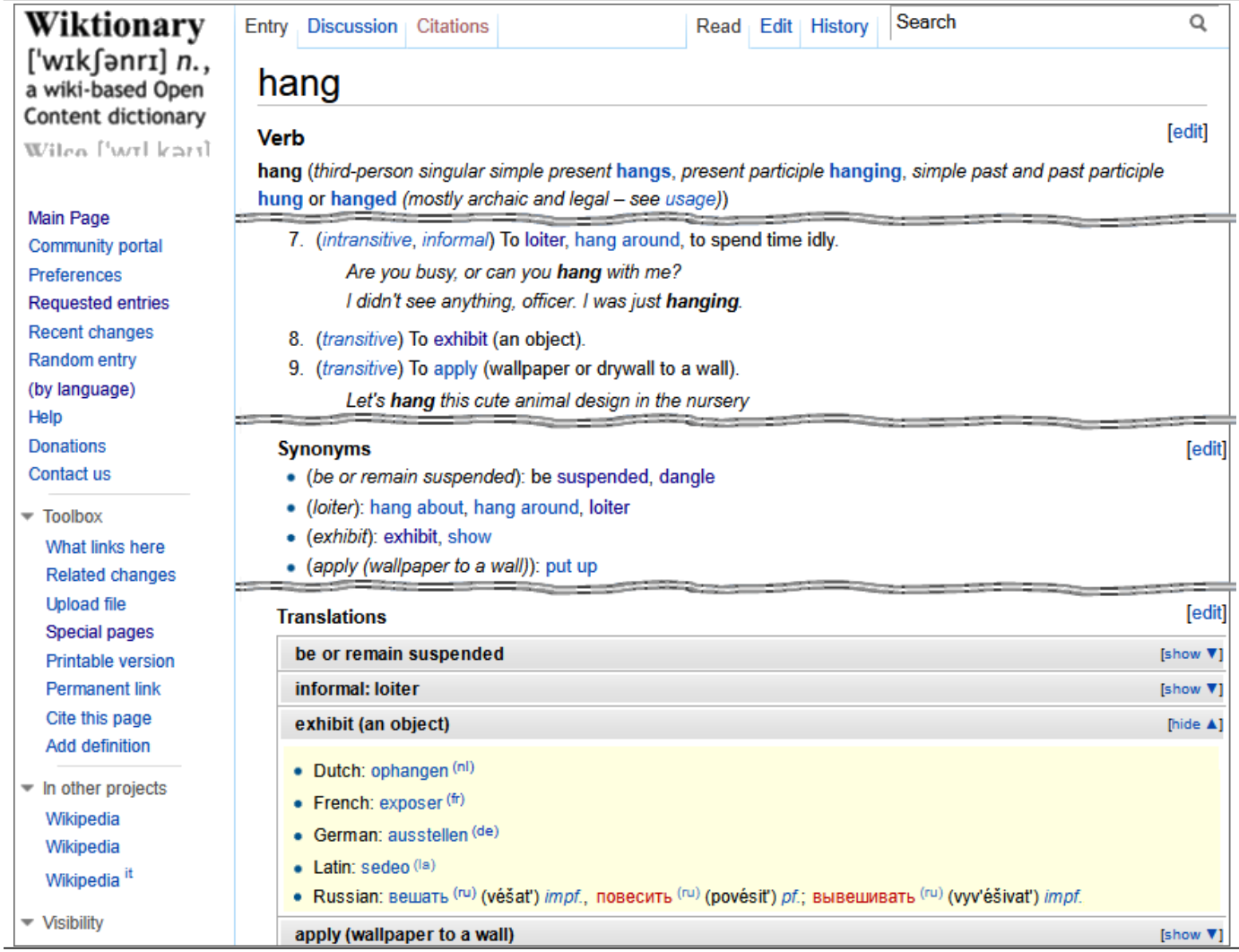
Contribution 2:

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



Wiktionary
['wɪkʃənrɪ] *n.*, a wiki-based Open Content dictionary

Main Page
Community portal
Preferences
Requested entries
Recent changes
Random entry
(by language)
Help
Donations
Contact us

Toolbox
What links here
Related changes
Upload file
Special pages
Printable version
Permanent link
Cite this page
Add definition

In other projects
Wikipedia
Wikipedia
Wikipedia it

Visibility

Entry | Discussion | Citations | Read | Edit | History | Search

hang

Verb [edit]

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

Synonyms [edit]

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

Translations [edit]

be or remain suspended [show ▼]

informal: loiter [show ▼]

exhibit (an object) [hide ▲]

- Dutch: ophangen (nl)
- French: exposer (fr)
- German: ausstellen (de)
- Latin: sedeo (la)
- Russian: вешать (ru) (vésat') *impf.*, повесить (ru) (povésit') *pf.*; вывешивать (ru) (vyv'éšivat') *impf.*

apply (wallpaper to a wall) [show ▼]

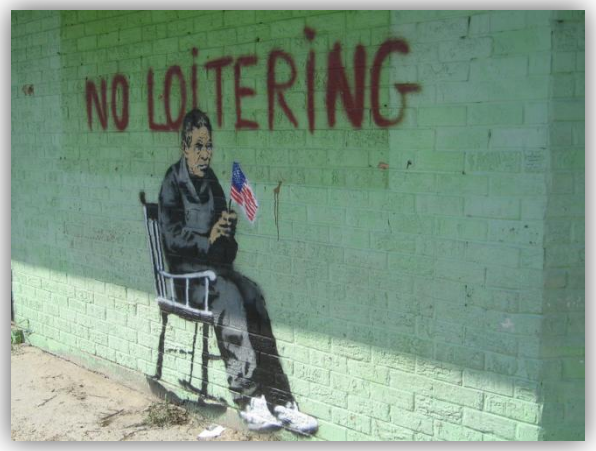
Disambiguation of Relations

Why Disambiguation Matters

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

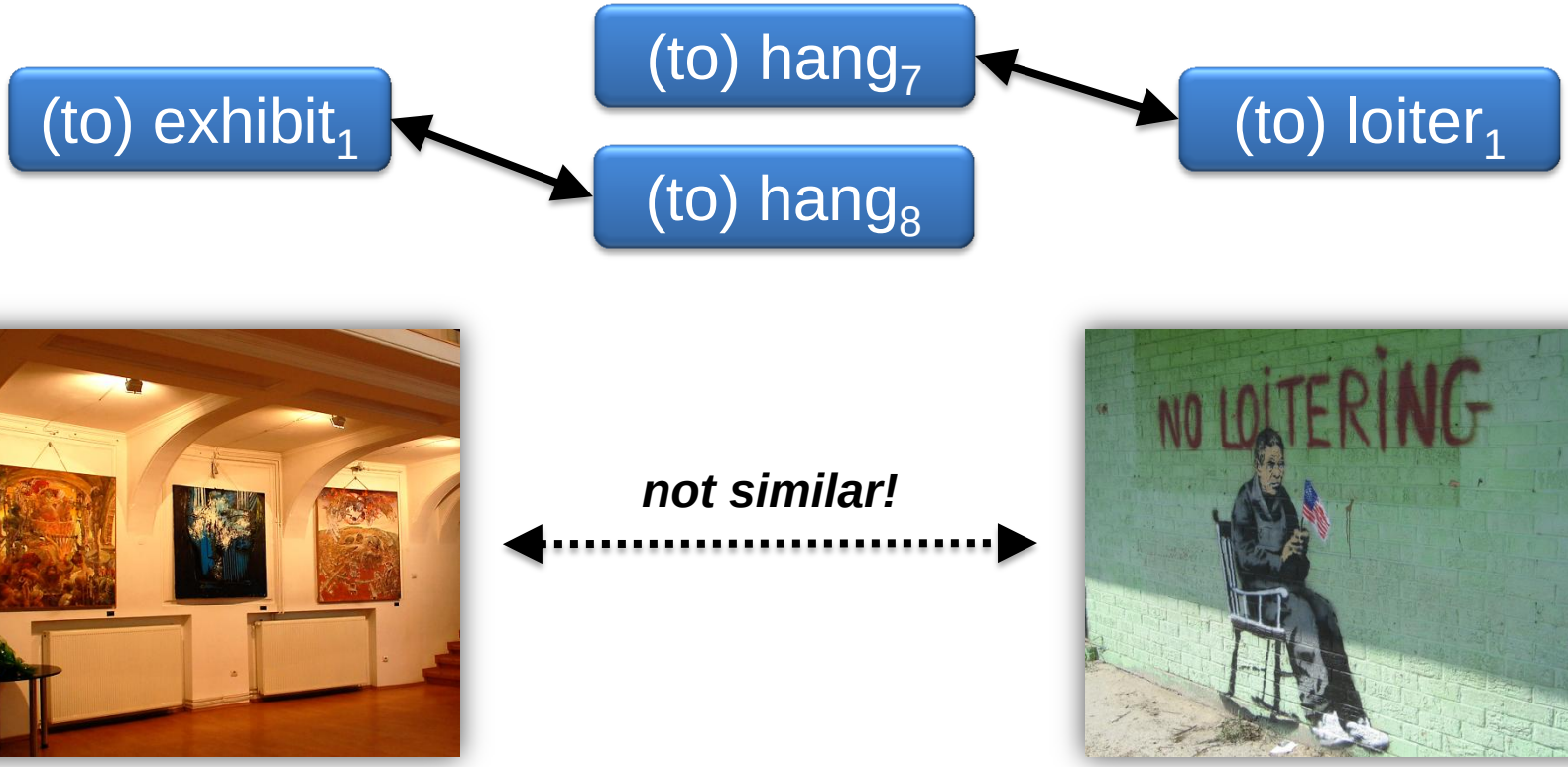
pretty similar?

shortest path contains only two edges



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

Why Disambiguation Matters



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

Monolingual Features

TECHNISCHE UNIVERSITÄT DARMSTADT

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 [illegible] 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, [illegible]

8. (*transitive*) To exhibit ([illegible]

9. (*transitive*) To apply (wallpaper or drywall to [illegible]

   Let's ***hang*** *this cute animal design in the* [illegible]

Definition

Linguistic labels

Lemma of the relation source

Synonyms

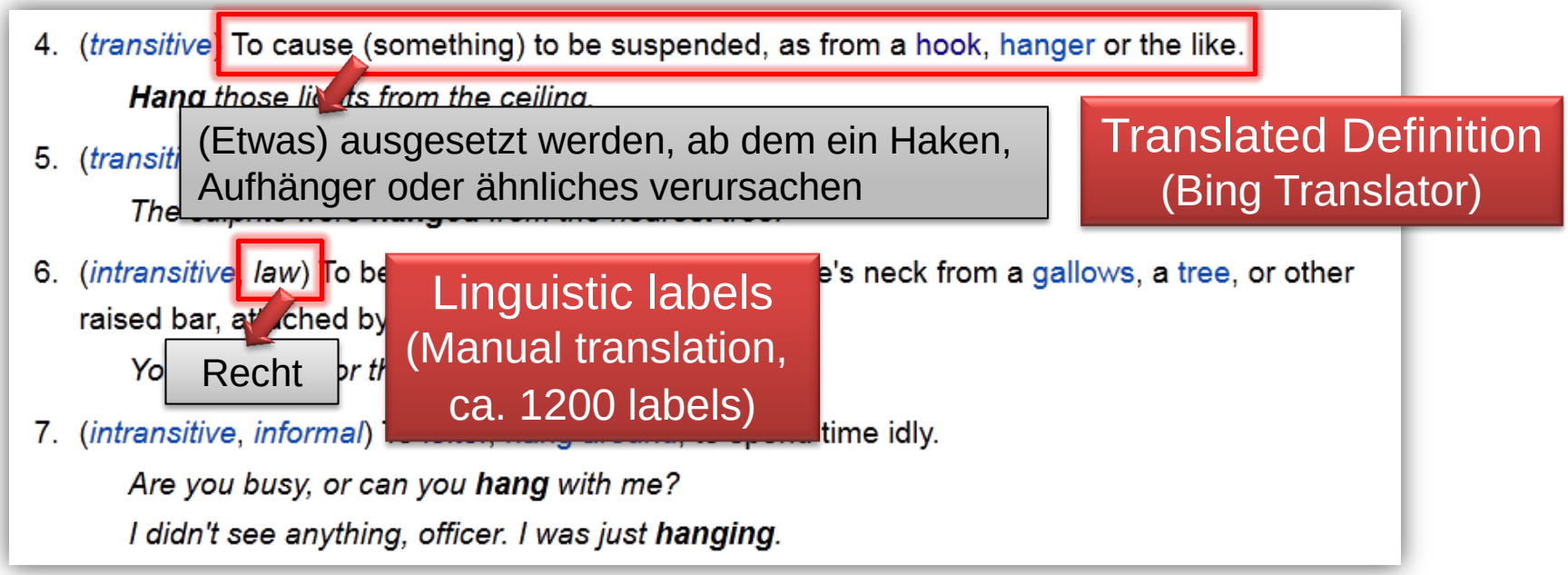
- (*display*): display, [illegible] al, exhibit
- (*indicate a fact to be true*): demonstrate, prove
- [illegible] arrive, show up

Inverse relation

Shared relations

UKP

Cross-lingual Features



Experimental Setup

TECHNISCHE UNIVERSITÄT DARMSTADT

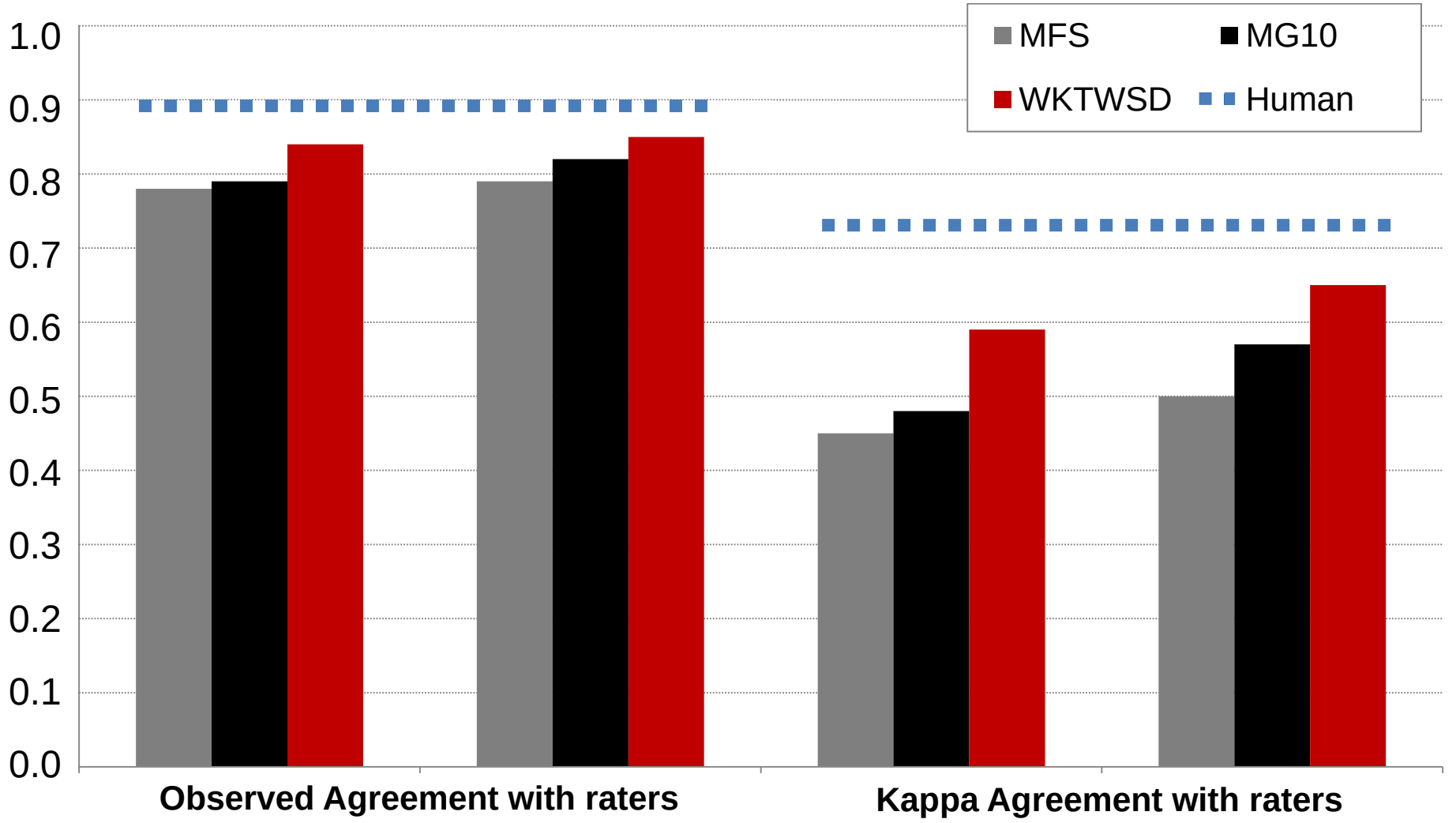
Disambiguation Method

- 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



Gold Standard Datasets

TECHNISCHE UNIVERSITÄT DARMSTADT

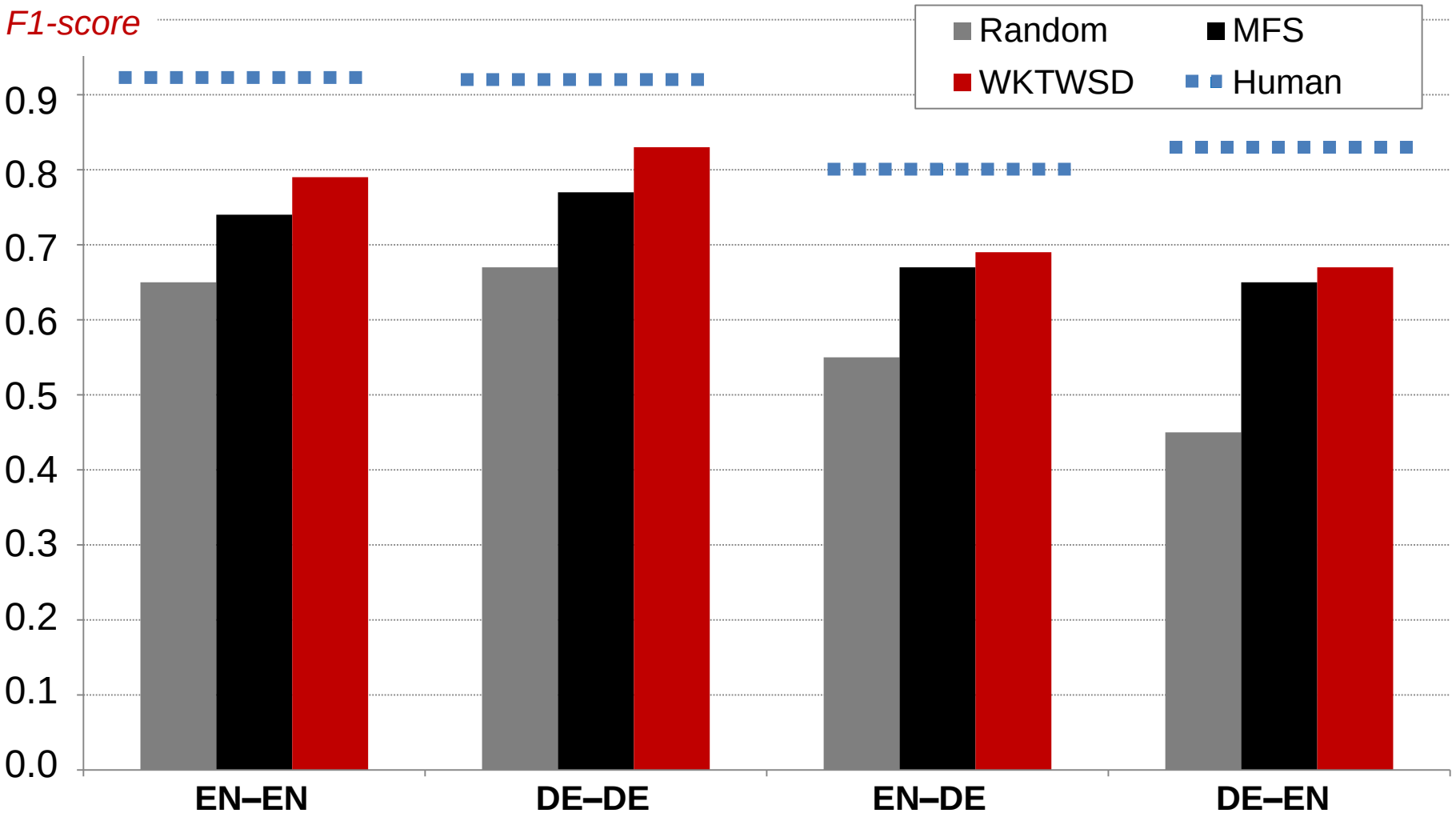
Four newly created monolingual and cross-lingual datasets

	EN–EN	DE–DE	EN–DE	DE–EN
Relations	394	459	204	204
Annotations	1,117	1,119	614	656
Agreement	.91	.92	.89	.90
Kappa	.82	.85	.73	.75

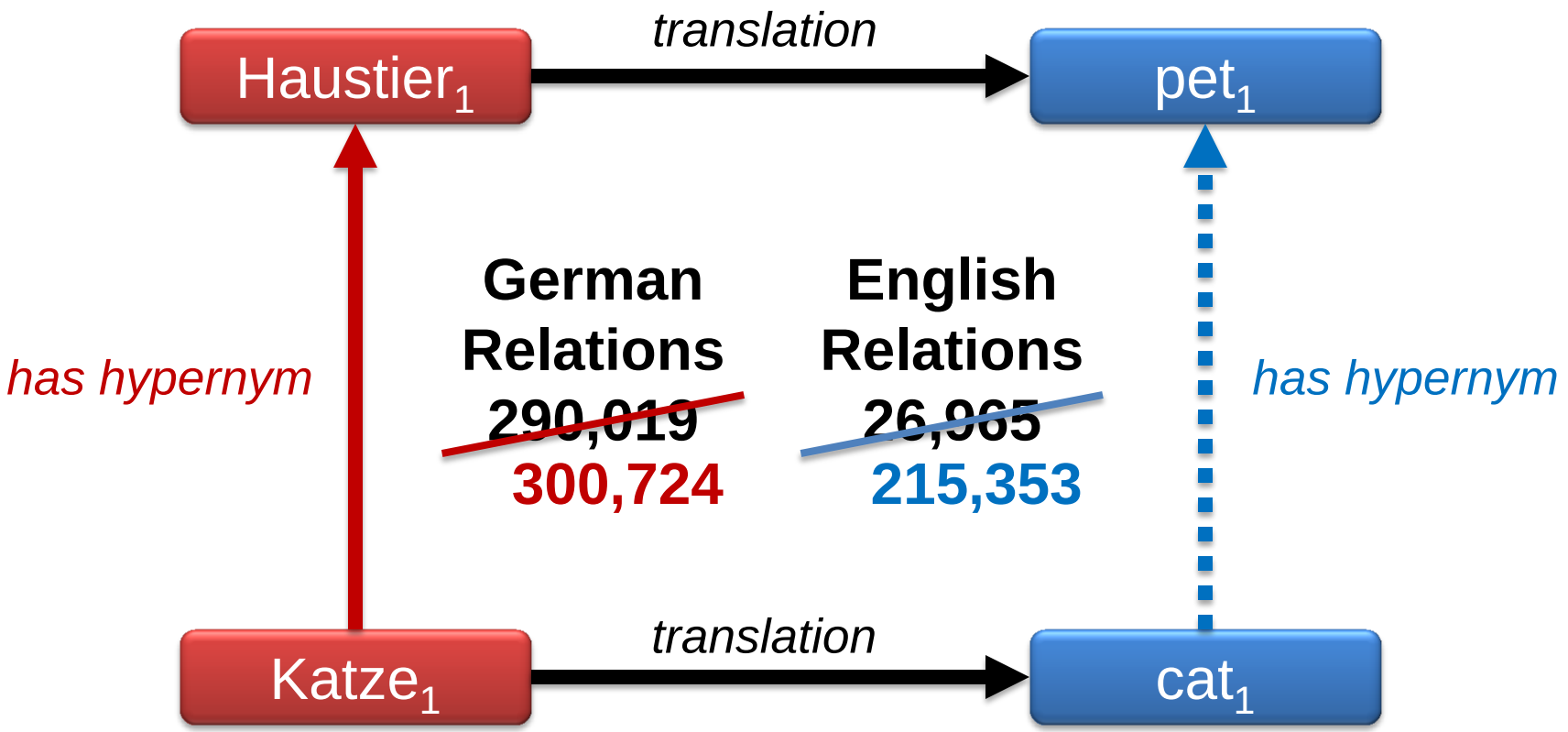
Annotation example:

s_i = *phenomenal*	D	t_j = *awesome*
(colloquial) Very remarkable; highly extra-ordinary; amazing.	0	Causing awe or terror; inspiring wonder or excitement.
(colloquial) Very remarkable; highly extra-ordinary; amazing.	1	(informal) Excellent, exciting, remarkable.

Comparison with Gold Standard



Inference of Semantic Relations



Final Resource

TECHNISCHE UNIVERSITÄT DARMSTADT

	Our Resource		Multilingual Wordnets	
	English	German	WordNet	GermaNet
Lexical entries	**379,694**	**85,574**	**156,584**	**85,257**
Word senses	**474,128**	**73,500**	**206,978**	**96,690**
Semantic relations	**215,353**	**300,724**	**1,398,868**	**512,653**
…Synonyms	*70,199*	*78,133*	*315,984*	*74,552*
…Antonyms	*35,291*	*33,391*	*7,979*	*3,359*
…Hypo-/Hypernyms	*54,494*	*87,246*	*658,804*	*397,335*
…Other types	*55,269*	*101,954*	*416,101*	*37,407*
Translations EN/DE	**79,382**		**16,347**	

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

Verb Similarity

TECHNISCHE UNIVERSITÄT DARMSTADT

Yang&Powers (2006)

- 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

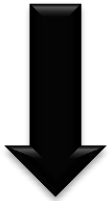
Cross-lingual Verb Similarity

TECHNISCHE UNIVERSITÄT DARMSTADT

Three new verb similarity datasets

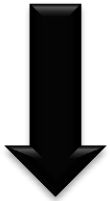
Publicly available!

EN–EN (approve, support) = 3



1. Manually translate verb pairs

DE–DE (annehmen, unterstützen) = 3

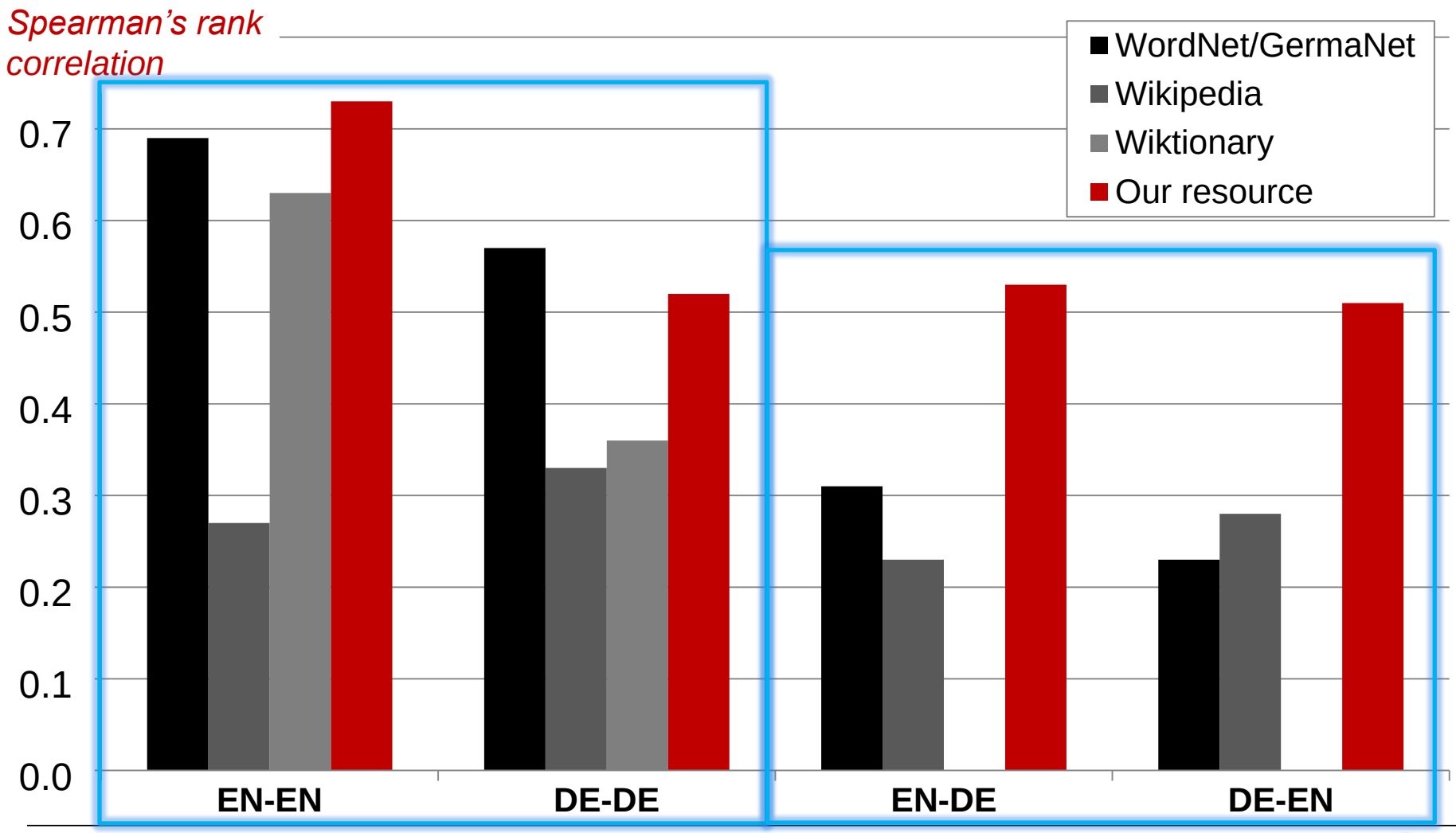


2. Mix EN–EN and DE–DE

EN–DE (approve, unterstützen) = 3

DE–EN (annehmen, support) = 3

Evaluation Results



Take-Home Message

TECHNISCHE UNIVERSITÄT DARMSTADT

Multilingual, Sense-disambiguated Resource based on Wiktionary data

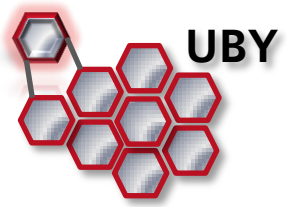
- Automatic disambiguation of semantic relations and translations
- Inference of new semantic relations
- Resource and evaluation datasets publicly available

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



Thank you for your attention!

TECHNISCHE UNIVERSITÄT DARMSTADT

Supplementary Online Material:
http://www.ukp.tu-darmstadt.de/data/lexical-resources/wiktionary/

References

Meyer, Christian M. & Gurevych, Iryna: **Worth its Weight in Gold or Yet Another Resource — A Comparative Study of Wiktionary, OpenThesaurus and GermaNet**, in Gelbukh, A. (Ed.): *Computational Linguistics and Intelligent Text Processing: 11th International Conference,* (= Lecture Notes in Computer Science 6008), pp. 38-49. Berlin/Heidelberg: Springer, March 2010.

Yang, Dongqiang & Powers, David M. W.: **Verb Similarity on the Taxonomy of WordNet**, in: *Proceedings of the Third International WordNet Conference,* pp. 121-128, January 2006. Jeju Island, Korea.

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

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, tradedmarks, 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.