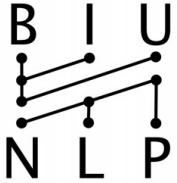
Factuality Prediction over Unified Datasets

Gabriel Stanovsky, Judith Eckle-Kohler, Yevgeniy Puzikov,

Ido Dagan and Iryna Gurevych

Bar-Ilan University, UKP - TU Darmstadt

ACL 2017



Factuality

Task Definition

Author's commitment towards a proposition

- **Factual**
  - It is *not surprising* that **the Cavaliers lost the championship**
- **Uncertain**
  - She still *has to check* whether **the experiment succeeded**
- **Counter-factual**
  - Don was *dishonest* when **he said he paid his taxes**

- Useful for
  - Knowledge base population
  - Question answering
  - Recognizing textual entailment

In this talk

- **Problem: Limited Generality**
  - Previous work focused on *specific* flavors of factuality
- **Approach**
  - Build a unified dataset
  - Train a new model
- **Contributions**
  - Normalized annotations
  - Large aggregated corpus
  - Improving performance across datasets

Problem:
Limited Generality

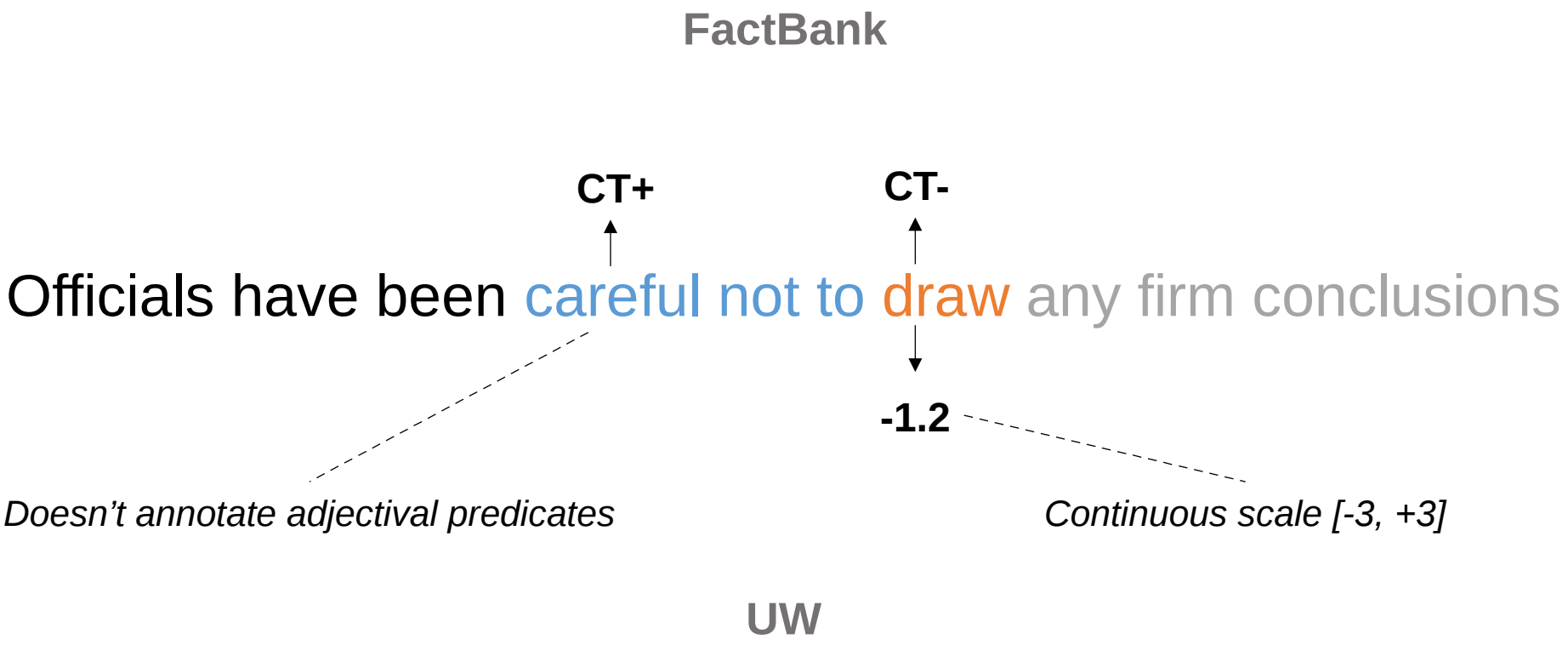
Datasets

- Many annotation efforts
  - FactBank (Saur´ı and Pustejovsky, 2009)
  - UW (Lee et al., 2015)
  - Meantime (Minard et al., 2016)
  - … and more

- Datasets differ in various aspects
  - **Discrete** vs. **continuous** values
  - **Expert** vs. **crowdsourced** annotation
  - **Point of view**

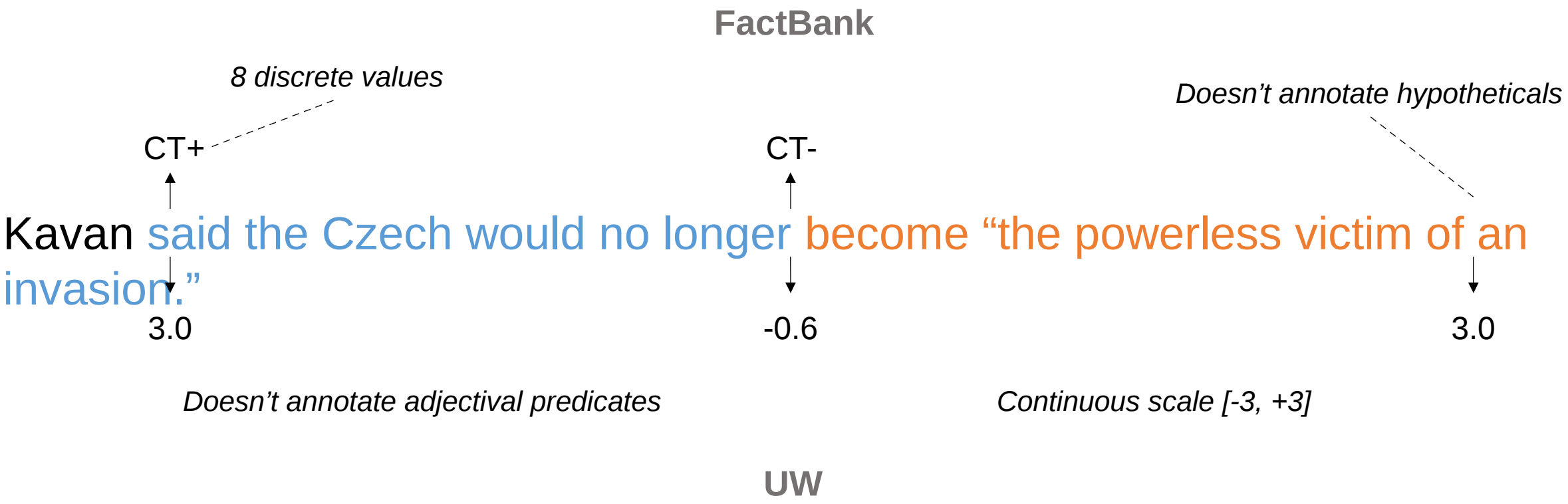
Annotated Examples

FactBank vs. UW



Annotated Examples

FactBank vs. UW



Previous Work: Factuality Prediction

- Models were designed and evaluated on *specific* datasets

- For example, Lee et al. (2015):
  - Used SVM on syntactic features
  - lemma, POS, dependency paths

  - Tested on the UW corpus

→ **Non-comparable results**

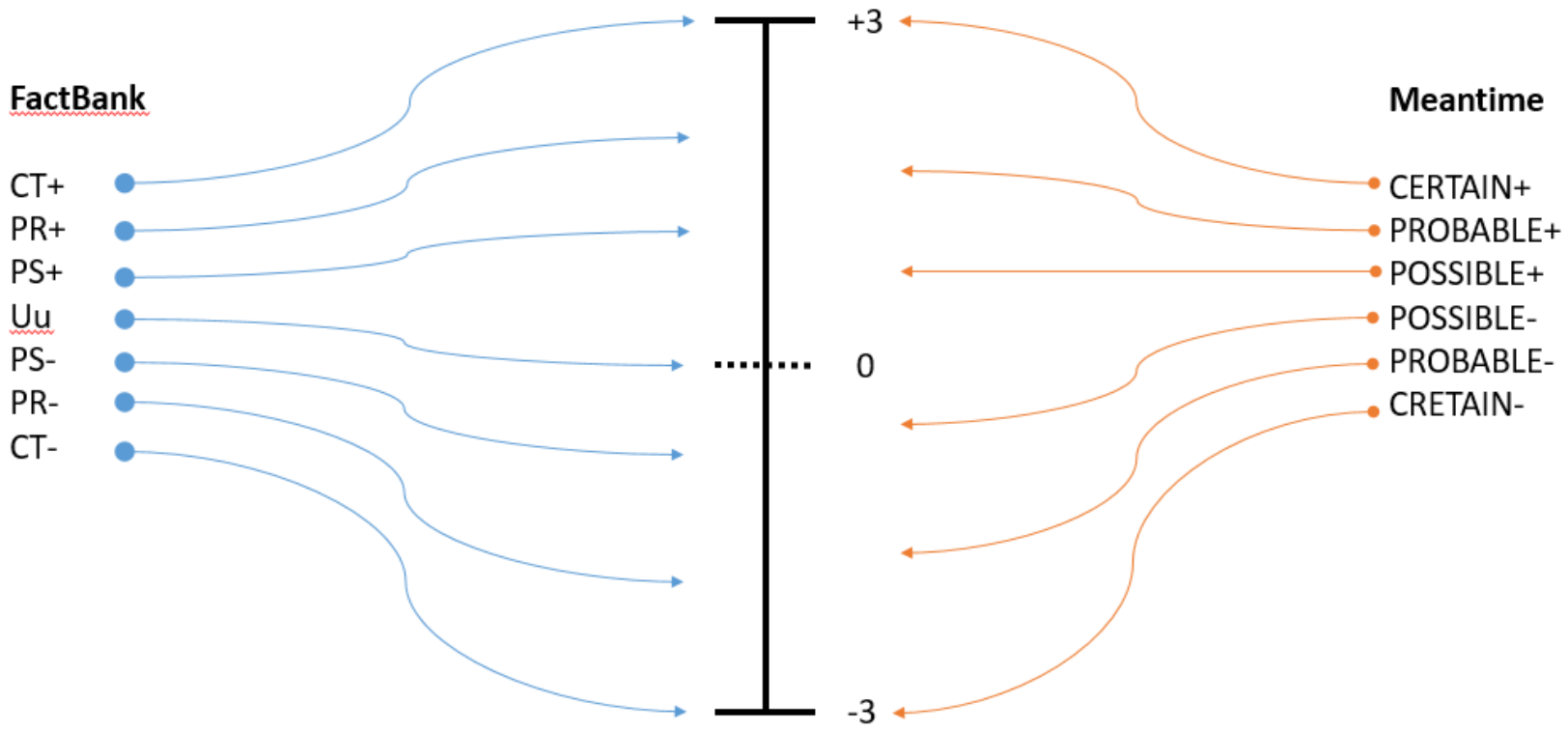
→ **Limited portability**

Solution:
Unified Corpus

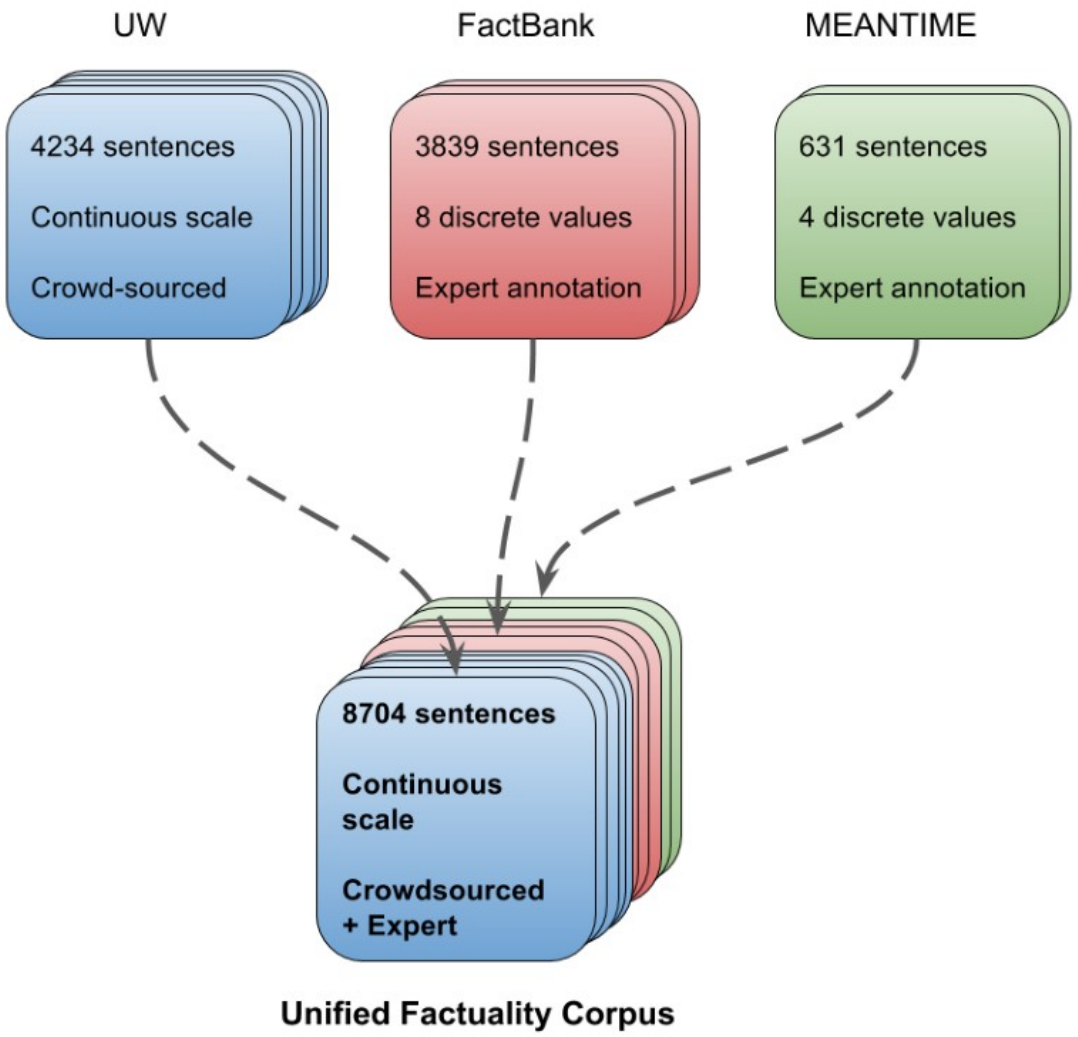
Extending TruthTeller Evaluation

Simple Normalization

- Mapping discrete values to the continuous UW scale
  - Simple mapping based on overlapping annotations

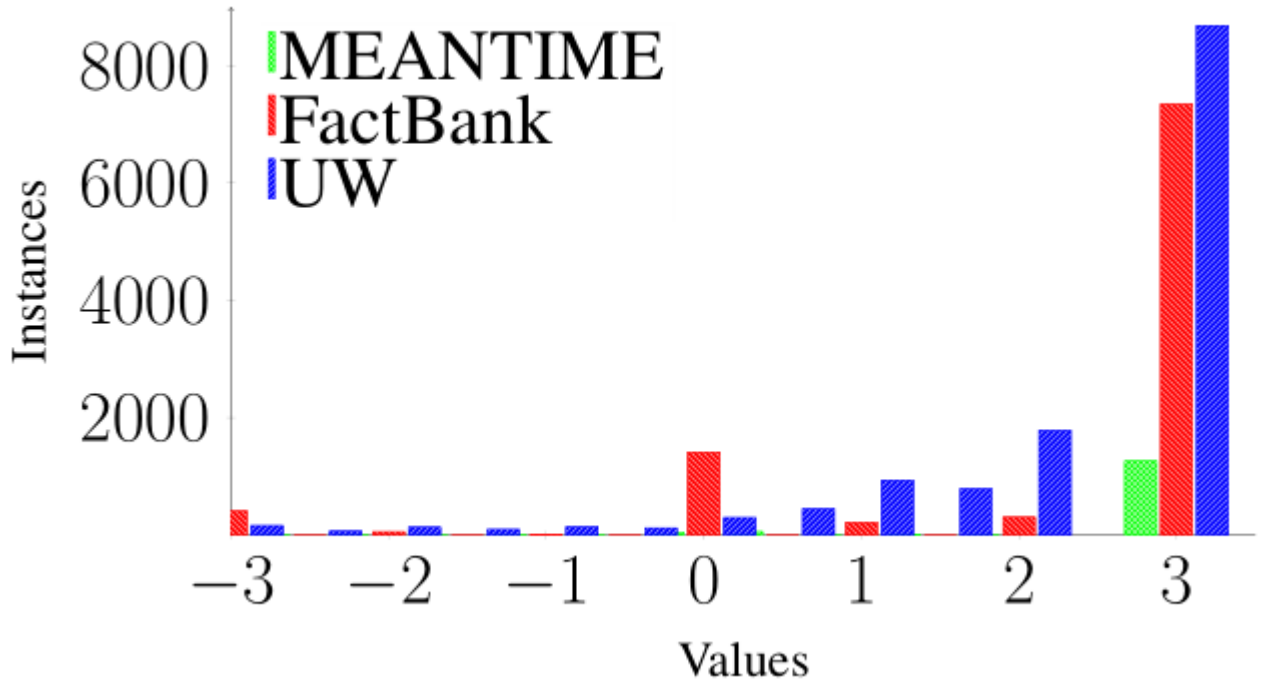


Unified Factuality Corpus



Biased Distribution

- Corpus skewed towards factual
- Inherent trait of the news domain?



Solution:

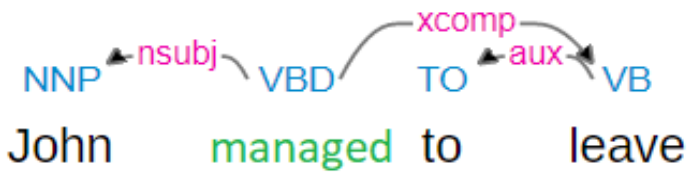
Unified Corpus

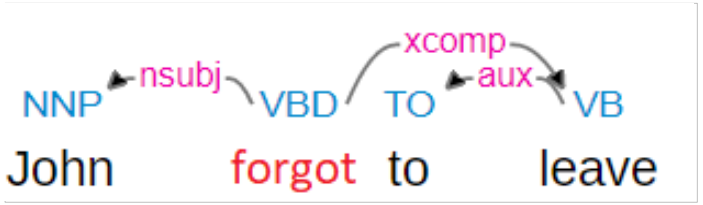
Model: **Extending TruthTeller**

Evaluation

TruthTeller (Lotan et al., 2013)

- Rule based approach on dependency trees
  - Karttunen implicative signatures
  - Syntactic cues (modality, negation, etc.)

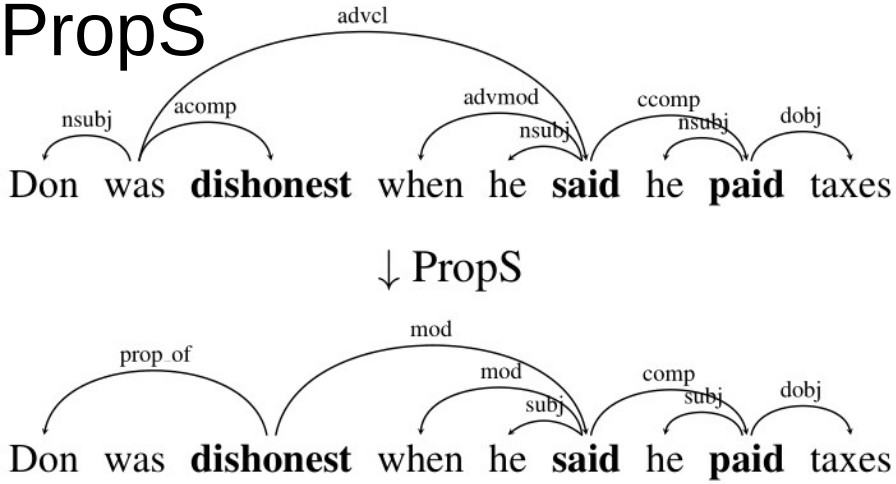




- Hand-written lexicon of 1,700 predicates

Extending TruthTeller

- Semi automatic extension of lexicon by 40%
  - Translated from German verb classes (Eckle-Kohler, ACL 2016)
- Supervised learning: **TruthTeller as signal**
- Application of implicative signatures on PropS



Solution:

Unified Corpus

Extending TruthTeller

Evaluation

Metrics (lee et al., 2015)

1. Mean Absolute Error
   - Range: [0, 6]
   - *Smaller is better!*

2. Pearson Correlation
   - How good is a system in recovering the variation
   - Well-suited for the biased news domain

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Marking all propositions as factual
Is a strong baseline on this dataset

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Dependency features correlate well

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Applying implicative signatures on AMR did not work well

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Hard coded rules aren't robust
Enough across datasets

Evaluations

Dataset	FactBank		UW		MEANTIME	
	MAE	r	MAE	r	MAE	r
All-factual	.80	0	.78	0	**.31**	0
UW feat.	.81	.66	.51	**.71**	.56	.33
AMR	.66	.66	.64	.58	.44	.30
Rule-based	.75	.62	.72	.63	.35	.23
Supervised	**.59**	**.71**	**.42**	.66	.34	**.47**

Our extension of TruthTeller gets good results across all datasets

Conclusions and Future Work

- Resources made publicly available
  - Unified Factuality corpus
  - Conversion code and trained models
- Future work
  - Annotate diverse domains
  - Integrate TruthTeller with more lexical-syntactic feats.
- Try our online demo: http://u.cs.biu.ac.il/~stanovg/factuality.html

Thanks for listening!