

# Coordination in English Enhanced UD: Analysis and Computational Modeling

Stefan Grünewald<sup>1,2</sup>

Prisca Piccirilli<sup>2</sup>

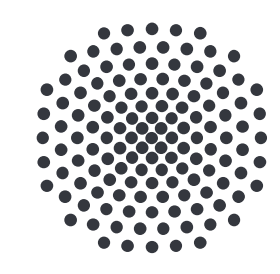
Annemarie Friedrich<sup>1</sup>

<sup>1</sup>Bosch Center for Artificial Intelligence

<sup>2</sup>University of Stuttgart, Germany



Invented for life



University of Stuttgart  
Germany

## Corpus Study and Manual Corrections: English Web Treebank (EWT)

- **EWT:** Enhanced annotations created using rule-based converter by Schuster & Manning (2016)
- We extract sentences containing **conjoined verb phrases**
- Modification/verification of enhanced links involved in (all) coordinate structures by expert annotator
- **New:** We also propagate non-core dependents: *obl*, *advcl*, *advmod*

	conj. sent	edited
train	1,926	999
dev	222	222
test	196	196
total	2,344	1,417

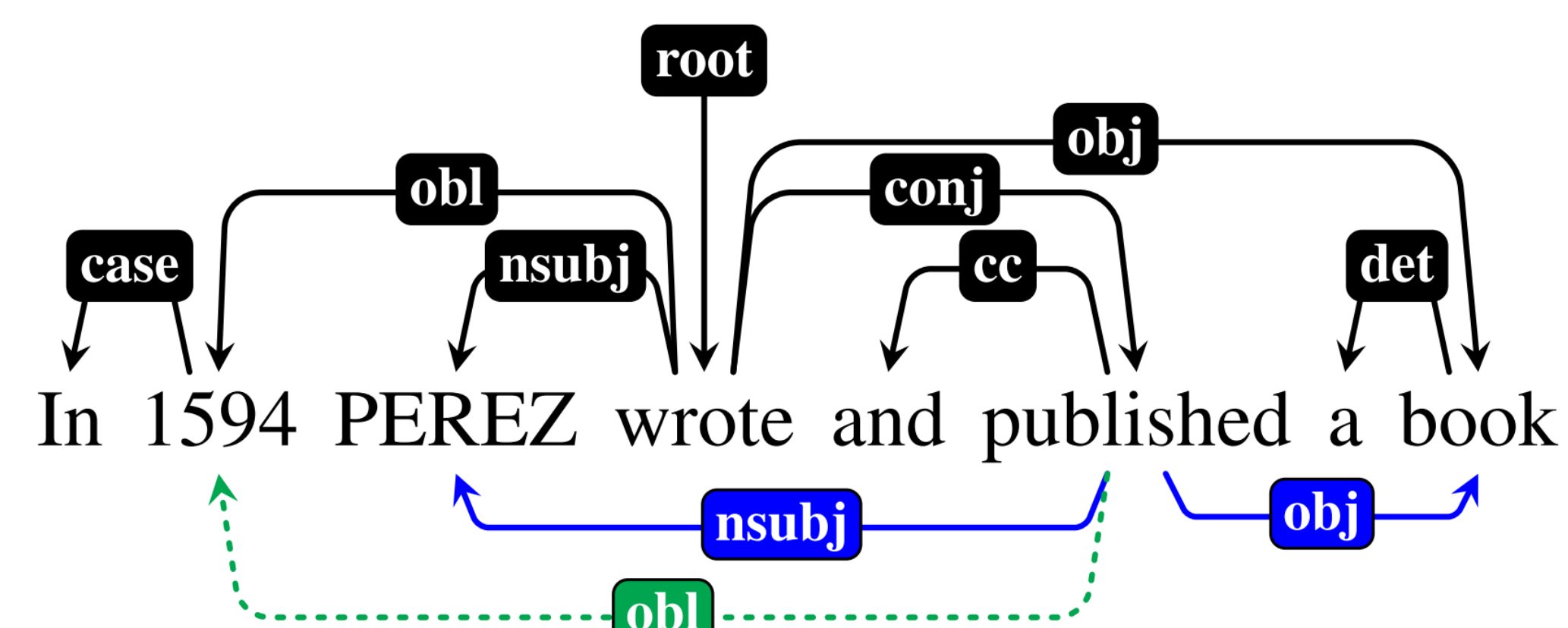
Dataset statistics

	A	B	C
A	-	90.1	94.9
B	95.2	-	97.2
C	80.5	77.9	-

Inter-annotator agreement study:  
100 sentences: precision/recall

### Our modifications:

- Frequently added: *nsubj*, *obl*, *nmod*, *advmod*
- Frequently removed: *nsubj:pass*, *nsubj*



**Black:** basic layer  
**Blue:** enhanced layer  
**Green:** proposed addition

### MOTIVATION:

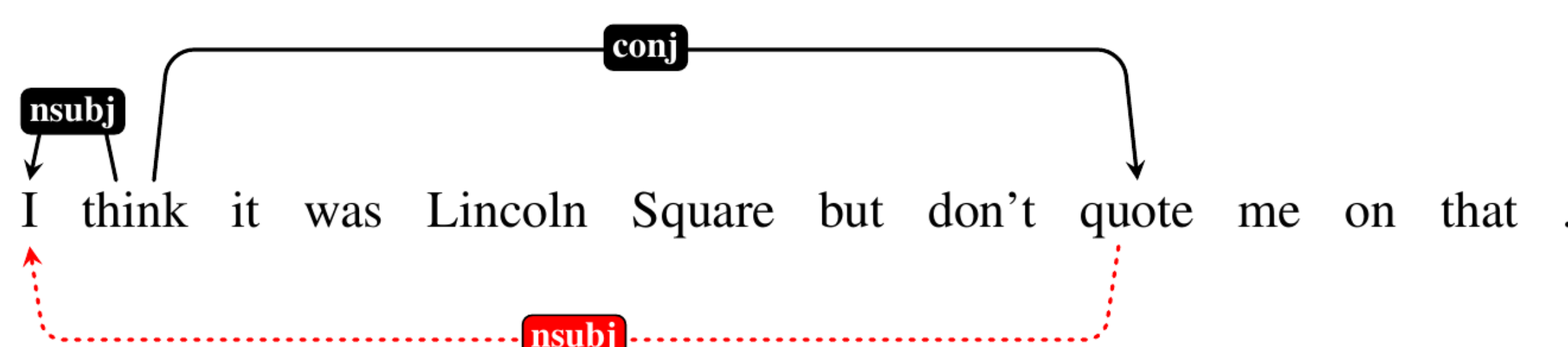
- Enhanced UD includes additional dependencies to propagate relations from the head of a conjunction to its other elements
- So far, English corpora such as EWT only contain enhanced links created via rule-based conversion (i.e., no human supervision)

### CONTRIBUTIONS:

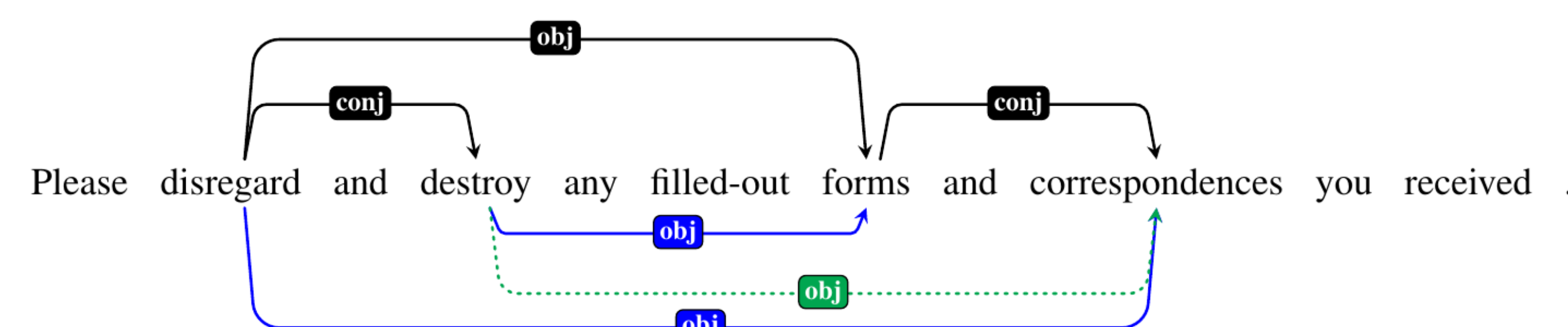
- We analyze and manually correct a large part of EWT w.r.t. conjunction propagation, identifying common converter errors
- Using our new dataset, we evaluate different methods of extracting propagated relations, finding that end-to-end enhanced parsing works best

## Corpus Study: Analysis of Modifications

- In addition to our propagation of non-core dependents, we identify common **error sources** in the automatically generated enhanced annotations, including:



Imperatives and passives



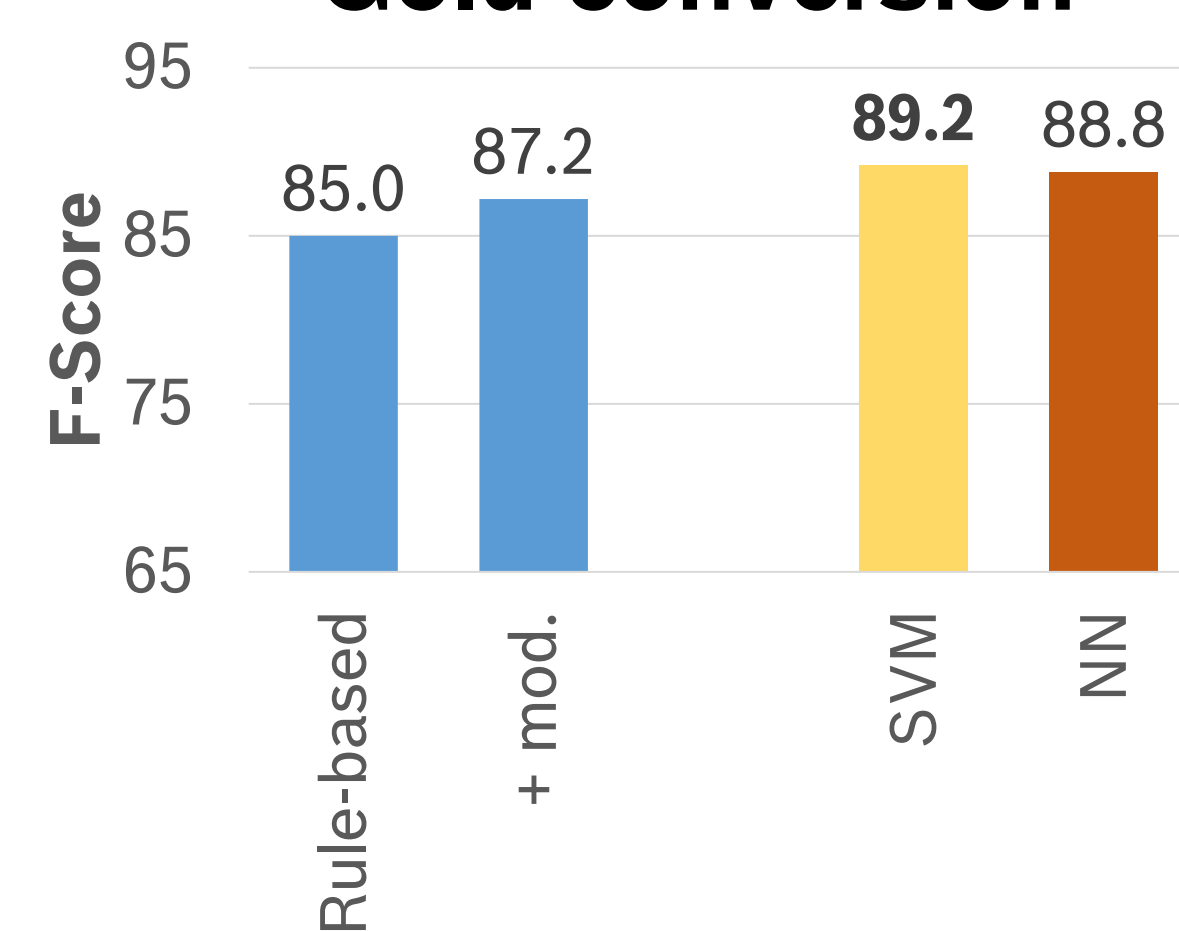
Multiple coordinations

- We propose changes to Schuster & Manning's (2016) rule-based converter to take these phenomena into account

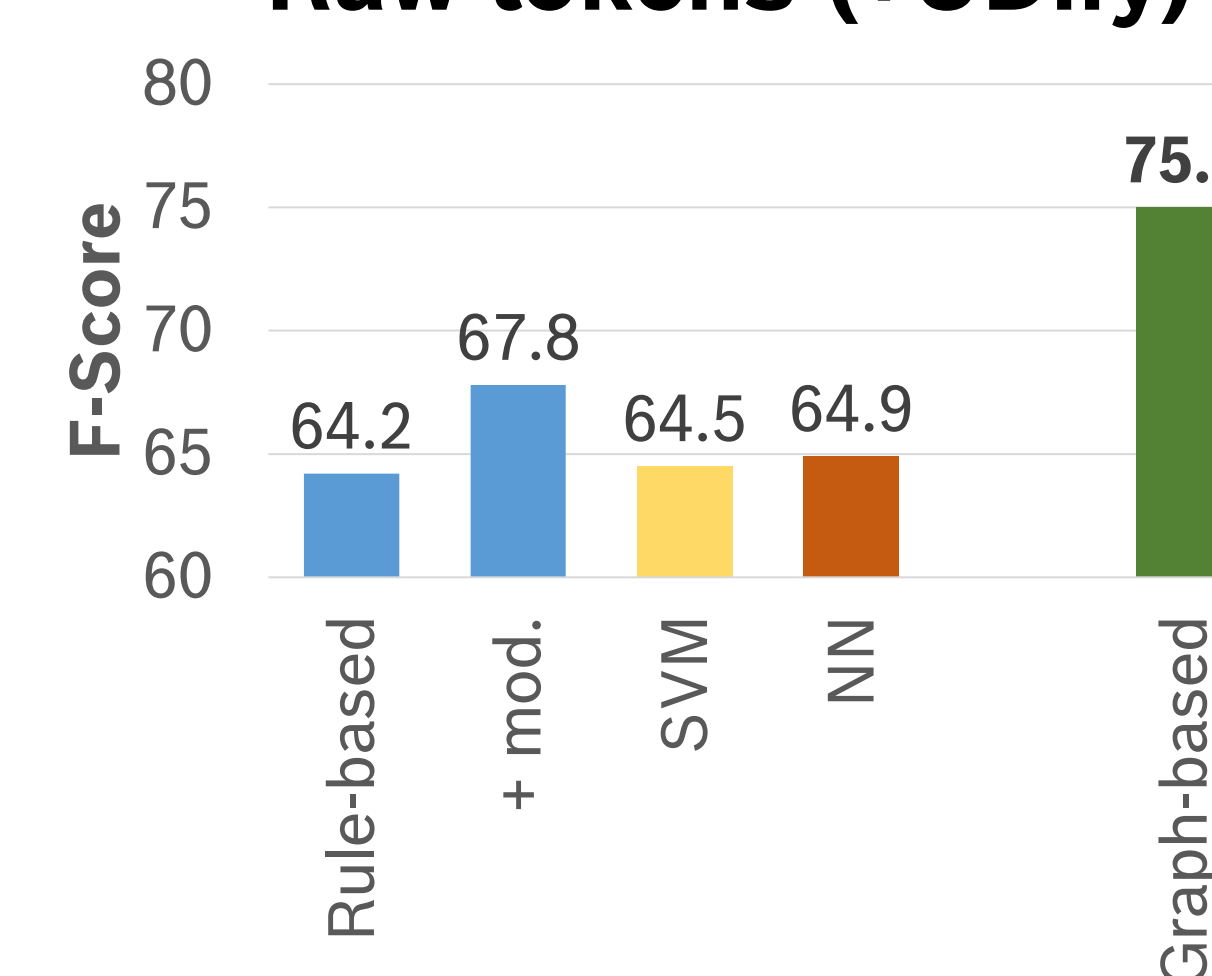
## Computational Modeling of Coordinate Structures

- We evaluate different methods of extracting propagated dependency links from either
  - a) gold-standard basic annotations, or
  - b) raw tokens only.
- We compare the following systems:
  - The **rule-based converter**, with or without our modifications;
  - Machine-learning-based converters using either an **SVM** (Nyblom et al., 2013) or an **NN** (own implementation)
  - A **graph-based parser** that predicts basic and enhanced dependencies simultaneously

### Gold conversion



### Raw tokens (+UDify)



- Machine learning models slightly outperform rule-based when extracting propagated dependencies from gold basic annotations
- On raw tokens, our graph-based parser achieves the best results

## Contributions and Conclusion

- We present the **first manually validated dataset** for coordination propagation in English Enhanced UD
- This new dataset allows for a **principled comparison of different methods for conjunction propagation** ranging from rule-based pipelines to an end-to-end neural system
- ML-based classifiers trained on original data work best to create **enhanced graphs from basic trees**
- A graph-based parser predicting basic and enhanced links simultaneously works best when only **raw tokens** are provided

Code and data are available on GitHub:

[github.com/boschresearch/coordinate\\_constructions\\_english\\_enhanced\\_ud\\_eacl2021](https://github.com/boschresearch/coordinate_constructions_english_enhanced_ud_eacl2021)

## References

- **Schuster & Manning (2016):** Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. *In: Proceedings of LREC 2016.*
- **Nyblom et al. (2013):** Predicting Conjoint Propagation and Other Extended Stanford Dependencies. *In: Proceedings of DepLing 2013.*