



# Combining Word Patterns and Discourse Markers for Paradigmatic Relation Classification

Michael Roth  
School of Informatics  
University of Edinburgh

Sabine Schulte im Walde  
Institute for Natural Language Processing  
University of Stuttgart



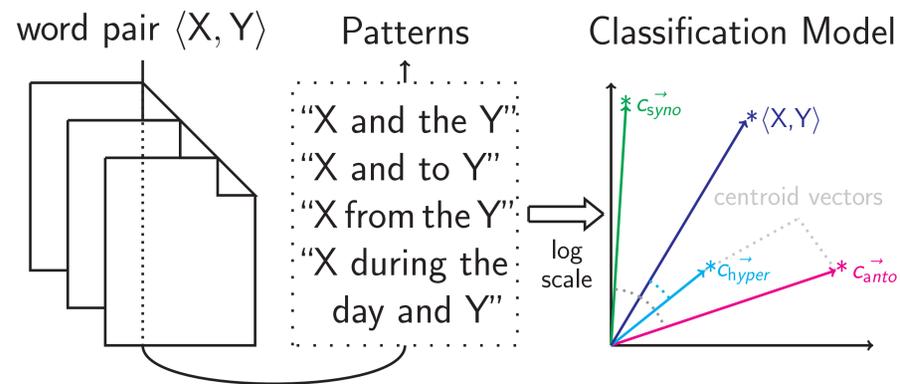
## Paradigmatic Relations

(cf. Murphy'03)

- ▶ Difficult to distinguish using distributional information (e.g., “The kid/child<sub>syno</sub> loves/hates<sub>anto</sub> his cat/pet<sub>hyper</sub>”)
  - ▶ Crucial for term expansion and inference-based tasks
- Approaches proposed in previous work:
- ▶ Purely supervised models based on thesauri, heuristics
  - ▶ Pattern-based models that can leverage unlabelled data

## Pattern-based Approach

(Schulte im Walde & Köper, 2013)



## Data Sets

(collected by Scheible & Schulte im Walde; Benotto & Lenci; Yap & Baldwin, 2009)

- ▶ 692 German and 648 English word pairs

	synonymy	antonymy	hypernymy
Noun	stone–rock	defeat–victory	thumb–finger
Verb	try–attempt	export–import	scribe–write
Adj.	unclean–dirty	left–right	historic–old

- ▶ 9,478 English noun pairs (50% unrelated)

Related <sub>syno</sub>	bend–turn	Unrelated	game–injury
-------------------------	-----------	-----------	-------------

## Discourse Markers and Relations

(Marcu and Echiabi, 2003; Prasad et al., 2008; inter alia)

### Motivation and Research Questions

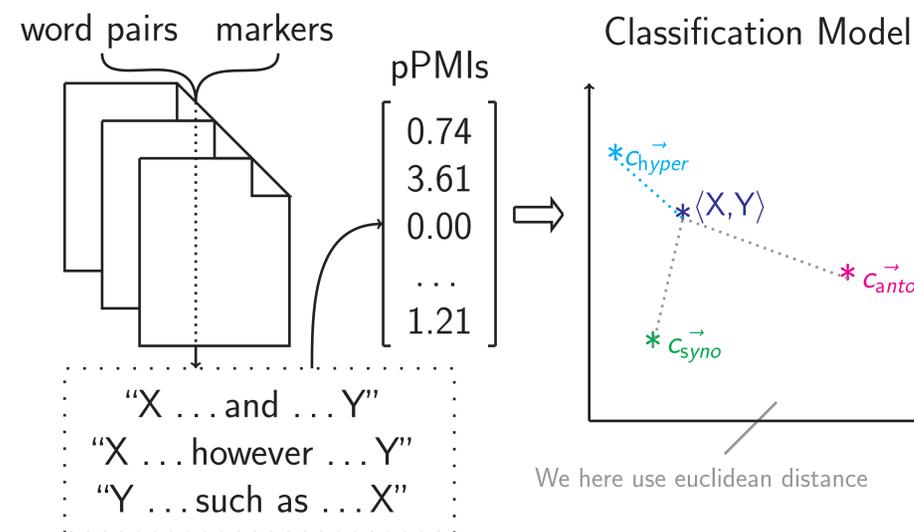
- ▶ Antonyms frequently indicate contrast relations
- ▶ Word pairs are generally good indicators for discourse relations

Can we apply these insight in reverse?  
Do discourse relations also indicate lexical relations?

### Markers as Proxies for Discourse Relations

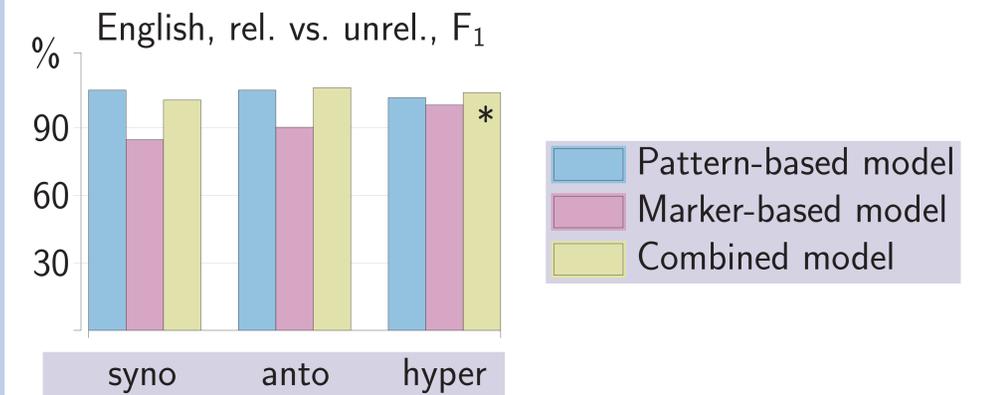
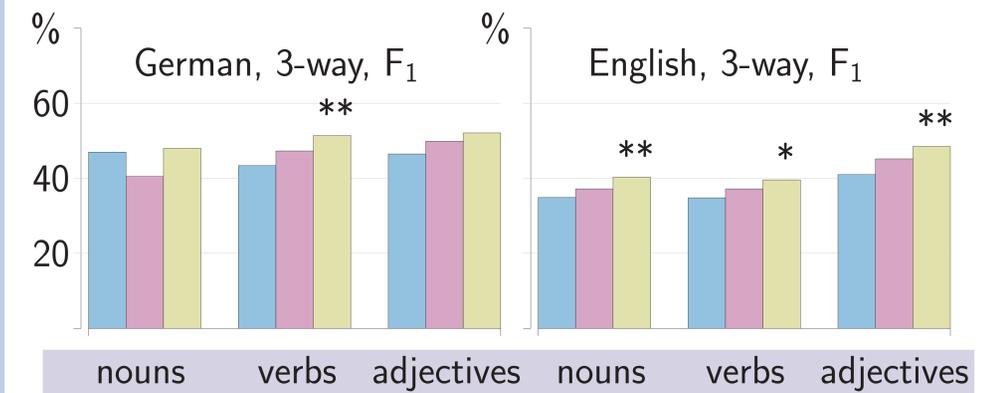
- + Exist in many different languages
- + Known to capture various semantic properties
- + Frequently found across genres
- + Definable as a small and fixed set

### Model



- ▶ Discourse markers from PDTB, translated via dict.cc
- ▶ Intra-sentential co-occurrences triples:  $\langle X, \text{MARKER}, Y \rangle$
- ▶ Allow for wild-cards between words and markers

## Results



## Conclusions

- ▶ Light-weight model: little memory and no parsing required
- ▶ Easily extendible to other languages (via translation)
- ▶ Higher recall and  $F_1$ -score than with  $>10,000$  word patterns
- ▶ Complementary strengths, best results in combination

## Acknowledgements

This research was funded by the DFG project “Distributional Approaches to Semantic Relatedness” and the DFG Heisenberg Fellowship SCHU-2580/1-1.