

# Porting an Open Information Extraction System from English to German



Bar-Ilan University



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Tobias Falke<sup>1</sup>, Gabriel Stanovsky<sup>2</sup>, Iryna Gurevych<sup>1</sup> and Ido Dagan<sup>2</sup>

<sup>1</sup> Research Training Group AIPHES / UKP Lab, Computer Science Department, Technische Universität Darmstadt, Germany

<sup>2</sup> Natural Language Processing Lab, Department of Computer Science, Bar-Ilan University, Israel

## Summary

### Open Information Extraction

- Extract minimal propositions as predicate-argument tuples from sentences
- Useful representation for semantic tasks
- Mostly extraction from dependency representations with hand-written patterns and extraction rules
- Lack of tools for non-English Open IE

### Research Question

- Can we port an existing system to a new language to obtain an Open IE system quickly and with reasonable effort?
- Our focus: Open IE with PropS and German as the target language

### Findings

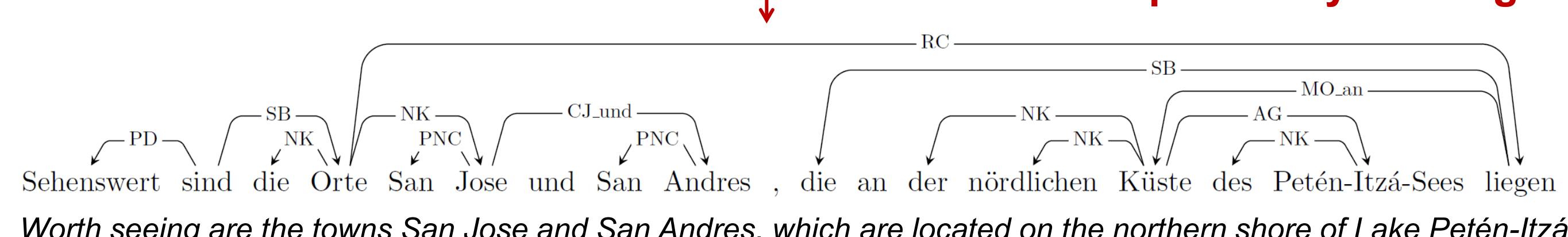
- Most rules of the English system can be ported directly or with small changes
- We could build a system for German, PropsDE, with 10% of the original effort
- PropsDE shows competitive performance

## Open IE with PropS

### 1. Sentence

Sehenswert sind die Orte San Jose und San Andres, die an der nördlichen Küste des Petén-Itzá-Sees liegen.

### 2. Dependency Parsing



### 4. Extractions

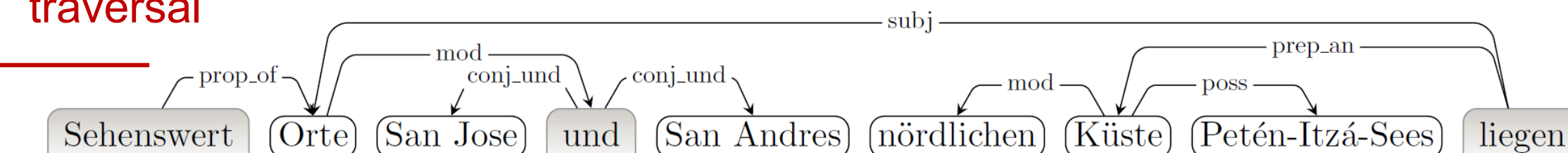
**liegen** ( die Orte San Jose und San Andres ;  
an der nördlichen Küste des Petén-Itzá-Sees )

Graph  
traversal

Rule-based  
conversion

### 3. PropS Graph

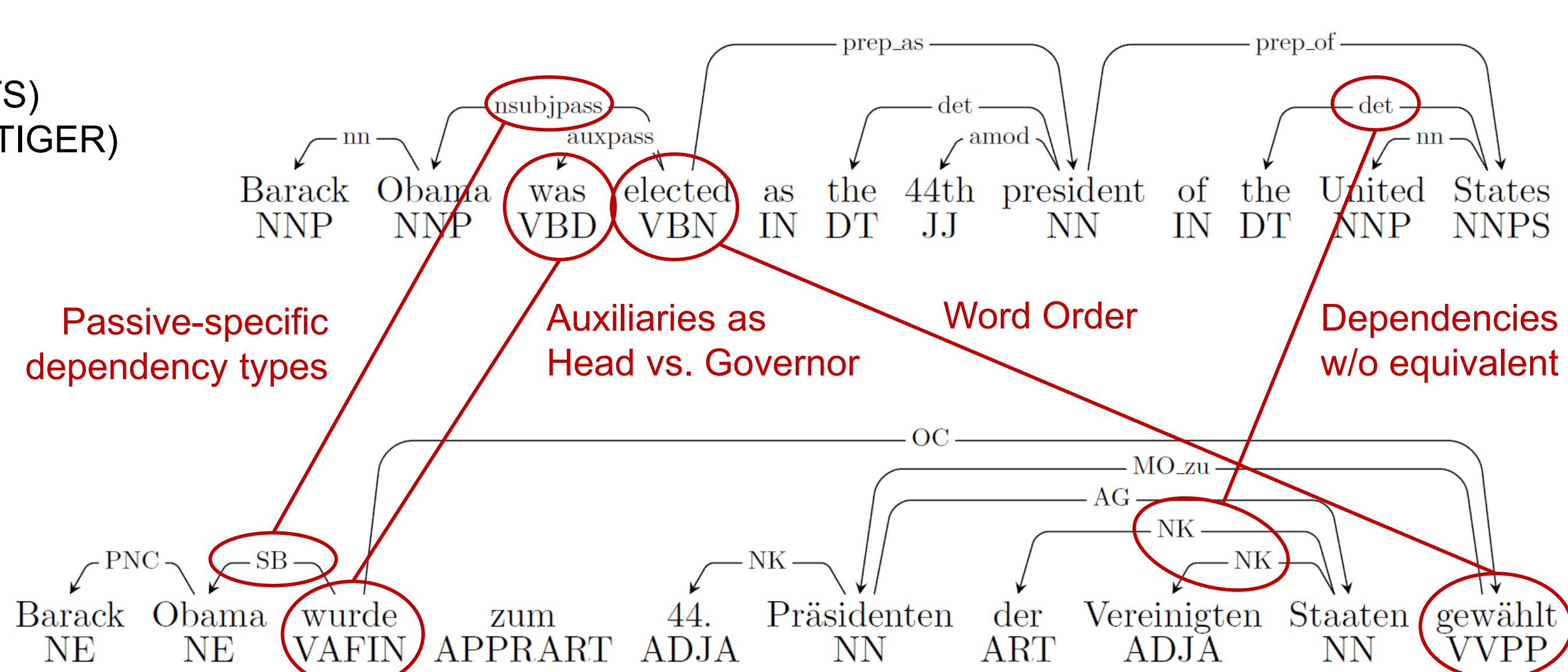
**sehenswert** ( die Orte San Jose und San Andres )



## Analysis of Portability

Different challenges circumvent an automatic translation of the PropS rule set:

- Different tag sets
  - Part of Speech (PTB vs. STTS)
  - Dependencies (Stanford vs. TIGER)
  - More or less fine-grained
  - Not directly mappable
- Different style
  - Auxiliaries
  - Conditionals
  - Relative clauses
- Language Specifics
  - Separated verb particles
  - Subjunctive mood



Portability	Required Effort	Rules
Directly	Simple translation of tags from English to German	38 %
Small changes	Translation of tags, small changes due to tag set and annotation style differences	35 %
Substantial changes	New or mostly changed rules for German to account for different representation	27 %

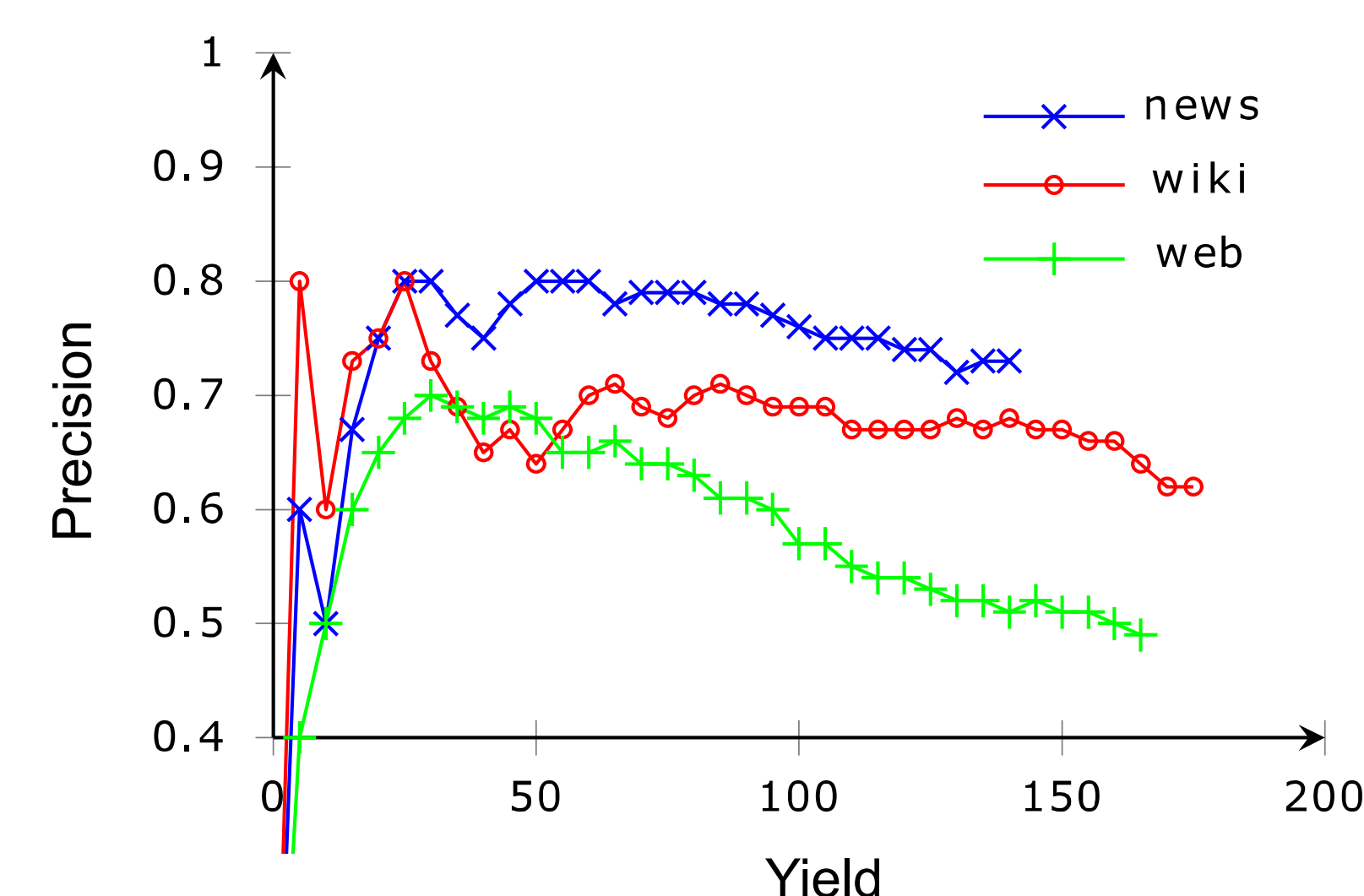
- Many rules could be translated with only few changes
- We could therefore build a system for German in a short amount of time with only 10% of the original effort

## Evaluation

- Open IE with PropsDE on 300 randomly sampled sentences from TIGER, Wikipedia and CommonCrawl
- All extractions evaluated for correctness by two annotators, agreement is 85% ( $\kappa=0.63$ )

- Results are comparable with Open IE for English

Genre	Sent.	Length	Yield	Precision
News (gold)	100	19.3	142	78.9
News	100	19.3	144	70.8
Wiki	100	21.4	178	61.8
Web	100	19.2	165	49.1
Total	300	20.0	487	60.2



<https://github.com/UKPLab/props-de>  
source code, online demo, examples

