NLP & Semantic Computing Group

# A Sentence Simplification System for Improving Open Relation Extraction

Christina Niklaus, Bernhard Bermeitinger, Siegfried Handschuh and André Freitas
Department of Computer Science and Mathematics - University of Passau
{christina.niklaus, bernhard.bermeitinger, siegfried.handschuh, andre.freitas}@uni-passau.de

## Introduction

- task of Open Relation Extraction (RE): recognizing the assertion of relationships between two or more entities in natural language text without requiring any relation-specific human input
- problem: In **syntactically complex sentences**, relations often span several clauses or are presented in a non-canonical form, **thus posing a challenge for current Open RE approaches**, which are prone to make incorrect extractions, while missing others, when operating on sentences with an intricate structure.
- objective of our work: improve the performance of state-of-the-art Open RE systems by simplifying the linguistic structure of NL text

## Method

- simplifies a sentence by separating out components that supply secondary information
- transforms complex sentences into simpler stand-alone context sentences
- the simplified set of sentences are easier to process for subsequently applied Open RE systems

**original sentence**

↓ set of hand-crafted simplification rules

**simplified version** consisting of
- **core sentence**: gist
- **context sentences**: background information

## Definition of Simplification Rules

- **linguistic analysis** of sentences from Wikipedia to identify syntactic elements that provide only secondary information:
  - non-restrictive relative clauses
  - non-restrictive and restrictive appositive phrases
  - participial phrases offset by commas
  - adjective and adverb phrases delimited by punctuation
  - particular prepositional phrases
  - lead noun phrases
  - intra-sentential attributions
  - parentheticals
  - conjoined clauses
- **specification of a set of hand-crafted grammar rules** for carrying out the simplification operations based on the syntactic features of those components

## Application of the Simplification Operations

1. identification of a **syntactic element** that is to be separated out of the source sentence
2. construction of a **context sentence**
3. reduction of the **main sentence**

"A few hours later, Matthias Goerne, a German baritone, offered an all-German program at the Frick Collection."

↓

**syntax-based sentence simplification**

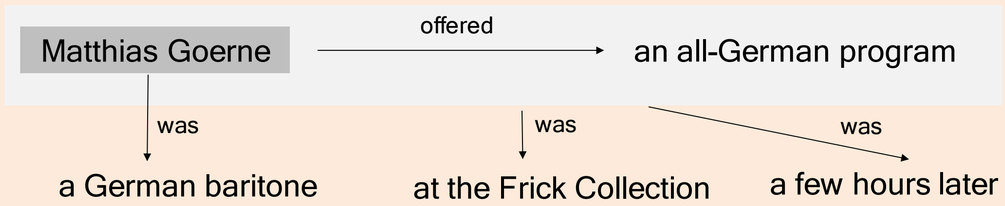
↓

- **core sentence:** Matthias Goerne offered an all German program.
- **context sentence:** Matthias Goerne was a German baritone.
- **context sentence:** This was a few hours later.
- **context sentence:** This was at the Frick Collection.

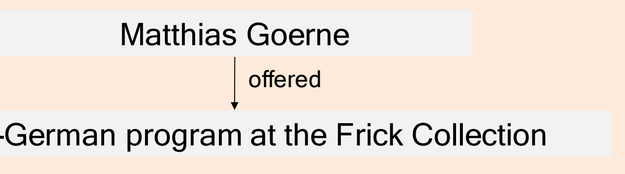
## Result of the Simplification Process

- transformation of a complex source sentence into a simplified **two-layered representation** in the form of core facts and accompanying contexts
- impact on Open RE:
  - complexity of determining intricate predicate-argument structures with **variable arity** and **nested structures** is removed
  - relations can be more easily extracted under a binary/ternary predicate-argument structure

Extracted relations when operating on the simplified sentences:




Result without a prior simplification step:




**accuracy** of the extracted relations:

| | precision | recall | F1 |
|---|---|---|---|
| ***without* simplification** | 84% | 22% | 35% |
| ***with* simplification** | 71% | 27% | 39% |

**information loss**:

| | noun chunks lost |
|---|---|
| ***without* simplification** | 60.4% |
| ***with* simplification** | 49.7% |

## Conclusions

State-of-the-art Open RE approaches obtain
- a **higher accuracy**
- a **lower information loss**

when operating on sentences that have been preprocessed with our text simplification framework.

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik
Lehrstuhl für Digital Libraries and Web Information Systems
Innstraße 43
94032 Passau, Germany