

Merlin Göttliger Lutz Schröder Philipp Heinrich Natalie Dykes Stefan Evert

Corpus Queries for Argument Mining

Reconstructing Arguments from Noisy Text (RANT)

Argument mining for social media

Challenges:

- ▷ Unmediated environment
- ▷ Non-traditional forms of argumentation
- ▷ Non-standard language, brevity, implicitness

Our approach uses **discourse analysis**, **corpus linguistics**, **argumentation theory**, and **logic**

Data and preprocessing

- ▷ 2,623,484 deduplicated English tweets
- ▷ Initial IDs collected by Milajevs by filtering for #Brexit in the run-up to the Brexit referendum
- ▷ Extended by tweets in the preceding discussion threads to capture arguments more clearly
- ▷ Off-the-shelf algorithms for tokenization and tagging
- ▷ Self-implemented lemmatizer based on a morphological analyser combined with heuristic rules

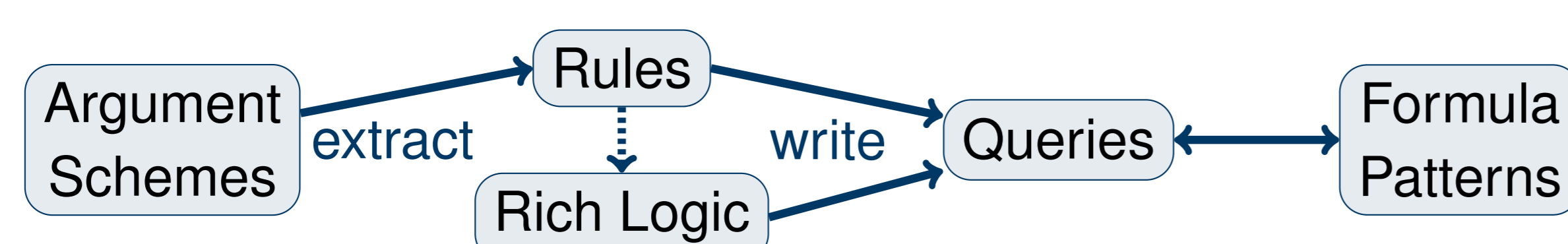
CQP queries

- ▷ Use macro facilities of the CQP query language to incorporate modular components similar to a context-free grammar
- ▷ Designed to capture argumentative and linguistic regularities in two different ways
- ▷ **Grammatical level:** Macros and particular arrangements of POS tags reflect regular lexico-grammatical patterns, which often correspond to phrases or other constructions above word level
- ▷ **Lexico-semantic categories:** summarised in word lists, which consist of near-synonyms or other semantically related lexical items
- ▷ Developed iteratively by alternating between generalisation and specialisation, guided by analysis of the corresponding concordances

Listing 1: CQP Query

```
[lemma = $nouns_finance] "could|might|to|will" [lemma = $verbs_increase] "if" /np[]+ /vp[]? [lemma = "brexit"]? /vp[] /np[]?;
```

Formula patterns



- ▷ Extract rules and logic from argument schemes
- ▷ Write queries for interpreting natural language to logic
- ▷ Queries correspond to formula patterns

Listing 2: Concordance and formula example tweet

The actual match is indicated by pointed brackets followed by its logical representation.

```
|206700547: saralbennett60 : prosecco <prices will increase if we brexit> :( #strongerin rt @canof  
|=> brexit ⇔ prices increase
```

where $a \Leftrightarrow \psi \equiv$ "were action a to be executed, ψ would be true as a consequence of it"

Work in progress

- ▷ Assemble formulas into structured argument maps connected via argument schemes
 - Enrich argument maps via reasoning capabilities through the logical representation
- ▷ Develop more expressive queries
 - Automatic word list extensions using word embeddings
 - Fine-grained POS tags, phrase chunking, NER, dependency parses
 - Develop a novel corpus query engine which supports querying of dependency parses
- ▷ Interactive web application

Conclusion

- ▷ Each query represents one argument in a particular linguistic context
- ▷ One argumentation scheme corresponds to several queries
- ▷ Expansion beyond purely lexical level through lexico-grammatical patterns
 - logically defined slots