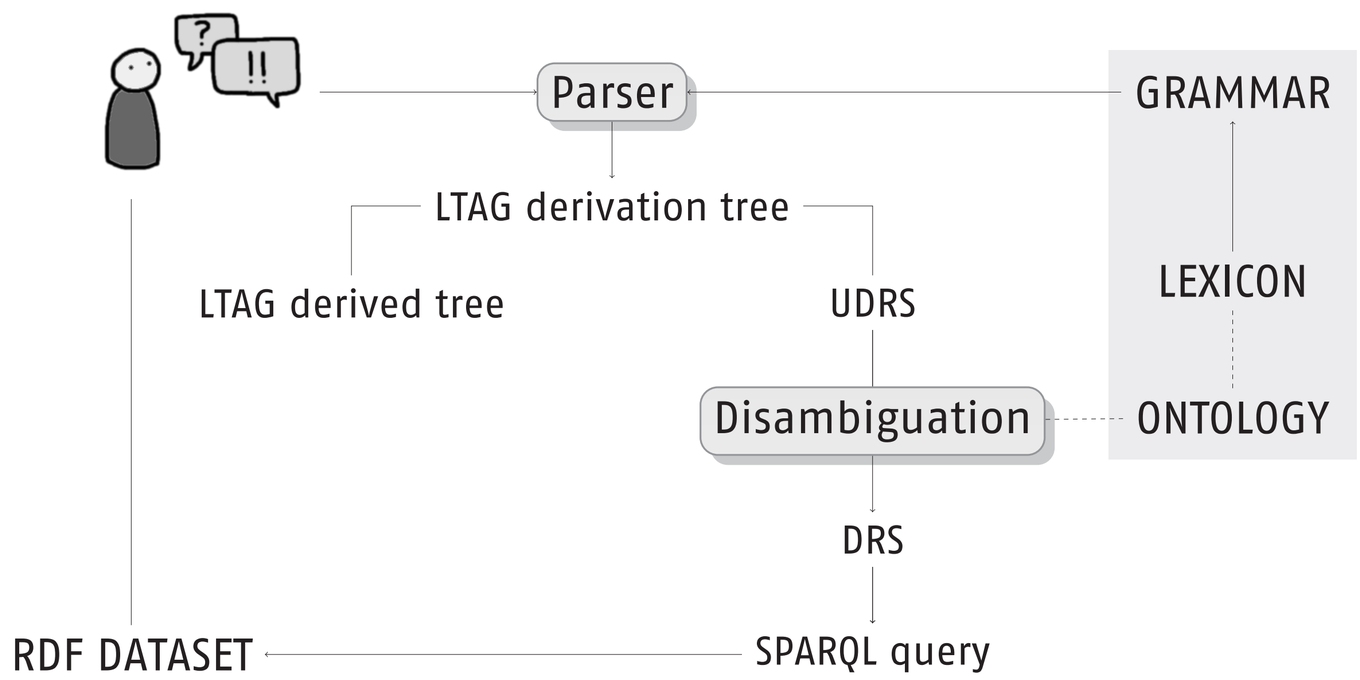
Pythia: Compositional meaning construction for question answering

Christina Unger, Philipp Cimiano
Semantic Computing Group · CITEC · Universität Bielefeld

ARCHITECTURE

Pythia is an ontology-based question answering system that translates natural language questions into formal queries, relying on:

- a compositional construction of general meaning representations
  - captures the semantic structure of the input question
  - covers also complex constructions, such as quantification, comparatives and superlatives, negation, etc.
- a domain-specific grammar generated from an ontology lexicon
  - resulting semantic representations are aligned to a specific ontology, thereby ensuring a precise and correct mapping of natural language terms to corresponding ontology concepts



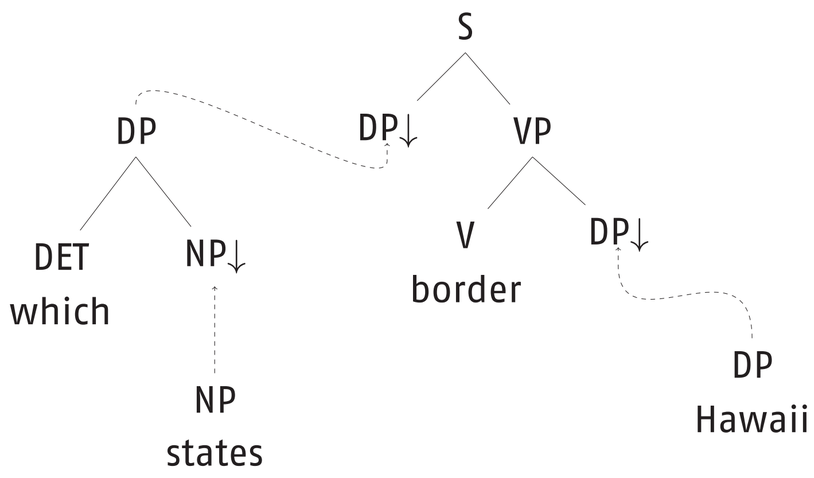
ONLINE PROCESSING STEPS

Parsing and interpretation

Pythia incorporates an Early-style parser for LTAG and a parallel meaning construction that returns underspecified DRSs.

Example

Which states border Hawaii?



?x y
geo:state(x) geo:border(x, y) y = geo:hawaii

```
PREFIX geo: ⟨http://www.geobase.org/⟩
SELECT ?x WHERE {
    ?x rdf:type geo:state .
    ?x geo:borders ?y .
    FILTER(?y==geo:hawaii) .
}
```

Disambiguation

The mapping between natural language expressions and ontology concepts need not be one-to-one.

Examples:

- biggest ⇝ max `geo:area`, max `geo:population`
- to have ⇝ `geo:inState`, `geo:flows_through`

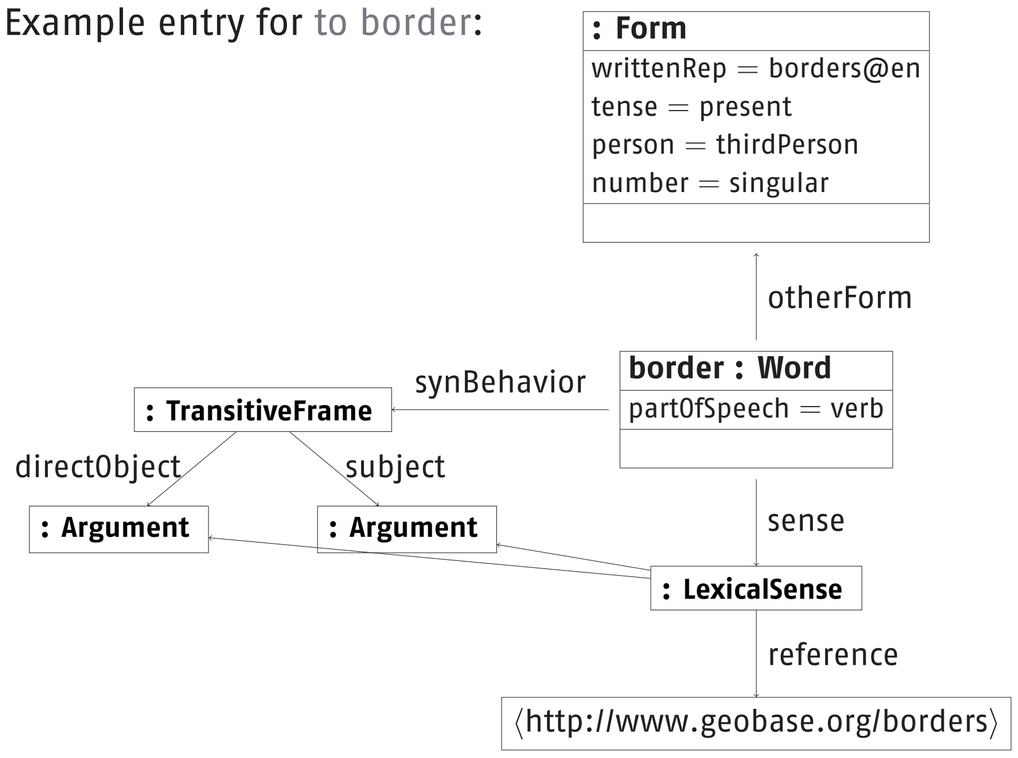
Pythia resolves ambiguities by means of ontological reasoning, exploiting sortal restrictions in meaning representations.

OFFLINE GRAMMAR CREATION

Ontology lexicon

A **lemon** lexicon specifies linguistic realisations of ontology concepts, in particular word forms, morphological properties, subcategoriziation frames and how syntactic and semantic arguments correspond to each other.

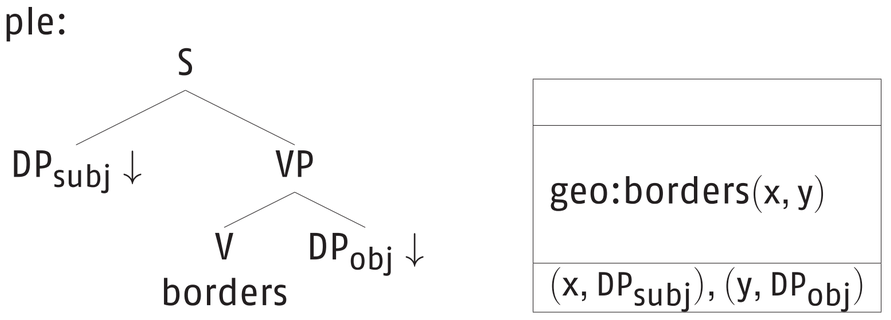
Example entry for to border:



Grammar generation

Grammar entries consist of a syntactic representation (LTAG tree) and a semantic representation (extended UDRT).

Example:



Grammars consist of a domain-independent part (wh-words, determiners, auxiliary verbs, etc.) and a domain-specific part, which is automatically generated from an ontology lexicon.