LOME: Large Ontology Multilingual Extraction

Patrick Xia*, Guanghui Qin*, Siddarth Vashishtha, Yunmo Chen, Tongfei Chen,
Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, Benjamin Van Durme

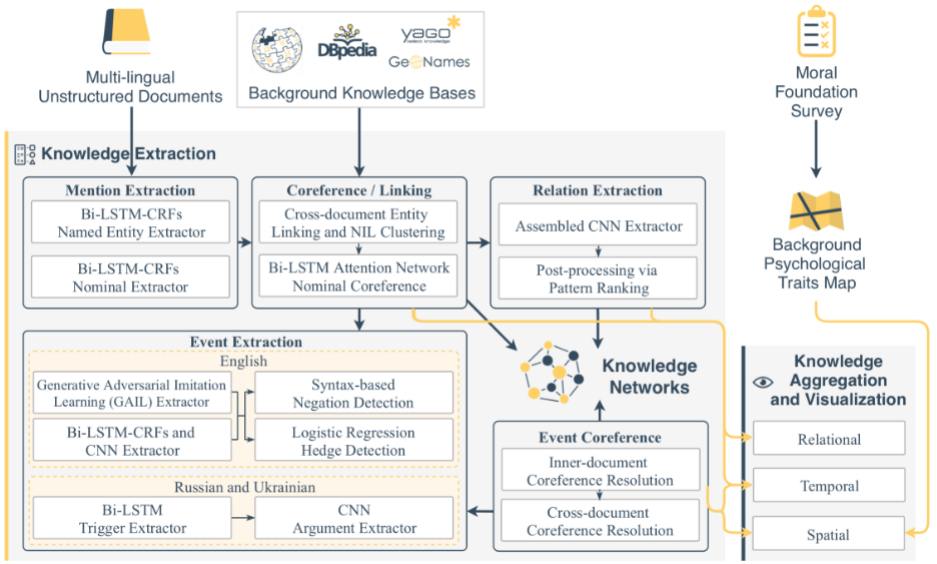
2021 EACL System Demonstrations

JOHNS HOPKINS UNIVERSITY

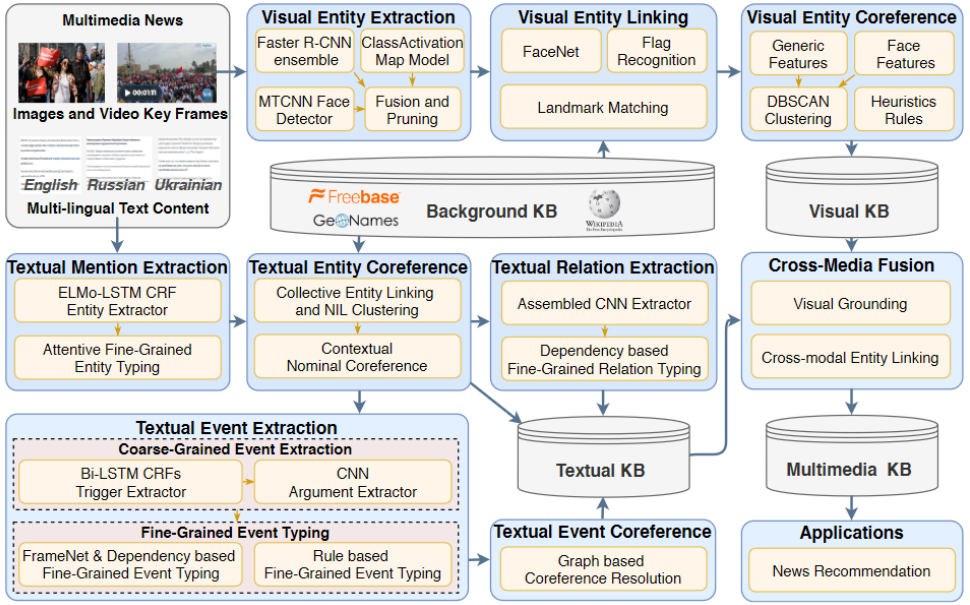
UNIVERSITY of ROCHESTER

Background

- Recent advances in IE →
  - More detailed and fine-grained predictions
  - Document-level > Sentence-level
  - Multilingual
- Previous multilingual IE limitations:
  - separate subsystems per language (→)
  - limited scope of the task (Pan et al., 2017)



Li et al., 2019

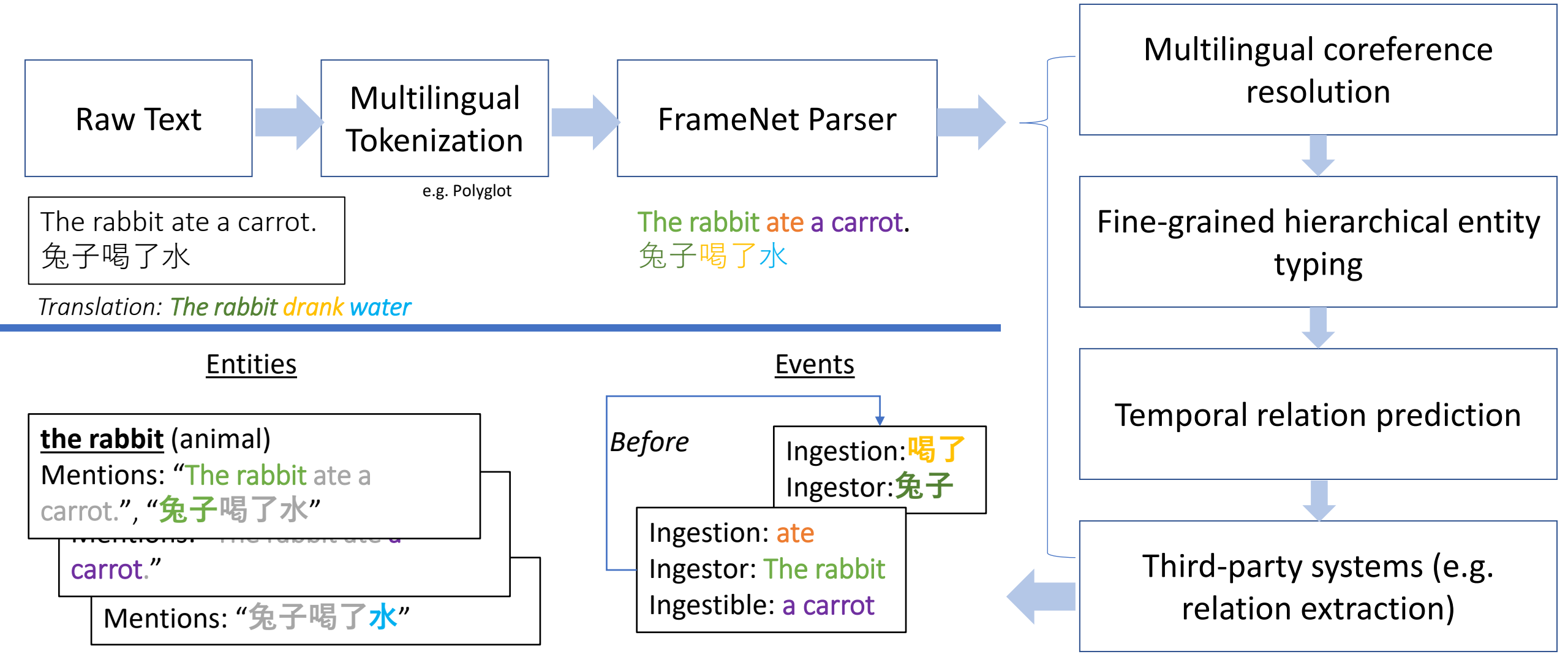


Li et al., 2020 (GAIA)

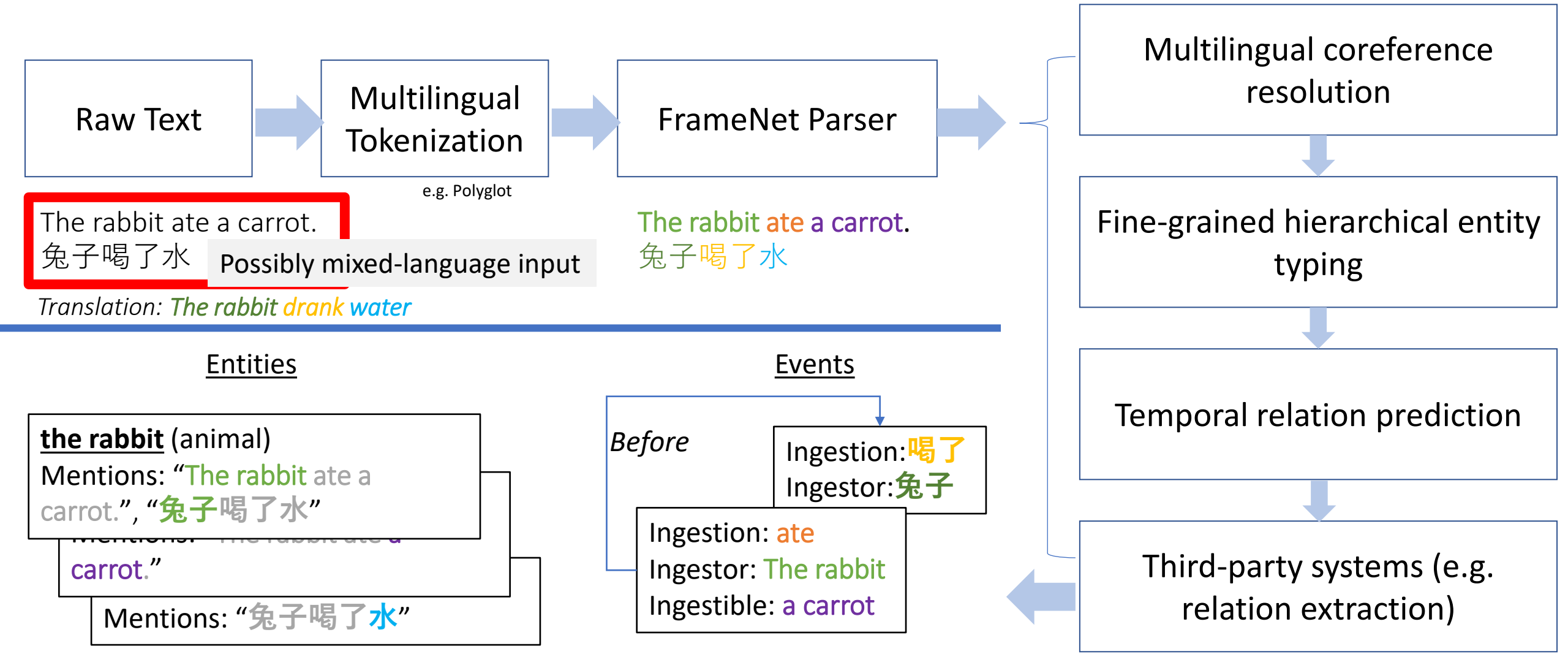
LOME: Large Ontology Multilingual Extraction

- A single, modularized multilingual system
- **Input**: document in any language (supported by XLM-R)
- **Output**:
  1. FrameNet parse (events and arguments)
  2. Coreference linking
  3. Entity typing (fine-grained, cluster-level)
  4. Temporal relation between events

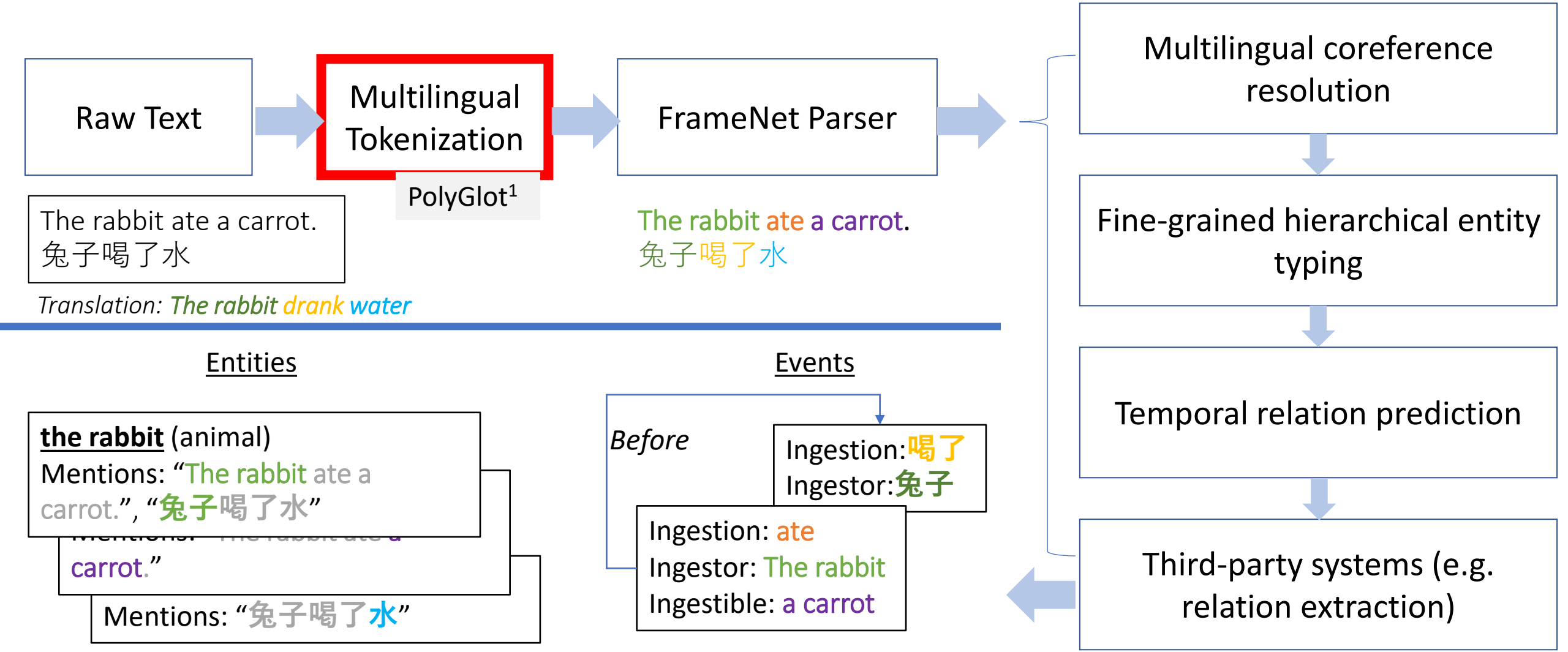
LOME Architecture



LOME Architecture

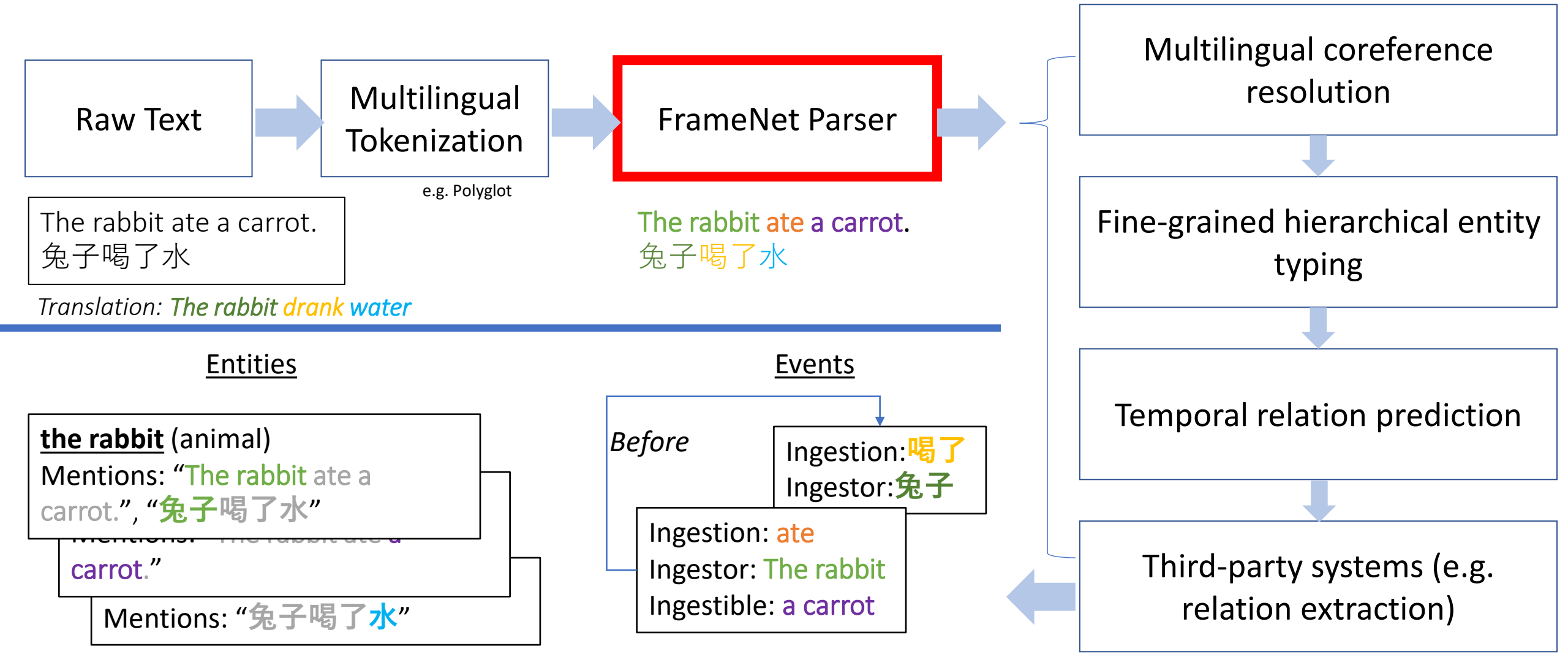


LOME Architecture



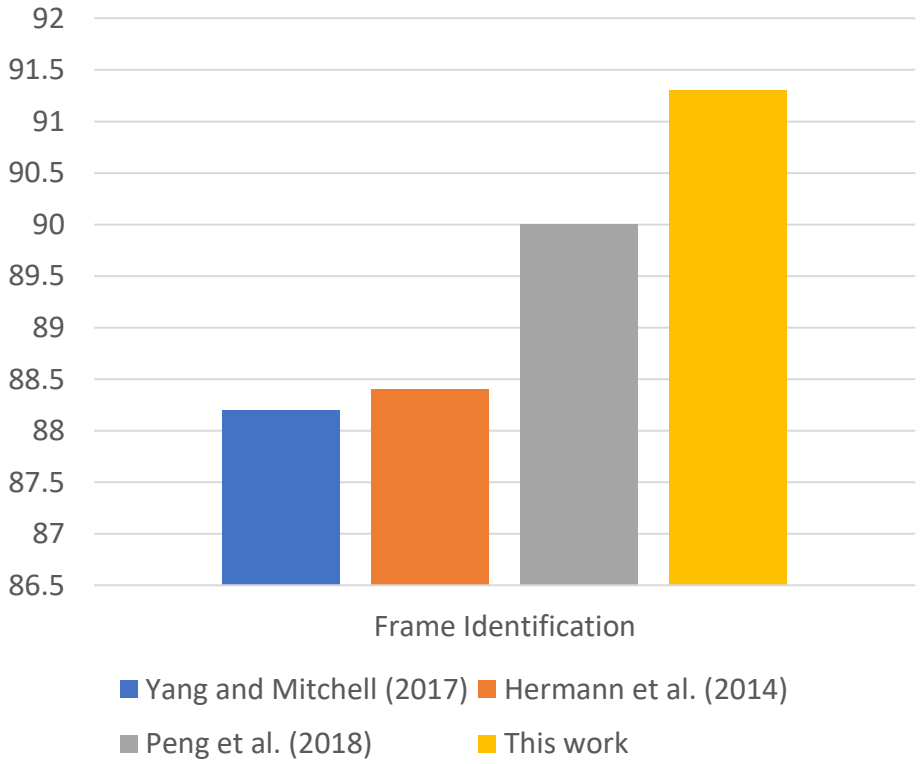
[1] https://github.com/aboSamoor/polyglot

LOME Architecture

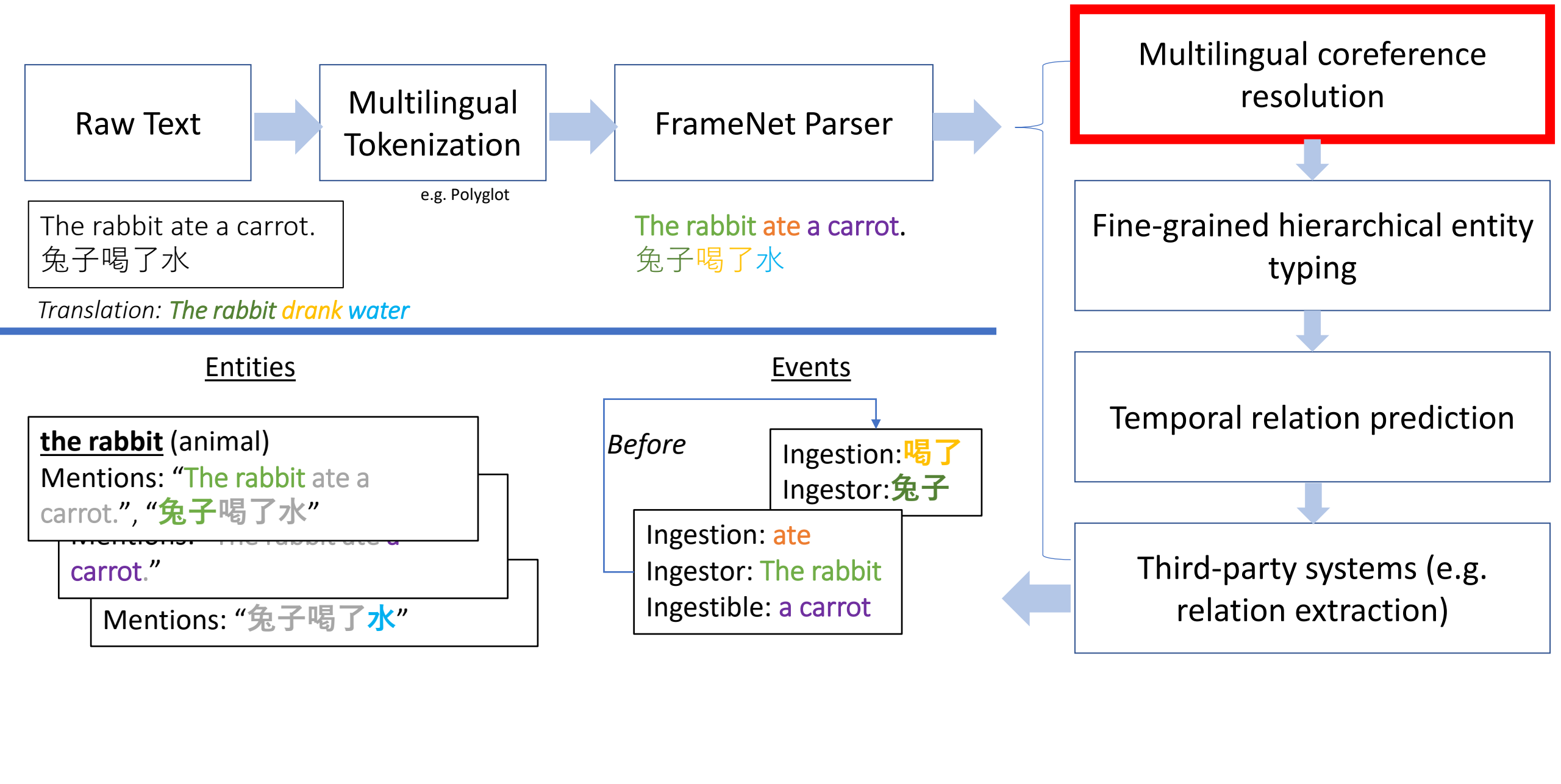


FrameNet Parser

- Goal: Find trigger spans and arguments, then label with FrameNet roles
- Model:
  - XLM-R encodings
  - BIO tagger to find trigger spans
  - Typing module for labeling spans
- Data: FrameNet v1.7
- SOTA on Frame ID
- First to report on full FrameNet parsing

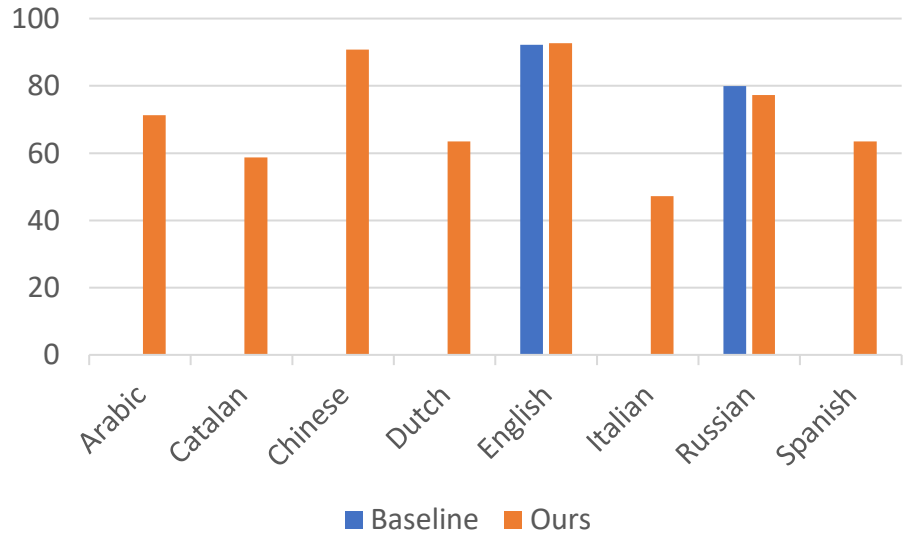


LOME Architecture

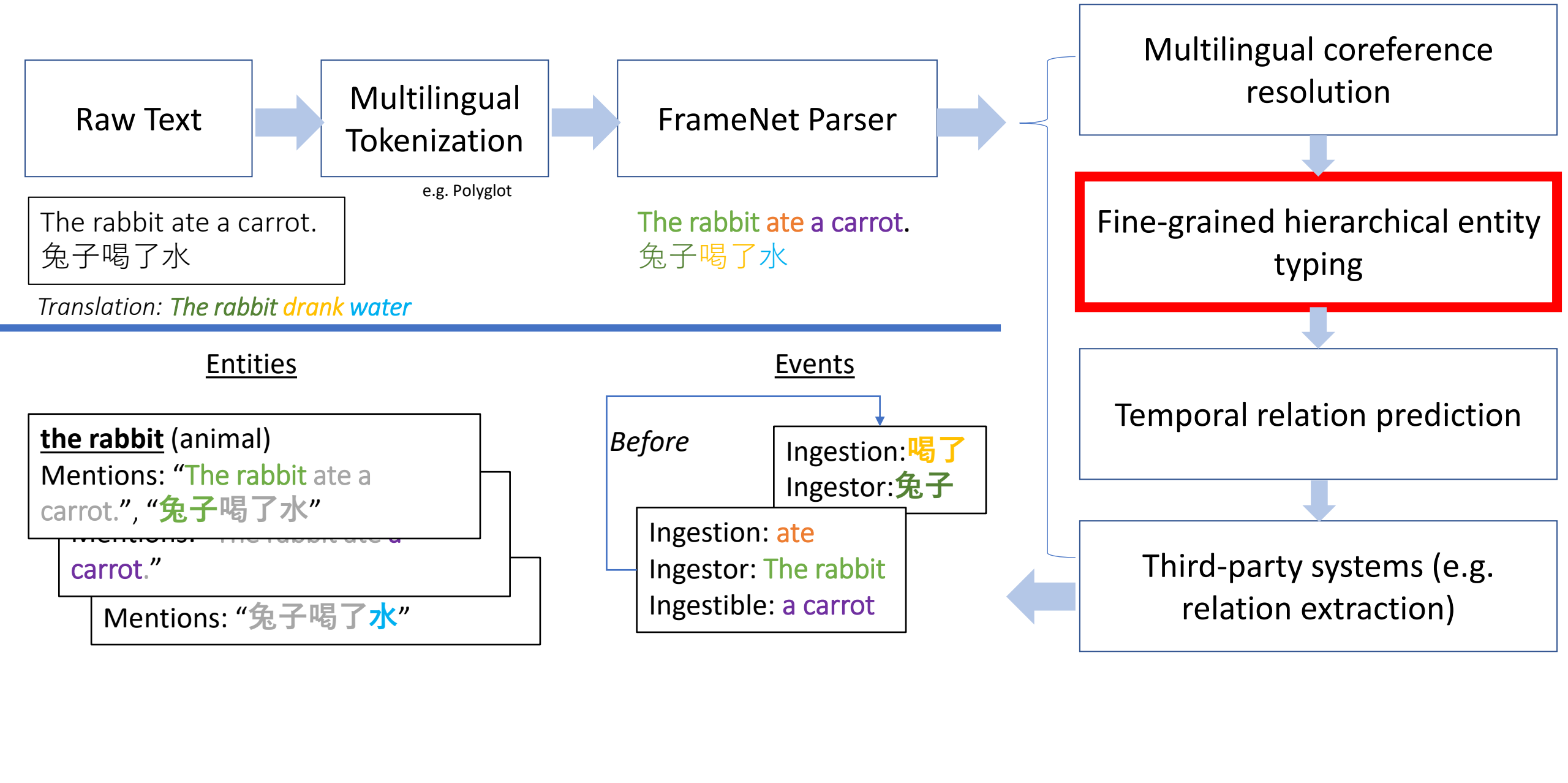


Coreference Resolution

- Goal: Determine which **given** mentions refer to the same entity
- Model: Modified Incremental Coreference Model (Xia et al., 2020) + XLM-R
- Data: OntoNotes 5.0, SemEval 2011 Task 1,
  Russian RuCor/AnCor
- Results: matches SpanBERT (Joshi et al., 2020) in English,
  comparable to RuBERT for Russian,
  sets neural baselines

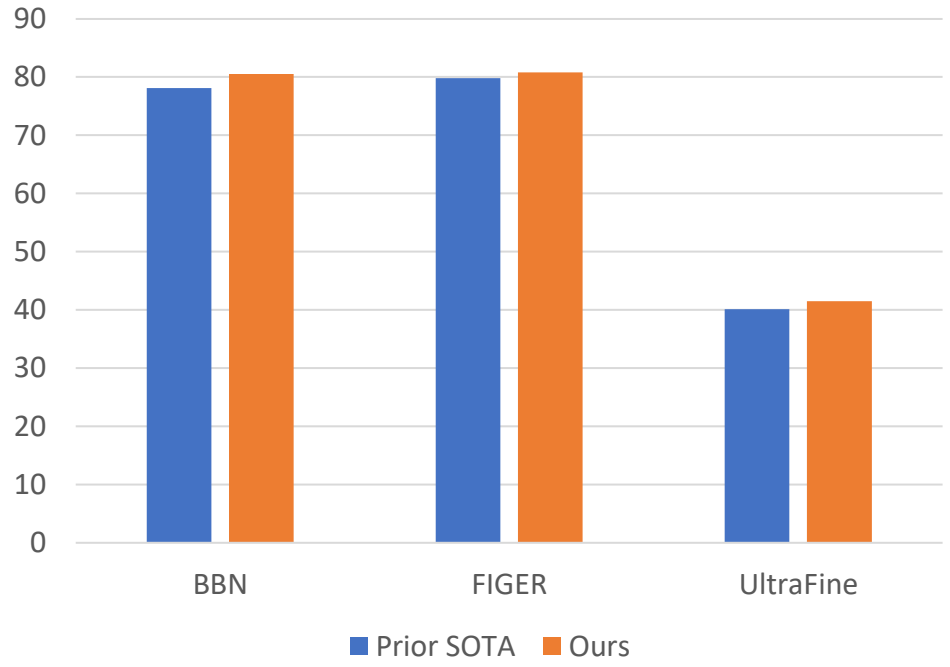


LOME Architecture

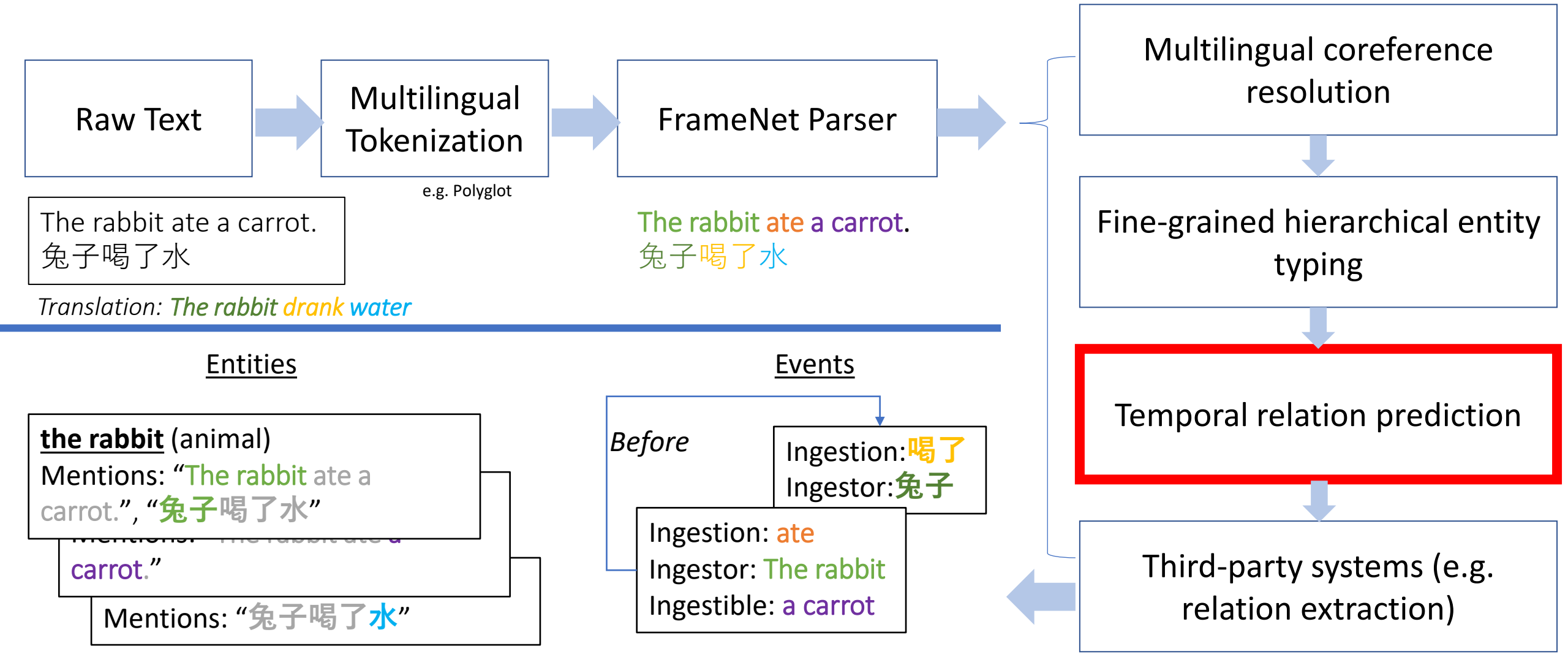


Hierarchical Entity Typing

- Goal: associate entities (*mention clusters*) with fine-grained hierarchical entity types
- Model: Coarse-to-fine decoder (Chen et al., 2020) + XLM-R + *Borda Voting*
- Data: various (AIDA Ontologies, BBN, FIGER, UltraFine)



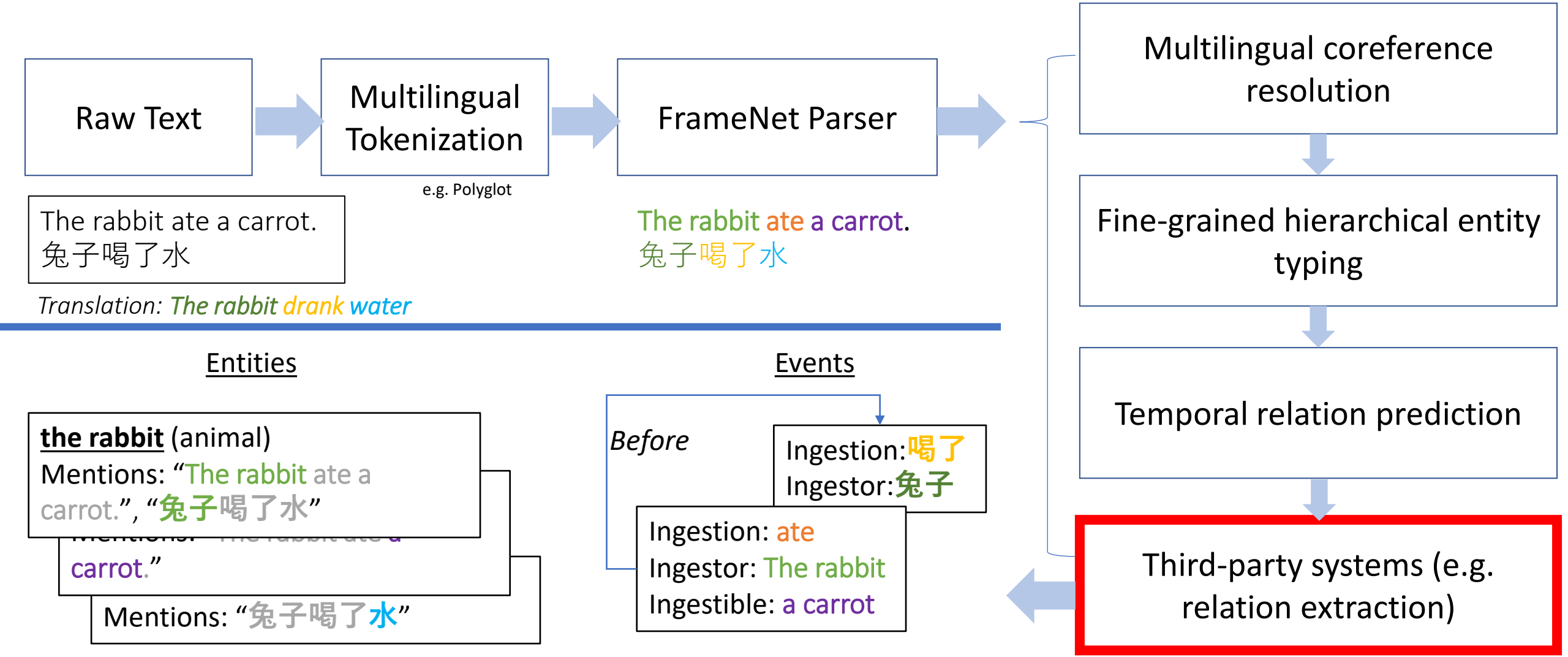
LOME Architecture



Temporal Relation Prediction

- Goal: given two events, predict their temporal relation (e.g. before, overlap, etc)
- Model: Real-valued event pairs (Vashishtha et al., 2019) + XLM-R
- Data: TimeBank Dense, TempEval3, Chinese corpus
- Cross-lingual performance

LOME Architecture



Applications

- TAC SM-KBP 2020 Task 1
  - Pipeline: GAIA (Li et al., 2020) → Coref → Entity Type → argument linking
- Schema inference tasks also needs relations
  - Attach OneIE (Lin et al., 2020) to output of LOME

LOME is modular: designed to be mix/matched with other systems

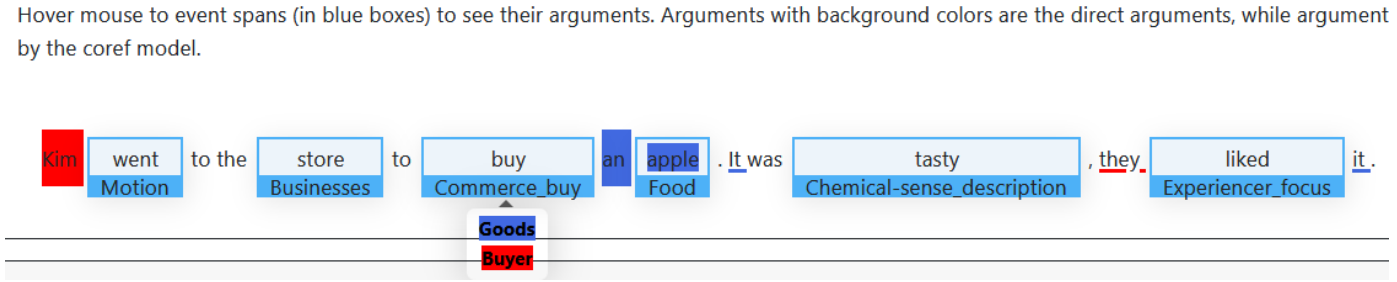
Website: https://nlp.jhu.edu/demos/

- Available on Docker Hub
- Concrete (Ferraro et al., 2014) format outputs

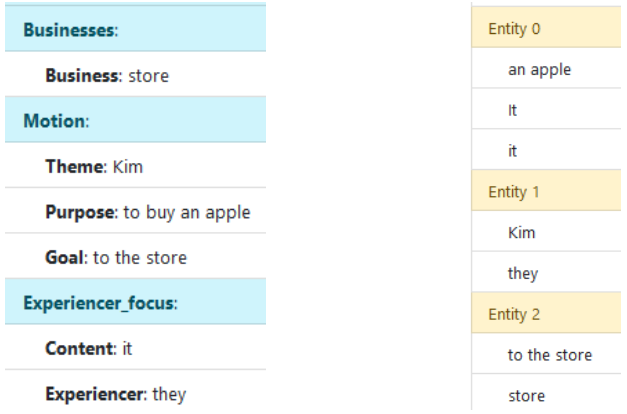
```
SituationMention 0-0:
        text:                   March
        situationType:          EVENT
        situationKind:          Calendric_unit
        Argument 0:
           None
            role:               Unit
            entityMention:      March
```

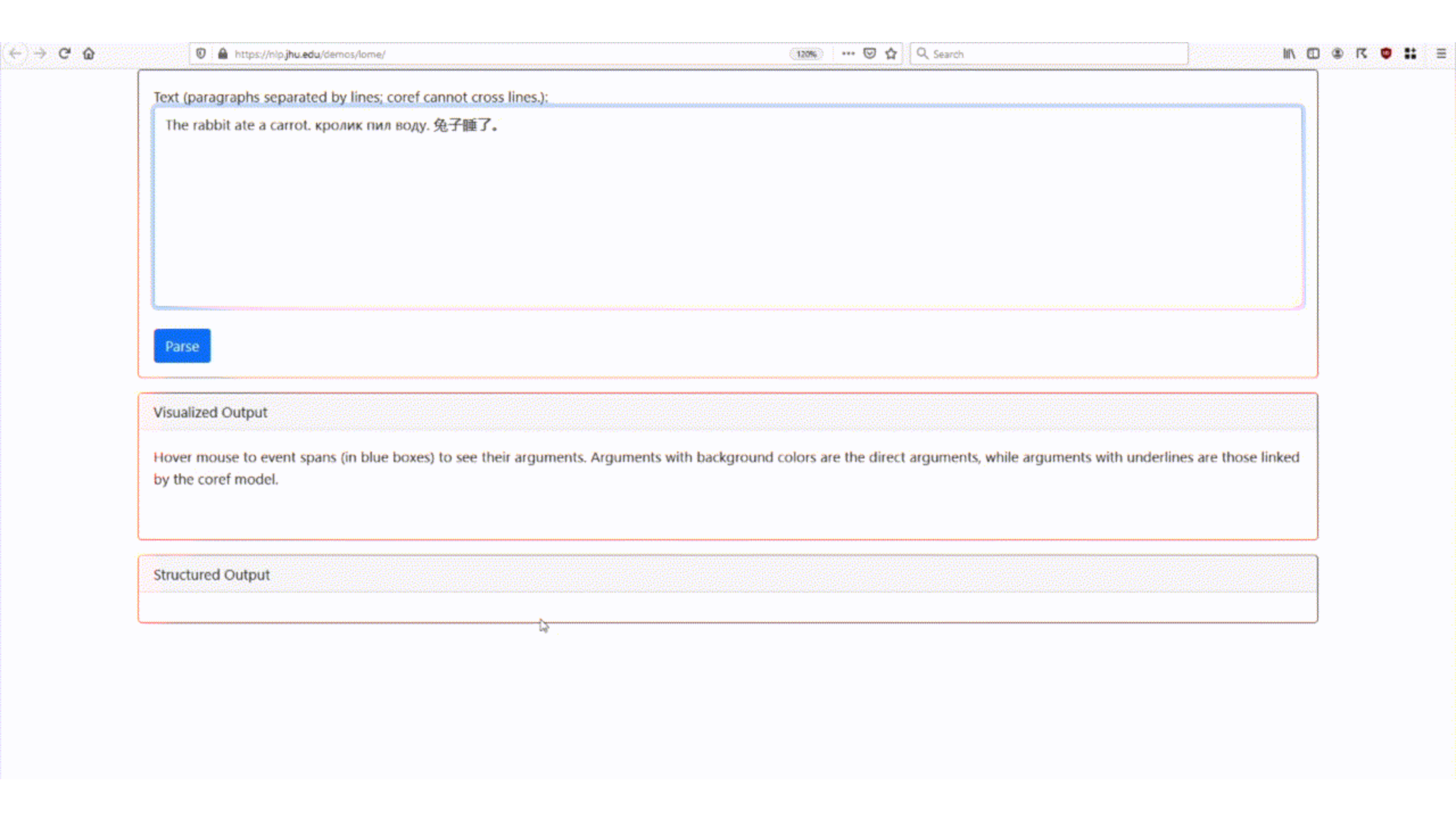
```
Entity 1-0:
 Entity Type: VAL.Number.Number
    EntityMention 1-0-0:
        tokens:     on March 10th , 2013
        text:       on March 10th , 2013
    EntityMention 1-0-1:
        tokens:     March
        text:       March
    EntityMention 1-0-2:
        tokens:     March 10th , 2013
        text:       March 10th , 2013
```

- Web Demo
- Interactive output visualization



- Structured Frame/entity output




https://nlp.jhu.edu/demos/lome/
120%
Search
Text (paragraphs separated by lines; coref cannot cross lines.):
The rabbit ate a carrot. кролик пил воду. 兔子睡了。
Parse
Visualized Output
Hover mouse to event spans (in blue boxes) to see their arguments. Arguments with background colors are the direct arguments, while arguments with underlines are those linked by the coref model.
Structured Output

Summary

- Single system for multilingual information extraction
  - Highly reliant on strong multilingual encoders and their cross-lingual transfer ability
- Available as a web demo and on Docker

Website: https://nlp.jhu.edu/demos/