Zero-Shot Information Extraction as a Unified Text-to-Triple Translation

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, Dawn Song

UC Berkeley   Tsinghua University

Overview

- Information extraction (IE) is crucial to many NLP applications
  - Limited transferability: there are many IE tasks with different task-specific pipelines
  - We need a unified IE approach
- Our approach: **Text-to-Triple Translation**
  - Unified framework that solves IE tasks
  - Zero-shot IE without the need of any task-specific training set
  - Generalization by transferring the latent knowledge that language models have
  - Better interpretability through the enhanced model transparency
  - State-of-the-art or competitive performance compared to fully supervised methods

Text-to-Triple Translation

- Generating: produces general information about the task via pre-trained language models
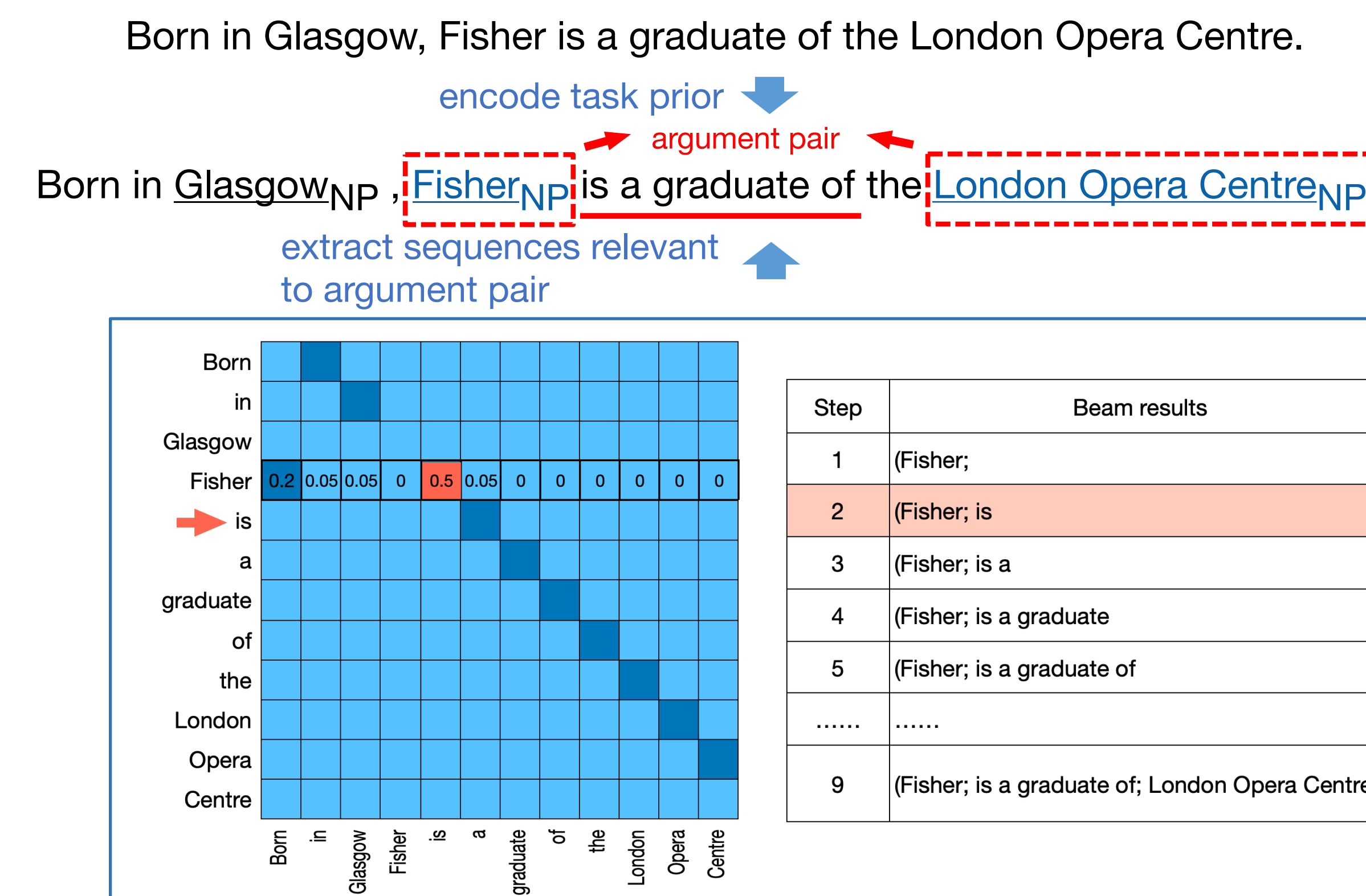
- Ranking: finds triples that are of interest to the task via a ranking model pre-trained on a task-agnostic relational corpus
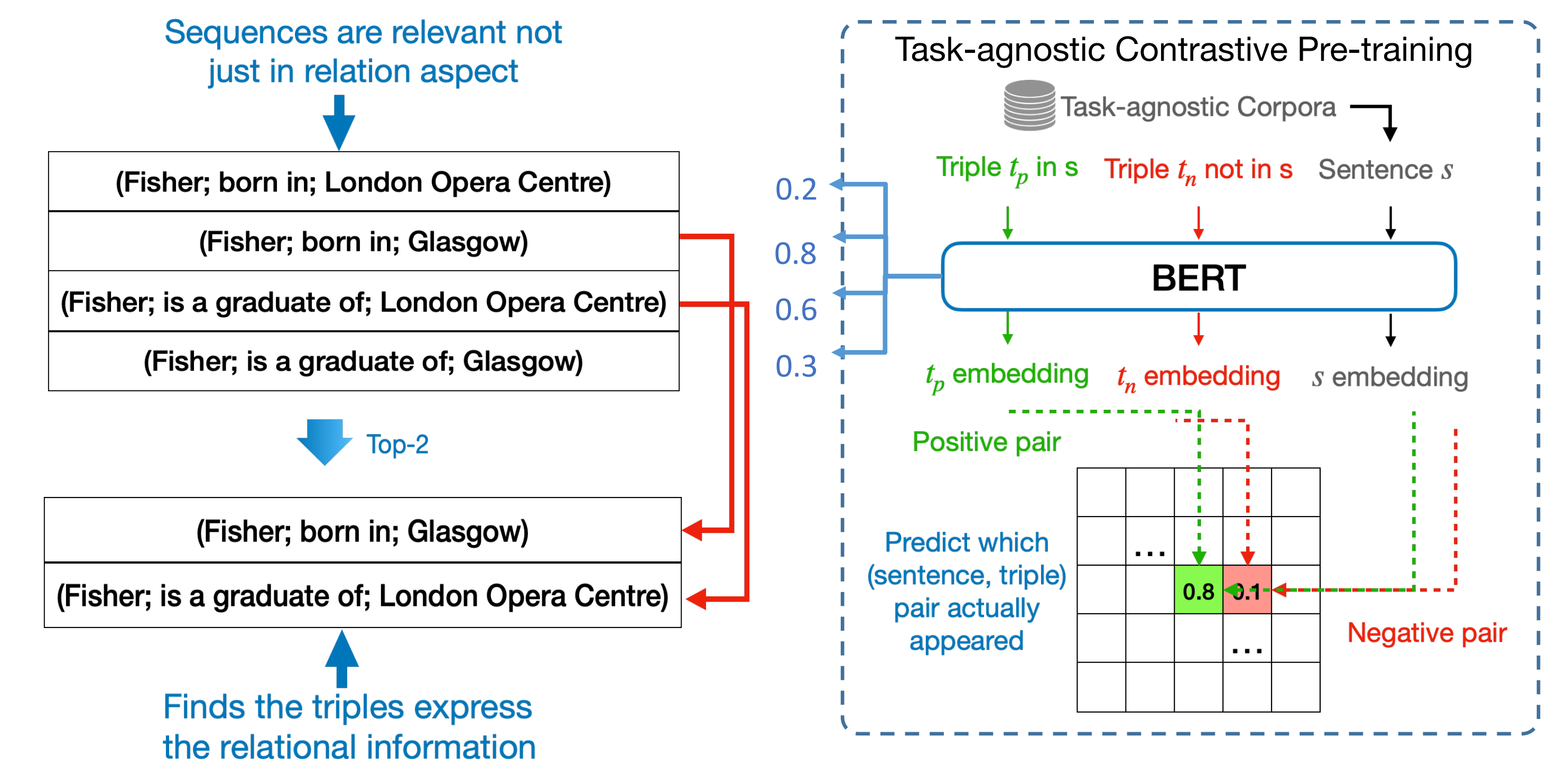
By leveraging the task priors encoded in the input, we enable the zero-shot transfer of the general knowledge that a pre-trained language model has about the task
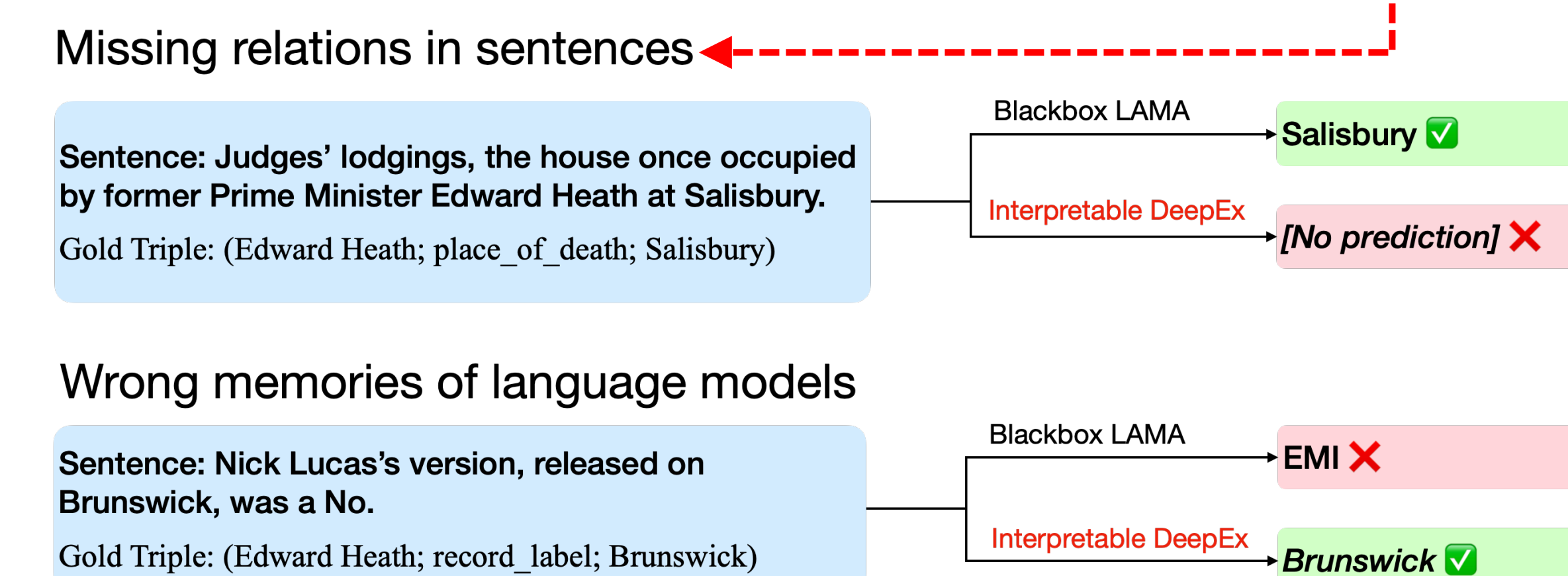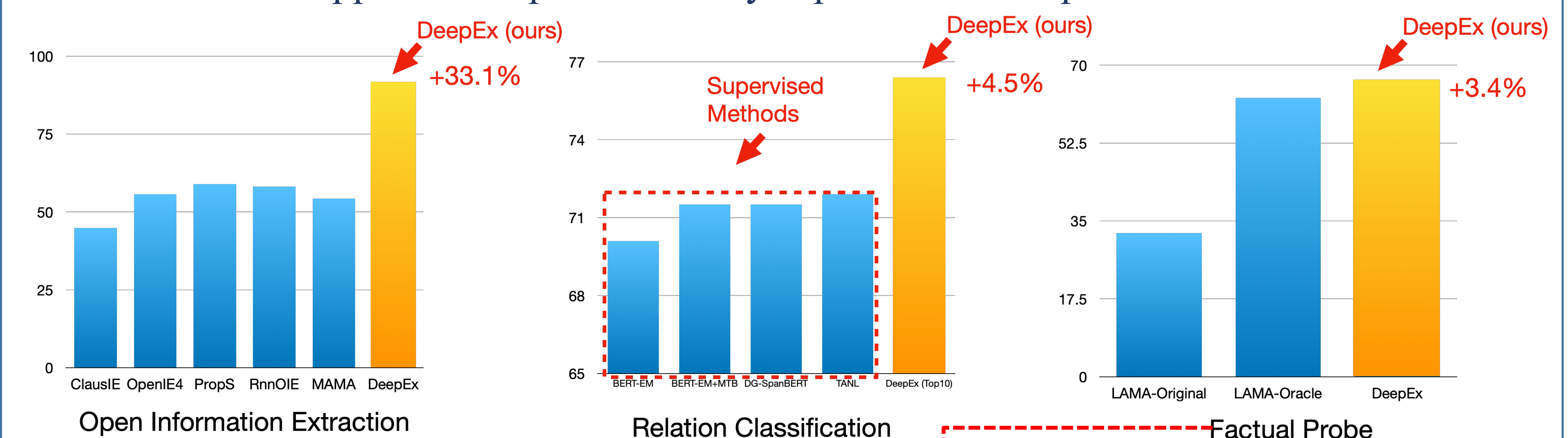
DeepEx

- Same text-to-triple translation is shared across tasks, the only difference is the input encoding
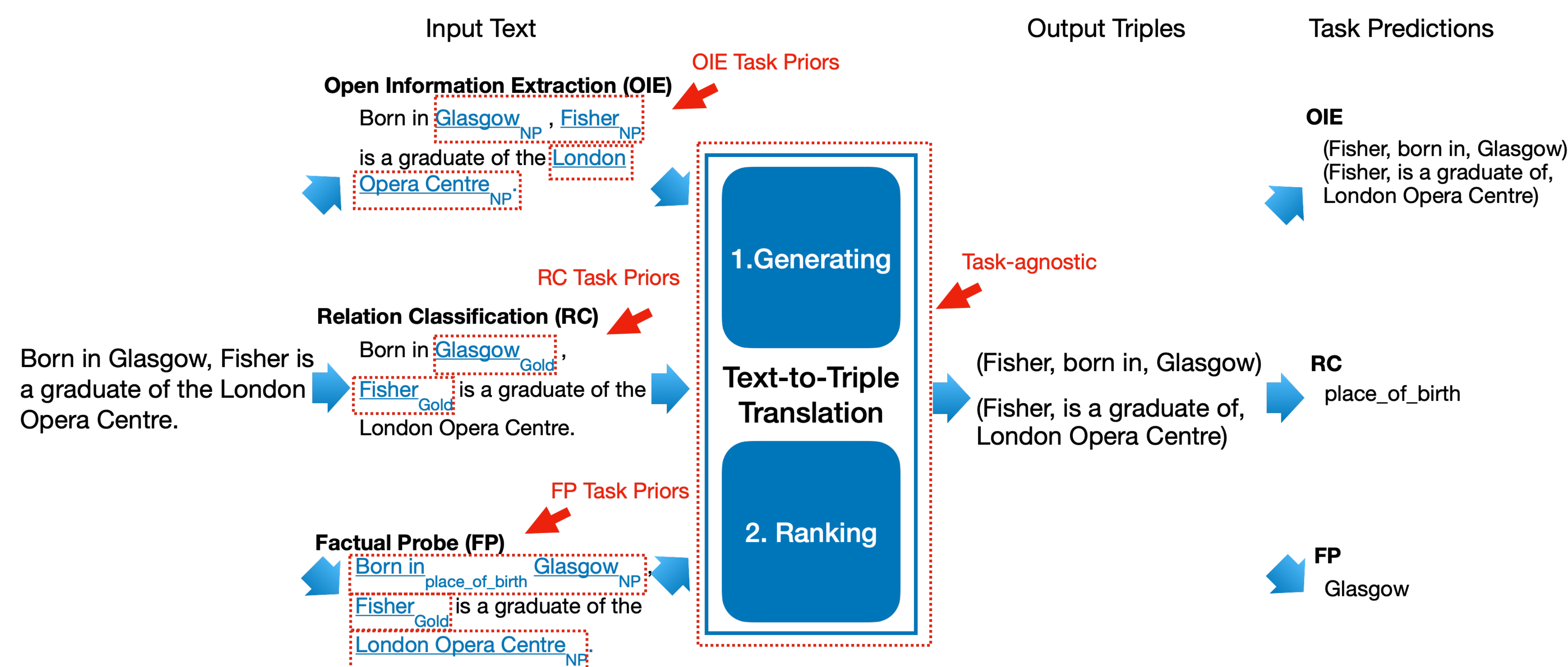
The basic idea: treat every information extraction problem as a "text-to-triple" problem, i.e., translating input text to output triples, then decoding into task predictions

Results

- Our unified approach achieves state-of-the-art or competitive results on all three tasks
- Our zero-shot approach outperforms fully supervised task-specific models

Open Information Extraction   Relation Classification   Factual Probe

Missing relations in sentences

Sentence: Judges' lodgings, the house once occupied by former Prime Minister Edward Heath at Salisbury.
Gold Triple: (Edward Heath; place_of_death; Salisbury)
Blackbox LAMA → Salisbury ✅
Interpretable DeepEx → [No prediction] ❌

Wrong memories of language models

Sentence: Nick Lucas's version, released on Brunswick, was a No.
Gold Triple: (Edward Heath; record_label; Brunswick)
Blackbox LAMA → EMI ❌
Interpretable DeepEx → Brunswick ✅

Get our code!

- Our method delivers more interpretable results thanks to the enhanced transparency