

The Role of Global and Local Context in Named Entity Recognition

Arthur Amalvy¹ Vincent Labatut¹ Richard Dufour²

¹Laboratoire Informatique d'Avignon {firstname.lastname}@univ-avignon.fr

²Laboratoire des Sciences du Numérique de Nantes {firstname.lastname}@univ-nantes.fr

June 26, 2023



Context in Named Entity Recognition (NER)

*"Raoden stood, and as he did, his eyes fell on **Elantris** again."*

Is **Elantris** a **PER** (person) or a **LOC** (location)? To answer, we need **context**.

- Two levels: local (close to the sentence) or global (document level)
- Issue: Range limitations of transformers-based models for global context
- Is it interesting to retrieve global/document-level context?

Context in Named Entity Recognition (NER)

*"Raoden stood, and as he did, his eyes fell on **Elantris** again."*

Is **Elantris** a **PER** (person) or a **LOC** (location)? To answer, we need **context**.

- Two levels: local (close to the sentence) or global (document level)
- Issue: Range limitations of transformers-based models for global context
- Is it interesting to retrieve global/document-level context?

A New Literary Dataset

To study the role of global context, we re-annotate an existing English NER literary dataset [DKE19].

- First chapter of 40 novels
- We fix PER dataset errors with a consistent annotation scheme
- We annotate LOC and ORG entities
- Freely available¹!

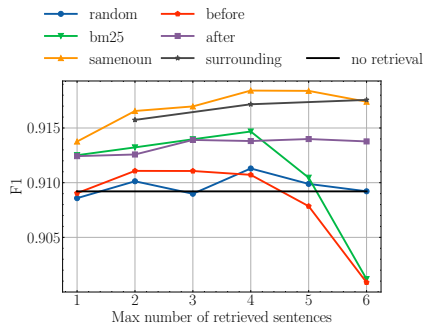
¹<https://github.com/CompNet/conivel/tree/ACL2023>

Experiment: Comparing Local and Global Context

Is it useful to retrieve global context? We compare:

- local heuristics: `before`, `after`, `surrounding`
- global heuristics: `random`, `bm25`, `samenoun`

Experiment: Comparing Local and Global Context

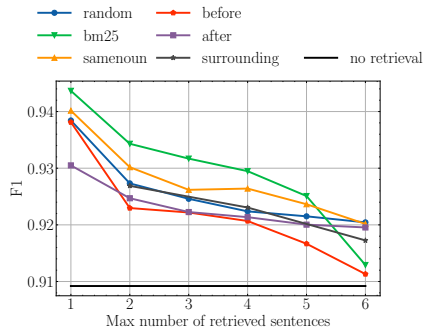


- Context leads to increased performance for all methods
- Global context is competitive even with simple heuristics
- Stronger methods may yield performance gains

Experiment: Evaluating Re-ranking Oracles

- We evaluate a "retrieve and rerank" procedure with a perfect oracle re-ranker
- Performance upper-bound

Experiment: Evaluating Re-ranking Oracles



- Oracle re-rankers perform best when ranking over global context
- This result prompts for further research in training re-ranker models

Thank you!

Code, data and paper at <https://github.com/CompNet/conivel/tree/ACL2023>