

Advances in Natural Language Processing: Bridging Text and Knowledge via Grounding and Innovative Applications

Edoardo Barba, Pere-Lluís Huguet Cabot, Riccardo Orlando, Roberto Navigli

Sapienza University of Rome

WG1 WG3 WG4

Problem

Lexical Ambiguity
Machine Translation

Solution

Code-Switching + Word Senses!

Reconstruction or Translation

Pretraining Noisification Strategy

Methodology

I like singing and dancing

Noisification

I like *cantare* and *bailar*

1) Reconstruction

I like *singing* and *dancing*

2) Translation

J'aime chanter et danser

Transfer learning via noisification and reconstruction of the original input sentence / translation in a target language

Dataset Creation

Dataset Creation:
 Word Sense Disambiguation + BabelNet

I like *singing* and *dancing*

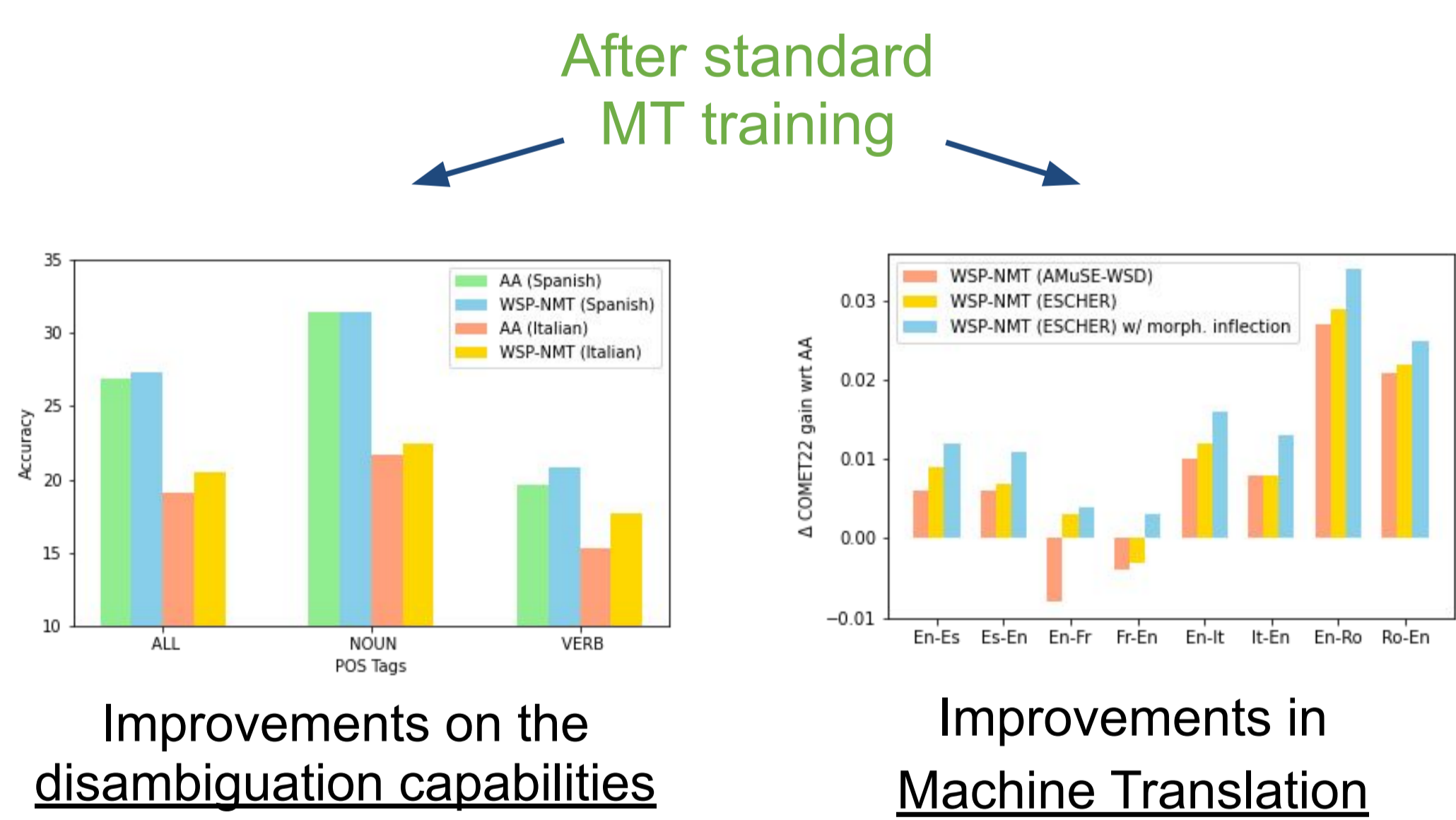
Word Sense Disambiguation

BabelNet

ballare
danzare
bailar
danser

cantare
canter
chanter

Results



Conclusion

- **Takeaways.** Our research shows that:
 - The integration of **explicit semantics** into LLMs improves the quality of AI systems
 - KGs are a rich, reliable source of information to **ground "black box" LLMs**
- **Ongoing and Future work:**
 - Improving generative AI systems (e.g., for MT) with **other explicit semantic signals** (e.g., entities retrieved from knowledge graphs)
 - Rethinking how we evaluate generative AI systems by **leveraging semantic structures**

Work In Progress

- **Unifying Word Sense Disambiguation Entity Linking and Relation Extraction**
- **Large Semantically annotated Dataset Creation for training of Semantically enhanced Large Language Models**
- **Common framework for knowledge-intensive span based classification**

