



WG3

# Probing Coreference Models of Romance Languages using Multiword Expressions

Evelin Amorim

INESC TEC, Porto, Portugal

## The Coreference Task

Who is the entity mentioned by some expression?

O usuários antigos eram os que mais reclamavam.

Old users were the ones who complained the most.

Glossary: Os usuários mais antigos/Old Users, eram/were, os/the, que/who, mais/the most, reclamavam/complained

## Challenges

- Pro-drop pronouns in Romance languages (except French)
- Address coreferential multiword expressions

## Research Questions

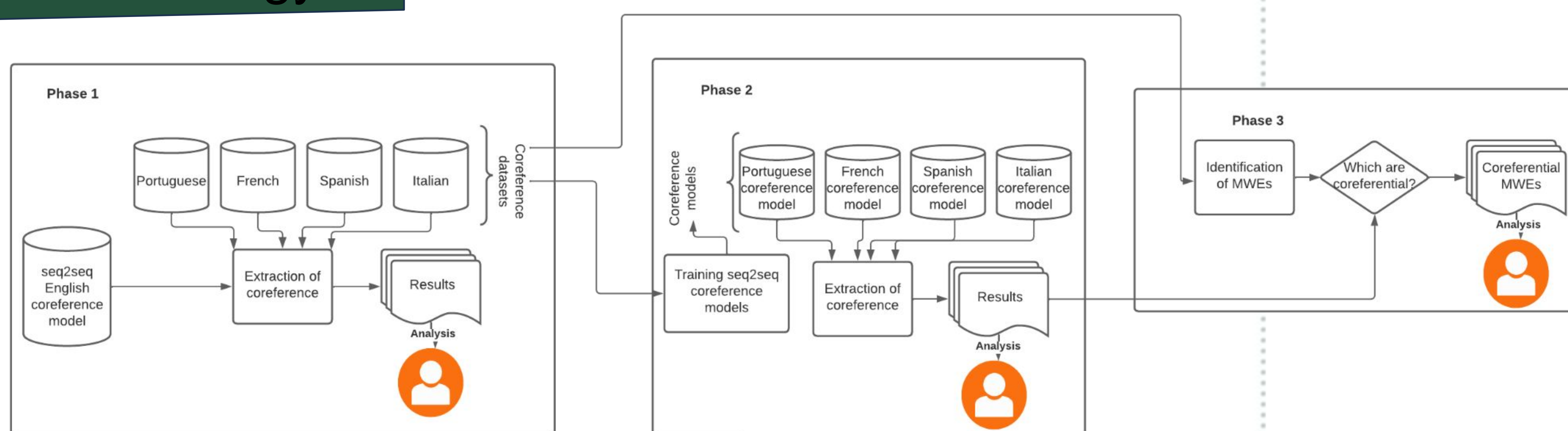
**RQ1:** Could state-of-the-art English coreference models, when used zero-shot in Romance languages, consistently produce similar errors?

**RQ2:** Do few-shot coreference models trained in Romance languages and tested in the same languages exhibit similar errors?

**RQ3:** If consistent errors occur across Romance languages in both zero-shot and few-shot coreference models, can these errors be categorized to offer insights for improving coreference models in these languages?

**RQ4:** To what extent can Multi-Word Expressions (MWEs) influence the prediction errors of coreference models, and can this influence be quantified in Romance Languages?

## Planned Methodology



The architecture for the coreference model will be based on the work "Seq2seq is all you need for coreference resolution.", Wenzheng Zhang, Sam Wiseman, and Karl Stratos. EMNLP 2023

## Expected Outcomes

1. Shedding light on the current state-of-art coreference models for romance languages;
2. Mapping the influence of MWE in the coreference models for romance languages;
3. Improvements in the coreference models for romance languages.

## Contacts:

