

## Responsible NLP Checklist

Paper title: *False Friends or Cognates? A Cross-lingual Semantic Ambiguity Evaluation for Galician, Portuguese and Spanish*

Authors: *Marta Vázquez Abun, Jose Camacho-Collados, Marcos Garcia*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

---

### A. Questions mandatory for all submissions.

- A1. Did you describe the limitations of your work?

*This paper has a Limitations section.*

- A2. Did you discuss any potential risks of your work?

*We discuss potential risks in the Limitations section, including the risk of data contamination (false-friends sources and contextual sentences might have been included in models' pre-training corpora) and the possibility of models hallucinating justifications while producing correct answers, which could mislead interpretations about their true understanding capabilities.*

### B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

- B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?

*Our validation process (Section 3 Data Construction) focused on verifying semantic appropriateness of target words. Wikipedia sentences were extracted automatically based on linguistic criteria and manually validated by linguistic annotators. Human-authored sentences followed the same semantic guidelines. No personally identifying information was included, as sentence selection was based on the presence of target words in appropriate semantic contexts.*

- B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?

*We report dataset statistics in Section 3 (Dataset Construction), including Table 2 which details the number of instances per language pair (704 word pairs, 1,408 sentences). More detailed statistics are provided in Appendix A.*

### C. Did you run computational experiments?

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

*Section 4 describes the experimental setup.*

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean,

---

*The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice. ACL 2026 used a subset of ARR checklist form.*

etc. or just a single run?

*Section 5 (Results) presents performance metrics across all the experiments.*

**D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

*Full instructions are not included. Instead, Section 3 (Annotation task subsection) summarizes the core task given to annotators: verify (i) sense assignment correctness and (ii) semantic relation validity.*

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

*Annotators were research staff with linguistic backgrounds conducting annotation as part of their duties, not separately compensated for this task.*

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

*Data was either from public dumps (Wikipedia) or created directly by annotators as part of their research roles.*

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

*Annotation was conducted by research staff as part of their research duties.*

**E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?**

- E1. If you used AI assistants, did you include information about their use?

*(left blank)*