

## Responsible NLP Checklist

Paper title: *Does Context Matter? A Prosodic Comparison of English and Spanish in Monolingual and Multilingual Discourse Settings*

Authors: *Debasmita Bhattacharya, David Sasu, Michela Marchini, Natalie Schluter, Julia Hirschberg*

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?

*Our work is mainly in the exploratory phase and thus has few associated risks at the current stage. Our work is not currently tied to any particular applications and we do not presently see any paths to negative applications.*

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

#### B1. Did you cite the creators of artifacts you used?

*We cite the creators of the datasets we use throughout the paper, particularly in the Introduction (Section 1) and Data (Section 3) sections. We also cite other artifacts like our prosodic feature extraction method and an off-the-shelf model used in Section 4.*

#### B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

*We state the license information of the data sets we use in Section 3 about the data. We comply with the conditions of use by referring to each corpus by name appropriately.*

#### B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?

*The data sets we use did not specify their intended use. Nonetheless, we use them only for research purposes. Similarly, we do not explicitly note the consistency of our use of other existing artifacts with creator intentions, since these were developed for research purposes such as ours. We do not publicly share the code we wrote, so we do not specify its intended use.*

#### 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?

*We state in the Ethics Statement that we did not access any information that uniquely identifies individuals in the data sets we used. We mention that the original authors had already de-identified each corpus, as outlined in the original data set documentation.*

---

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.

B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?  
*We state the coverage of domains and languages represented in our work in Section 3 about the data. We also refer to the demographic groups represented in Section 1 and Section 3, with additional detail in the Limitations section.*

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 relevant statistics about the data sets in Table 1 and information about train/test splits in Section 4 about the methodology.*

**C. Did you run computational experiments?**

C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?  
*We report the number of parameters in each model, the total computational budget, and computing infrastructure used in Section 4 about the method.*

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?  
*We describe the experimental setup in detail in the Method (Section 4). This includes information about hyperparameter search, best-found values, and model selection.*

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, etc. or just a single run?  
*It is transparent in Tables 2 and 3 and Section 5.2 that we are reporting performance metrics from a single run. We qualify these model results by reporting statistically significant p-values from comparisons via z-tests of proportions, as stated in Section 5.2.*

C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?  
*We provide links to the Disvoice and Optuna packages used in the Method section (Section 4). These do not have explicit version numbers, so we are unable to report them.*

**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.?  
*(left blank)*

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)?  
*(left blank)*

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)?  
*(left blank)*

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

D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?  
*(left blank)*

**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)