A For all submissions:

A1. Did you discuss the limitations of your work? (*)
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in `Limitations' section.

A2. Did you discuss any potential risks of your work? (*)
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in `Ethics Statement & Risks' section.

A3. Do the abstract and introduction summarize the paper's main claims? (*)
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes.

A4. Have you used AI writing assistants when working on this paper?
See our blog for the discussion of some frequent cases. (*)
If you answer Yes, please describe what system it was, what tasks it performed and for which sections of the paper.

No.

B. For submissions using or creating scientific artifacts:

B. Did you use or create scientific artifacts? (*)
If you answer Yes, provide the section number; if you answer No, you can skip the rest of this section.

Yes, in Section 3, Section 4, Section A and Section C.

B1. Did you cite the creators of artifacts you used?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section 3, Section 4, Section A and Section C.

B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, Section A. And as far as we know, there is no explicit license for PBC dataset.

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)?

If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in `Ethics Statement & Risks' section.

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?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in `Ethics Statement & Risks' section.

B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section 3, Section 4, Section A and Section C.

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?

Yes, in Section 3, Section 4, Section A and Section C.

Field description
C For papers running computational experiments:

C. Did you run computational experiments? (*)
If you answer Yes, provide the section number; if you answer No, you can skip the rest of this section.

Yes, in Section 3, Section 4.

C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section3, Section 4, Section B, Section C and `Limitation' section.

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section 3, Section 4 and Section B.

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?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section 3, Section 4 and Section C.

C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
If you answer Yes, provide the section number; if you answer No, provide a justification.

Yes, in Section D.

D. For papers involving human annotation or research with human subjects:

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects? (*)
If you answer Yes, provide the section number; if you answer No, you can skip the rest of this section.

Yes, our human annotator is one the coauthors who is from linguistic background and manually evaluate the results of our method and classify each language to a category. Details in Section 4 and Section C.

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.?
If you answer Yes, provide the section number; if you answer No, provide a justification.

No, not relevant in our case.

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)?
If you answer Yes, provide the section number; if you answer No, provide a justification.

No, not relevant in our case.

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)?
If you answer Yes, provide the section number; if you answer No, provide a justification.

No, not relevant in our case.

D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
If you answer Yes, provide the section number; if you answer No, provide a justification.

No, not relevant in our case.

D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

If you answer Yes, provide the section number; if you answer No, provide a justification.

No, not relevant in our case.