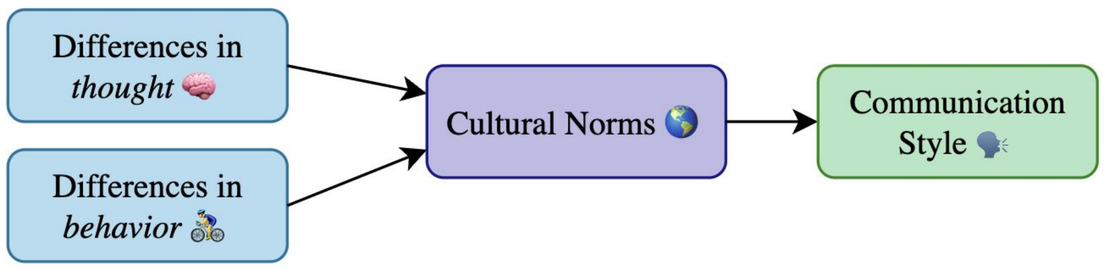
Comparing Styles Across Languages

Penn Engineering

Shreya Havaldar, Matthew Pressimone, Eric Wong, Lyle Ungar

Style varies cross-culturally

- Communication practices, specifically *linguistic styles* (like politeness), vary across cultures.



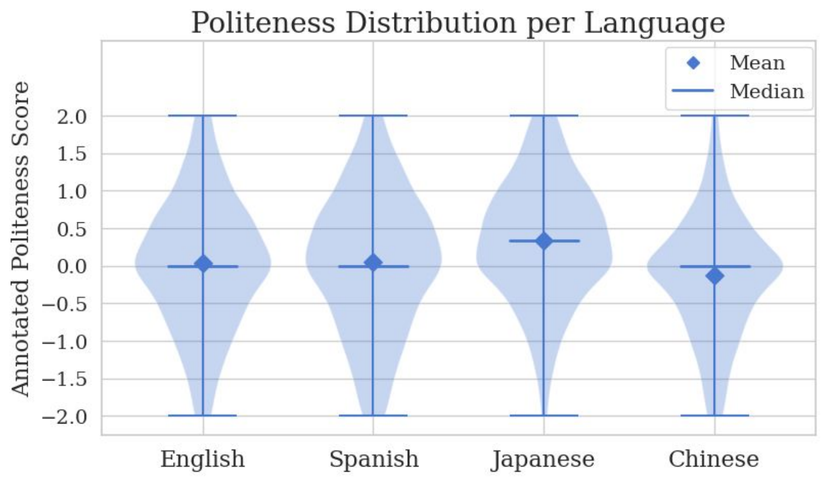
- But, multilingual LMs struggle to generate stylistically appropriate language in non-English languages.

We seek to understand how styles, like politeness, differ across languages.

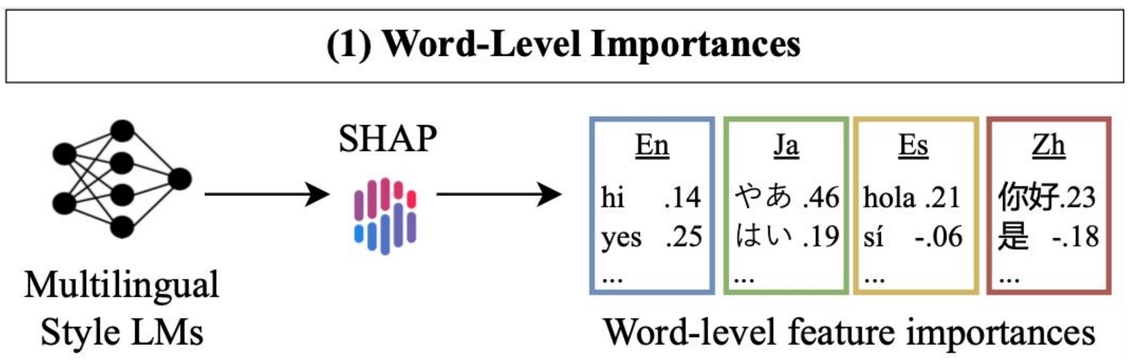
Creating a holistic politeness dataset

Using Wikipedia editor talk pages, we create the first multilingual politeness dataset to:

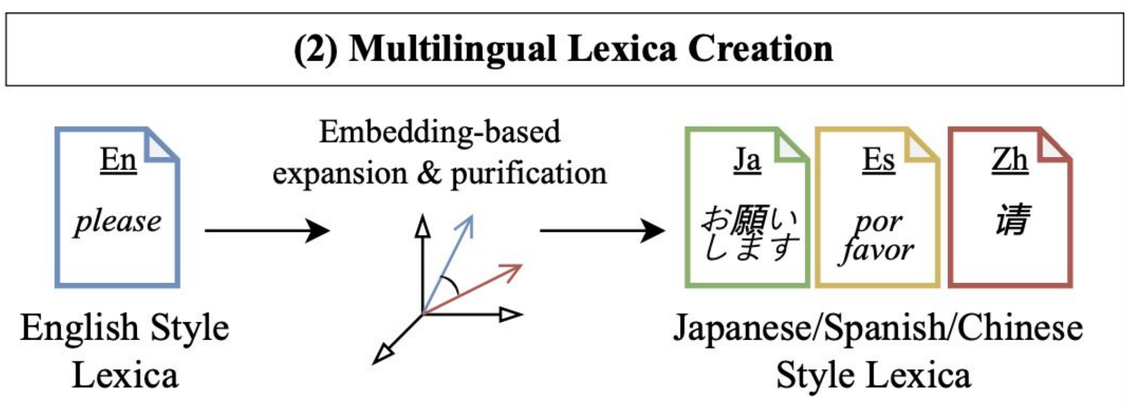
1. Cover *all dialog acts,* not just questions
2. Include *all annotated data*, neutrality along with politeness and rudeness
3. Treat politeness prediction as a *regression task*, not a binary classification



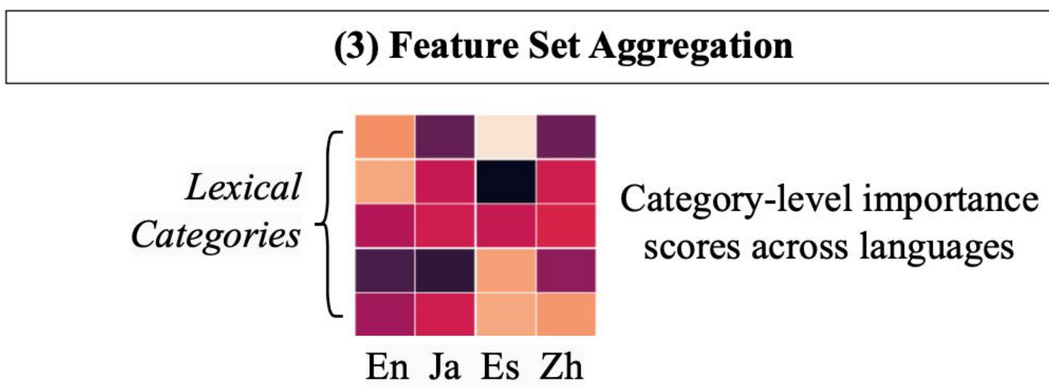
A framework for style comparison



We fine-tune four XLM-RoBERTa models (one per language) on our holistic politeness dataset and extract token-level Shapley values for every utterance. We then aggregate the token-level Shapley values to word-level importance scores.

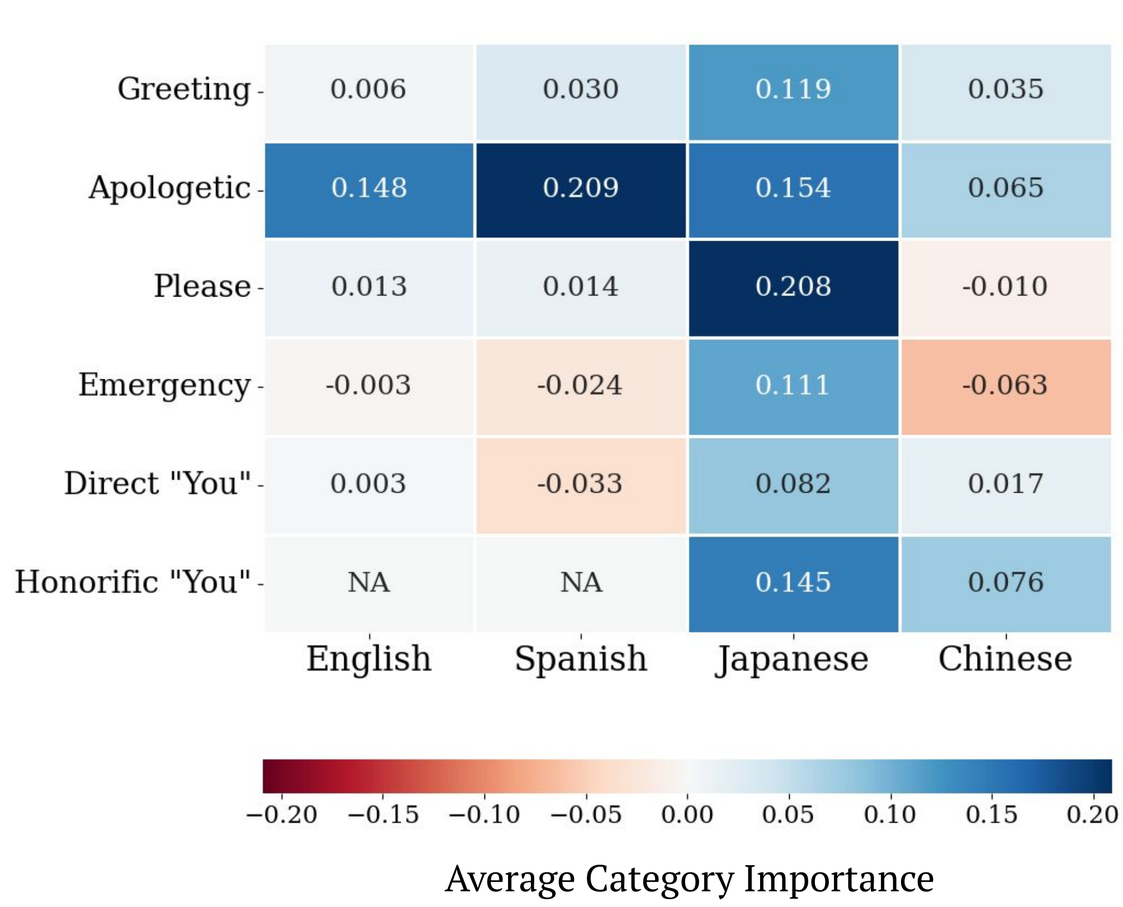


Style is subjective and expressed differently across languages, so standard 1:1 translation of style lexica is flawed. We use embedding-based expansion and purification techniques to improve lexica translation.



Step (2) gives us parallel lexical categories, like "*Greeting*" or "*Apologetic*", across languages. We aggregate word-level importance scores from (1) into the lexical categories from (2) to get a comparison of how style differs multilingually.

Visualizing differences in politeness



Takeaways + future work

- Our framework provides an explanation of how style differs across languages that is *faithful* as well as *interpretable.*

- Future work can use our methodology to:
  - Inform culturally-adaptable LMs
  - Help people learning a second language to understand stylistic nuances and improve fluency

Our paper

Code + Data

My website