

## Overview

- Language captures cultural differences through style
  - Hofstede's Cultural Dimensions
  - Low vs. High Context
  - Ideal Affect

## Issues with Translation

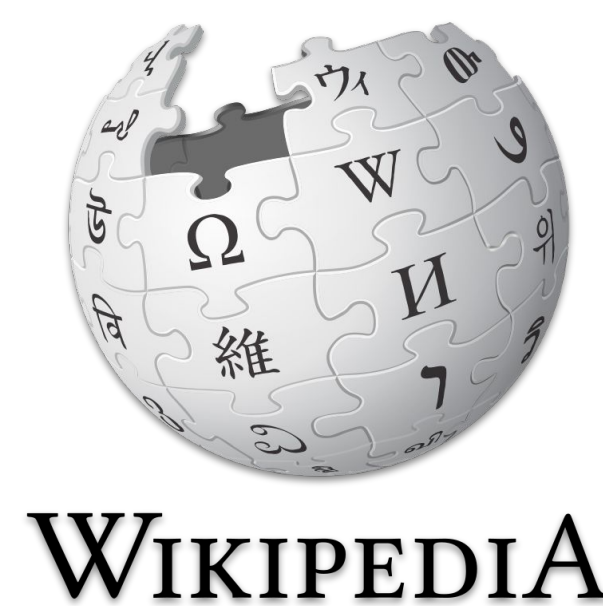
- Google Translate preserves content, but fails to preserve style

### 1. English Phrase → Chinese Phrase

### 2. Same Chinese Phrase → English Phrase

## Dataset

- 57,800 Wikipedia Talk Pages
  - Conversations between Wikipedia editors proposing page edits
- Pages in English, Spanish, Japanese, and Chinese



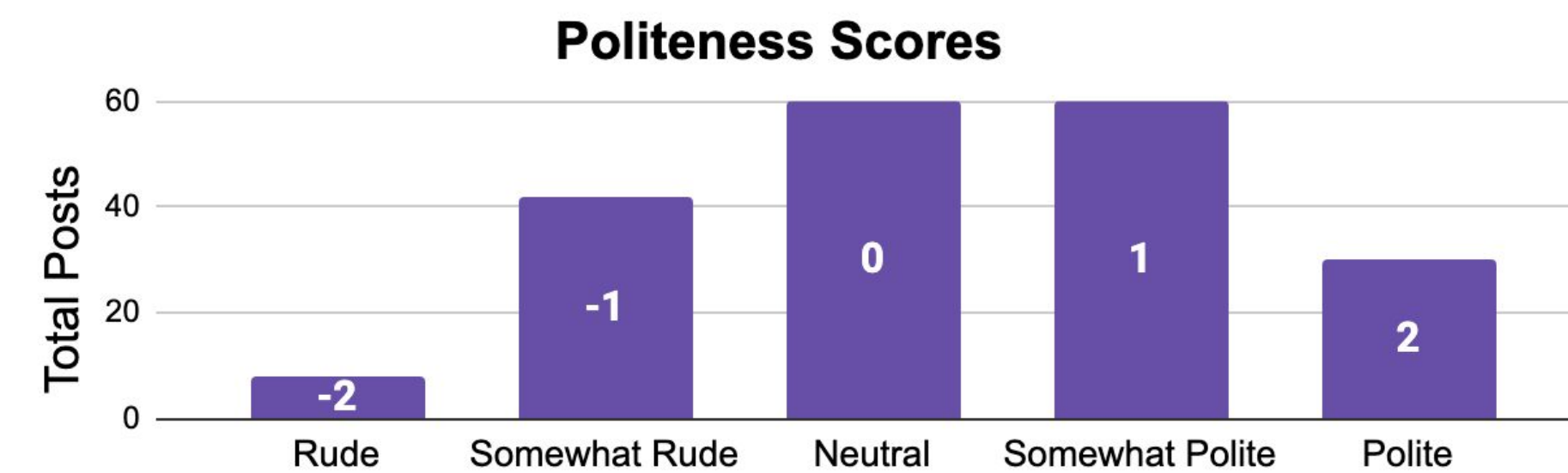
## Task Formulation

### Analyzing politeness in English vs. Chinese

- Past research identifies politeness/rudeness markers in text. Examples:

Politeness Strategy	Example Sentence
Gratitude	I really <b>appreciate</b> that you've done them
Indirect (btw)	<b>By the way</b> , where did you find...
Counterfactual modal	<b>Could/Would</b> you...
Rudeness Strategy	Example Sentence
Direct start	<b>So</b> can you retrieve it or not?
Factuality	<b>In fact</b> you did link...
Direct question	<b>What</b> is your native language?

- Annotaters rate 200 turns from Chinese Wikipedia Talk pages (scaled -2→2)



## Embedding-based Lexica Translation

### Politeness doesn't translate.

### So how do we measure it in Chinese?

- Framework to translate English lexica → other languages using available word embeddings

#### Translation

Use words with a 1:1 translation as category "seed" words. For words with many possible translations, select the word with an embedding closest to the category's average seed embedding

#### Purification

Discard all words that have an embedding greater than  $x_1$  distance from the category's average seed embedding

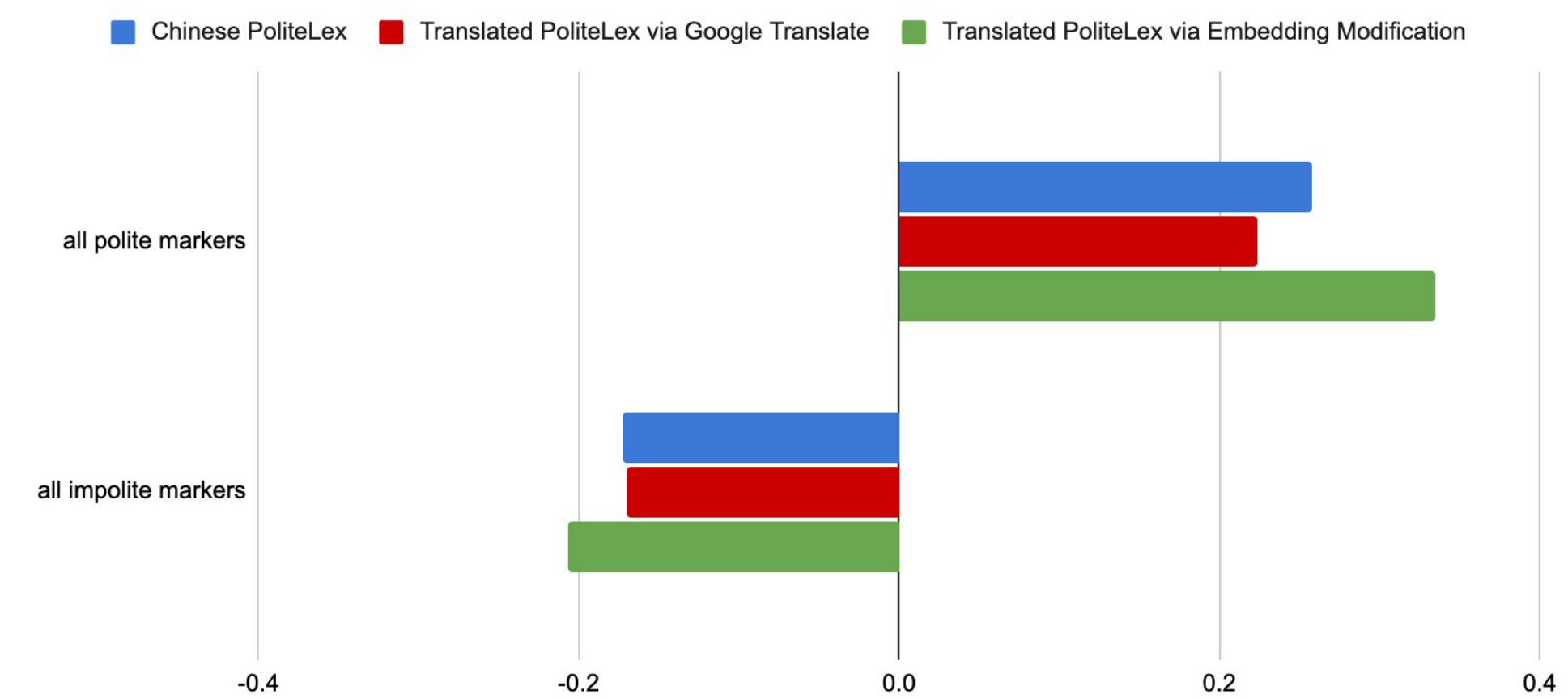
#### Expansion

For each category, add all Chinese words that are up to  $x_2$  distance from the category's average seed embedding

- We translate an English politeness lexicon, "PoliteLex" (Li et. al., 2016) and compare our results to:
  - Chinese PoliteLex (developed by native speakers)
  - English PoliteLex Google Translated to Chinese

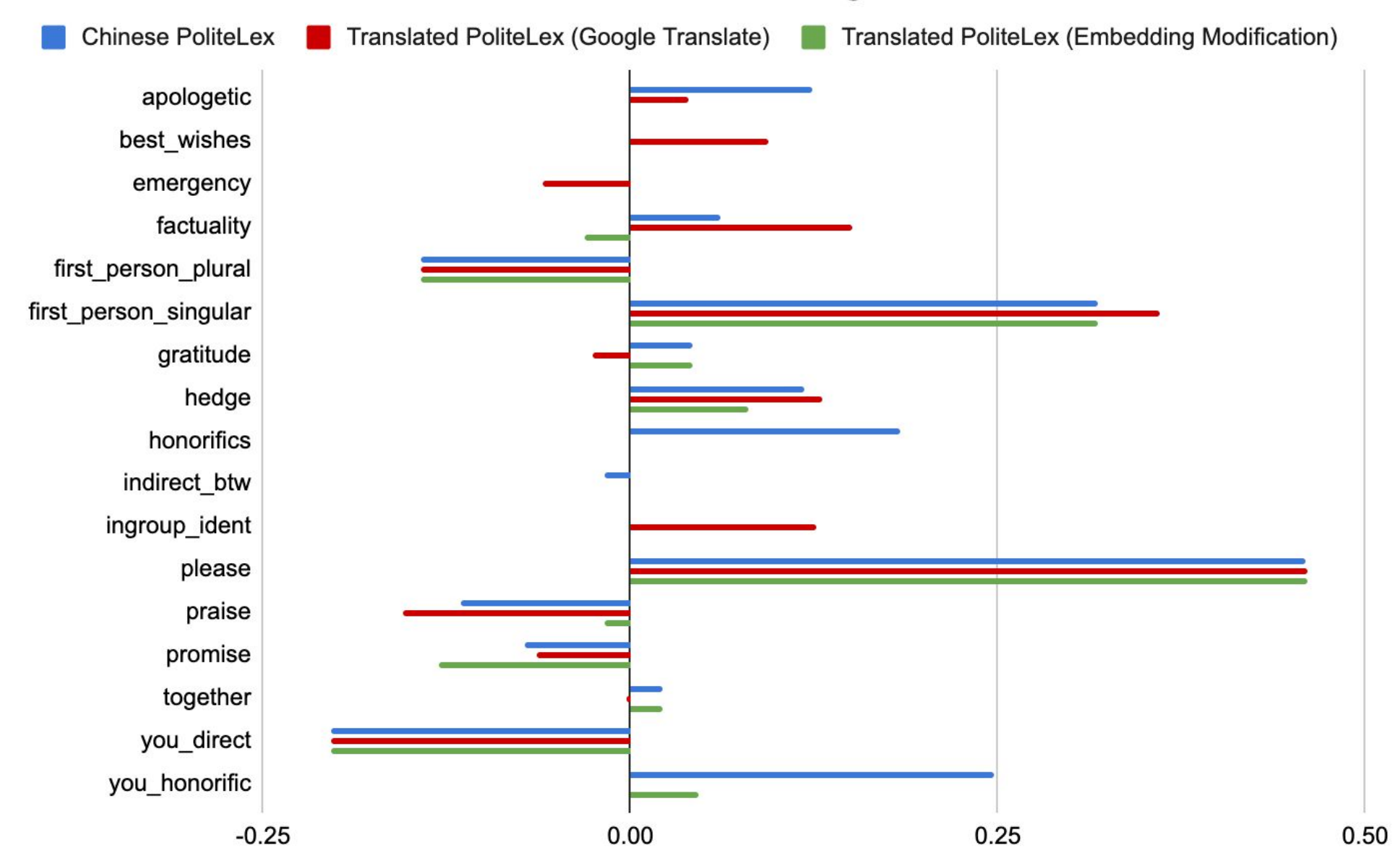
## Overall Results

Pearson Correlations between Polite/Impolite Marker Groups & Politeness Scores



## Lexicon Category Results

Pearson Correlations between Lexica Categories & Politeness Scores



## Applications & Future Work

- Reduce the amount of human engineering (annotations, lexica development) needed to create multilingual style models
- Analyze politeness as a marker of Hofstede's Power Distance
- Use translated lexica as features for deep learning language models
- Create a model that translates both content and style