# Cross-Linguistic Processing of Emotion and Abstraction from English to Chinese

Josephine Kaminaga*, Jennie Wu*, Daniel Yeung*, & Simon Todd

Department of Linguistics

University of California, Santa Barbara

*= Equal Contribution. Authors listed in alphabetical order.

## Introduction

Emotionality and abstraction, and their impact on language processing, have long been important topics of analysis in psycholinguistics. However, most studies have yielded mixed results on the effects of emotionality/abstraction on cross-linguistic tasks. To clarify this relationship between translation, emotionality, and abstraction, we used a neural network to model a simultaneous bilingual mapping within an English-Mandarin semantic space.

### Research Questions and Hypotheses

- Does higher emotionality and/or concreteness facilitate word processing?
- Are the advantages/disadvantages of these measures conveyed in words encoded in the contexts they are used?
- What does word translatability by emotionality and abstraction imply about the structure of semantic spaces cross-linguistically?

## Data

- 38,000 pairs of English words and corresponding Chinese translation equivalents were obtained.

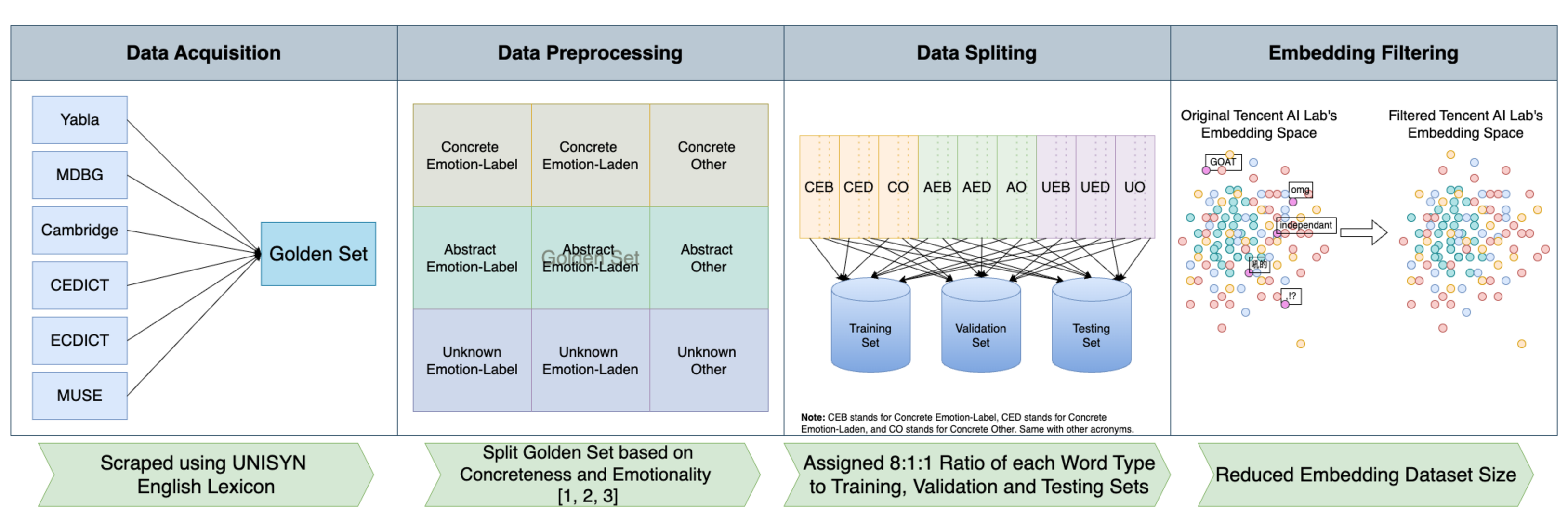



**Figure 1** (above): Data Pipeline

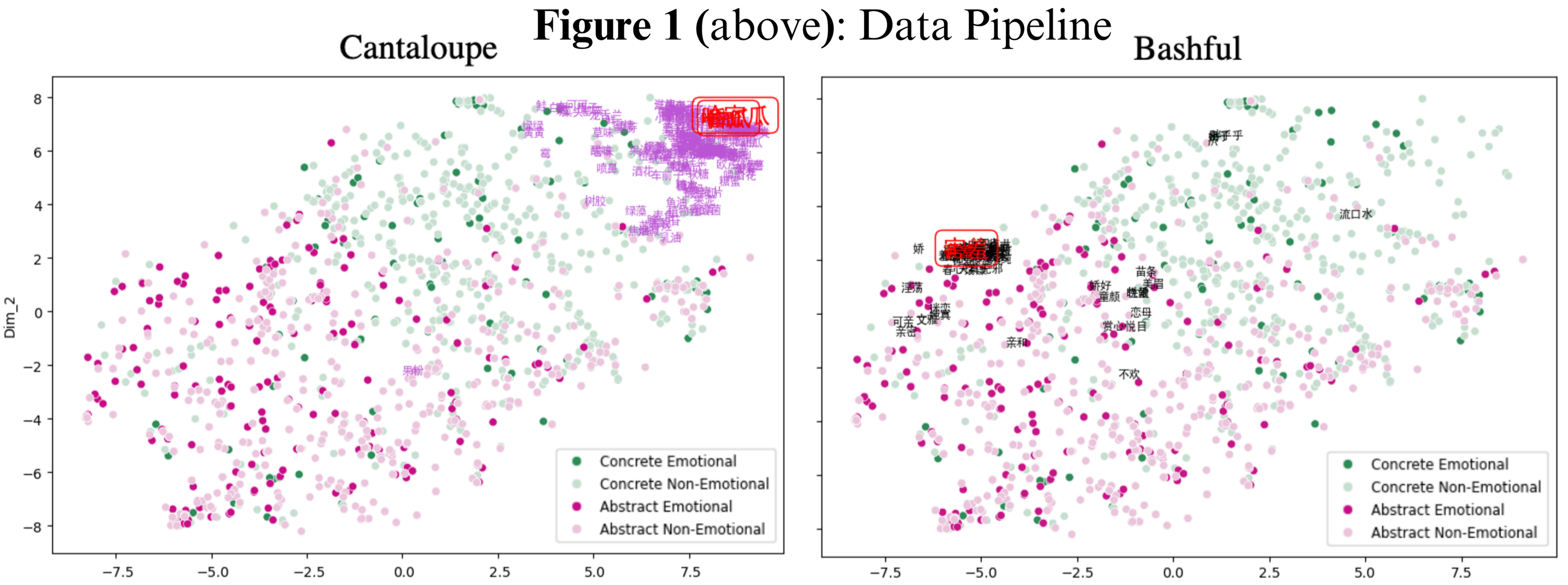
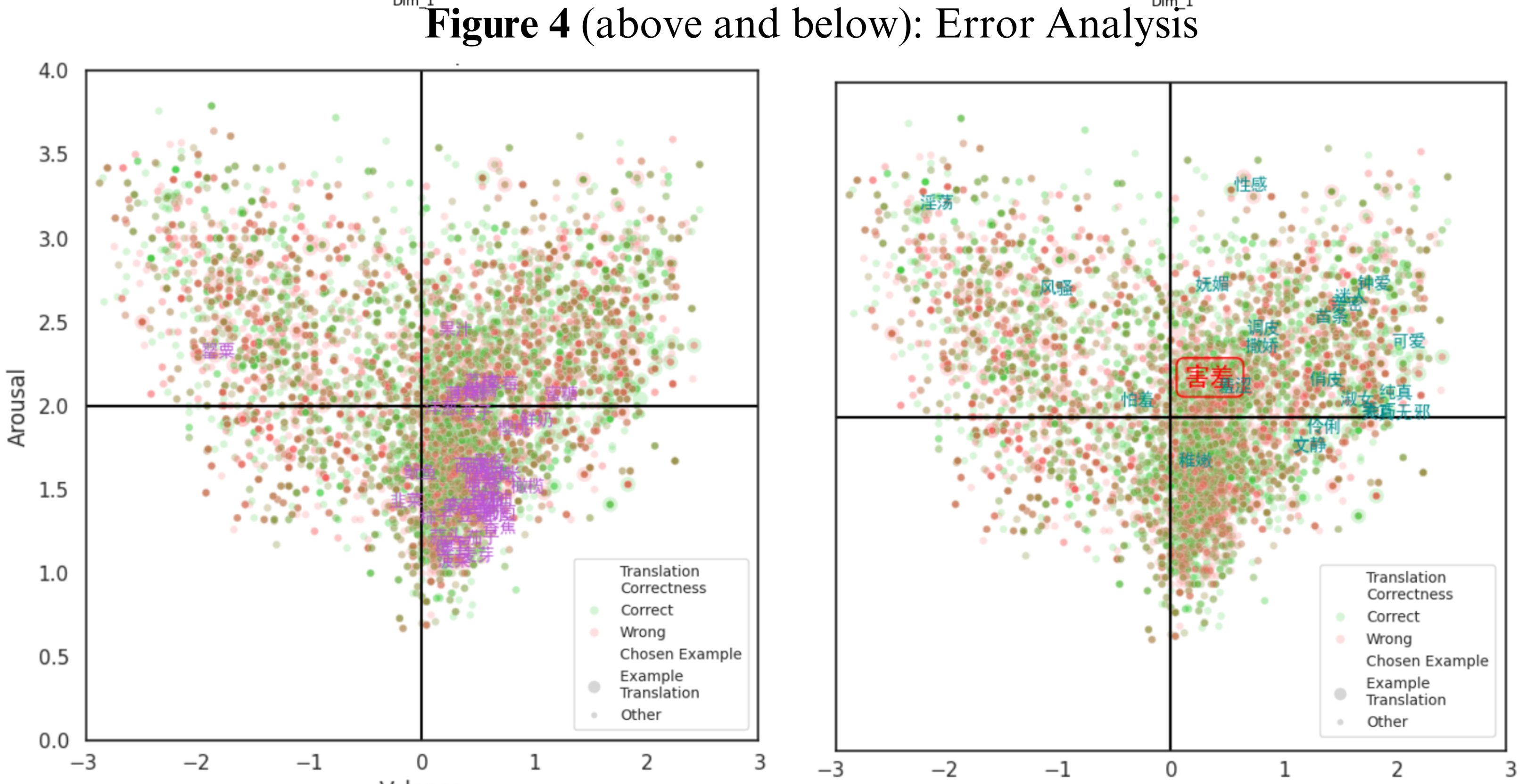



**Figure 4** (above and below): Error Analysis



## Methods

- The autoencoder translates an English word to its Mandarin Chinese equivalent.
- The model uses Tencent AI Lab's pretrained English and Chinese embeddings to learn a semantic space mapping between the two languages.
- We utilized relative embeddings to reduce the variability of our latent spaces [4].
- Model outputs were compared against our sets of dictionary-validated "correct" translations to determine successful outputs.

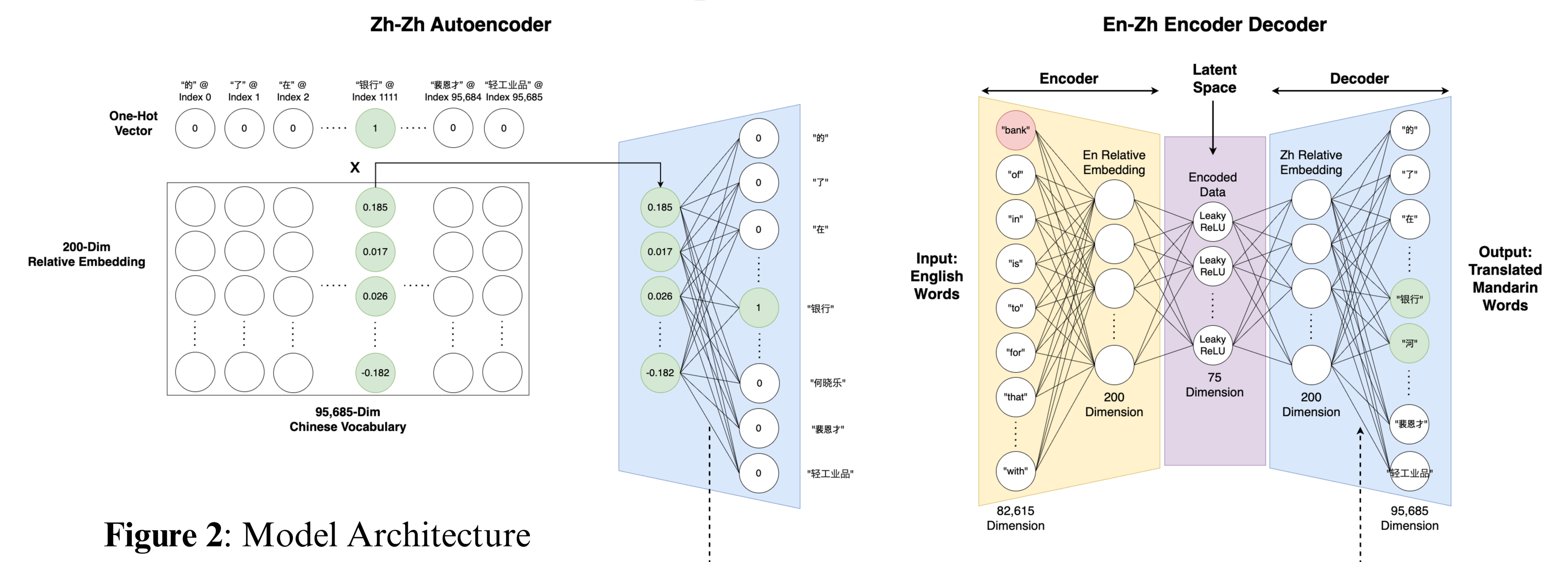



**Figure 2**: Model Architecture

## Results

| Word Class | Emotion Class | Size | Translation Accuracy | Example |
|---|---|---|---|---|
| **Concrete** | Emotional | 195 | **14.36%** | grave, sweet |
| | Non-Emotional | 684 | 8.48% | scallion, raincoat |
| **Abstract** | Emotional | 299 | 5.69% | improve, depressed |
| | Non-Emotional | 536 | 4.66% | control, overall |
| **Unknown Abstraction** | Emotional | 42 | 4.76% | committed, bothering |
| | Non-Emotional | 914 | 3.39% | biking, roadbed |

**Table 1**: Model Performance

- The focus is on analyzing the model's relative performance across these word categories
- Significant difference in translation accuracy: concrete vs. abstract ($p < 0.001$), concrete vs. unknown ($p < 0.001$), and emotional vs. nonemotional ($p < 0.005$).
- Despite task difficulty, the model achieved an F1 score of 0.004 on the test set, which is 40% of its training F1 score (0.01).

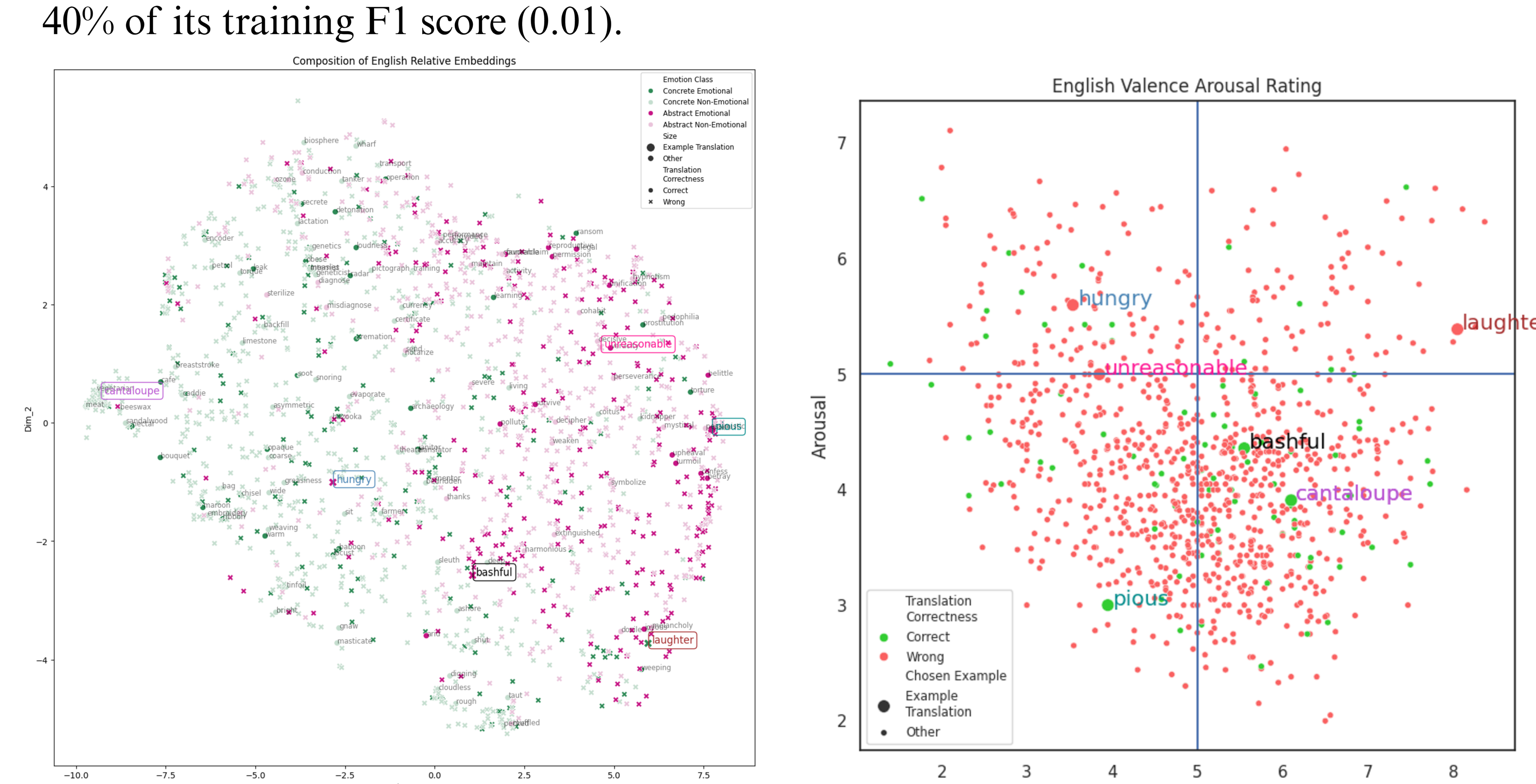



**Figure 3**: Embedding Spaces

## Results & Interpretations

- The better translation performance for concrete and emotional words parallels the ones of previous literature regarding human participants [5].
- The result suggests that what facilitates the processing of emotional and concrete words in humans is encoded in the context of the words' use.
- Despite low translation accuracy, the model outputs words that were highly related to the target translations, indicating the model's success in learning an approximate mapping between the lexical semantic spaces.
- When the model's outputs differ greatly in emotionality from their input word, it could indicate that cultural factors strongly influence the cross-linguistic mapping of language (e.g. bashful being related to Chinese words that comment on attractiveness like "Sexy").

## Limitations

- Our definition of correctness was overly strict: restricting "correct" outputs to lists of dictionary-validated translation counterparts to given input words. Instead, the metric could be more nuanced, such as measures of semantic distance from a target output.
- Polysemy is a major limitation, as differing levels of emotionality and abstraction in the different meanings of polysemous words can lower accuracy.
- The model utilized only one hidden layer, restricting the amount of complex information it can learn.

## Acknowledgement

We would like to thank Mary Bucholtz and the LING 193 thesis seminar class for their support!

Funded by UCSB URCA Grant FR-2023-918

## References

[1] Brysbaert et al., 2014
[2] Zupan et al., 2023; Mohammad and Turney, 2013
[3] Warriner et al., 2013
[4] Moschella et al., 2022
[5] Guasch and Ferré, 2021; Ferré et al., 2017