# Lost in Translation: Using Language Models to Investigate Cultural Adaptation of Chinese Immigrants from Parallel Chinese Restaurant Names in the United States

MIT center for constructive communication

Hang Jiang, Nanxi Liu, Jad Kabbara, Deb Roy | {hjian42, nanxiliu, jkabbara, dkroy} @mit.edu



#### Introduction

American Chinese restaurants (ACR) are central to the dining scene in the United States and have reflected how Chinese immigrants have culturally adapted into American life. As a way of studying this cultural accommodation, previous research has found Chinese restaurants tend to use common and stereotypical words (e.g., "dragon," "china," and "lucky") in their English names. However, prior work paid little attention to the disparities between the parallel Chinese and English names.





#### Contributions

- Conducted a study on 4,030 American Chinese restaurants (12% of the total) sampled across the USA.
- Annotated these parallel Chinese and English names by enrolling crowdworkers to review Google Map photos.
- Designed 7-dimension Name Frames to capture the encoded meanings in names including personal names, specialty, superlatives & other positive modifiers, culture, location, ambiance, romanization, and creative pun.
- Proposed a simple and scalable LLM-based approach to quantify the differences between the parallel names with both frame-based and embedding-based metrics.
- Uncovered systematic nuances in parallel names in ACRs and showed interesting correlation between the meaning nuances and socio-contextual factors.
- Interviewed ACR owners in Greater Boston to understand the motivation for using parallel ACR names.

#### **Human-LLM Annotation Workflow**

- ➤ Prolific crowdworkers double-transcribed Chinese names on 4,030 names.
- > Three authors annotated the name frames on 1,000 pairs of parallel names. Great IIA w/ pairwise Cohen's Kappa scores ranging between 0.79 to 0.93.
- > LLMs are instructed with different prompts to detect name frames.



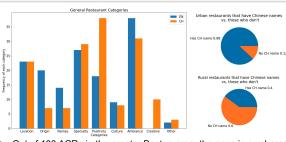
**LLM Annotation** 

## LLM Performance on Name Frame Detection (NFD)

Name Frame	English		Chinese	
	Accuracy	Weighted F1	Accuracy	Weighted F1
	(	GPT-3.5 Turbo		
Personal Name	88.54	84.02	89.74	86.78
Speciality	37.23	37.40	61.05	51.08
Positivity	84.73	84.96	43.95	50.16
Culture	39.38	45.50	41.84	47.50
Location	77.09	77.14	68.16	70.00
Ambiance	73.03	73.69	77.63	76.56
Creative Pun	55.61	68.90	87.11	91.10
Romanization	79.24	80.76	-	_

- The NFD task presents challenges to LLMs, there is room to improve the classification performance. Plan to experiment with multi-label classification as well as smaller models as comparison.
- Zero-shot long prompts with detailed descriptions perform the best out of 10 prompt variants.

### **Empirical Findings**



- Out of 130 ACRs in the greater Boston area, the ones in rural areas are found to be 50% less likely to have a Chinese name compared to those in urban areas.
- Creativity and positivity are more prominent in Chinese names.
- > Name, location & ambiance are more prominent in English names
- > The average embedding similarity between the original and literally translated English names is less than 0.45 (substantial disparities).
- > We showed systematic discrepancies between parallel names.

#### **Future Work**

- > Study the relationship between name meaning change and various social-contextual factors or restaurant popularity/satisfaction.
- > Finish in-depth 10+ interviews with ACR owners to hear their voices.
- ightharpoonup Build an ACR naming toolkit with cultural awareness to help owners.

#### **Acknowledgement & References**

- We would love to acknowledge great advice from Diyi Yang, Rukun Zhang, Cassandra Lee, Artemisia Luk, and Dan Jurafsky.
- Chen, Lindsey NH. "Of authenticity and assimilation: Names of American Chinese restaurants." Names 66. no. 1 (2018): 3-13.
- Ferdman, R. A., and Christopher Ingraham. "We analyzed the names of almost every Chinese restaurant in America. This is what we learned." Washington Post (2016).
- Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. "Can Large Language Models Transform Computational Social Science?." arXiv preprint arXiv:2305.03514 (2023).

## **Codebook** → **Instructions**

Please decide if a restaurant name contains any personal name.

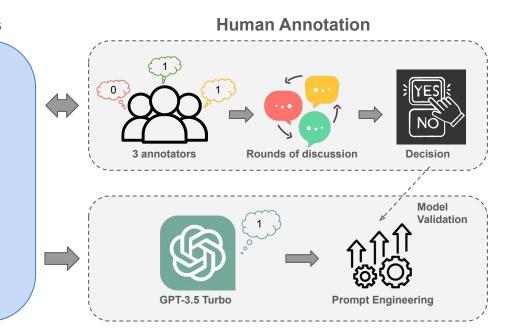
#### Definition

Personal names are usually, but not necessarily, surnames or first names.

## Here are some common examples:

- Surnames with a possessive: Qing's Kitchen, Hoy's Wok
- Names without a possessive: China Lee, Hunan Mao, House of Louie

If the name of the restaurant "Sumiao Hunan Kitchen" contains any personal name, return 1; if not, return 0.



# **LLM Annotation**