Transformations in Chinese Greeting Expressions: A Diachronic Analysis of Question and Non-Question Forms

Zhen Li

1 Introduction

The questions “你吃了吗？”(pinyin: nǐ chī le mā? English: Have you eaten?)” and “你好吗？”(pinyin: nǐ hǎo mā? English: How are you?) are often compared to “How are you doing?” when studying greetings in Chinese. But in fact, questions are very commonly used in the United States, not common in greetings between Chinese, especially strangers. This paper will take a statistic approch to compare the usage of those 3 greetings, and analyze the differences in their usage and pragmatics, trying to provide a diverse and insightful perspective on the evolution, pragmatics, and sociocultural implications of these greetings.

2 Literature Review

2.1 Greetings

2.2 Previous Research on Chinese Greetings

The literature on the comparative study of greetings in Chinese and English languages, especially focusing on expressions such as “你吃了吗？(Have you eaten?)”, “你好吗？(How are you?)” in Chinese, and “How are you doing?” in English, offers a diverse and insightful perspective on the evolution, pragmatics, and sociocultural implications of these greetings.

(Xia et al., 2023)’s study provides a historical context, tracing the evolution of greeting culture in China from the 17th to the 20th century. It reveals a shift towards impersonalization in greetings, increased semantic informativeness, and a departure from traditional politeness norms of self-denigration and other-elevation. This historical perspective is crucial for understanding the current state of Chinese greeting practices.

(Duranti, 1997) and (House & Kádár, 2021) offer frameworks for analyzing greetings across languages and the relationships between conventional expressions and speech acts. Duranti’s six criteria and Juliane’s categorization system for alternative uses of daily expressions like greetings are instrumental in understanding the substantive and social purposes of these linguistic interactions.

(Liu, 2016) and (House et al., 2022) focus specifically on the comparison between English and Chinese greetings. Liu underscores the significance of greetings in social identity and cross-cultural communication, while House addresses the challenges Chinese learners of English face due to pragmatic differences in greeting conventions. House’s inclusion of empirical studies adds depth to our understanding of these challenges.

(Gumperz, 2015) and (Boxer, 2002) delve into the broader field of sociolinguistics and interactional studies, providing foundational concepts and methodologies such as

interactional sociolinguistics and conversation analysis that are essential for analyzing greeting behaviors in different social settings.

(Bobgan, 2000) and (曲卫国 & 陈流芳, 2001) contribute to the understanding of demographic and linguistic specifics. Bobgan explores the influence of age and gender on responses to greetings in English, while 曲卫国 offers a detailed analysis of the linguistic form, topic, and pragmatic constraints of Chinese greetings, highlighting their openness, convertibility, and diversity.

Overall, these works collectively offer a comprehensive view of the sociolinguistic and pragmatic aspects of greetings in Chinese and English, emphasizing historical evolution, comparative analysis, and the influence of social and cultural factors on these everyday linguistic practices.

3 Data and Methodology

3.1 Research questions

In this paper, we analyze 2 representative dataset for Chinese greetings. The first dataset is ***MAGICDATA Mandarin Chinese Conversational Speech Corpus*** (Yang et al., 2022). The second dataset is a set of movie subtitles for Chinese movies, ranging from 1960s to 2020s. They are obtained from Srtku, which is a website for downloading movie and TV show subtitles in various languages. These two datasets are chosen because they contains a representative set of daily conversations. The first dataset consists of 219,325 lines of speach and the second dataset consists of about 50 movies for each decade, enabling us to analyze the diachronic changes of Chinese greetings.

By comparing the usage of the greetings in the 2 datasets, we aim to answer the following three questions:

1. Are there any diachronic changes of greetings used in Chinese conversations?
2. If yes, what are the differences ?
3. What are the socio-historical motivations underlying these differences?

3.2 Data Collection

In this study, we count the frequency of greetings from the subtitles with regular expressions (Table 1) below and analyze the differences in their usage and context.

Regular expression	Matched target
`r'((?<!\p{Han})(你|您)好[啊]?(?!\p{Han}))'`	“你好” with no Hanzi characters before or after
`r'((?<!\p{Han})(上午|下午|晚上|中午|早上)好[啊]?(?!\p{Han}))'`	“你好吗” with no Hanzi characters before or after
`r'((?<!\p{Han})(你|您)(最近)?好[吗么没嘛啊](?!\p{Han}))'`	
`r'(吃[过]?[饭]?了[吗么没嘛啊](?!\p{Han}))'`	“吃了吗” with no Hanzi characters after

* Full width puctuations, such as `。` and `？` are not matched by `\p{script=Han}`.

Table 1: Regular expressions for matching greetings

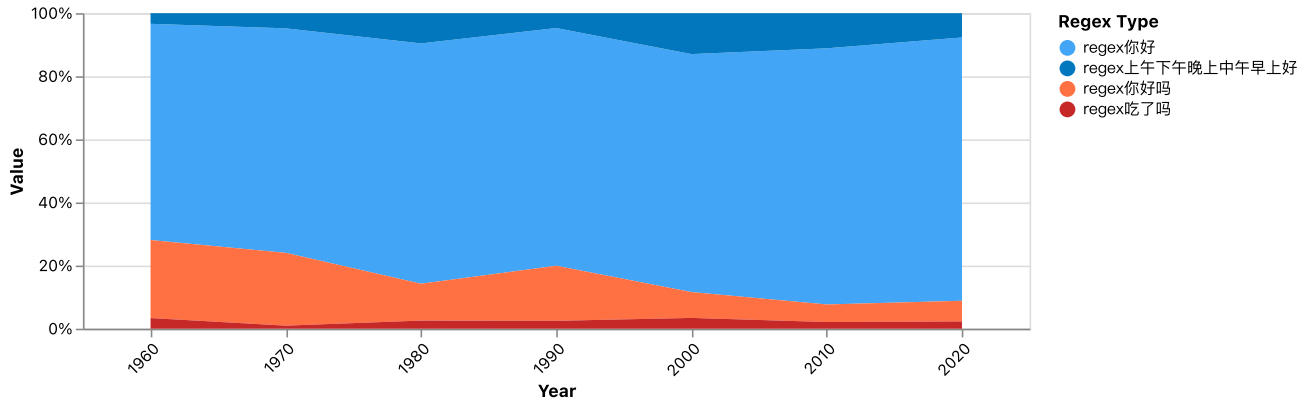

Figure 1: Visualization of the data

4 Result and Discussion

Greeting	Matches
你好	
你好吗 or 吃了吗	

Table 2: Regular expressions for matching greetings

5 Conclusion

Bibliography

Bobgan, J. E. (2000). *A sociolinguistic investigation of the response variable to the question "How are you doing?"* (p. 86). https://www.proquest.com/dissertations-theses/sociolinguistic-investigation-response-variable/docview/231356778/se-2

Boxer, D. (2002). *Applying sociolinguistics: Domains and face-to-face interaction* (Vol. 15). John Benjamins Publishing.

Duranti, A. (1997). Universal and Culture-Specific Properties of Greetings. *Journal of Linguistic Anthropology*, *7*(1), 63–97. https://doi.org/https://doi.org/10.1525/jlin.1997.7.1.63

Gumperz, J. J. (2015). Interactional Sociolinguistics A personal perspective. *The Handbook of Discourse Analysis*, 309–323.

House, J., & Kádár, D. Z. (2021). Altered speech act indication: A contrastive pragmatic study of English and Chinese Thank and Greet expressions. *Lingua*, *264*, 103162. https://doi.org/https://doi.org/10.1016/j.lingua.2021.103162

House, J., Kádár, D. Z., Liu, F., & Liu, S. (2022). Greeting in English as a Foreign Language: A Problem for Speakers of Chinese. *Applied Linguistics*, *44*(2), 189–216. https://doi.org/10.1093/applin/amac031

Liu, L. (2016). Different Cultures and Social Patterns Matter in English and Chinese Greetings. *Literacy Information and Computer Education Journal*, *7*. https://doi.org/10.20533/licej.2040.2589.2016.0310

Xia, D., Huang, L., & Yang, T. (2023). A diachronic study of Chinese greetings between new acquaintances. *Journal of Pragmatics*, *217*, 156–171. https://doi.org/https://doi.org/10.1016/j.pragma.2023.08.007

Yang, Z., Chen, Y., Luo, L., Yang, R., Ye, L., Cheng, G., Xu, J., Jin, Y., Zhang, Q., Zhang, P., & others. (2022). Open Source MagicData-RAMC: A Rich Annotated Mandarin Conversational (RAMC) Speech Dataset. *Arxiv Preprint Arxiv:2203.16844*.

曲卫国, & 陈流芳. (2001). 汉语招呼分析.

Appendix

All the code and data used in this paper can be found at the repository sociolinguistics-greeting-analysis.

Obtaining and Analyzing Movie Subtitle Data

1. Collect the movie names from Douban Explore to filter the movies. Paste the scripts in the browser console and it will automatically click the expand button for 10 times and print 200 ~ 300 movie names in the console.
2. Run `movie_caption_crawl.ipynb` to crawl the movie subtitles. Switch IP if blocked by the website.
   - The movie names are only fuzzy matched in the search box of zimuku.net. Some manual work is needed to remove the wrong matches.
3. Run `clean.ipynb` to clean the downloaded subtitles.
4. Run `encoding.ipynb` to convert the subtitles from `UTF-8 with BOM` / `GB2312` to `UTF-8` encoding.
5. Run `stat.ipynb` to count the frequency of greetings by regex matches in the subtitles.
6. Run `eda.ipynb` to aggregate the results.
7. Copy the saved `regex_sum.json` to `visualization.html` and open it in a browser to visualize the results.

Data Result