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

Zhen Li
zhenli.craig@gatech.edu

1 Introduction

Greetings are a type of speech act that serves to establish and maintain social relationships. They are a fundamental part of everyday communication, and their usage and form are influenced by social and cultural factors. This project investigates the evolving nature of Chinese greetings, focusing on the use of questions like "你吃了吗?" (nǐ chī le mā, *Have you eaten?*) and "你好吗?" (nǐ hǎo mā, *How are you?*), compared to non-interrogative forms, "你好!" (nǐ hǎo, *Hello!*) and "早上好!" (zǎo shàng hǎo, *Good morning!*). Contrary to the commonly used greeting "How are you doing?" in English, there's an observation that such interrogative forms are less prevalent in Chinese greetings, especially among strangers.

To unveil the underlying reasons for this observation, in this project, we examine the diachronic changes in the use of interrogative and non-interrogative forms in Chinese greetings. By analyzing a conversational corpus alongside a set of diachronically categorized movie subtitles, we aim to understand how these greetings have changed over time and what these changes reveal about Chinese society and culture. This examination can be part of a broader inquiry into the nature of language as an evolving entity that mirrors social dynamics. In addition to a statistical approach, we also attempt to delve into the pragmatics and sociocultural significance of these greetings, exploring the relationship between changing Chinese greating patterns and the social context.

2 Literature Review

2.1 Greetings

Duranti proposed a operational definition for analyzing greetings across languages (Duranti, 1997). Duranti's six criteria for alternative uses of daily expressions like greetings are instrumental in understanding the substantive and social purposes of these linguistic interactions.

Gumperz and Boxer 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 (Boxer, 2002; Gumperz, 2015).

2.2 Previous Research on Chinese Greetings

Xia's study provides a historical context, tracing the evolution of greeting culture in China from the 17th to the 20th century (Xia et al., 2023). 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.

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

Qu offers a detailed analysis of the linguistic form, topic, and pragmatic constraints of Chinese greetings, highlighting their openness, convertibility, and diversity (Qu & Chen, 2001).

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 project, we analyze 2 representative datasets 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. These two datasets are chosen because they contain a representative set of daily conversations. The first dataset consists of 219,325 lines of speech 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 social motivations behind these changes?

3.2 Data Collection

In this study, we intend to extract the greetings from the conversation a corpus and movie subtitles and count their frequency of appearance to analyze the differences in their usage and context.

The corpus, MAGICDATA Mandarin Chinese Conversational Speech Corpus, includes 180 hours of Mandarin Chinese speech from 633 speakers. All the transcripts are joined into a 219,325 line file for further analysis.

We use the movie lists from Douban, an online movie database that provides rich filters for movie search. This website provides recommendations for movies from different decades, which enabled us to perform a diachronic analysis over a representative set of movies. The movie subtitles are crawled from Srtku. For each decade from 1960s to 2020s, we select the

recommended 200~300 movies and download their subtitles from Srtku. After cleaning, the final dataset contains 820 readable movie subtitles in total:

1960s	1970s	1980s	1990s	2000s	2010s	2020s
40	101	201	169	90	142	77

Table 1: Number of movies for each decade

While the movie dialogues primarily capture the perceptions of scriptwriters and directors regarding daily conversations, or perhaps even only prestige forms adopted by the movie creators, it can reflect real-world conversational patterns and offer valuable insights into broader linguistic trends in greetings.

3.3 Greeting Extraction

We use a naïve regular expression approach to extract the greetings from the corpus and movie subtitles. For each category of greetings (non-interrogative forms and interrogative forms), 2 regular expressions are designed to match the most frequently used greetings in daily conversations. The regular expressions are listed in Listing 1.

Category	Regular expression and captured sentences
Non-interrogative forms	`r'((?<!\p{Han})(你|您)好[啊]?(?!\p{Han}))'` Informal or formal *"Hello!"*
	`r'((?<!\p{Han})(上午|下午|晚上|中午|早上)好[啊]?(?!\p{Han}))'` *"Good morning/afternon/evening/noon/(early) morning!"*
Interrogative forms	`r'((?<!\p{Han})(你|您)(最近)?好[么么没嘛啊](?!\p{Han}))'` Informal or formal *"How are you doing?"*
	`r'((?<!\p{Han})(你|您)?(已经)?吃[过]?[饭]?了[吗么没嘛啊](?!\p{Han}))'` Informal or formal *"Have you eaten?"*

Listing 1: Regular expressions for matching greetings

Take the first regular expression as an example, there are 5 major parts:

1. `(?<!\p{Han})` : matches the position where the previous character is not a Hanzi character. The `\p{Han}` represents any character in the Unicode Han script.
2. `(你|您)` : matches either 你 (nǐ, *you*) or 您 (nín, formal *you*).
3. 好 : matches 好 (hǎo), which translates to *"good"* or *"well"* in English.
4. `[啊]?` : matches one or zero times of 啊 (ā, an exclamation or modal particle used for emphasis or mood).
5. `(?!\p{Han})` : matches the position where the next character is not a Hanzi character.

The leading and trailing `(?<!\p{Han})` and `(?!\p{Han})` are used to avoid disambiguation from the sentences like "你好棒" (*You are so amazing*). All possible sentences matched by regexes used in this project represents a sentence with similar meanings, but in some scenario they might

convey more of a contextual meaning and hence not a standard "greeting" defined by (Duranti, 1997). Detailed explanation for other 3 regular expressions can be found in Section F.

After matching the greetings with regular expressions, the occurrences of greetings in the corpus and movie subtitles are counted respectively and displayed in Table 2. The results are visualized as the normalized stacked area chart shown in Figure 1.

Decade	(你|您)好[啊]? *Hello!*	(上午|下午|晚上|中午|早上)好[啊]? *Good morning/ afternoon/...!*	(你|您)(最近)?好[吗么没嘛啊] *How are you?*	(你|您)?(已经)?吃[过]?[饭]?了[吗么没嘛啊] *Have you eaten?*
1960	43	0	10	2
1970	74	5	24	1
1980	318	38	50	2
1990	263	17	49	5
2000	216	38	24	8
2010	417	58	28	7
2020	217	20	17	3
MAGICDATA Corpus	8	0	0	1

Table 2: Occurences of greetings in the MAGICDATA Mandarin Chinese Conversational Speech Corpus and movies
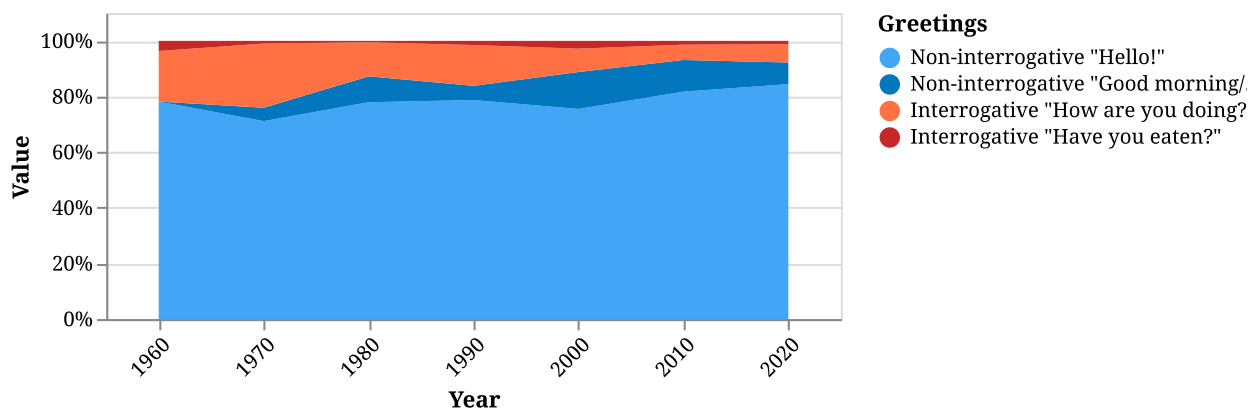

Figure 1: Changes of the frequency of greetings in the movie subtitles from 1960s to 2020s

4 Result and Discussion

From Figure 1, we can observe an obvious decline of interrogative forms and a rise of non-interrogative forms. Specifically, the usage of （你|您)(最近)?好[吗么没嘛啊] (*How are you?*) has been a major part of decline in interrogative forms.

To further explore the reason behind the decline in the use of interrogative greetings, we specifically analyze dialogues from *Rickshaw Boy*, *A Better Tomorrow II* and *A Journey Through*

Time with Antony, as highlighted in Figure 2, Figure 3, Figure 4. These instances demonstrate a strong correlation between the use of interrogative greetings and the conversation's context.

- 吃饭了吗
  Have you eaten?

  　　　　　　　　　　　　　　　　　　　　　　　　- 还没拉上座儿呢
  　　　　　　　　　　　　　　　　　　　　　　　　No business yet.[1]

  　　- 我这有钱　先拿去吃饭吧
  I have some money here, take it and go eat first.

Figure 2: A dialogue from the movie *Rickshaw Boy*
(骆驼祥子), China mainland, 1982

- 胡 Sir　你还是不要退休了　还有很多事情等着你做
  Hu Sir, you shouldn't retire. There are still many things waiting for you to do.

  　　　　　　　　　　　　　　　　- Ken 哥　我们很挂念你　你好吗?
  　　　　　　　　　　　　　　　　Ken, we miss you so much. **How are you doing?**

Figure 3: A dialogue from the movie *A Better Tomorrow II*
(英雄本色 II), Hong Kong, 1987

- 我们在澳洲　就算是陌生人见面　也不说「嗨」　而是说「你好吗」
  When we meet strangers in Australia, we don't say **"hi".** *We say* **"How are you?"**

  　　　　　　　　　- 你已经养成习惯了　每次给我发讯息都说　「你好吗?」
  You've adapted to this custom. You say **"How are you?"** *every time you send me a message.*

Figure 4: A dialogue from the movie *A Journey Through Time with Antony*
(陪安东尼度过漫长岁月), China mainland, 2015

In *Rickshaw Boy* (Figure 2), the greeting *"Have you eaten?"* is closely tied to the context where the listener is a rickshaw puller leading a difficult life, struggling to make ends meet. This greeting is a query about the need for food mixed with concern for the other's well-being, reflecting the socio-economic status of the speakers and emphasizing basic needs and communal care.

In the second example (Figure 3) from *A Better Tomorrow II*, the greeting *"How are you doing?"* is set against the backdrop of a reunion. This greeting conveys a care close to its literal meaning and reflects the speaker's genuine concern following a long period of separation. Distinct from the common usage of this phrase in English, where it often functions as a conversational norm, the Chinese version serves more as an expression of genuine concern and longing, marking the re-establishment of a connection after a period of separation.

In the third conversation (Figure 4) from *A Journey Through Time with Antony*, the two speakers reunited after a long separation across different countries. The second speaker's

---

[1]I didn't find an English version of subtitle for *Richshaw Boy*. The translation might be inaccurate. The speaker is a rickshaw puller and he says the seat is not taken yet, which implies he hasn't got any earnings for a meal.

comment on the first speaker's greeting underscores the notable difference between the Chinese and English versions of this greeting. The Chinese version of "*How are you doing?*" is not a commonly used greeting and often implies more personalized concern and intimacy, which is the reason why the second speaker highlighted the first speaker's use of this greeting.

The close relationship between context and the usage of interrogative greetings resonates with the analysis of (Qu & Chen, 2001). Due to the openness of the topic, Chinese greetings are highly context-dependent. Their specific usage is influenced by various pragmatic factors. Movie dialogue examples further highlight how the function and significance of interrogative greetings in Chinese are intricately linked with speakers' social dynamics and situational context. In contrast, non-interrogative greetings like "Hello!" are more impersonal and less reliant on context, hence emerging as a more prevalent form of greeting.

The evolving patterns of greetings mirror changes in contemporary Chinese society. These patterns, as seen in movie subtitles, align with findings from (Xia et al., 2023)'s study. By examining classic Chinese literature, they observed a trend toward "impersonalization" in greetings between new acquaintances in the 20th century, both in topic selection and sentence structure. This trend is the most notable diachronic change in interrogative greetings, corresponding with their decline in our study. As the economy develops and the focus shifts away from basic survival, the frequency of traditional greetings like "吃了吗?(*Have you eaten?*)" decreases, as noted by (Xia et al., 2023). In a more stable and fast-paced society, the prevalence of non-interrogative greetings like "你好!(*Hello!*)" increases. Technological advancements, especially the widespread use of social media and instant messaging, have also transformed communication. The platforms often encourage quicker and more straightforward interactions, possibly favoring more standardized greetings. As noted in the movie conversations (Figure 4), the speakers get more used to impersonalized greetings like "*Hi*" and English-styled "*How are you doing?*" becomes a somehow quirky variation of greeting.

5 Conclusion

In conclusion, this project has provided a comprehensive analysis of the evolution of Chinese greetings through a corpus and a set of movie subtitles, examining their transition from question forms like "你吃了吗?" (*Have you eaten?*) and "你好吗?" (*How are you?*) to non-interrogative expressions such as "你好!" (*Hello!*) and "早上好!" (*Good morning!*). Through an quantitative analysis of conversational corpora and movie subtitles, we have traced the diachronic changes in greeting practices and explored their sociocultural implications.

Our findings indicate a shift in greeting practices within Chinese society. The decline in the use of interrogative greetings and the rise of non-interrogative forms reflect broader sociocultural changes, including shifts towards greater impersonalization, and alterations in politeness norms. These trends mirror the ongoing transformations in Chinese social dynamics and cultural practices.

Bibliography

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., 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

Qu, W., & Chen, L. (2001). 汉语招呼分析 *(An Analysis of Chinese Greetings)*.

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*.

Appendix

All the data, codes, and graphs used in this project can be found at the GitHub repository [sociolinguistics-greeting-analysis](#).

F Explanation of the Regular Expressions

Constraints for avoiding surruounding Hanzi characters are omitted below.

F.a Non-interrogative form:
`(?<!\p{Han})(上午|下午|晚上|中午|早上)好[啊]?(?!\p{Han})` (*Good morning / evening / ... !*)

- `(上午|下午|晚上|中午|早上)` : matches either `上午` (shàngwǔ, *morning*), `下午` (xiàwǔ, *afternoon*), `晚上` (wǎnshàng, *evening*), `中午` (zhōngwǔ, *noon*), or `早上` (zǎoshàng, *early morning*).
- `好` : matches `好` (hǎo), which translates to *"good"* or *"well"* in English.

F.b Interrogative form: `(?<!\p{Han})(你|您)(最近)?好[吗么没嘛啊](?!\p{Han})` (*How are you?*)

- `(你|您)` : matches either `你` (nǐ, *you*) or `您` (nín, formal *you*).
- `(最近)?` : matches one or zero times of `最近` (zuìjìn, *recently*).
- `好` : matches `好` (hǎo), which translates to *"good"* or *"well"* in English.
- `[吗么没嘛啊]` : matches either `吗` (ma, question particle), `么` (me, question particle), `没` (méi, negative particle), `嘛` (ma, modal particle), or `啊` (ā, an exclamation or modal particle used for emphasis or mood). In this greeting, they are used as modal particles.

F.c Interrogative form:

`(?<!\p{Han})(你|您)?(已经)?吃[过]?[饭]?了[吗么没嘛啊](?!\p{Han})` (*Have you eaten?*)

- `(你|您)?` : matches 你 (nǐ, *you*), 您 (nín, formal *you*) or nothing.

- `(已经)?` : matches one or zero times of 已经 (yǐjīng, *already*).

- `吃` : matches 吃 (chī, *eat*).

- `[过]?` : matches one or zero times of 过 (guò, a mark indicating the completion of an action).

- `[饭]?` : matches one or zero times of 饭 (fàn, *meal*).

- `了` : matches 了 (le, perfective particle).

- `[吗么没嘛啊]` : matches either 吗 (ma, question particle), 么 (me, question particle), 没 (méi, negative particle), 嘛 (ma, modal particle), or 啊 (ā, an exclamation or modal particle used for emphasis or mood). In this greeting, they are used as modal particles.

---

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

* The standard library `re` in Python doesn't support Unicode property escapes like `\p{Han}`. I used the [regex](#) library instead.

G Obtaining Movie Subtitles

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 SrtKu. Some manual work is needed to remove the wrong matches.

   - Due to limited time, the data cleaning process was conducted with a basic approach. This might have led to the inadvertent inclusion of some non-Chinese movies in the dataset.

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.

H Analyzing Conversational Corpus and Movie Data

1. Run `count_regex.ipynb` to count the appearances of greetings in subtitles by regex matches.

   - Run `eda.ipynb` to aggregate the results.

- Copy the generated `regex_sum.json` to `visualization.html` and open it in a browser to visualize the results.

2. Run `count_magic.ipynb` to count the appearances of greetings in the corpus by regex matches.

All subtitles can be found in the `caption` folder.

Results can be found in the root folder.